CENTROMERE SIZE AND ITS IMPACT ON HAPLOID INDUCTION IN MAIZE

by

NA WANG

(Under the Direction of R. KELLY DAWE)

ABSTRACT

Centromeres are specified by the histone variant CENH3, which can interact with centromeric retroelements and tandem repeat arrays to determine the locations and sizes of active centromeres. Centromere mediated haploid induction occurs when crossing wild-type plants with plants expressing CENH3 variants. Centromere size variation has been proposed to underlie uniparental genome elimination that leads to haploid formation. In this hypothesis, the centromere size dimorphism between the wild type and CENH3 variant genomes in the zygote causes the removal of the smaller centromeres and uniparental genome elimination. In this study, we employed multiple approaches to explore centromere size and its impact on haploid induction in maize. We first investigated the centromere size across 26 maize inbred lines and demonstrated that centromere size is not influenced by centromere sequence but positively correlated with chromosome size and genome size. We then confirmed the relationship between genome size and centromere size by introducing maize centromeres into two larger genome backgrounds Oaxaca and *Zea luxurians*. Our results suggest that maize centromeres are expanded in the larger genome background. Literature from other species showed that overexpression of CENH3 could cause the ectopic formation of functional centromeres. However, we found threefold overexpression of CENH3 in maize did not significantly increase

centromere size. Taken together, these results suggest that centromere size is scalable with genome size and controlled by multiple limiting factors but not exclusively to CENH3.

To test the centromere size model for haploid induction, we utilized CRISPR/Cas9 to create a maize *cenh3* null mutant that is homozygous lethal but poorly transmissible through both male and female gametophytes. We found that haploids were formed at high frequency when wild-type plants are crossed with *cenh3* heterozygous mutants either as female (haploid induction ratio 5%) or male (0.5%). Our results are consistent with the hypothesis that diminished/small centromeres induce haploid formation. Genotyping endosperm of the seeds that gave rise to haploid plants showed that all were fertilized by *cenh3* pollen. Since CENH3 is present in all plants, this method may have potential to induce haploid across diverse species.

INDEX WORDS: centromere, CENH3, centromere size, genome size, haploid induction

CENTROMERE SIZE AND ITS IMPACT ON HAPLOID INDUCTION IN MAIZE

by

NA WANG

B.S., Henan Normal University, P.R.China, 2009

M.S., Sun Yat-Sen University, P.R.China, 2012

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

2020

CENTROMERE SIZE AND ITS IMPACT ON HAPLOID INDUCTION IN MAIZE

by

NA WANG

Major Professor:	R. Kelly Dawe
Committee:	James Leebens-Mack Wolfgang Lukowitz Robert Schmitz

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
August 2020

ACKNOWLEDGEMENTS

I would like to first express my sincere gratitude and appreciation to my advisor R. Kelly Dawe. His consistent guidance, dedicated support, and encouragement have made my PhD study an inspiring experience for me. He has taught me not only how to do science but also how to be a good advisor. I would like to thank my committee James Leebens-Mack, Wolfgang Lukowitz, Robert Schmitz and former committee Xiaoyu Zhang for their critical comments and invaluable advice on my dissertation projects.

I would like to thank all past and current lab members of the Dawe lab for their contributions, help and suggestions to my project. In particular, I would like to thank Jonathan Gent for his contribution to my training and mentorship. I would like to express my special thanks to Jianing Liu who helped me a lot with the centromere size analysis. I thank Bill Ricci for his great help on western blot and flow cytometry analysis. I would like to thank my undergraduate worker Tindall Smith and Sydney Greenlaw for all the tedious work on genotyping.

I thank all the members of the Plant Functional Genomics Center and Centromere Group for their support and valuable suggestions. I would like to thank the plant biology greenhouse staff Mike Boyd and former greenhouse staff Kevin Tarner for taking great care of my plants.

I thank my family, especially my mother and mother-in-law. They traveled from China to the US to help me take care of my daughter. I thank my husband, Shaopeng Li, for his unconditional love and support. Finally, I would like to thank my sweetest daughter, Isabella Li, who brings me so much happiness and joy.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

In the great majority of all plants and animals, centromeres are defined not by sequence but by the presence of a specialized histone H3 variant known as CENP-A/CENH3 (Warburton et al., 1997; Zhong et al., 2002; Allshire and Karpen, 2008). Evidence supporting the fundamental importance of CENH3 derives from the fungal and animal literature where there is an unequivocal biochemical connection between the presence of CENH3 and the localization and function of the overlying proteins that make up the kinetochore (Stoler et al., 1995; Buchwitz et al., 1999; Howman et al., 2000; Maehara et al., 2010; Sekulic and Black, 2012). While the genetic and biochemical experiments required for studying cell division are more challenging in plants, RNAi-mediated *CENH3* knockdown in *Arabidopsis* resulted in a reduced number of mitotic divisions and abnormal meiosis, consistent with the view that CENH3 is essential for chromosome segregation (Lermontova et al., 2011a).

CENH3 differs from H3 by a divergent N-terminal tail and specialized sequence in the C-terminal histone fold domain. The sequence of the N-terminal region is highly variable in plants and is thought to be involved in the loading of CENH3 during meiosis (Lermontova et al., 2006; Ravi et al., 2011; Maheshwari et al., 2015) and may have other functions in regulating cell division (Zhang et al., 2005). Sequences in the histone fold domain are sufficient for centromeric localization in animals (Sullivan et al., 1994; Vermaak et al., 2002; Black et al., 2004; Morey et al., 2004), *Arabidopsis*, and other non-plant models (Morey et al., 2004; Lermontova et al., 2006). Specifically, a region known as the CENP-A centromere-targeting domain (CATD),

which includes loop1 and the α2 helix that are exposed to the outside of the folded protein, is essential for centromere targeting (Vermaak et al., 2002; Black et al., 2004; Lermontova et al., 2006). In animals, CENH3 nucleosomes bind directly to two inner kinetochore proteins known as CENP-C and KNL2 (Maddox et al., 2007; Gopalakrishnan et al., 2009; Shono et al., 2015; Hori et al., 2017; Sandmann et al., 2017). Both CENP-C and KNL2 are also found in plants and appear to have similar properties (Dawe et al., 1999; Lermontova et al., 2013; Sandmann et al., 2017). These proteins and others form the base of the large multi-protein kinetochore complex that binds to microtubules and regulates chromosome segregation (Figure 1.1). While CENH3-containing nucleosomes bind to CENP-C and KNL2, CENP-C interacts with the MIS12 complex, which associates with the NDC80 complex that ultimately interacts directly with microtubules (Cheeseman et al., 2006). The major components, including KNL2, CENP-C, MIS12, and NDC80 are conserved in plants and most other species (Dawe et al., 1999; Talbert et al., 2004; Sato et al., 2005; Du and Kelly Dawe, 2007; Li and Kelly Dawe, 2009; Lermontova et al., 2013). In addition, other proteins such as *Arabidopsis* γ-tubulin complex protein 3-interacting proteins (GIPs) may also contribute to CENH3 deposition and/or stability (Batzenschlager et al., 2013, 2015). Generally speaking, the removal of CENH3 by mutation abolishes the localization of CENP-C and all other kinetochore proteins (Howman et al., 2000; Régnier et al., 2005). However, in some insects, CENH3 and CENP-C are absent with the outer kinetochore proteins MIS12 and NDC80 retained (Drinnenberg et al., 2014).

Plant Centromere Sequences

In plants, CENH3-containing nucleosomes localize to a multiplicity of highly divergent and difficult to categorize sequences. *Arabidopsis* typifies the simplest form of centromeric

organization, which is based on largely uninterrupted arrays of simple nucleosome-sized repeat (180 bp, plus or minus) sequences known as satellites (Murata et al., 1994). In rice and maize, centromeres are composed of two different types of repeats, satellite repeats (rice: CentO, 155 bp; maize: CentC, 156 bp) and centromere-specific retrotransposons (CR elements) (Jiang et al., 2003). However, these components can vary tremendously. The centromeres of the rice relative *Oryza brachyantha* does not have canonical CR elements and instead contains a novel Ty3-gypsy retrotransposon and a satellite repeat called CentO-F that has no sequence similarity to CentO (Lee et al., 2005; Gao et al., 2009). In potato, there are several unrelated types of satellite arrays on different chromosomes (Gong et al., 2012; Zhang et al., 2014), and a closely related wild species *Solanum verrucosum* does not share these repeats (Gong et al., 2012). Genes may occur within centromeres, but they are rare and have generally weaker association with CENH3 than repetitive DNA in centromeres (Yan et al., 2008). Taken together, the available data suggest that while most plant centromeres have satellite arrays and retroelements, there are few obvious limitations on the sequence makeup of centromeres.

Replication of CENH3 Position

The process of CENH3 loading to centromeres is thought to involve three main steps: initiation, deposition, and maintenance (De Rop et al., 2012). The CENH3 initiation process produces the epigenetic context for CENH3 assembly by the Mis18 complex, which includes the key protein Mis18BP1/KNL2. The Mis18 complex loads onto centromeres before CENH3 assembly in animals (De Rop et al., 2012). The deposition process is dependent on chaperones such as the human holiday junction recognition protein (HJURP), the *Drosophila* chromosome alignment defect 1 protein (CAL1), and the yeast suppressor of chromosome missegregation

protein 3 (SCM3) (Mizuguchi et al., 2007; Dunleavy et al., 2009; Foltz et al., 2009; Mellone et al., 2011). Homology among these three chaperones is barely detectable at either the sequence or structural level (Phansalkar et al., 2012) and the functionally equivalent protein(s) in plants is not yet known. The CENH3 maintenance process in humans is thought to involve ATP-dependent chromatin remodelers and spacing factors that stabilize the newly incorporated CENH3 (Obuse et al., 2004; Izuta et al., 2006).

Different organisms recruit CENH3 to centromeres at different cell-cycle stages. The loading of CENH3 in *Drosophila* starts at late telophase and continues during G1 with a low level of CENH3 turnover at other cell-cycle stages (Schuh et al., 2007). In human, CENH3 is mainly loaded during G1 with no detectable turnover at other stages (Jansen et al., 2007) and budding yeast CENH3 incorporation occurs during S phase with a low level of turnover at telophase and G1 (Pearson et al., 2004). In contrast, in several plants tested, CENH3 is loaded primarily during late G2 (Nagaki et al., 2005; Lermontova et al., 2007) with a low level of turnover during G1 to the early G2 (Lermontova et al., 2011b), and CENH3 is naturally divided equally between the replicated DNA strands at S phase and replenished later in G2 (Lermontova et al., 2006).

The molecular mechanism of CENH3 deposition is an area of active study in many species. We know from the timing of deposition that it cannot be coupled to DNA replication such that information is transferred from one strand to the other while the newly replicated strands are still in close proximity. Instead, CENH3 replication must make use of existing marks that are left over after the CENH3 nucleosomes have been depleted by half during replication. Recent data strongly suggest that both KNL2, CENP-C and NASP^{SIM} take part in this process (Hori et al., 2017; Sandmann et al., 2017; Le Goff et al., 2020). CENP-C and KNL2 share a

region known as the CENPC-k motif, which is homologous to a CENP-C domain in animals that binds directly to the CATD of CENH3 nucleosomes (Sandmann et al., 2017). In addition, both KNL2 and CENP-C bind to DNA (Du et al., 2010; Sandmann et al., 2017). Critically, binding to both CENH3 and flanking DNA would enable these proteins to serve as molecular tethers that bridge the non-conserved DNA substrate to the location of the CENH3 protein. Both KNL2 and CENP-C interact with the mammalian HJURP chaperone to recruit new CENH3 nucleosomes (French et al., 2017) presumably to positions in vicinity to existing CENH3 or to locations where CENH3 recently resided. CENP-C serves a particularly important role because it binds to CENH3 and its chaperone, to DNA, and to MIS12 and the outer kinetochore (Screpanti et al., 2011; Klare et al., 2015). The histone H3 chaperone $NASP^{SIM}$ interacts with both N-terminal part and histone fold domain of CENH3, but it's not closely associated with chromatin, indicating it may function in escorting non-nucleosomal CENH3 histones for CENH3 deposition (Le Goff et al., 2020). KNL2, CENP-C and $NASP^{SIM}$ may be the primary and only heritable marks for centromere position during plant female gametogenesis and zygote formation; in *Arabidopsis*, GFP-tagged CENH3 is apparently absent from egg nuclei and not visible at all in young embryos until 2-8 h after fertilization (Ingouff et al., 2010), although this GFP-tagged gene does not complement a null mutation (De Storme et al., 2016).

Centromere replication is, however, not a precise process. When maize centromere positions were assayed in sibling plants from the B73 inbred, there was evidence of subtle shifting from side to side along the DNA (Gent et al., 2015, 2017), and we have observed a similar centromere drifting among individuals from the W22 inbred (Figure 1.2A). Such lateral movement suggests that centromeres might stochastically and then heritably shift to nearby positions. However, the data show that centromeres rarely move by chance alone and instead

gravitate back to the location of the population average (Gent et al., 2015, 2017) (Figure 1.2B). In contrast, over evolutionary timescales, or in response to genomic disturbance, centromeres can move to entirely new locations in a process called neocentromere formation (Birchler et al., 2011). Neocentromeres have been observed following interspecies crosses, chromosomal breakage, and centromere deletion (Nasuda et al., 2005; Topp et al., 2009; Fu et al., 2013; Zhang et al., 2013; Liu et al., 2015b). Once a neocentromere has formed, the new location is again quite stable.

Regulation of Centromere Size

Centromere size and position are generally measured by mapping CENH3 footprints using chromatin immunoprecipitation followed by sequencing (ChIP-seq). Such data demonstrate that centromere shapes and sizes are highly variable among different species. Most plant chromosomes have localized ‘‘regional’’ centromeres that encompass several kilobases to megabases of repeat sequences (Liu et al., 2015c). There are also forms known as polycentric centromeres, including metapoly centromeres and holocentromeres, that include multiple centromeric domains distributed over large genomic areas. Metapolycentric centromeres with several CENH3-containing domains per chromosome are found in Pisum and Lathyrus species (Neumann et al., 2015), and holocentromeres that traverse entire chromosomes in discontinuous fashion are found in *Luzula* and *Cuscuta* (Pazy and Plitmann, 1995; Nagaki et al., 2005; Heckmann et al., 2013).

Using electron microscopy to measure subnuclear structures in multiple plant species, Bennett et al. (1981) demonstrated that total centromere volume is directly proportional to nuclear volume and genome size. Zhang and Dawe (2012) confirmed the centromere volume-to-

nuclear volume correlation in 10 grass species using CENH3 immunostaining and noted that the chromosomes within a species have similar centromere sizes. They argued that the size of any given centromere is roughly a function of overall genome size divided by the chromosome number (Zhang and Dawe, 2012). For instance, in two genomes of similar size, the one with fewer chromosomes would have larger centromeres. To confirm this relationship, Wang et al. (2014) used ChIP-seq to analyze centromere sizes in oat-maize additional lines, in which maize chromosomes had been transferred into oat. The oat genome (11300 Mb) is approximately 4-fold larger than the maize genome (2300 Mb) but the chromosome number of oat is only twice that of maize (42 versus 20). Consistent with expectations, seven maize chromosomes transferred into oat proved to have centromeres that were twice as large as their original size in maize (Wang et al., 2014).

Centromere size is perhaps best viewed in the framework of a quantitative architecture where only a small fraction of the cytoplasmic pool is present on centromeres and average centromere size is determined by a mass-action mechanism (Bodor et al., 2014). Under this model, one potential regulator of centromere size might be CENH3 itself. In human, CENH3 overexpression leads to increased incorporation at non-centromeric regions but without the formation of extended or new kinetochores (Van Hooser et al., 2001; Gascoigne et al., 2011; Lacoste et al., 2014). However, in *Drosophila* cultured cells, overexpression of CENH3 induced the formation of ectopic centromeres in noncentromeric regions (Heun et al., 2006). New CENH3 deposition in both human and *Drosophila* occurred over regions that lack centromere repeats (Van Hooser et al., 2001; Gascoigne et al., 2011), similar to stable neocentromeres in plants and other species (Nasuda et al., 2005; Topp et al., 2009; Fu et al., 2013; Zhang et al., 2013; Liu et al., 2015b). CENH3 depletion mechanisms may also be involved in regulating

centromere size. In yeast, mislocalized CENH3 is degraded by ubiquitin-mediated proteolysis, whereas CENH3 within functional centromeres is stable (Ranjitkar et al., 2010; Deyter and Biggins, 2014). The degradation of yeast CENH3 is mediated through the CATD domain (Ranjitkar et al., 2010), raising the possibility that CENP-C or KNL2, by binding to the CATD, serve a stabilizing role. In *Drosophila*, proteasome-mediated degradation of CENH3 can reduce the mislocalized CENH3 and regulate the CENH3 levels in the cell (Moreno-Moreno et al., 2006). In plants, such protease might also prevent mislocation of CENH3 via binding to CATD domain and act as a negative regulator.

A Brief Overview of Haploid Induction in Plants

Haploids, which only have one set of the chromosomes, can produce pure homozygous lines in a single generation by genome doubling. It is much more efficient than the long-term traditional inbreeding process which needs to take 6-7 generations to produce the inbred lines. The spontaneous occurrence of haploids have been reported in many crop species, although they are extremely rare (Dunwell, 2010; Dwivedi et al., 2015). The first discovery of a natural haploid in plants was in *Datura stramonium* (Blakeslee et al., 1922). Later, a variety of protocols have been developed to generate haploids, including anther culture, pollen irradiation, seed selection with twin embryos, sparse pollination, alien cytoplasm (Touraev et al., 2008; Kasha and Maluszynski, 2003; Wang et al., 2019b). Across these approaches, anther culture is most promising. The anther culture and its derived haploid induction protocols have been widely applied in more than 250 plants, although the induction frequency is extremely low, and many protocols are highly genotype dependent (Premvaranon et al., 2011; Maluszynska, 2003). In some species, haploids have been induced via interspecies and intraspecies hybridization.

Although the above haploid methods have proved useful for many years, they are restricted to few crop species or varieties. CENH3-mediated haploid induction has a broad application impact since CENH3 is present in most plant species thus it can be applied to a variety of species. In this introduction, I will focus on the haploid induction systems including intraspecific hybridization, wide species crosses and CENH3-based methods.

Haploid Induction via Intraspecific Hybridization

Double haploid technology has been widely applied in maize. There are two major intraspecific hybridization haploid induction systems: the *indeterminate gametophyte1* (*ig1*) mutant and stock6-derived lines. The *ig1* mutant which was first identified in the inbred line W23 as a spontaneous mutation, can lead to ~3% paternal haploids when *ig1* is crossed as a female parent (Kermicle, 1969). The haploids have the genome of the pollen donor and cytoplasm of the female inducer. Fine mapping studies showed that *IG1* is on chromosome 3 and encodes a LATERAL ORGAN BOUNDARIES (LOB) domain protein (Evans, 2007). The mutation of this gene caused defects in embryo sac development and leaf morphology. Because of the abnormal structure, many defective *ig1* embryo sac lead to abnormal kernels or early seed abortion (Evans, 2007; Kermicle, 1969). However, the exact mechanism of how the *ig1* mutant can induce haploid is unclear, which also restricts the application of IG1-system to other crop species.

Stock6 is the progenitor haploid inducer in maize that can induce maternal haploid at a rate about 2.3-3.2% when crossed as a male (Coe, 1959). This line was later introduced to different genetic backgrounds and the haploid induction ratio (HIR) has been improved to 7-16% (Hu et al., 2016; Prigge et al., 2012). Eight quantitative trait loci on six chromosomes associated

with haploid induction have been identified, suggesting that the haploid induction ability in Stock6-derived lines is controlled by several genes (Prigge et al., 2012; Dong et al., 2013; Hu et al., 2016; Lashermes and Beckert, 1988; Barret et al., 2008). Among them, *qhir1* and *qhir8* are the two major quantitative trait loci on chromosome 1 and 9 explaining 66% and 20% of the genetic variance respectively, indicating that *qhir1* is modulated by other loci which are not able to induce haploid by themselves (Prigge et al., 2012; Liu et al., 2015a). The gene on *qhir1* was identified as *Matrilineal* (*Matl*)/*PHOSPHOLIPASEA1*(*PLA1*)/*NOT LIKE DAD* (*NLD*), which is a patatin-like phospholipase and expressed primarily in pollen (Kelliher et al., 2017; Liu et al., 2017a; Gilles et al., 2017). A 4-bp insertion in the fourth exon of the *MTL*/*PLA1*/*NLD* gene was identified in the inducer lines, which result in a frameshift and a premature transcription termination. As a result, the protein is truncated by 29 amino acids. The *mtl*/*pla1*/*nld* mutant was confirmed to induce haploids *in vivo* and also lead to kernel abortion and segregation distortion (Kelliher et al., 2017; Liu et al., 2017a; Gilles et al., 2017). It's noteworthy that the 4-bp insertion is absent from the maize ancestor teosinte but distinct to the inducer lines, indicating this mutation happened after the maize domestication (Liu et al., 2017b). However, the mechanism that how the C-terminal protein truncation affects the function of MTL/PLA1/NLD and how the mtl/*pla1*/*nld* mutant causes haploid formation are still unclear. The *MATL*/*PLA1*/*NLD* gene was also identified in other monocot species such as rice and the CRISPR-Cas9 technology generated mutant was able to induce haploid at a rate 2-6% (Yao et al., 2018). Later, the gene on locus *qhir8* was identified as *ZmDMP*. When *ZmDMP* knockout was present in the mutant of *mtl*/*pla1*/*nld*, the haploid induction ratio was enhanced and increased by 5-6 fold (Zhong et al., 2019). In *Arabidopsis*, the *ZmDMP-like* genes *AtDMP8* and *AtDMP9* were characterized and the

loss-of-function mutations can induce material haploids at an average HIR 1-3.2% (Zhong et al., 2020).

Haploid Induction via Wide Crosses

Haploid plants can be induced by wide-species crosses. The first case of haploid induction via wide hybridization was reported in barley, when *Hordeum vulgare* was crossed with *Hordeum bulbosum*, the chromosomes of *Hordeum vulgare* were frequently lost and *Hordeum bulbosum* haploids were obtained via embryo rescue (Kasha and Kao, 1970). Haploid formation was also demonstrated when wheat cross with maize, sorghum, barley, teosinte, and pearl millet (Ohkawa et al., 1992; Riera-Lizarazu and Mujeeb-Kazi, 1993; Laurie and O'Donoughue, 1994), and when oat cross with maize (Kynast et al., 2001). The limitation of this method is that it's only applicable to those few crop species.

Centromere size is probably an important factor in the success or failure of interspecies crosses. A case in point is the cross between maize and oat, which can be successful with the aid of embryo rescue, but usually results in haploid oat plants where the entire maize genome has been lost (Riera-Lizarazu et al., 1996). The fact that maize chromosomes survive in oat at low frequencies (Kynast et al., 2001) may reflect the low likelihood that a centromere will expand to the necessary size in the early stages of embryogenesis (or other genomic imbalances). Some maize chromosomes are rarely recovered in oat-maize hybrids, indicating that centromere expansion may regularly fail on those chromosomes. The failure of small centromeres to expand in a larger genome environment may be one of the main reasons for genome elimination in the oat by maize cross (Wang et al., 2014). Genome elimination following wide species crosses to form haploids has also been observed in many other interspecies crosses. Ishii et al. (Ishii et al.,

2016) identify over 100 documented cases. In many cases, the genome sizes of the parents are similar, and centromere size disparity alone is an unlikely explanation. However, many crosses that result in haploids involve wide disparities in predicted centromere sizes, and in these cases, the genome with the smaller centromeres is lost. Using average chromosome size as a proxy for centromere size, we identified 24 examples of genome elimination that may be attributable to centromere size dimorphism. These include crosses of oat, barley, and wheat (large centromere species) by sorghum, maize, pearl millet, adlay millet, and perennial ryegrass (small-centromere species). The tabulated data, along with examples that cannot be attributable to differences in centromere size are well summarized in Wang and Dawe (2018).

A classic case of genome elimination involves the cross between *Hordeum vulgare* and *H. bulbosum* (Kasha and Kao, 1970). The parents of this cross have similar genome sizes and predicted centromere sizes. Nevertheless, cytological analysis demonstrated that chromosome loss in this hybrid is associated with loss of CENH3 from the *H. bulbosum* chromosomes (Sanei et al., 2011). The authors proposed that differences in natural timing of DNA replication between the two species may have an impact on the timing of CENH3 deposition and impair loading on the *H. bulbosum* chromosomes. Other potential explanations relate to the possibility that the entire genome of H. bulbosum may be inactivated at the epigenetic level, including centromeres. These results highlight the fact that there are multiple potential mechanisms for CENH3 loss and genome loss, particularly in wide crosses where general incompatibilities are expected.

Centromere-mediated Haploid Induction

In a landmark paper, Ravi and Chan (Ravi and Chan, 2010) found that a *cenh3*-/- *Arabidopsis* null mutant, when complemented with a modified version of CENH3 called

‘‘tailswap-CENH3’’ can induce haploids at a very high frequency (25%-45%). In such haploids, the genome derived from the tailswap parent is lost, leaving only the genome from the parent that contributed normal CENH3. The tailswap-CENH3 construct involved two changes to CENH3, the replacement of the native CENH3 tail domain with the tail of the H3.3 histone variant and the addition of GFP to the N terminus of the new protein (GFP-H3.3 tail-CENH3). Other forms of tailswap involving CENH3 genes from different species with or without GFP induce haploids at different frequencies (Britt and Kuppu, 2016). This observation has been extended to show that the same effect can be obtained with much smaller changes than the addition of GFP to the N terminus. Point mutations that confer single amino acid changes in the histone fold domain of CENH3 can induce haploids at a lower frequency (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015), as does the replacement of *Arabidopsis* CENH3 with CENH3 genes from related species (Maheshwari et al., 2015).

Additional data from *Arabidopsis*, barley, beet and cucumber suggest that the mechanism underlying centromere-mediated haploid induction may involve the stability of the modified or mutant CENH3 proteins (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2020); (VAN Dun Cornelis et al., 2017). Diploid barley (*Hordeum vulgare*) has two CENH3 variants, αCENH3 and βCENH3. An EMS-induced barley βCENH3mutant L92F carried a substitution in the CATD that caused a loss of visible centromeric βCENH3 loading in interphase cells. The authors then created corresponding mutations in *Arabidopsis* (L130F) and a sugar beet (L106F) and observed strongly impaired, but not abolished, CENH3 loading. When *Arabidopsis* plants carrying the L130F mutation were crossed with wild-type plants, 4.8% of the progeny were haploid. Comparative western blot analysis demonstrated that lines carrying the L130F mutation had less total CENH3 than wild-type plants. In addition, CRISPR/Cas9-mediated in-frame deletions in

CENH3 can also induce haploid in *Arabidopsis* (Kuppu et al., 2020), and a cucumber haploid inducer (HIR 1%) was induced by a frameshift mutation at position 102 of the 154 amino acid protein (VAN Dun Cornelis et al., 2017). The above studies indicated that haploid induction is correlated with the amount of *CENH3* or the defective *CENH3*. Although similar CENH3 stability assays have not been carried out on the original tailswap-CENH3 lines, the tailswap construct is known to cause meiotic abnormalities. Tailswap-CENH3 is virtually absent from the chromosomes of male meiotic cells (Ravi et al., 2011) and only a low percentage of pollen survives. Characterization of centromeric dynamics in tailswap-CENH3 progeny has revealed centromere size dimorphism with the genome from the tailswap parent displaying the smaller centromere signal (M. Marimuthu and L. Comai, U.C. Davis, personal communication). A parsimonious explanation for the remarkable haploid-inducing properties of the tailswap-CENH3 line and by extension, other mutants that cause reduced accuracy of CENH3 deposition (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2020), is that they transmit small or defective centromeres that do not compete with larger fully functional centromeres of the crossing partner.

Centromere Size Model for Haploid Induction

Based on the haploid induction via wide crosses and centromere manipulation, we proposed a centromere size model to explain the relationship between centromere size and haploid induction. When a line with small or defective centromeres is crossed with a line with larger or normal centromeres, the smaller or defective centromeres are preferentially degraded, leading to genome elimination from that line (Figure 1.3).

The degradation of centromere might be determined by the cell-to-centromere ratio. Centromeres adopt a roughly spherical shape in the cell (Bennett et al., 1981), smaller

centromeres should have a higher surface to volume ratio than large centromeres. Further, the gain and loss of centromere factors is most likely to occur at the surface. If changes in centromere size are driven primarily by a net addition of proteins, smaller centromeres will tend to enlarge more quickly over the same time period, whereas if change is driven primarily by degradation (Collins et al., 2004; Deyter et al., 2017), smaller centromeres will shrink more rapidly and have a higher likelihood of being removed. The centromere size model rests on the assumption that during early embryo formation, in a subset of progeny from crosses involving parents with different centromere sizes, the balance between addition and degradation of centromere factors can be weighted toward degradation.

While we lack a comprehensive understanding of the proteins and chaperones present at the plant centromere/kinetochore interface, solid evidence already suggests that KNL2, NASP^{SIM} and CENP-C have critical roles in recruiting and stabilizing CENH3 (Dawe et al., 1999; Lermontova et al., 2013; Le Goff et al., 2020). CENH3 requires KNL2, NASP^{SIM} and CENP-C for accurate recruitment, but KNL2, NASP^{SIM}and CENP-C cannot bind to centromeres without CENH3. Centromeres transmitted from a small-centromere parent are likely to contain less KNL2 and NASP^{SIM}; this will in turn recruit less CENH3, which can bind to less CENP-C. Expansion of a centromere must involve new deposition of one or all of these proteins, perhaps through the interaction with outer kinetochore proteins such as MIS12 and NDC80. As there are mechanisms to remove misplaced CENH3 (and KNL2; (Lermontova et al., 2013)), expansion is presumably an inefficient process. Nevertheless, small centromeres must either rapidly expand in a short time frame to match the average size of the other centromeres or be degraded by failing to compete for limiting factors. Ultimately, degraded centromeres will fail to align on the mitotic spindle, where the chromosome will be lost in the cytoplasm and degraded (Sanei et al., 2011).

Importantly, the simple centromere size model explains the confusing observation that tailswap-CENH3 and other inducer lines in *Arabidopsis* are self-fertile and do not produce haploids (Ravi and Chan, 2010; Kuppu et al., 2015; Maheshwari et al., 2015). Hordeum vulgare, sugar beet, and *Arabidopsis* lines carrying mutations in the CATD reduce or destabilize CENH3 nucleosomes (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015, 2020). In these and other haploid inducer lines, the centromeres are competing on an equal basis for the same (defective) factors. From the same cytoplasmic pool, fewer CENH3 nucleosomes are incorporated, resulting in smaller average centromere sizes. The smaller centromeres in haploid inducer lines are expected to align on the metaphase plate less accurately, and a higher rate of mitotic error would be expected. However, plants are more tolerant of errors than animals; while the components of the spindle checkpoint pathway are present in plants, they function primarily to regulate endoreduplication, meiosis, and flowering time (Komaki and Schnittger, 2016). Further, there is ample evidence from animals that normal centromere size is larger than necessary for spindle attachment and chromosome movement (a single microtubule is more than sufficient to move a large chromosome (Rago and Cheeseman, 2013)). In human cells, centromeres containing as little as 10% of the normal amount of CENP-A are sufficient to assemble functional kinetochores (Rago and Cheeseman, 2013; Liu et al., 2006). Therefore, while haploid inducer lines are expected to maintain a smaller population of functional CENH3-containing nucleosomes and may have higher levels of stochastic chromosome loss, we expect this phenotype to be faithfully transmitted. In a self-cross, both male and female will transmit centromeres of roughly the same size, and uniparental genome elimination is not expected.

Key predictions of the centromere size model are that any mutation that reduces the deposition of CENH3, whether it be on CENH3 itself or its chaperones or binding partners, can

have the effect of reducing centromere size. Such mutations are likely to reduce the total area of centromeres as measured by ChIP-seq. Further, as long as CENH3 deposition mutations support plant growth, they have a high probability of inducing aneuploidy and haploidy when crossed to wild-type lines of the same species or to species with larger average centromere size.

Application of Centromere-mediated Haploid Induction in Crops

Haploids have many potential applications to both breeding and basic research. Haploids can be used to generate homozygous genome-edited germlines. Two studies have reported the utilization of maize haploid inducer lines with CRISPR/Cas9 technology for genome editing (Kelliher et al., 2019; Wang et al., 2019a). In this system, the haploid inducer lines carrying the CRISPR-Cas9 cassette are used to cross with the elite lines, the genome from the inducer line is lost after the formation of haploids, only the edited genome is left. The edited haploids can then be doubled to generate germlines with homozygous mutations. This technique is versatile because they are direct germline editing, without tissue culture and genotype independent (Kelliher et al., 2019; Wang et al., 2019a). As in the centromere-mediated haploid induction system, the maternal cytoplasm is maintained but the maternal nuclear genome is lost, haploids can facilitate the generation of any desired combination of nuclear and cytoplasmic genome. Besides, a lower ploidy level can be induced by crossing the polyploid with a haploid inducer. Due to the fact that sex-specific lethal mutants can be propagated through the unaffected sex, haploid can be used to generate homozygotes for gametophyte lethal mutations (Ravi et al., 2014).

Double haploids can be applied to rapidly construct mapping populations (Seymour et al., 2012), creating parents for reverse breeding (Wijnker et al., 2014) and generating clonal

seeds (Marimuthu et al., 2011). It takes 6-8 generations for the traditional approach to generate homozygous recombinant inbred line populations. However, by creating doubled haploid lines from the heterozygous F1 hybrids, the whole process is shortened to two generations (Seymour et al., 2012). Reverse breeding is a tool that is designed to generate parental homozygous lines for a hybrid individual. The key of this method is to suppress meiotic recombination to ensure the transmission of non-recombinant chromosomes to haploid gametes. By applying centromere-mediated haploid induction, the gametes were converted to haploid plants and subsequently into doubled haploid plants which are completely homozygous individuals. Intercrossing these genetically complementary and homozygous parent lines enable the creation of an elite hybrids and chromosome substitution lines (Wijnker et al., 2012, 2014). The MiMe system is where the meiosis was replaced by mitosis, causing the ploidy level to be increased at each generation. This drawback can be overcame by combining the haploid induction with the MiMe-system and thus enabling apomixis engineering and producing clonal seeds (Marimuthu et al., 2011).

In crop species, an efficient centromere-mediated haploid induction technology is not yet available. Although several species have applied a CENH3-based method to induce haploid, the HIR is relatively low (0.2-1.5%), these works have been well summarized by Kalinowska et al. (2019). In maize, Kelliher et al. (2016) used AcGREEN-tailswap-CENH3 transgenes to complement a *Robertsons Mutator* (*Mu*) induced *cenh3* mutant, and produced a haploid inducer. However, they observed an average of haploid induction rates ~0.86% when the line was crossed as male and no haploids when crossed as a female. In tomato and rice, EMS-induced point mutations in the N-tail of CENH3 were used to induce haploid and the ratio was up to 4% and 1%, relatively (Op Den Camp et al. 2017). In cucumber, a heterozygous CENH3 frameshift mutant was crossed with wild type plants, which induced 1% haploids (Van Dan et al. 2017). In

melon, plants carrying homozygous mutations in the histone fold domain can induce 1.5% of haploid progeny.

The essential functions of CENH3 are highly conserved across eukaryotes. It has been identified and characterized in many species, e.g., rice (Hirsch et al., 2009), tobacco (Nagaki et al., 2009), soybean, *Brassica oleracea* (Wang et al., 2011), potato (Gong et al., 2012), onion and garlic (Nagaki et al., 2012), pea (Neumann et al., 2012), common bean (Iwata et al., 2013), carrot (Dunemann et al., 2014), barley (Sanei et al., 2011), wheat (Yuan et al., 2015) and rye (Evtushenko et al., 2017). But the CENH3-based haploid induction protocol is not well established in most of the above species. Additionally, the CENH3-based haploid inducer can induce haploids as both female and male (Ravi and Chan, 2010). The improvement of centromere-mediated haploid induction technique is of significant importance to the crop breeding.

PURPOSE OF THIS STUDY

Centromeres are the key feature of chromosomes, which controls chromosome segregation during cell division by attaching to spindles (Allshire, 1997; Nicklas, 1997; Farr, 2004). Previous studies have found that centromere size is highly correlated with genome size across different species (Zhang and Dawe, 2012; Wang et al., 2014), although it is not clear whether natural genome size variation within a species corresponds to changes in centromere size, and how plant centromeres respond to the overexpression of CENH3 is unknown. The main purpose of **Chapter 2** is to explore the limiting factors on centromere size, we first examined natural variation on centromere size across 26 maize inbred lines and found that centromere size is positively correlated with genome size and chromosome size, while it has no relationship with

centromere sequence. We then transferred maize centromeres into larger genome background Oaxaca and *Zea luxurians* to detect whether maize centromeres expanded according to the larger genome. Further, we tested the effect of overexpression of CENH3 on centromere size, the result suggests that centromere size is unchanged with CENH3 overexpression in maize.

Haploid plants are important for breeding as they can be simply doubled to generate a true breeding line which has the desired homozygosity in a single generation. There are two major haploid induction systems. One is the Stock6 derived Matrilineal system, which has been applied in maize and other monocots. The other is the centromere mediated haploid induction which only has high haploid induction ratio in *Arabidopsis*. One drawback of the stock6-derived haploid lines is that they can only produce haploids with maternal genome. However, generating a female inducer line is necessary and useful.

We proposed centromere size variation might be the reason for uniparental genome elimination that leads to haploidy (Wang and Dawe, 2018). When a line with small or weak centromeres is crossed to a line with larger or normal centromeres, the small/weak centromeres are selectively degraded or unable to be maintained, leading genome elimination from the small centromere parent. Uniparental genome elimination has been observed following interspecies crosses and with lines carrying modified or mutated CENH3 proteins. However, there is no direct evidence that such centromere size differences underlie the process of chromosome elimination. To test this hypothesis, in **Chapter 3** of the dissertation, the centromere mediated haploid induction is investigated in maize and the role of centromere size may play during chromosome elimination is discussed. Using the CRISPR/Cas9 system, we created a *cenh3* null mutant, which is homozygous lethal but transmissible through both male and female gametophytes. Haploids are formed when *cenh3* heterozygous mutants are crossed with wild

type plants, and the haploid induction ratio is higher when *cenh3* heterozygotes are female (5%) than male (0.5%) in the crosses.

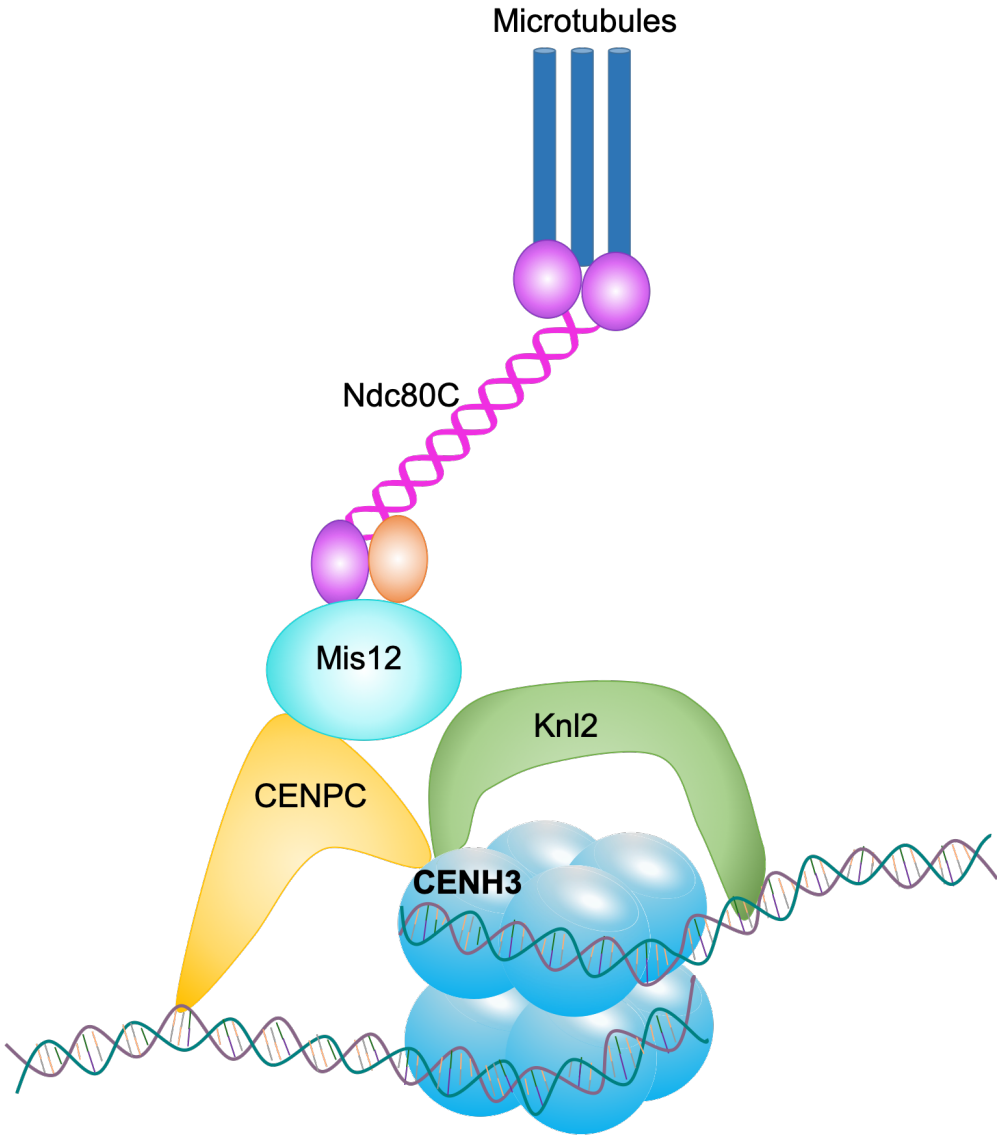


Figure 1.1 Known Kinetochore Complex Proteins and Their Structural Organization in Plants.

CENH3 lies within the nucleosome and binds to CENP-C and KNL2. CENP-C interacts with the MIS12 complex, which interacts with the NDC80 complex and ultimately microtubules.

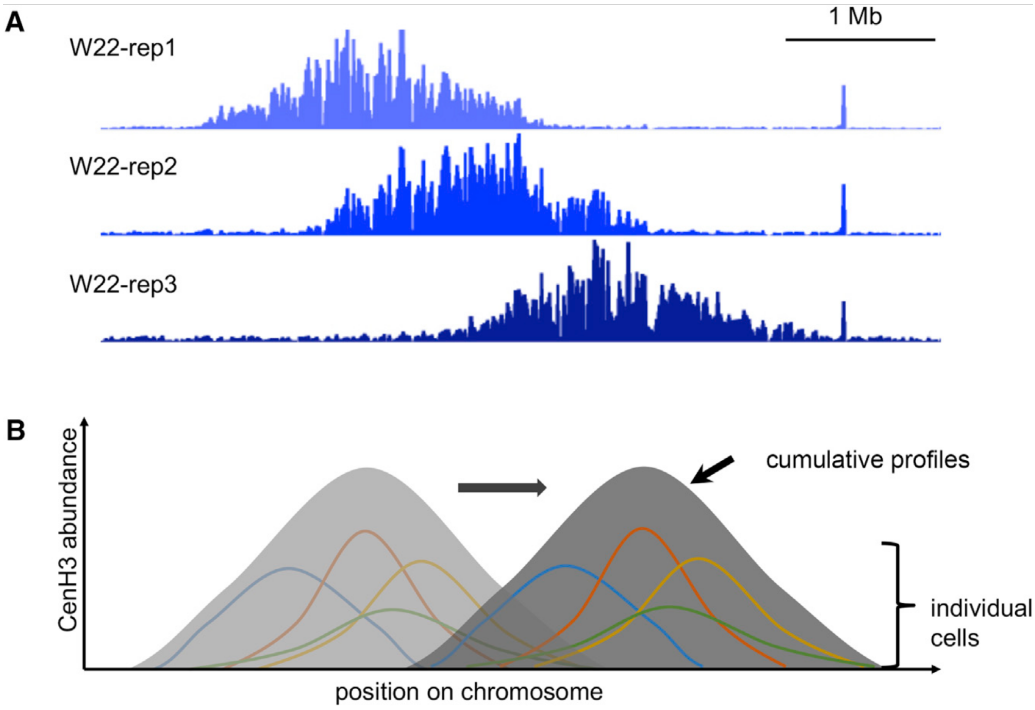


Figure 1.2 Centromere Shift in the Maize W22 Inbred.

(A) CENH3 ChIP-seq was performed on individual seedlings of the W22 inbred and the sequence reads aligned to the W22 reference genome (maizegdb.org). Obvious differences in the position of centromere 5 were observed in three different W22 individuals. These data are not published. (B) A model explaining centromere shift was proposed by Gent et al. (2015). According to this view, centromere positions observed by ChIP-seq are the cumulative result of multiple CENH3 profiles in the collection of cells from an individual plant. The cumulative profiles (black arrows) may shift although the average position in a population remains stable.

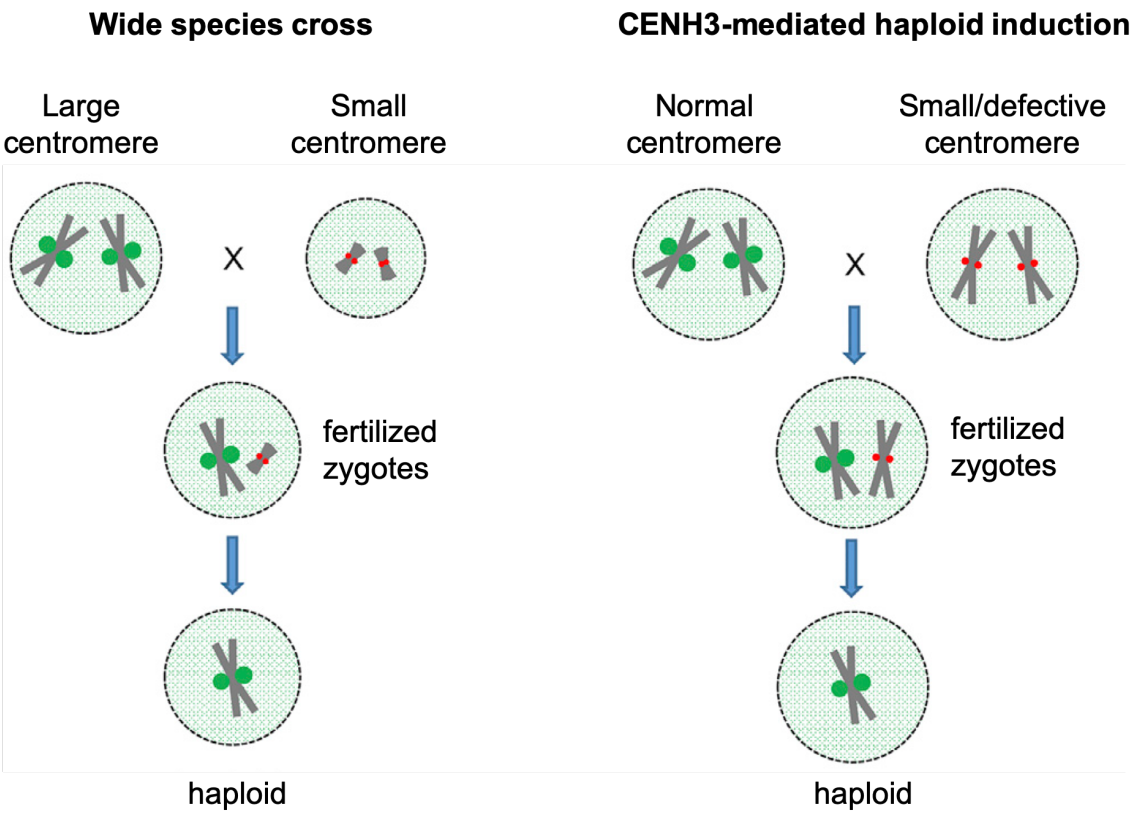


Figure 1.3 Centromere Size Model for Haploid Induction.

As proposed by Zhang and Dawe (2012), centromere/kinetochore size is roughly equal among chromosomes from the same species but depends on the average chromosome size. The hybrid between a large-centromere species and a small-centromere species is expected to show clear centromere dimorphism, and often results in haploids where the small centromere chromosomes are lost (left). Likewise, CENH3-mediated haploid induction involves crosses between wild-type lines and haploid inducer lines with small or defective centromeres, and the chromosomes with smaller centromeres are frequently lost to form haploids (right).

REFERENCES

Allshire, R.C. (1997). Centromeres, checkpoints and chromatid cohesion. Curr. Opin. Genet. Dev. **7**: 264–273.

Allshire, R.C. and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat. Rev. Genet. **9**: 923–937.

Barret, P., Brinkmann, M., and Beckert, M. (2008). A major locus expressed in the male gametophyte with incomplete penetrance is responsible for in situ gynogenesis in maize. Theoretical and Applied Genetics **117**: 581–594.

Batzenschlager, M. et al. (2015). Arabidopsis MZT1 homologs GIP1 and GIP2 are essential for centromere architecture. Proc. Natl. Acad. Sci. U. S. A. **112**: 8656–8660.

Batzenschlager, M., Masoud, K., Janski, N., Houlné, G., Herzog, E., Evrard, J.-L., Baumberger, N., Erhardt, M., Nominé, Y., Kieffer, B., Schmit, A.-C., and Chabouté, M.-E. (2013). The GIP gamma-tubulin complex-associated proteins are involved in nuclear architecture in Arabidopsis thaliana. Front. Plant Sci. **4**: 480.

Bennett, M.D., Smith, J.B., Ward, J., and Jenkins, G. (1981). The relationship between nuclear DNA content and centromere volume in higher plants. J. Cell Sci. **47**: 91–115.

Birchler, J.A., Gao, Z., Sharma, A., Presting, G.G., and Han, F. (2011). Epigenetic aspects of centromere function in plants. Curr. Opin. Plant Biol. **14**: 217–222.

Black, B.E. and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Curr. Opin. Cell Biol. **20**: 91–100.

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

Blakeslee, A.F., Belling, J., Farnham, M.E., and Bergner, A.D. (1922). A HAPLOID MUTANT IN THE JIMSON WEED, "DATURA STRAMONIUM." Science **55**: 646–647.

Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T., Cleveland, D.W., Black, B.E., Shah, J.V., and Jansen, L.E. (2014). The quantitative architecture of centromeric chromatin. Elife **3**: e02137.

Britt, A.B. and Kuppu, S. (2016). Cenh3: An Emerging Player in Haploid Induction Technology. Front. Plant Sci. **7**: 357.

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

Cheeseman, I.M., Chappie, J.S., Wilson-Kubalek, E.M., and Desai, A. (2006). The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. Cell **127**: 983–997.

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

Coe, E.H. (1959). A Line of Maize with High Haploid Frequency. The American Naturalist **93**: 381–382.

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

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

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

De Storme, N., Keçeli, B.N., Zamariola, L., Angenon, G., and Geelen, D. (2016). CENH3-GFP: a visual marker for gametophytic and somatic ploidy determination in Arabidopsis thaliana. BMC Plant Biol. **16**: 1.

Deyter, G.M.R. and Biggins, S. (2014). The FACT complex interacts with the E3 ubiquitin ligase Psh1 to prevent ectopic localization of CENP-A. Genes Dev. **28**: 1815–1826.

Deyter, G.M.R., Hildebrand, E.M., Barber, A.D., and Biggins, S. (2017). Histone H4 Facilitates the Proteolysis of the Budding Yeast CENP-ACse4 Centromeric Histone Variant. Genetics **205**: 113–124.

Dong, X., Xu, X., Miao, J., Li, L., Zhang, D., Mi, X., Liu, C., Tian, X., Melchinger, A.E., and Chen, S. (2013). Fine mapping of qhir1 influencing in vivo haploid induction in maize. Theoretical and Applied Genetics **126**: 1713–1720.

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

Dunemann, F., Schrader, O., Budahn, H., and Houben, A. (2014). Characterization of centromeric histone H3 (CENH3) variants in cultivated and wild carrots (Daucus sp.). PLoS One **9**: e98504.

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

Dunwell, J.M. (2010). Haploids in flowering plants: origins and exploitation. Plant Biotechnol. J. **8**: 377–424.

Du, Y. and Kelly Dawe, R. (2007). Maize NDC80 is a constitutive feature of the central kinetochore. Chromosome Research **15**: 767–775.

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

Dwivedi, S.L., Britt, A.B., Tripathi, L., Sharma, S., Upadhyaya, H.D., and Ortiz, R. (2015). Haploids: Constraints and opportunities in plant breeding. Biotechnol. Adv. **33**: 812–829.

Evans, M.M.S. (2007). The indeterminate gametophyte1 gene of maize encodes a LOB domain protein required for embryo Sac and leaf development. Plant Cell **19**: 46–62.

Evtushenko, E.V., Elisafenko, E.A., Gatzkaya, S.S., Lipikhina, Y.A., Houben, A., and Vershinin, A.V. (2017). Conserved molecular structure of the centromeric histone CENH3 in Secale and its phylogenetic relationships. Sci. Rep. **7**: 17628.

Farr, C.J. (2004). Centromeres, kinetochores and the segregation of chromosomes. Foreword. Chromosome Res. **12**: 517–520.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. Cell **137**: 472–484.

French, B.T., Westhorpe, F.G., Limouse, C., and Straight, A.F. (2017). Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. Dev. Cell **42**: 190–199.e10.

Fulcher, N., Teubenbacher, A., Kerdaffrec, E., Farlow, A., Nordborg, M., and Riha, K. (2015). Genetic architecture of natural variation of telomere length in Arabidopsis thaliana. Genetics **199**: 625–635.

Fu, S., Lv, Z., Gao, Z., Wu, H., Pang, J., Zhang, B., Dong, Q., Guo, X., Wang, X.-J., Birchler, J.A., and Han, F. (2013). De novo centromere formation on a chromosome fragment in maize. Proc. Natl. Acad. Sci. U. S. A. **110**: 6033–6036.

Gao, D. et al. (2009). A lineage-specific centromere retrotransposon in Oryza brachyantha. The Plant Journal **60**: 820–831.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. Cell **145**: 410–422.

Gent, J.I., Wang, K., Jiang, J., and Dawe, R.K. (2015). Stable Patterns of CENH3 Occupancy Through Maize Lineages Containing Genetically Similar Centromeres. Genetics **200**: 1105–1116.

Gent, J.I., Wang, N., and Dawe, R.K. (2017). Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. Genome Biol. **18**: 121.

Gilles, L.M. et al. (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J. **36**: 707–717.

Gong, Z., Wu, Y., Koblížková, A., Torres, G.A., Wang, K., Iovene, M., Neumann, P., Zhang, W., Novák, P., Robin Buell, C., Macas, J., and Jiang, J. (2012). Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. The Plant Cell **24**: 3559–3574.

Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., and Robertson, K.D. (2009). DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum. Mol. Genet. **18**: 3178–3193.

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., Novák, P., Neumann, P., Taudien, S., Platzer, M., and Houben, A. (2013). The holocentric species Luzula elegans shows interplay between centromere and large-scale genome organization. Plant J. **73**: 555–565.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the Drosophila Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. Developmental Cell **10**: 303–315.

Hirsch, C.D., Wu, Y., Yan, H., and Jiang, J. (2009). Lineage-specific adaptive evolution of the centromeric protein CENH3 in diploid and allotetraploid Oryza species. Mol. Biol. Evol. **26**: 2877–2885.

Hori, T., Shang, W.-H., Hara, M., Ariyoshi, M., Arimura, Y., Fujita, R., Kurumizaka, H., and Fukagawa, T. (2017). Association of M18BP1/KNL2 with CENP-A Nucleosome Is Essential for Centromere Formation in Non-mammalian Vertebrates. Dev. Cell **42**: 181–189.e3.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc. Natl. Acad. Sci. U. S. A. **97**: 1148–1153.

Hu, H. et al. (2016). The Genetic Basis of Haploid Induction in Maize Identified with a Novel Genome-Wide Association Method. Genetics **202**: 1267–1276.

Ingouff, M., Rademacher, S., Holec, S., Soljić, L., Xin, N., Readshaw, A., Foo, S.H., Lahouze, B., Sprunck, S., and Berger, F. (2010). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. Curr. Biol. **20**: 2137–2143.

Ishii, T., Karimi-Ashtiyani, R., and Houben, A. (2016). Haploidization via Chromosome Elimination: Means and Mechanisms. Annu. Rev. Plant Biol. **67**: 421–438.

Iwata, A. et al. (2013). Identification and characterization of functional centromeres of the common bean. Plant J. **76**: 47–60.

Izuta, H., Ikeno, M., Suzuki, N., Tomonaga, T., Nozaki, N., Obuse, C., Kisu, Y., Goshima, N., Nomura, F., Nomura, N., and Yoda, K. (2006). Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. Genes Cells **11**: 673–684.

Jansen, L.E.T., Black, B.E., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. Journal of Cell Biology **176**: 795–805.

Jiang, J., Birchler, J.A., Parrott, W.A., and Dawe, R.K. (2003). A molecular view of plant centromeres. Trends Plant Sci. **8**: 570–575.

Kalinowska, K., Chamas, S., Unkel, K., Demidov, D., Lermontova, I., Dresselhaus, T., Kumlehn, J., Dunemann, F., and Houben, A. (2019). State-of-the-art and novel developments of in vivo haploid technologies. Theor. Appl. Genet. **132**: 593–605.

Karanja, L., Kinyua, M., and Maling'a, J. (2002). Development of a doubled haploid system for wheat through wheat x maize crosses in Kenya. African Crop Science Journal **10**.

Karimi-Ashtiyani, R. et al. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proc. Natl. Acad. Sci. U. S. A. **112**: 11211–11216.

Kasha, K.J. and Kao, K.N. (1970). High frequency haploid production in barley (Hordeum vulgare L.). Nature **225**: 874–876.

Kasha, K.J. and Maluszynski, M. (2003). Production of doubled haploids in crop plants. An introduction. Doubled Haploid Production in Crop Plants: 1–4.

Kelliher, T. et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature **542**: 105–109.

Kelliher, T. et al. (2019). One-step genome editing of elite crop germplasm during haploid induction. Nat. Biotechnol. **37**: 287–292.

Kermicle, J.L. (1969). Androgenesis conditioned by a mutation in maize. Science **166**: 1422–1424.

Klare, K., Weir, J.R., Basilico, F., Zimniak, T., Massimiliano, L., Ludwigs, N., Herzog, F., and Musacchio, A. (2015). CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. J. Cell Biol. **210**: 11–22.

Komaki, S. and Schnittger, A. (2016). The spindle checkpoint in plants — a green variation over a conserved theme? Current Opinion in Plant Biology **34**: 84–91.

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., and Britt, A.B. (2020). A variety of changes, including CRISPR/Cas9 mediated deletions, in CENH3 lead to haploid induction on outcrossing. Plant Biotechnol. J.

Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B. (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. PLoS Genet. **11**: e1005494.

Kynast, R.G. et al. (2001). A complete set of maize individual chromosome additions to the oat genome. Plant Physiol. **125**: 1216–1227.

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A.V., Barth, T., Cantaloube, S., Kurumizaka, H., Imhof, A., and Almouzni, G. (2014). Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. Mol. Cell **53**: 631–644.

Lashermes, P. and Beckert, M. (1988). Genetic control of maternal haploidy in maize (Zea mays L.) and selection of haploid inducing lines. Theor. Appl. Genet. **76**: 405–410.

Laurie, D.A. and O'Donoughue, L.S. (1994). Wheat × Maize Crosses for the Production of Wheat Haploids. Maize: 102–118.

Lee, H.-R., Zhang, W., Langdon, T., Jin, W., Yan, H., Cheng, Z., and Jiang, J. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc. Natl. Acad. Sci. U. S. A. **102**: 11793–11798.

Le Goff, S. et al. (2020). The H3 histone chaperone NASP escorts CenH3 in Arabidopsis. Plant J. **101**: 71–86.

Lermontova, I., Fuchs, J., Schubert, V., and Schubert, I. (2007). Loading time of the centromeric histone H3 variant differs between plants and animals. Chromosoma **116**: 507–510.

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

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., Demidov, D., Schubert, V., and Schubert, I. (2013). Arabidopsis kinetochore null2 is an upstream component for centromeric histone H3 variant cenH3 deposition at centromeres. Plant Cell **25**: 3389–3404.

Lermontova, I., Rutten, T., and Schubert, I. (2011b). Deposition, turnover, and release of CENH3 at Arabidopsis centromeres. Chromosoma **120**: 633–640.

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

Liu, C. et al. (2017). A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. Mol. Plant **10**: 520–522.

Liu, C., Li, W., Zhong, Y., Dong, X., Hu, H., Tian, X., Wang, L., Chen, B., Chen, C., Melchinger, A.E., and Chen, S. (2015a). Fine mapping of qhir8 affecting in vivo haploid induction in maize. Theor. Appl. Genet. **128**: 2507–2515.

Liu, S.-T., Rattner, J.B., Jablonski, S.A., and Yen, T.J. (2006). Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. J. Cell Biol. **175**: 41–53.

Liu, Y., Su, H., Pang, J., Gao, Z., Wang, X.-J., Birchler, J.A., and Han, F. (2015b). Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. Proc. Natl. Acad. Sci. U. S. A. **112**: E1263–71.

Liu, Y., Su, H., Zhang, J., Liu, Y., Han, F., and Birchler, J.A. (2015c). Dynamic epigenetic states of maize centromeres. Front. Plant Sci. **6**: 904.

Li, X. and Kelly Dawe, R. (2009). Fused sister kinetochores initiate the reductional division in meiosis I. Nature Cell Biology **11**: 1103–1108.

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J. Cell Biol. **176**: 757–763.

Maehara, K., Takahashi, K., and Saitoh, S. (2010). CENP-A reduction induces a p53-dependent cellular senescence response to protect cells from executing defective mitoses. Mol. Cell. Biol. **30**: 2090–2104.

Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L. (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. **11**: e1004970.

Maluszynska, J. (2003). Cytogenetic tests for ploidy level analyses — chromosome counting. Doubled Haploid Production in Crop Plants: 391–395.

Marimuthu, M.P.A., Jolivet, S., Ravi, M., Pereira, L., Davda, J.N., Cromer, L., Wang, L., Nogué, F., Chan, S.W.L., Siddiqi, I., and Mercier, R. (2011). Synthetic clonal reproduction through seeds. Science **331**: 876.

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

Mizuguchi, G., Xiao, H., Wisniewski, J., Mitchell Smith, M., and Wu, C. (2007). Nonhistone Scm3 and Histones CenH3-H4 Assemble the Core of Centromere-Specific Nucleosomes. Cell **129**: 1153–1164.

Moreno-Moreno, O., Torras-Llort, M., and Azorín, F. (2006). Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres. Nucleic Acids Res. **34**: 6247–6255.

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

Murata, M., Ogura, Y., and Motoyoshi, F. (1994). Centromeric repetitive sequences in Arabidopsis thaliana. Genes & Genetic Systems **69**: 361–370.

Nagaki, K., Kashihara, K., and Murata, M. (2009). A centromeric DNA sequence colocalized

with a centromere-specific histone H3 in tobacco. Chromosoma **118**: 249–257.

Nagaki, K., Kashihara, K., and Murata, M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant Luzula nivea. Plant Cell **17**: 1886–1893.

Nagaki, K., Yamamoto, M., Yamaji, N., Mukai, Y., and Murata, M. (2012). Chromosome dynamics visualized with an anti-centromeric histone H3 antibody in Allium. PLoS One **7**: e51315.

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

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., and Macas, J. (2012). Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS Genet. **8**: e1002777.

Neumann, P., Pavlíková, Z., Koblížková, A., Fuková, I., Jedličková, V., Novák, P., and Macas, J. (2015). Centromeres Off the Hook: Massive Changes in Centromere Size and Structure Following Duplication of CenH3 Gene in Fabeae Species. Mol. Biol. Evol. **32**: 1862–1879.

Nicklas, R.B. (1997). How cells get the right chromosomes. Science **275**: 632–637.

Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (2004). Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. Genes to Cells **9**: 105–120.

Ohkawa, Y., Suenaga, K., and Ogawa, T. (1992). Production of Haploid Wheat Plants through Pollination of Sorghum Pollen. Ikushugaku zasshi **42**: 891–894.

Pazy, B. and Plitmann, U. (1995). Chromosome divergence in the genusCuscutaand its systematic implications. Caryologia **48**: 173–180.

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

Phansalkar, R., Lapierre, P., and Mellone, B.G. (2012). Evolutionary insights into the role of the essential centromere protein CAL1 in Drosophila. Chromosome Res. **20**: 493–504.

Premvaranon, P., Vearasilp, S., Thanapornpoonpong, S.-N., Karladee, D., and Gorinstein, S. (2011). In vitro studies to produce double haploid in Indica hybrid rice. Biologia **66**.

Prigge, V., Xu, X., Li, L., Babu, R., Chen, S., Atlin, G.N., and Melchinger, A.E. (2012). New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. Genetics **190**: 781–793.

Rago, F. and Cheeseman, I.M. (2013). The functions and consequences of force at kinetochores. The Journal of Cell Biology **200**: 557–565.

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

Ravi, M. et al. (2014). A haploid genetics toolbox for Arabidopsis thaliana. Nat. Commun. **5**: 5334.

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

Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M., and Chan, S.W.L. (2011). Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana. PLoS Genet. **7**: e1002121.

Régnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W., and Brown, W. (2005). CENP-A Is Required for Accurate Chromosome Segregation and Sustained Kinetochore Association of BubR1. Molecular and Cellular Biology **25**: 3967–3981.

Riera-Lizarazu, O. and Mujeeb-Kazi, A. (1993). Polyhaploid Production in the Triticeae: Wheat × Tripsacum Crosses. Crop Science **33**: 973–976.

Riera-Lizarazu, O., Rines, H.W., and Phillips, R.L. (1996). Cytological and molecular characterization of oat x maize partial hybrids. Theor. Appl. Genet. **93**: 123–135.

Sandmann, M., Talbert, P., Demidov, D., Kuhlmann, M., Rutten, T., Conrad, U., and Lermontova, I. (2017). Targeting of Arabidopsis KNL2 to Centromeres Depends on the Conserved CENPC-k Motif in Its C Terminus. Plant Cell **29**: 144–155.

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

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

Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr. Biol. **17**: 237–243.

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

Sekulic, N. and Black, B.E. (2012). Molecular underpinnings of centromere identity and maintenance. Trends Biochem. Sci. **37**: 220–229.

Shono, N., Ohzeki, J.-I., Otake, K., Martins, N.M.C., Nagase, T., Kimura, H., Larionov, V., Earnshaw, W.C., and Masumoto, H. (2015). CENP-C and CENP-I are key connecting factors for kinetochore and CENP-A assembly. J. Cell Sci. **128**: 4572–4587.

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

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J. Cell Biol. **127**: 581–592.

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

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L., and Dawe, R.K. (2009). Identification of a Maize Neocentromere in an Oat-Maize Addition Line. Cytogenetic and Genome Research **124**: 228–238.

Touraev, A., Forster, B.P., and Jain, S.M. (2008). Advances in Haploid Production in Higher Plants (Springer Science & Business Media).

Van Dun Cornelis, Clara, L.C.L., and Movahedi, S. (2017). NON-TRANSGENIC HAPLOID INDUCER LINES IN CUCURBITS. World Patent.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J. Cell Sci. **114**: 3529–3542.

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

Wang, B., Zhu, L., Zhao, B., Zhao, Y., Xie, Y., Zheng, Z., Li, Y., Sun, J., and Wang, H. (2019a). Development of a Haploid-Inducer Mediated Genome Editing System for Accelerating Maize Breeding. Mol. Plant **12**: 597–602.

Wang, G., He, Q., Liu, F., Cheng, Z., Talbert, P.B., and Jin, W. (2011). Characterization of CENH3 proteins and centromere-associated DNA sequences in diploid and allotetraploid Brassica species. Chromosoma **120**: 353–365.

Wang, K., Wu, Y., Zhang, W., Dawe, R.K., and Jiang, J. (2014). Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res. **24**: 107–116.

Wang, S., Jin, W., and Wang, K. (2019b). Centromere histone H3- and phospholipase-mediated haploid induction in plants. Plant Methods **15**: 42.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., Poirier, G.G., and Earnshaw, W.C.

(1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr. Biol. **7**: 901–904.

Wijnker, E. et al. (2014). Hybrid recreation by reverse breeding in Arabidopsis thaliana. Nat. Protoc. **9**: 761–772.

Yan, H., Talbert, P.B., Lee, H.-R., Jett, J., Henikoff, S., Chen, F., and Jiang, J. (2008). Intergenic Locations of Rice Centromeric Chromatin. PLoS Biology **6**: e286.

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., and Kelliher, T. (2018). OsMATL mutation induces haploid seed formation in indica rice. Nat Plants **4**: 530–533.

Yuan, J., Guo, X., Hu, J., Lv, Z., and Han, F. (2015). Characterization of two CENH3 genes and their roles in wheat evolution. New Phytol. **206**: 839–851.

Zhang, B., Lv, Z., Pang, J., Liu, Y., Guo, X., Fu, S., Li, J., Dong, Q., Wu, H.-J., Gao, Z., Wang, X.-J., and Han, F. (2013). Formation of a functional maize centromere after loss of centromeric sequences and gain of ectopic sequences. Plant Cell **25**: 1979–1989.

Zhang, H. and Dawe, R.K. (2012). Total centromere size and genome size are strongly correlated in ten grass species. Chromosome Res. **20**: 403–412.

Zhang, H., Koblížková, A., Wang, K., Gong, Z., Oliveira, L., Torres, G.A., Wu, Y., Zhang, W., Novák, P., Robin Buell, C., Macas, J., and Jiang, J. (2014). Boom-Bust Turnovers of Megabase-Sized Centromeric DNA in Solanum Species: Rapid Evolution of DNA Sequences Associated with Centromeres. The Plant Cell **26**: 1436–1447.

Zhang, X., Li, X., Marshall, J.B., Zhong, C.X., and Dawe, R.K. (2005). Phosphoserines on maize CENTROMERIC HISTONE H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. Plant Cell **17**: 572–583.

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

Zhong, Y. et al. (2020). A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous Arabidopsis. Nat Plants **6**: 466–472.

Zhong, Y. et al. (2019). Mutation of ZmDMP enhances haploid induction in maize. Nat Plants **5**: 575–580.

CHAPTER 2

MAIZE CENTROMERIC CHROMATIN SCALES WITH GENOME SIZE BUT NOT EXCLUSIVELY WITH CENH3 LEVELS [1]

[1] Na Wang, Jianing Liu, William A. Ricci, Jonathan I. Gent, and R. Kelly Dawe. Manuscript to be submitted to *Genetics*.

Abstract

Centromeres are defined by the location of Centromeric Histone H3 (CENP-A/CENH3) which interacts with DNA to define the locations and sizes of functional centromeres. An analysis of 26 maize genomes including 113 fully assembled centromeric regions revealed positive relationships between centromere size and chromosome size and genome size. Differences in centromere size are not correlated with variation in the amounts of the major centromeric satellite sequence CentC. To further investigate the relationship between centromere size and genome size, we backcrossed known centromeres into two different lines with larger genomes and observed consistent increases in functional centromere sizes for multiple centromeres. These effects are likely to involve changes in CENH3 levels, however, stably overexpressing CENH3 by threefold did not result in significantly increased centromere size. Literature from other fields demonstrate that changes in genome size affect protein levels, organelle size and cell size. Our data demonstrate that centromere size is among these scalable features, and that multiple limiting factors together contribute to a stable centromere size equilibrium.

Introduction

While kinetochores have a well-defined function, our interpretations of kinetochore structure have undergone many changes since their original descriptions (Schrader, 1936). In the early literature the kinetochore was often described as an organelle (O'Connell et al., 2012) based on the fact that it is an easily distinguishable and consistent feature of the cell. However, when kinetochore proteins were identified and placed in context with centromeres, the kinetochore was reframed as an extension of a specialized chromatin environment based around centromere repeats. From this perspective it was natural to posit that centromeric DNA may

define the location and size of the kinetochore, and this proved to be at least partially true in species where centromere sequence has a strong influence the location of kinetochore proteins (such as some yeasts, human and mouse centromeres) (Roy and Sanyal, 2011; Aldrup-MacDonald et al., 2016; Iwata-Otsubo et al., 2017). However, a more common form of centromere specification is epigenetic, where neither the sequence type or abundance has a strong effect on centromere location (Dawe and Henikoff, 2006; Allshire and Karpen, 2008).

Epigenetic determination is the dominant mechanism in maize, where centromeric repeats and transposons do not correlate in any obvious way to kinetochore size or location (Gent et al., 2017, 2015). The positions of maize centromeres measurably shift from one individual to another while maintaining an average location over many generations (Gent et al., 2017, 2015). The observed plasticity fits well with a quantitative architecture where centromeres are highly dynamic and regulated by the concentration of kinetochore proteins through a mass action mechanism (Bodor et al., 2014). An analysis of multiple grass species demonstrated that the sum of all kinetochore sizes in a cell scales linearly with genome size (Bennett et al., 1981; Zhang and Dawe, 2012), suggesting that the amount of kinetochore proteins is at least partially dependent on genome size and cell volume (Zhang and Dawe, 2012). As a test, maize chromosomes were introduced into an oat genomic background, which has a two-fold larger genome, and centromere size measured by the amount of DNA occupied by CENH3 as interpreted by ChIP-seq. The maize centromeres increased in size by two-fold in the oat background as predicted (Wang et al., 2014). These results mirror a variety of studies showing that subcellular structures frequently scale with cell size (Price et al., 1973; Gregory, 2001; Robinson et al., 2018; Cavalier-Smith, 2005; Gillooly et al., 2015).

As outlined by Marshall (Marshall, 2016), cellular scaling could occur by several mechanisms. The simplest is the limiting precursor model, where the amount of a key component increases with cell size and directly contributes to the size of the structure of interest. In the case of centromeres, a likely candidate is CENH3/CENP-A. Prior data from *Drosophila* and human cell lines have shown that overexpression of CENP-A causes ectopic centromere formation in non-centromeric regions (Heun et al., 2006; Shrestha et al., 2017). Other likely limiting components are those involved in CENP-A deposition. CENP-A loading involves licensing factors such as Kinetochore Null 2 (KNL2) (Lermontova et al., 2013; Sandmann et al., 2017; Boudichevskaia et al., 2019), specific chaperones (Sanchez-Pulido et al., 2009; Chen et al., 2014) and interactions with other kinetochore proteins such as Centromere Protein C (CENP-C) (French et al., 2017; Sandmann et al., 2017). CENP-A, its licensing factors, chaperones, and other inner kinetochore proteins may directly or indirectly regulate centromere size either alone or in combination.

In the current work, we tested the idea that maize centromeres are scalable by analyzing recent genome assemblies of multiple inbreds, experimentally manipulating genome size using genetic crosses, and overexpressing CENH3. We find no consistent association between specific sequences and centromere size, and no change in centromere size after overexpressing *CENH3* by threefold. However, we observe strong evidence of centromere scaling with genome size both among inbreds that naturally vary in genome size and in lines with experimentally manipulated genome sizes. The data support the conclusion that centromere size is not controlled by DNA sequence or by CENH3 alone, but by a mass-action mechanism that is sensitive to cell volume and regulated by the concentration of multiple precursors.

Results

Methodology and reproducibility of ChIP-seq centromere size measurements

Maize centromeres are composed primarily of retrotransposons and arrays of the 156 bp tandem repeat CentC (Wolfgruber et al., 2009). While both components are repetitive, their sequences are degraded to the point that many maize centromeres have been fully assembled, and a surprising number of short reads can align to the assembled centromeres uniquely (Gent et al., 2012, 2015, 2017). For instance, seven B73 centromeres (2, 3, 4, 5, 8, 9 and 10) were assembled gaplessly in the recent B73-Ab10 assembly (Liu et al., 2020). This makes it possible to identify the sequence occupied by centromeric nucleosomes by aligning CENH3 ChIP-seq data from each inbred to the subset of assembled centromeres from that inbred. Functional centromere sizes can be estimated by identifying regions where the depth of ChIP-seq reads exceed an enrichment threshold, and enforcing a minimal peak size and maximal distance between peaks (see Methods).

Like all biochemical purifications, the efficiency of CENH3 chromatin immunoprecipitation varies from day to day and sample to sample. The results are generally assessed after the experiment is over, when the average read depth within centromere cores is compared to the average read depth over chromosome arms. We observed previously (Gent et al., 2015), as well as in the current datasets, that measured centromere sizes vary with ChIP efficiency. This is because CENH3 ChIP-seq profiles take on the shape of bell-shaped curves. When there is low efficiency, a smaller profile of the curve exceeds the 4-fold enrichment cutoff used to define edges of the centromere (Figure 2.1B). To ameliorate this effect, we normalized data across samples by proportionally allocating the ChIP-seq signals from chromosome arms to the CENH3 core, and consequently sharpening the centromere curve for samples with low

efficiency (Figure 2.1). Using these methods we compared centromere sizes across four different IL14H biological replicates. While the ChIP enrichment ranged from 7.36 to 13 (fold increase over background in the 5Mb core region), their measured centromere sizes were similar (Figure 2.1). Throughout this study, we only analyzed centromeres that were fully scaffolded where sequence gaps (if any) were of known size.

No apparent impact of sequence on centromere location or size

To assess natural centromere variation among a variety of maize lines, we measured centromere size in the 26 NAM founder inbreds (McMullen et al., 2009), for which high quality de novo genome assemblies have recently been completed (www.maizegdb.org/NAM_project). CENH3 ChIP-seq data are available for a subset of the NAM inbreds (Schneider et al., 2016), however the data for eight of the lines were either absent or showed poor ChIP efficiencies. We completed ChIP-seq on the NAM lines with missing data and replaced the poor-quality data with new ChIP-seq data with better enrichment. For several inbreds, we performed ChIP-seq in duplicate or triplicate (Table 2.6). Alignment of the ChIP-seq data to the assemblies revealed that of the 260 centromeres present, 113 centromeres were fully scaffolded. Of these, 88 were assembled gaplessly and 25 contained one or more gaps of known size (Figure 2.2A). These new data were used reassess earlier conclusions that DNA sequence has little or no role in CENH3 recruitment in maize (Birchler and Han, 2009; Liu et al., 2020)

Centromeres composed primarily of CentC arrays were rarely fully scaffolded. However, many of the assembled centromeres have mixtures of retroelements and long CentC arrays, making it possible to test whether CentC has an inordinate impact on the size or location of CENH3. This is the case in human centromeres, where the alpha satellite actively recruits CENP-

A and directly contributes to centromere size (Aldrup-MacDonald et al., 2016; Iwata-Otsubo et al., 2017). In maize, however, we found no positive trend between CentC and measured centromere size among the assembled centromeres (Figure 2.2B). If CentC were functioning as a preferred CENH3 substrate, we would expect most of the CentC in the genome to lie within functional centromere regions. However, the assemblies revealed that a large number of CentC arrays lie outside of the functional centromere regions (51.2%) in flanking pericentromeric areas.

Centromeres and the flanking pericentromeric regions also contain many common transposons that are not unique to those areas, but found throughout the genome. The most common are the Gypsy retroelements Flip, Prem, Cinful-Zeon and Gyma (Figure 2.2A). In addition, maize centromeres contain a class of centromere-specific transposons called CRM elements that specifically target active centromeres (Schneider et al., 2016). Flip, Prem, Cinful-Zeon and Gyma can be thought of as neutral centromere substrates that interact with CENH3 when they insert by chance into centromeric regions, while CRM elements actively target centromeres once CENH3 is present (Schneider et al., 2016). When present in centromere cores, each of these transposons have a similar likelihood of interacting directly with CENH3 (Liu et al., 2020) (Figure 2.3). Larger centromeres have larger numbers of all five classes of these transposons (Figure 2.2A). These positive correlations are not because transposons determine centromere size, but because longer centromeres have more transposons.

Positive correlation between centromere size and genome and chromosome sizes in NAM inbreds

The NAM founder inbreds sample a wide genetic and demographic diversity, ranging from tropical locales to northern regions, and include traditional cornbelt varieties as well as

popcorn and sweetcorn (McMullen et al., 2009). The genome sizes among the NAM founder inbreds vary from 2.09 to 2.50 Gb where most of the differences in genome size can be attributed to differences in the amount of tandem repeat arrays within heterochromatic regions known as knobs (Chia et al., 2012).

We plotted the 113 measured centromere sizes against genome size and chromosome size. We observed a clear positive correlation between centromere size and genome size (Figure 2.4A), supporting the prior cross-species comparisons that came to the same conclusion (Zhang and Dawe, 2012). Although we and others have speculated that the centromeres within a species are roughly the same size (Zhang and Dawe, 2012; Moens, 1979), our diverse collection of assembled centromeres revealed that larger chromosomes accommodate larger centromeres (Figure 2.4B), consistent with an earlier report from human cells showing a minor correlation between chromosome size and the number of attached microtubules (McEwen et al., 1998).

Centromeres expand when introduced into larger genomes

Prior data supporting a correlation between centromere size and genome size come from comparisons among inbreds (Figure 2.4) or species (Zhang and Dawe, 2012), and the results from an unnatural wide cross between maize and oat (Wang et al., 2014). We sought to empirically confirm these results using natural crosses between the B73 inbred and two different backgrounds: a maize landrace from Oaxaca Mexico with a genome about 1.3 times the size of the B73 genome, and the intercrossing species *Zea luxurians* with a genome size about 1.6 times the size of B73. The genomes of these two accessions are larger primarily because they contain more heterochromatic knob repeats (Bilinski et al., 2018), although *Zea luxurians* also contains a larger proportion of retroelements (Tenaillon et al., 2011). Figure 2.5 shows the basic crossing

schemes. We first crossed B73 with either Oaxaca or *Zea luxurians* to create F1s, which were self crossed to create F2s and crossed again to the larger-genome parent to obtain BC1 lines. The BC1 lines were then self crossed to create BC1F2 segregating for B73 centromeres. The genome sizes were measured for each cross using flow cytometry (Figure 2.5B). For the Oaxaca crosses, we found that the genomes of F2 progeny were 1.15 times larger than the B73 genome and the BC1F2 progeny 1.2 times larger. For the *Zea luxurians* crosses, the genomes of F2 progeny were 1.31 times larger than the B73 genome and the BC1F2 progeny 1.47 times larger. The seven fully assembled B73 centromeres segregate in these progenies, providing the opportunity to measure changes in CENH3 area as a function of genome size.

B73 X Oaxaca hybrids

While the Oaxaca genome is not sequenced, ChIP-seq revealed that multiple centromeres from Oaxaca are similar to those in B73 (Figure 2.6A). To avoid complexities associated with mapping two centromeres onto a single reference, we developed PCR markers to identify Oaxaca centromeres and focused our analysis entirely on B73 centromeres that were homozygous in the F2 or BC1F2 progeny. Analysis of the data revealed that B73 centromeres 2, 4, 8, 9 and 10 were significantly larger in the Oaxaca background than in their original smaller-genome context, and that the observed increases were strongly correlated with genome size (Figure 2.7, Table 2.1, Table 2.2).

B73 X Zea luxurians hybrids

The centromeres in *Zea luxurians* are known to contain long CentC arrays on every centromere (Albert et al., 2010). ChIP data from *Zea luxurians* consistently yielded high

enrichment for CentC, but when these data were mapped to the B73 reference, there were no clear peaks. This is because the CentC repeat arrays that are present in the B73 assembly do not align well to CentC repeats from *Zea luxurians* (Figure 2.6B). The absence of significant ChIP-seq read alignment from *Zea luxurians* centromeres allowed us to assess B73 centromere size in both the heterozygous and homozygous conditions.

The data reveal that in the first generation F1 hybrids between B73 and *Zea luxurians*, there was little or no change in centromeres size. We observed only minor increases in the size of Cen3 and Cen4 in F1 progeny and no obvious change in Cen2, Cen5, Cen8, Cen9 and Cen10 (Figure 2.8B). In the F2 progeny, however, all seven centromeres showed significant increases in the measured centromere area (Figure 2.8, Table 2.1, Table 2.3). A case in point is centromere 5, which is 1.86 Mb in the B73 inbred but expands to 2.4 Mb in the B73 X *Zea luxurians* F2 progeny. Although there were only three B73 centromeres segregating in the BC1F2 population (Cen5, 9 and 10), these three centromeres were also significantly expanded, confirming the trends observed in the F2 progeny. Taken together, the data from NAM centromere comparisons and Oaxaca and *Zea luxurians* crosses indicate that centromere size is positively correlated with genome size.

Threefold overexpression of CENH3 does not affect centromere size

It is possible that centromere size is determined by the total amount of CENH3 that is available to bind to centromeric DNA. This hypothesis is supported by early work in *Drosophila* showing that overexpression of CENP-A/Cid caused a spreading of centromere locations to ectopic sites (Heun et al., 2006). A recent study of maize lines overexpressing of a YFP-tagged version of CENH3 described subtle shifting of centromere locations, partially supporting this

view, although the fusion protein was not sufficient to complement a strong hypomorphic mutation (Feng et al., 2019).

To test this hypothesis in a purposeful way, we first created a true null mutation of *cenh3* using a CRISPR strategy (Wang et al, submitted). The mutation causes a stop codon in the N-terminal tail and is homozygous inviable. As a means to propagate the null, we also introduced a complete genomic copy of the transgene that differs from wild type by five silent nucleotide changes. RNA-seq on one of the transgenic lines (CENH3-Ox-1) showed approximately four-fold higher expression of the transgene than the wild type copy of *CENH3* (Figure 2.9B). Quantitative PCR analysis of genomic DNA from the CENH3-Ox-1 line indicated that the high *CENH3* expression in this line is caused by multiple transgene insertions, which is a frequent occurrence in *Agrobacterium* transformants (Shou et al., 2004; Jupe et al., 2018). Data from the companion study (Wang et al, submitted) demonstrate that these four transgenes are sufficient to fully complement the *cenh3* null mutation.

Analysis of leaf and root protein revealed that CENH3-Ox-1 lines have approximately three-fold higher nuclear CENH3 levels than wild type lines (Figure 2.9A, 2.9B), providing an excellent resource to test whether altered CENH3 levels change centromere size. The transgenic lines have a mixed genetic background but centromeres 4 and 10 are identical to those in B73. CENH3 ChIP-seq analysis of these two centromeres revealed no significant size differences between CENH3-Ox-1 lines and wild type siblings or the B73 inbred (Figure 2.9C, 2.9D, 2.9E and 2.9F), indicating that centromere size is not affected by a threefold increase in CENH3 protein levels.

Discussion

Here we combine data from a large collection of centromere sequences, a novel CENH3 overexpression line, and empirical manipulations of genome size to test how maize cells determine centromere size. Data from many sources suggest that at least in plants, DNA sequence alone does not control centromere location. Prior analysis of the maize B73 inbred revealed that centromeres differ tremendously in sequence makeup, and that no particular sequence is more likely to bind to CENH3 than any other sequence (Liu et al., 2020). Here, viewed from a total centromere size perspective, we again find that centromere sequence appears to have no meaningful impact on the location or distribution of CENH3. This differs from human and mouse, where sequence composition does influence centromere size (Sullivan et al., 2011; Iwata-Otsubo et al., 2017). However, we did observe some larger trends. Somewhat surprisingly, we observed a correlation between chromosome size and centromere size. This trend was not predicted from prior work, and was not apparent in maize until multiple centromeres from different genomes were compared (Figure 2.4B). Another trend predicted by earlier work (Zhang and Dawe, 2012; Wang et al., 2014) is a positive correlation between total genome size and centromere size (Figure 2.4A).

CENP-A/CENH3 binds directly to DNA and is widely interpreted as a limiting factor for centromere establishment. Early work in *Drosophila* demonstrated that excess CENP-A was inappropriately distributed along chromosome arms, where it was sufficient to recruit all overlying kinetochore proteins and activate spurious centromeres (Heun et al., 2006). Similarly, overexpression of human CENP-A causes a mislocalization of centromeric proteins and is associated with chromosome instability (Van Hooser et al., 2001; Shrestha et al., 2017). However, we find that in maize, overexpression of the native CENH3 protein by threefold has no

discernible effect on the size or distribution of CENH3 as assayed by ChIP (Figure 2.9). These results suggest that other factors limit the incorporation of excess CENH3 in maize, with likely candidates being the histone chaperones that direct CENH3 to centromeric locations (Dunleavy et al., 2009; Mizuguchi et al., 2007; Foltz et al., 2009; Chen et al., 2014). Among these are KNL2, which is required for CENH3 deposition in Arabidopsis (Lermontova et al., 2013) and NASP^{SIM3}, which modulates soluble CENH3 levels (Le Goff et al., 2020). Another potential limiting factor is CENP-C, a key inner centromere protein that has been implicated in multiple aspects of centromere specification and stability (Du et al., 2010; Musacchio and Desai, 2017).

There is a clear positive correlation between genome size and total centromere size in flowering plants (Bennett et al., 1981; Zhang and Dawe, 2012). When maize centromeres were transferred into the oat genome, their sizes increased roughly two-fold, in line with expectations based on the difference in genome size (Wang et al., 2014). However, the wide oat-maize cross rarely succeeds and does not result in a stable hybrid (Kynast et al., 2001). Here we took a different approach of making natural crosses within *Zea* and tracking changes in centromere size over several generations. The results demonstrate that B73 centromeres increase in size when crossed into the larger Oaxaca and *Zea luxurians* backgrounds. Changes were clearly evident in F2 as well as BC1F2 progeny, revealing heritable increases (Figure 2.7 and 2.8). However, the size increase was not apparent in the first generation B73 X *Zea luxurians* individuals, consistent with prior results with the same F1 cross (Gent et al., 2017). Multiple cellular generations or passage through meiosis may be required for CENH3 and associated binding partners and redistribute along chromosomes and reach a new centromere size equilibrium. Taken together, our results indicate that centromere size is scalable-free to expand over flanking sequences with few or no limitations imposed by DNA sequence.

The impact of genome size on centromere size can be explained as a general cellular scaling process. Many types of evidence from multiple species show strong correlations between genome size, nuclear size, and cell size (Price et al., 1973; Gregory, 2001; Robinson et al., 2018; Cavalier-Smith, 2005; Gillooly et al., 2015). With remarkably few exceptions, the entire cellular system scales in response to changes in genome size (Gregory, 2001; Schmoller and Skotheim, 2015; Amodeo and Skotheim, 2016). A study of plant sepal cells revealed that changes in ploidy result in a corresponding increase in total mRNA abundance (Robinson et al., 2018). Studies in yeast demonstrated that increased cell size caused increased rates of transcription while retaining a consistent ratio between RNA and protein (Zhurinsky et al., 2010). The collective data indicate that changes in genome size result in more protein per cell and more and larger macromolecular structures such as mitochondria, microtubules, and ribosomes (Schmoller and Skotheim, 2015). Given that the number of centromeres is constrained by the number of chromosomes, any increases in centromere size will be manifested as extensions of existing centromeres spread over larger chromosomal areas.

The scaling model not only requires scalable centromeres, but a deposition mechanism that is responsive to the amount of soluble precursors. A prior study of CENP-A dynamics in human cells provides support for the view that a mass-action mechanism regulates the number of CENP-A molecules bound to DNA (Bodor et al., 2014). The authors showed that about 4% of total CENP-A binds to centromeres over a range of natural expression variation, implying that centromere sizes vary with the amount of CENP-A available to bind (Bodor et al., 2014). They also overexpressed human CENP-A by approximately ~2.5 fold but did not observe corresponding changes in the amounts of the conserved kinetochore proteins CENP-C or NDC80, suggest that these and/or other key kinetochore proteins are limiting and help to buffer

the effects of CENP-A overexpression. The available information from both human and maize show that while centromere sizes are malleable, moderate overexpression of CENP-A/CENH3 alone does not alter the size of the functional centromere domain, consistent with the view that multiple limiting factors together contribute to a stable centromere size equilibrium.

Materials and Methods

Plant materials and crossing

The plant materials used in this study were obtained from the Germplasm Resources Information Network (GRIN), Ames, Iowa. The lines were B73 (PI 550473), a domesticated landrace from Oaxaca, Mexico (PI 628470) and *Zea luxurians* (PI 462368). Crosses among lines were made over several years in the UGA Plant Biology greenhouses or an adjoining outdoor field site.

NAM genome assemblies

Methodology for the PacBio assembly of NAM genomes is described in (Liu et al., 2020), with the exception that Nanopore data were not used. The descriptions and interpretations of these data are not yet published, but fully assemblies and annotations are freely accessible at www.maizegdb.org/NAM_project.

ChIP-seq

Whole seedlings were collected and CENH3 ChIP conducted following a published native ChIP protocol (Gent et al., 2017). CENH3 antibodies raised against maize CENH3 (Zhong et al., 2002) and rice CENH3 (Nagaki et al., 2004) were used for ChIP experiments. An Illumina

sequencing library was prepared following the protocol provided by Illumina (KAPA hyper prep kit #KK8500). Multiple adapters were used for pooling libraries (KAPA Single-Indexed Adapter Kit KK8700 and NEBNext® Multiplex Oligos for Illumina NEB #E7535S/L). The DNA samples were sequenced using the Illumina NextSeq500 platform and 150-nucleotide single-end reads were generated. ChIP datasets generated as a part of this study are listed in Table 2.6. Sequence Read Archive run IDs for all the ChIP data used in this study are listed in Table 2.7.

Measuring centromere size

We designed a custom workflow to minimize the impact of different CENH3 ChIP efficiencies and genomic background on centromere size analysis. The steps were as follows: 1) Normalize data across samples and relative to input samples. 2) Normalize ChIP efficiency by allocating background signals to centromere cores; 3) Identify mapping gaps and chain ChIP islands; and 4) Compare islands among replicates to remove outliers.

1) *Input data normalization*. PE150 genomic input reads of all NAM lines were subsampled to 30x with seqtk (v1.2, https://github.com/lh3/seqtk) relative to assembly size. The CENH3 ChIP data in the form of PE100 reads were downloaded from SRP067358, converted to single-end data, and subsampled to 5 million reads using seqtk (v1.2). The SE150 ChIP reads generated in this study were subsampled to 3.33 million. Subsampled ChIP and genomic data were subjected to adapter removal with trimglore (v0.4.5, https://github.com/FelixKrueger/TrimGalore/) and mapped to corresponding genomes with bwa-mem (v0.7.17) at default parameters (Li and Durbin, 2009). PCR Duplicates were removed from bam files using piccard (v2.16) and alignments with a mapping quality higher than 20 were extracted with samtools (v1.9) (Li et al., 2009). The resulting CENH3 ChIP bam files were then

normalized against input with deeptools (v3.2.1) (Ramírez et al., 2014) using the RPKM method with 5Kb non-overlapping windows (--binSize 5000 --normalizeUsing RPKM --scaleFactorsMethod None). Regions with an enrichment higher than 5 were extracted and merged into islands with bedtools (v2.28) (Quinlan, 2014).

2) *Normalization of ChIP efficiency*. Centromeres were located manually and placed into 5Mb windows, and all remaining genomic space classified as background. The sum of ChIP RPKM values (>=0) in the 5 Mb centromere regions and background areas were calculated for each chromosome, and ratios between these two values obtained. The core/background ratios were then modified using the formula (1+ core/background) x ChIP RPKM enrichment. This scaling step reapportioned dispersed background ChIP reads to the core, and increased the ChIP enrichment for each sample. The resulting ChIP bedgraph files exhibited more pronounced curves compared with that before scaling (Figure 2.1C).

3) *Chaining ChIP islands separated by mapping gaps*. Due to the high density of CentC and CRM transpons, alignment gaps were frequently observed in ChIP and input files. Short-read alignment gaps were identified by extracting regions (>100bp) with lower than 2 or higher than 101 reads mapped using bedtools (v2.28). Prior to chaining, ChIP island density was calculated with bedtools (v2.28) using a window size of 100Kb and a step size of 50Kb. ChIP islands located in areas with low ChIP islands density (<0.2) were filtered out. Adjacent ChIP islands separated by mapping gaps were then merged using bedtools merge (v2.28; -d 15000). After the chaining step, islands smaller than 15Kb were removed and remaining islands with enrichments higher than 3 were merged with a 50Kb interval using bedtools merge (v2.28; -d 50000).

4) *Using replicates to remove outliers*. For lines with ChIP replicates, final coordinates were determined and centromere sizes were calculated for each replicate. Centromere sizes were compared among replicates, and outliers were removed using the mean absolute deviation (MAD) method. The mean centromere sizes among replicates were calculated after outlier removal.

Genome size measurements of Oaxaca and *Zea luxurians* lines

Genome sizes were estimated by flow cytometry. Young leaf samples from single plants were sent to Plant Cytometry Services (Schijndel, the Netherlands) for flow cytometry measurements using *Vinca major* (2C = 4.2 pg) as an internal standard. We also included the reference maize inbred B73 in each batch as a second internal control to reduce technical error. Genome sizes were calculated as a ratio over the reference B73 (where B73 was assigned a genome size value of 1.0). The genome size measurements for each individual are listed in Table 2.4.

Centromere genotyping

Mapping ChIP-seq reads from Oaxaca to B73 revealed that Oaxaca centromeres 2, 3, 8 and 9 have similar locations as the centromeres in B73 (Figure 2.6). However, the centromere sequences are not identical and can be differentiated by multiple SNPs. Using ChIP-seq reads, we identified SNPs with GATK HaplotypeCaller (v3.8-1) at default parameters (Poplin et al., 2017). SNP2CAPS software (Thiel et al., 2004) was used to design Cleaved Amplified Polymorphic Sequence (CAPS) markers that distinguish B73 centromeres 2, 3, 8, 9 from those in the Oaxaca line (Table 2.5). DNA extractions were carried out on leaf tissue using a CTAB

protocol (Clarke, 2009). PCR reactions were carried out in 25 µl with Promega PCR Master Mix under the following conditions: 95℃, 3 min; 35X [95℃, 30s; annealing Tm, 30s; 72℃, 30s]; 72℃, 5 min. PCR products were then digested with restriction enzymes (Table 2.5) to identify lines homozygous for the B73 centromeres.

B73-Oaxaca, B73-luxurians hybrids and *CENH3* transgenic lines centromere analysis

CENH3 ChIP-seq SE150 data from the Oaxaca-B73, *Zea luxurians*-B73 and *CENH3* overexpression lines described here were subsampled to 3.33 million. Additional ChIP and input samples from the parental Oaxaca and *Zea luxurians* lines were downloaded from SRP105290 (Gent et al., 2017). The subsampled reads were trimmed with TrimGalore (version 0.4.5) and then mapped to Zm-B73-REFERENCE-NAM-5.0 (https://nam-genomes.org/) with BWA-mem (version 0.7.17) at default parameters (Li and Durbin, 2009). Centromere sizes were determined using the same methods used for NAM centromere analysis, except that B73 30X genomic Illumina data were used as input reads for all samples. After the merging steps, small islands less than 100Kb were manually removed. The results were visualized using IGVTools (version 2.3.98) at coverage calculated on 5kb intervals (Thorvaldsdóttir et al., 2013).

Reads were mapped to Zm-B73-REFERENCE-NAM-5.0 (https://nam-genomes.org/) with BWA-mem (version 0.7.17) at default parameters (Li and Durbin, 2009). Only uniquely mapped reads (defined with MAPQ scores of at least 20) and no more than 1 mismatch (1 mismatch on each read) were used for peak calling. To identify B73 homozygous centromeres in the Oaxaca-B73 F2 or BC1F2 progeny, 1 mismatch reads from the Oaxaca-B73 F2 or BC1F2 progeny were used for SNP calling with GATK HaplotypeCaller at default parameters. If no

SNPs were present in the centromeric region, the corresponding centromere was classified as homozygous for the B73 centromere.

Measuring *CENH3* copy number and expression in overexpression lines

Quantitative PCR was used to determine *CenH3* copy number in wild type and overexpression lines. Young leaf DNA was prepared from three biological replicates from wild type and *CENH3*-overexpressed transgenic lines. qPCR was carried out using a BioRad CFX96 Real-Time PCR system using a SYBR Green qPCR kit (Thermo Fisher Scientific). The single copy *Adh1* gene (Osterman and Dennis, 1989) was used as an internal control gene, primers are listed in Table 2.5.

For RNA-seq, mRNA was prepared from young leaves of three wild type and three overexpression lines using a plant total RNA kit (IBI Scientific IB47342). 800 ng of total RNA was used for library construction with a mRNA-seq kit (KAPA mRNA hyper prep kit #KK8580). RNA-seq reads were trimmed with Trimmomatic at the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (Bolger et al., 2014), then the trimmed reads were mapped to Zm-B73-REFERENCE-NAM-5.0 with hisat2 at the following parameters: --min-intronlen 20, --max-intronlen 500000, --rna-strandness R (Kim et al., 2019). The alignments were converted to BAM files and sorted with SAMtools. Stringtie was used to compute the gene expression level (Kim et al., 2019; Pertea et al., 2015). Transcripts Per Kilobase Million (TPM) = 1 was used as the cutoff to determine expression.

Nuclear protein isolation and protein blotting

Approximately 2g of flash-frozen leaves or roots were collected and chopped into 1.5 ml pre-chilled nuclei extraction buffer (1mM EDTA, 1x cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail, 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.2% NP-40, 5 mM 2-mercaptoethanol, 0.1 mM PMSF). The mixture was poured through miracloth and filtered through a 40 µm cell strainer. Then, 30 µl of the filtered sample was stained with 4,6-diamidino-2-phenylindole and nuclei counted using fluorescence microscopy. Nuclei concentrations were normalized based on these measurements. The nuclei were centrifuged at 5000 g for 5 mins and the pellets flash-frozen and stored at -80°C until used for protein blots. Nuclei were resuspended in Laemmli buffer and loaded into 4-20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Cat #4561093). SDS-PAGE and protein blotting were performed according to (Dawe et al., 2018). CENH3 was detected with anti-CENH3 antibody (Zhong et al., 2002) (1:1000 dilution) and normalized to total H4 histones revealed by an anti-H4 antibody (1:1000 dilution, Abcam, ab7311). Primary antibodies were detected using anti-rabbit secondary antibodies (1:5000 dilution, Anti-Rabbit IgG HRP Linked Whole Ab Sigma Cat# GENA934-1ML). The band intensities were quantified with Image J (Schneider et al., 2012).

Statistical analysis

A single factor analysis of variance (ANOVA) tests were performed to determine whether genome size and centromere size were significantly different across a variety of subgroups. Tukey's HSD test was subsequently used for pairwise comparisons among different subgroups with the available R package (http://www.r-project.org/). Significance was set at $P < 0.05$. A linear relationship between genome size and centromere size is fitted with linear regression (Rosner, 2015).

Data Availability

ChIP reads can be obtained from the NCBI Sequence Read Archive (ncbi.nlm.nih.gov/sra) under project PRJNA639705.

Table 2.1 Centromere sizes in six B73 replicates. The unit of centromere size is bp.

ID	centromere2	centromere3	centromere4	centromere5	centromere8	centromere9	centromere10
B73_1	2070767	1976279	1866041	1785000	1840000	1591001	1642503
B73_2	2085767	1940287	1924125	1843429	1765418	1625440	1807706
B73_3	2145000	1922637	1898658	1890000	1701113	1655000	1737649
B73_4	2145000	1993243	1780569	1905283	1895000	1610000	1883320
B73_5	2011087	1935287	1918658	1730464	1779595	1455000	1715152
B73_6	2104227	1940287	1895433	1985000	1784798	1627141	1715152

Table 2.2 The genome and centromere sizes in Oaxaca X B73 hybrids. Genome sizes are averages of all individuals measured from the indicated group, expressed as a ratio over the B73 reference genome. The unit of centromere size is bp.

ID	genome size	centromere2	centromere3	centromere4	centromere8	centromere9	centromere10
OaxB73_F2_1	1.15	2345000	NA	NA	NA	NA	NA
OaxB73_F2_2	1.15	2420000	2219160	NA	1998403	NA	NA
OaxB73_F2_3	1.15	NA	NA	2320000	NA	1870000	NA
OaxB73_F2_4	1.15	2415000	NA	NA	NA	NA	1939465
OaxB73_F2_5	1.15	NA	NA	NA	NA	1936000	NA
OaxB73_F2_6	1.15	2491357	NA	NA	NA	NA	NA
OaxB73_F2_7	1.15	NA	NA	2299125	NA	NA	NA
OaxB73_F2_8	1.15	NA	NA	NA	2047653	1776712	NA
OaxB73_F2_9	1.15	2240000	NA	2393658	2140354	1600000	1910817
OaxB73_BC1F2_1	1.20	2732740	2578497	2159125	NA	NA	2015000
OaxB73_BC1F2_2	1.20	2299486	NA	NA	2456832	NA	NA
OaxB73_BC1F2_4	1.20	NA	NA	NA	NA	NA	NA
OaxB73_BC1F2_6	1.20	NA	2395000	NA	2460000	NA	NA
OaxB73_BC1F2_8	1.20	NA	2500000	2664999	NA	NA	NA
OaxB73_BC1F2_9	1.20	2625000	2379160	2520000	NA	NA	NA
OaxB73_BC1F2_10	1.20	NA	NA	NA	NA	NA	NA
OaxB73_BC1F2_11	1.20	NA	2310000	2315000	2186832	1961778	NA
OaxB73_BC1F2_12	1.20	NA	NA	2650000	2160821	1781712	2088648
OaxB73_BC1F2_13	1.20	NA	NA	NA	2456832	NA	NA
OaxB73_BC1F2_15	1.20	NA	2225287	NA	2441068	NA	2073648
OaxB73_BC1F2_16	1.20	NA	2375000	NA	2900780	1986778	NA
OaxB73_BC1F2_17	1.20	2520000	NA	2320000	NA	2036778	2038648
OaxB73_BC1F2_18	1.20	2391577	NA	NA	NA	NA	NA
OaxB73_BC1F2_19	1.20	2538991	NA	NA	NA	NA	NA

Table 2.3 The genome and centromere sizes in B73 X luxurians hybrids. Genome sizes are averages of all individuals measured from the indicated group, expressed as a ratio over the B73 reference genome. The unit of centromere size is bp.

ID	genome size	centromere 2	centromere3	centromere4	centromere5	centromere8	centromere9	centromere10
B73lux_F1_1	1.31	2445000	2100638	2200000	2265464	1965000	1765000	1752649
B73lux_F1_2	1.31	2230000	2140287	2148658	2200505	1955000	1655000	1752649
B73lux_F1_3	1.31	2150000	2214447	2154125	2315000	2020000	1795000	1905817
B73lux_F2_1	1.31	2765000	2114450	NA	2303903	2445000	1776712	2275000
B73lux_F2_2	1.31	2160482	2385000	NA	2457934	2175000	1760000	NA
B73lux_F2_3	1.31	2426577	2520000	NA	2488026	1936832	1800000	1920817
B73lux_F2_4	1.31	2620000	2180287	2240992	NA	2100000	2036778	NA
B73lux_F2_5	1.31	2633991	2245000	NA	NA	2302376	2200000	2407741
B73lux_F2_6	1.31	2381577	2379160	2269125	2313940	NA	NA	2008246
B73lux_BC1F2_1	1.47	NA	NA	NA	3280000	NA	2296778	NA
B73lux_BC1F2_2	1.47	NA	NA	NA	3125000	NA	2301778	2347913
B73lux_BC1F2_6	1.47	NA	NA	NA	NA	NA	2055000	2435700
B73lux_BC1F2_11	1.47	NA	NA	NA	3105000	NA	2021778	2295000
B73lux_BC1F2_13	1.47	NA	NA	NA	2275000	NA	NA	2078648
B73lux_BC1F2_15	1.47	NA	NA	NA	2937934	NA	2186778	2132913
B73lux_BC1F2_16	1.47	NA	NA	NA	2597302	NA	NA	2005000
B73lux_BC1F2_17	1.47	NA	NA	NA	3309471	NA	2035000	2435000
B73lux_BC1F2_19	1.47	NA	NA	NA	3220000	NA	2065000	2095000

Table 2.4 Genome size measurement in this study. Data are expressed as ratios over the B73 reference genome.

ID	DNA content pg/2C	Genome size	mean of genome size	STDEV
B73_1	5.25	1.00	1.00	0.00
B73_2	5.21	0.99		
B73_3	5.25	1.00		
B73_4	5.17	0.99		
B73_5	5.33	1.02		
B73_6	5.17	0.99		
Oaxaca_1	6.72	1.28	1.30	0.02
Oaxaca_2	6.89	1.32		
Oaxaca_3	6.89	1.32		
OaxB73_F2_1	5.67	1.08	1.15	0.05
OaxB73_F2_2	6.34	1.21		
OaxB73_F2_3	6.13	1.17		
OaxB73_F2_4	5.92	1.13		
OaxB73_F2_5	6.38	1.22		
OaxB73_F2_6	5.96	1.14		
OaxB73_F2_7	5.59	1.06		
OaxB73_F2_8	5.96	1.14		
OaxB73_F2_9	6.17	1.18		
OaxB73_F2_10	6.17	1.18		
OaxB73_F2_11	6.01	1.14		
OaxB73_BC1F2_1	6.30	1.20	1.20	0.01
OaxB73_BC1F2_2	6.17	1.18		
OaxB73_BC1F2_3	6.22	1.18		
OaxB73_BC1F2_4	6.34	1.21		
OaxB73_BC1F2_5	6.34	1.21		
OaxB73_BC1F2_6	6.26	1.19		
OaxB73_BC1F2_7	6.22	1.18		
OaxB73_BC1F2_8	6.30	1.20		
OaxB73_BC1F2_9	6.34	1.21		
OaxB73_BC1F2_10	6.26	1.20		
luxurians_1	8.44	1.61	1.61	0.02
luxurians_2	8.36	1.60		
luxurians_3	8.53	1.63		
B73lux_F1_1	6.85	1.30	1.31	0.01
B73lux_F1_2	6.82	1.30		
B73lux_F1_3	6.90	1.31		
B73lux_F2_1	7.01	1.34	1.31	0.04
B73lux_F2_2	7.18	1.37		
B73lux_F2_3	6.97	1.33		
B73lux_F2_4	7.14	1.36		
B73lux_F2_5	6.64	1.26		
B73lux_F2_6	7.01	1.34		
B73lux_F2_7	6.89	1.31		
B73lux_F2_8	6.85	1.30		
B73lux_F2_9	6.64	1.26		
B73lux_F2_10	6.68	1.27		
B73lux_BC1F2_1	7.83	1.49	1.47	0.03
B73lux_BC1F2_2	8.03	1.53		
B73lux_BC1F2_3	7.73	1.47		
B73lux_BC1F2_4	7.78	1.48		
B73lux_BC1F2_5	7.58	1.44		
B73lux_BC1F2_6	7.54	1.44		
B73lux_BC1F2_7	7.41	1.41		
B73lux_BC1F2_8	7.78	1.48		
B73lux_BC1F2_9	7.73	1.47		

Table 2.5 Primers used in this study.

Primer name	Primer sequence	Enzyme	Purpose
Chr2_1F	TCAAGAGGAGCCACACAT	AciI	distinguish SNPs on Chr2
Chr2_1R	CTGTTGAACTTGTGTTCACTTG		distinguish SNPs on Chr2
Chr2_2F	GGAGCCGTCTGAAGAGA	DraI	distinguish SNPs on Chr2
Chr2_2R	CTGCCAAACGAGCTCTAAA		distinguish SNPs on Chr2
Chr3_1F	GGTCATCTTCCGCCAAAT	BccI	distinguish SNPs on Chr3
Chr3_1R	CCTAGGGAACCACATCAAAG		distinguish SNPs on Chr3
Chr3_2F	CAAGATCGTGACAGCCATAG	BsmAI	distinguish SNPs on Chr3
Chr3_2R	ACAGTGAAGGTGAGGTCT		distinguish SNPs on Chr3
Chr8_1F	AGGCTATGAGTTGGCATTG	BsrI	distinguish SNPs on Chr8
Chr8_1R	GAGACATCCTTACCAAACTAGCA		distinguish SNPs on Chr8
Chr8_2F	AGCTCTAGTCTCTTCTCCTTC	BccI	distinguish SNPs on Chr8
Chr8_2R	GGTGGTTCACCTTCTTCACA		distinguish SNPs on Chr8
Chr9_1F	CAGATCCAAGCGAATGGTG	HpyCH4III	distinguish SNPs on Chr9
Chr9_1R	AGAGCTGAAATGGGAGTGG		distinguish SNPs on Chr9
CENH3-F1	GGCTGCTCTTACTTGCTTGC	NA	Genotyping on *ImmuneCENH3*
CENH3-R1	TACTTCCTGATCTCGCGCAGGGCG	NA	Genotyping on *ImmuneCENH3*
CENH3-F2	GAAGTCCACTGAACCGCTCAT	NA	Copy number analysis on *ImmuneCENH3*
CENH3-R2	CGCTCTACTTTCCCGTTTGTTAC	NA	Copy number analysis on *ImmuneCENH3*
Adh1-F	GGCTGGAGGTATGTTCTATTC	NA	qRT-PCR on *Adh1*
Adh1-R	GACCACCAACCATACCCATAAA	NA	qRT-PCR on *Adh1*

Table 2.6 ChIP datasets generated in this study.

Sample_name	Species/subspecies	Cultivar	Tissue	Genome size (GB)	# of replicate
B73	Zea mays mays	B73	seedlings	2.30	1
CML103	Zea mays mays	CML103	seedlings	2.31	2
CML277	Zea mays mays	CML277	seedlings	2.34	2
CML333	Zea mays mays	CML333	seedlings	2.40	3
HP301	Zea mays mays	HP301	seedlings	2.21	3
IL14H	Zea mays mays	IL14H	seedlings	2.10	3
Ki11	Zea mays mays	Ki11	seedlings	2.38	3
Ki3	Zea mays mays	Ki3	seedlings	2.35	3
NC350	Zea mays mays	NC350	seedlings	2.40	3
Oh7b	Zea mays mays	Oh7b	seedlings	2.20	2
P39	Zea mays mays	P39	seedlings	2.10	3
Tzi8	Zea mays mays	Tzi8	seedlings	2.35	3
OaxB73_F2	Zea mays mays	B73 X PI 628470	seedlings	1.15	9
OaxB73_BC1F2	Zea mays mays	B73 X PI 628470	seedlings	1.20	15
B73lux_F1	Zea mays mays x Zea luxurians	B73 x PI 422162	seedlings	1.31	3
B73lux_F2	Zea mays mays x Zea luxurians	B73 x PI 422162	seedlings	1.31	6
B73lux_BC1F2	Zea mays mays x Zea luxurians	B73 x PI 422162	seedlings	1.47	9
trangenic_WT	Zea mays mays	mixed	seedlings	NA	3
trangenic_OX	Zea mays mays	mixed	seedlings	NA	3

Table 2.7 Sequence Read Archive run IDs for all the ChIP data used in this study.

Sample_name	Species/ subspecies	Cultivar	Accession	ChIP SRA run	input SRA run	Library layout	Tissue	Reference
CML103_1	Zea mays	CML103	PRJNA639705	SRR12023802	NA	paired	seedling	this study
CML103_2	Zea mays	CML103	PRJNA639705	SRR12023791	NA	paired	seedling	this study
CML277_1	Zea mays	CML277	PRJNA639705	SRR12023780	NA	paired	seedling	this study
CML277_2	Zea mays	CML277	PRJNA639705	SRR12023769	NA	paired	seedling	this study
CML333_1	Zea mays	CML333	PRJNA639705	SRR12023758	NA	single	seedling	this study
CML333_2	Zea mays	CML333	PRJNA639705	SRR12023747	NA	single	seedling	this study
CML333_3	Zea mays	CML333	PRJNA639705	SRR12023736	NA	single	seedling	this study
HP301_1	Zea mays	HP301	PRJNA639705	SRR12023730	NA	paired	seedling	this study
HP301_2	Zea mays	HP301	PRJNA639705	SRR12023729	NA	paired	seedling	this study
IL14H_1	Zea mays	IL14H	PRJNA639705	SRR12023801	NA	paired	seedling	this study
IL14H_2	Zea mays	IL14H	PRJNA639705	SRR12023800	NA	paired	seedling	this study
IL14H_3	Zea mays	IL14H	PRJNA639705	SRR12023799	NA	paired	seedling	this study
Ki11_1	Zea mays	Ki11	PRJNA639705	SRR12023798	NA	single	seedling	this study
Ki11_2	Zea mays	Ki11	PRJNA639705	SRR12023797	NA	single	seedling	this study
Ki11_3	Zea mays	Ki11	PRJNA639705	SRR12023796	NA	single	seedling	this study
Ki3_1	Zea mays	Ki3	PRJNA639705	SRR12023795	NA	paired	seedling	this study
Ki3_2	Zea mays	Ki3	PRJNA639705	SRR12023794	NA	paired	seedling	this study
NC350_1	Zea mays	NC350	PRJNA639705	SRR12023793	NA	paired	seedling	this study
NC350_2	Zea mays	NC350	PRJNA639705	SRR12023792	NA	paired	seedling	this study
Oh7b_1	Zea mays	Oh7b	PRJNA639705	SRR12023790	NA	paired	seedling	this study
Oh7b_2	Zea mays	Oh7b	PRJNA639705	SRR12023789	NA	paired	seedling	this study
P39_1	Zea mays	P39	PRJNA639705	SRR12023788	NA	single	seedling	this study
P39_2	Zea mays	P39	PRJNA639705	SRR12023787	NA	single	seedling	this study
P39_3	Zea mays	P39	PRJNA639705	SRR12023786	NA	single	seedling	this study
Tzi8_1	Zea mays	Tzi8	PRJNA639705	SRR12023785	NA	paired	seedling	this study
Tzi8_2	Zea mays	Tzi8	PRJNA639705	SRR12023784	NA	paired	seedling	this study
B73_4	Zea mays	B73	PRJNA639705	SRR12023803	NA	single	seedling	this study
OaxB73_F2_1	Zea mays	OaxB73_F2	PRJNA639705	SRR12023783	NA	single	seedling	this study
OaxB73_F2_2	Zea mays	OaxB73_F2	PRJNA639705	SRR12023782	NA	single	seedling	this study
OaxB73_F2_3	Zea mays	OaxB73_F2	PRJNA639705	SRR12023781	NA	single	seedling	this study
OaxB73_F2_4	Zea mays	OaxB73_F2	PRJNA639705	SRR12023779	NA	single	seedling	this study
OaxB73_F2_5	Zea mays	OaxB73_F2	PRJNA639705	SRR12023778	NA	single	seedling	this study
OaxB73_F2_6	Zea mays	OaxB73_F2	PRJNA639705	SRR12023777	NA	single	seedling	this study
OaxB73_F2_7	Zea mays	OaxB73_F2	PRJNA639705	SRR12023776	NA	single	seedling	this study
OaxB73_F2_8	Zea mays	OaxB73_F2	PRJNA639705	SRR12023775	NA	single	seedling	this study
OaxB73_F2_9	Zea mays	OaxB73_F2	PRJNA639705	SRR12023774	NA	single	seedling	this study

OaxB73_BC1F2_1	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023773	NA	single	seedling	this study
OaxB73_BC1F2_2	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023772	NA	single	seedling	this study
OaxB73_BC1F2_4	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023771	NA	single	seedling	this study
OaxB73_BC1F2_6	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023770	NA	single	seedling	this study
OaxB73_BC1F2_8	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023768	NA	single	seedling	this study
OaxB73_BC1F2_9	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023767	NA	single	seedling	this study
OaxB73_BC1F2_10	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023766	NA	single	seedling	this study
OaxB73_BC1F2_11	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023765	NA	single	seedling	this study
OaxB73_BC1F2_12	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023764	NA	single	seedling	this study
OaxB73_BC1F2_13	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023763	NA	single	seedling	this study
OaxB73_BC1F2_15	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023762	NA	single	seedling	this study
OaxB73_BC1F2_16	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023761	NA	single	seedling	this study
OaxB73_BC1F2_17	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023760	NA	single	seedling	this study
OaxB73_BC1F2_18	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023759	NA	single	seedling	this study
OaxB73_BC1F2_19	Zea mays	OaxB73_BC1F2	PRJNA639705	SRR12023757	NA	single	seedling	this study
B73lux_F1_1	Zea mays subsp. mays x Zea luxurians	B73lux_F1	PRJNA639705	SRR12023756	NA	single	seedling	this study
B73lux_F1_2	Zea mays subsp. mays x Zea luxurians	B73lux_F1	PRJNA639705	SRR12023755	NA	single	seedling	this study
B73lux_F1_3	Zea mays subsp. mays x Zea luxurians	B73lux_F1	PRJNA639705	SRR12023754	NA	single	seedling	this study
B73lux_F2_1	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023753	NA	single	seedling	this study
B73lux_F2_2	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023752	NA	single	seedling	this study
B73lux_F2_3	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023751	NA	single	seedling	this study
B73lux_F2_4	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023750	NA	single	seedling	this study
B73lux_F2_5	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023749	NA	single	seedling	this study
B73lux_F2_6	Zea mays subsp. mays x Zea luxurians	B73lux_F2	PRJNA639705	SRR12023748	NA	single	seedling	this study
B73lux_BC1F2_1	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023746	NA	single	seedling	this study
B73lux_BC1F2_2	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023745	NA	single	seedling	this study
B73lux_BC1F2_6	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023744	NA	single	seedling	this study
B73lux_BC1F2_11	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023743	NA	single	seedling	this study
B73lux_BC1F2_13	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023742	NA	single	seedling	this study
B73lux_BC1F2_15	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023741	NA	single	seedling	this study
B73lux_BC1F2_16	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023740	NA	single	seedling	this study
B73lux_BC1F2_17	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023739	NA	single	seedling	this study
B73lux_BC1F2_19	Zea mays subsp. mays x Zea luxurians	B73lux_BC1F2	PRJNA639705	SRR12023738	NA	single	seedling	this study
trangenic_WT_1	Zea mays	mixed	PRJNA639705	SRR12023737	NA	single	seedling	this study
trangenic_WT_2	Zea mays	mixed	PRJNA639705	SRR12023735	NA	single	seedling	this study
trangenic_WT_3	Zea mays	mixed	PRJNA639705	SRR12023734	NA	single	seedling	this study
trangenic_Ox_1	Zea mays	mixed	PRJNA639705	SRR12023733	NA	single	seedling	this study
trangenic_Ox_2	Zea mays	mixed	PRJNA639705	SRR12023732	NA	single	seedling	this study
trangenic_Ox_3	Zea mays	mixed	PRJNA639705	SRR12023731	NA	single	seedling	this study

MS71	Zea mays	MS71	PRJNA305893	SRR2994641	NA	paired	leaves	Schneider et al., 2016
Mo18W	Zea mays	Mo18W	PRJNA305893	SRR3018349	NA	paired	leaves	Schneider et al., 2016
B97	Zea mays	B97	PRJNA305893	SRR3018373	NA	paired	leaves	Schneider et al., 2016
CML333	Zea mays	CML333	PRJNA305893	SRR3018392	NA	paired	leaves	Schneider et al., 2016
P39	Zea mays	P39	PRJNA305893	SRR3018404	NA	paired	leaves	Schneider et al., 2016
Il14H	Zea mays	Il14H	PRJNA305893	SRR3018410	NA	paired	leaves	Schneider et al., 2016
KI11	Zea mays	KI11	PRJNA305893	SRR3018575	NA	paired	leaves	Schneider et al., 2016
Ky21	Zea mays	Ky21	PRJNA305893	SRR3018597	NA	paired	leaves	Schneider et al., 2016
Tx303	Zea mays	Tx303	PRJNA305893	SRR3018625	NA	paired	leaves	Schneider et al., 2016
NC358	Zea mays	NC358	PRJNA305893	SRR3018741	NA	paired	leaves	Schneider et al., 2016
KI3	Zea mays	KI3	PRJNA305893	SRR3018808	NA	paired	leaves	Schneider et al., 2016
Hp301	Zea mays	Hp301	PRJNA305893	SRR3018811	NA	paired	leaves	Schneider et al., 2016
CML69	Zea mays	CML69	PRJNA305893	SRR3018813	NA	paired	leaves	Schneider et al., 2016
CML247	Zea mays	CML247	PRJNA305893	SRR3018814	NA	paired	leaves	Schneider et al., 2016
TZI8	Zea mays	TZI8	PRJNA305893	SRR3018816	NA	paired	leaves	Schneider et al., 2016
M162W	Zea mays	M162W	PRJNA305893	SRR3018819	NA	paired	leaves	Schneider et al., 2016
CML103	Zea mays	CML103	PRJNA305893	SRR3018820	NA	paired	leaves	Schneider et al., 2016
OH7B	Zea mays	OH7B	PRJNA305893	SRR3018821	NA	paired	leaves	Schneider et al., 2016
OH43	Zea mays	OH43	PRJNA305893	SRR3018822	NA	paired	leaves	Schneider et al., 2016
CML322	Zea mays	CML322	PRJNA305893	SRR3018825	NA	paired	leaves	Schneider et al., 2016
CML52	Zea mays	CML52	PRJNA305893	SRR3018826	NA	paired	leaves	Schneider et al., 2016
CML228	Zea mays	CML228	PRJNA305893	SRR3018827	NA	paired	leaves	Schneider et al., 2016
M37W	Zea mays	M37W	PRJNA305893	SRR3018833	NA	paired	leaves	Schneider et al., 2016
B73_1	Zea mays	B73	SRP105290	SRR5466739	NA	single	seedling	Gent et al., 2017
B73_2	Zea mays	B73	SRP105290	SRR5466555	NA	single	seedling	Gent et al., 2017
B73_3	Zea mays	B73	SRP105290	SRR5466556	NA	single	seedling	Gent et al., 2017
B73_5	Zea mays	B73	SRP105290	SRR5466737	NA	single	seedling	Gent et al., 2017
B73_6	Zea mays	B73	SRP105290	SRR5466390	NA	single	seedling	Gent et al., 2017
J178-3	Zea luxurians	PI 422162	SRP105290	SRR5466387	SRR5466389	single	seedling	Gent et al., 2017
Oax-2	Zea mays mays	PI 628470	SRP105290	SRR5466588	SRR5466710	single	seedling	Gent et al., 2017

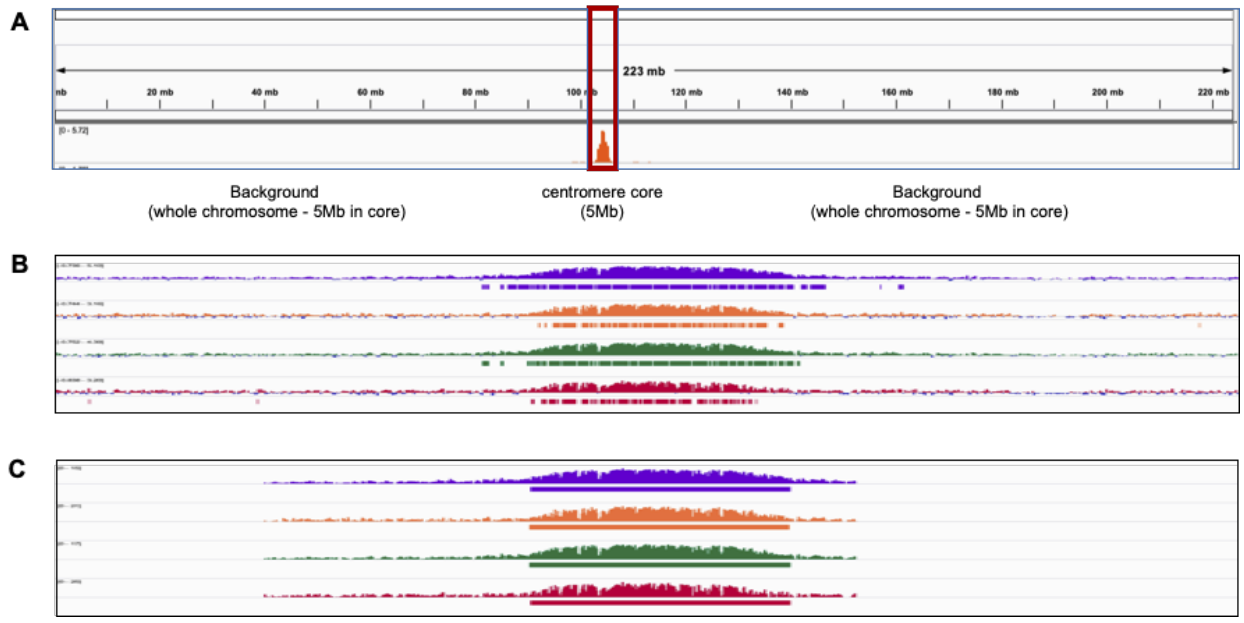


Figure 2.1 Impact of ChIP normalization on centromere size measurements.

(A) The 5 Mb centromere core region for centromere 5 in the inbred IL14H. All regions around this 5 Mb domain were defined as background. (B) CENH3-ChIP profiles of four IL14H biological replicates of centromere 5 before scaling. (C) The same four biological replicates after scaling. Window size: 10Mb.

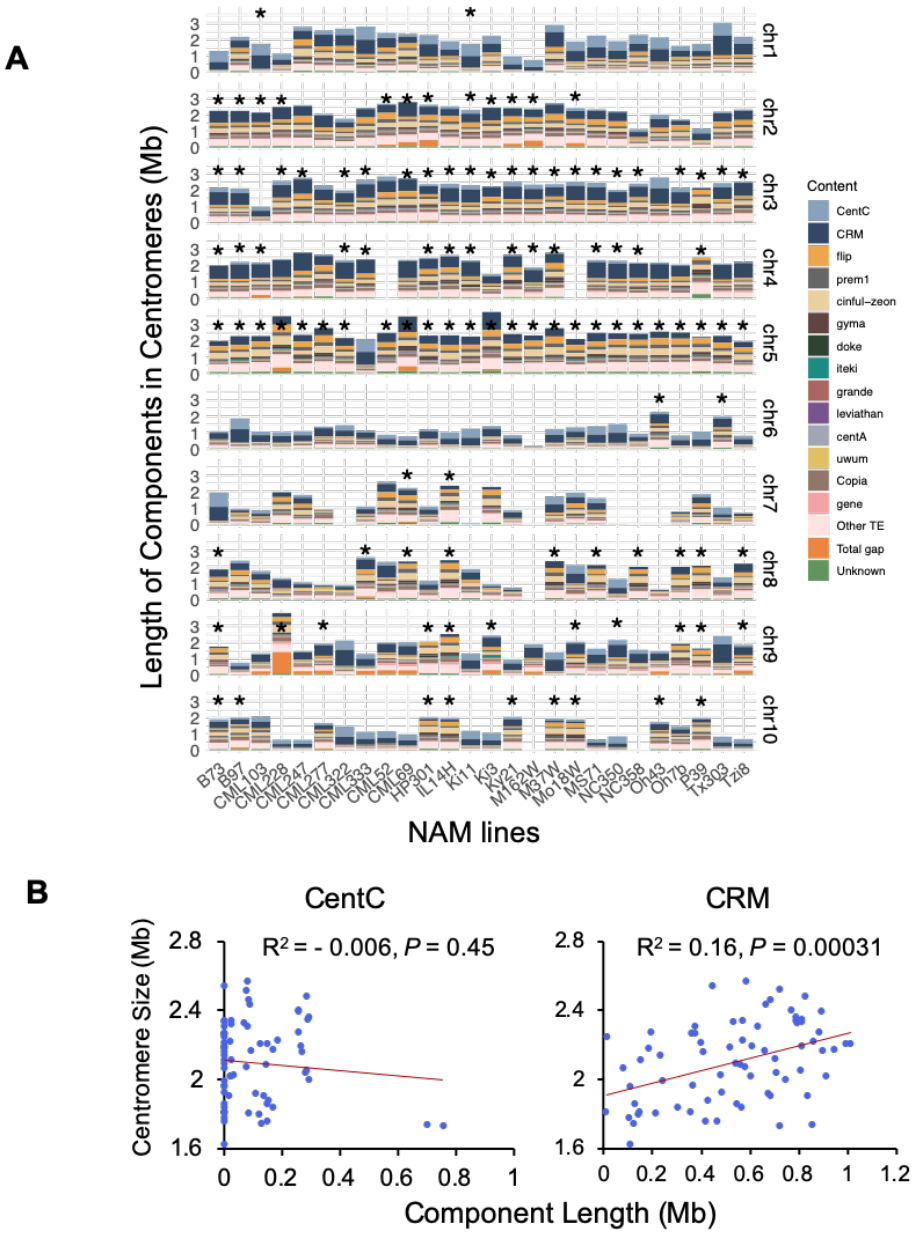


Figure 2.2 Centromeric repeats and their relationship to centromere size in the 26 NAM inbred lines.

(A) The relative abundance of major repeats in the centromeres of each inbred line. Stars indicate fully assembled centromeres. (B) Correlations between centromere size and the abundance of CentC and CRM.

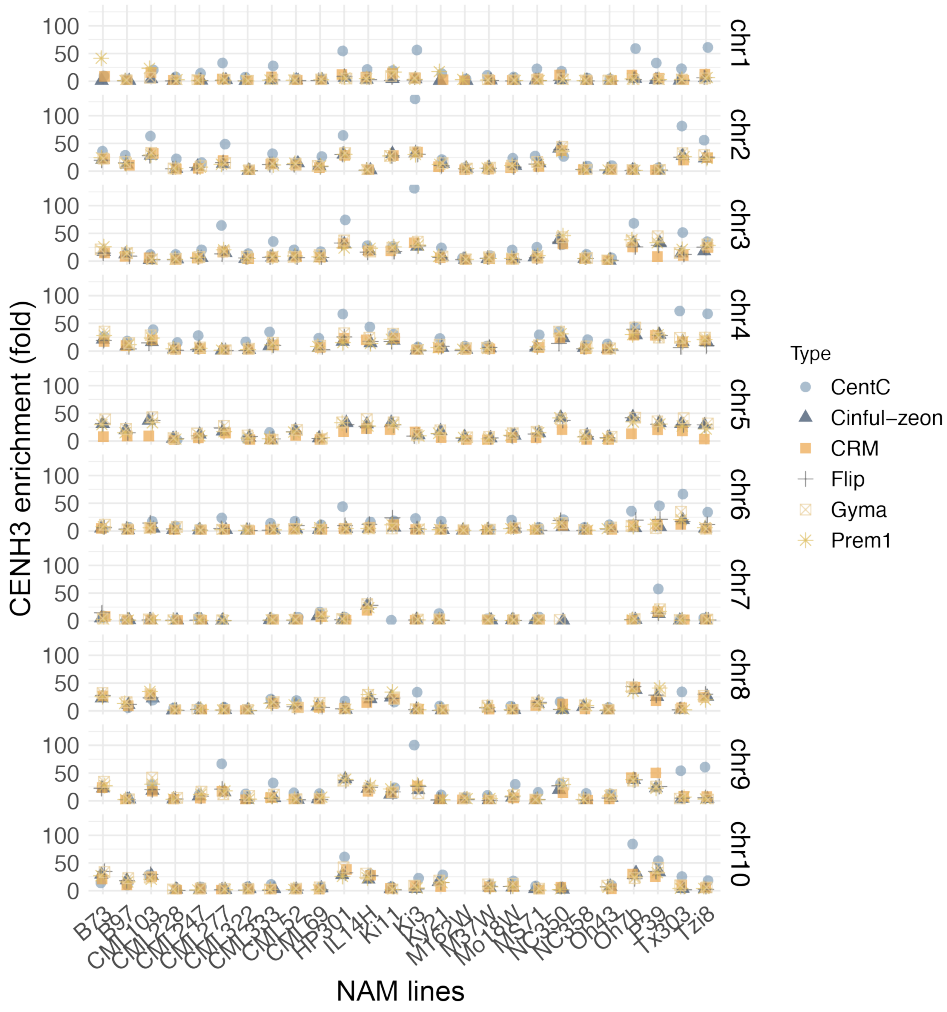


Figure 2.3 The ChIP enrichment for six classes of retrotransposons in centromeres of the NAM inbreds.

CRM, Flip, Prem, Cinful-Zeon and Gyma have a similar likelihood of interacting with CENH3. The apparently higher enrichment of CentC may be a result of artifactual alignments to the small fraction of total CentC in the assemblies.

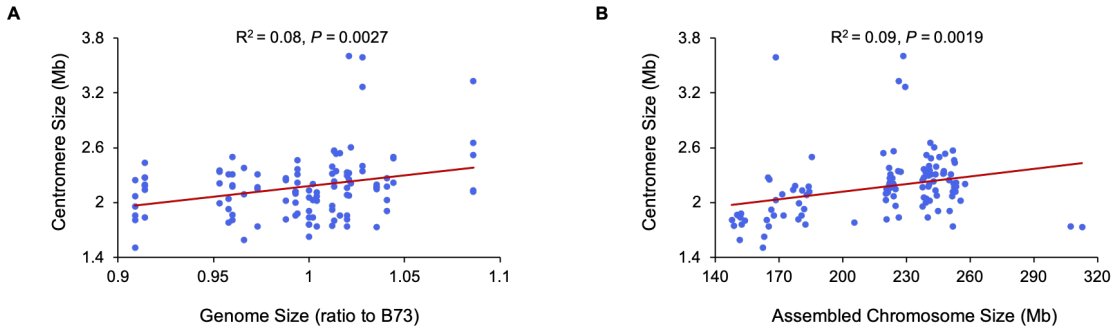
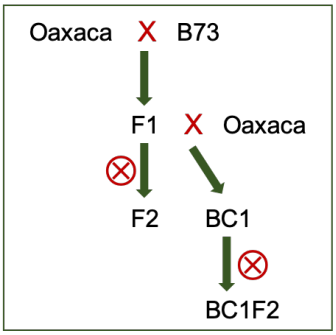
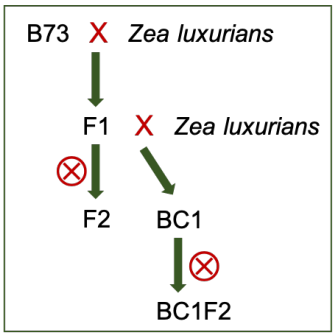


Figure 2.4 Centromere size is correlated with genome size and chromosome size.

The graphs show data from 113 fully assembled centromeres in the 26 NAM genomes. (A) Correlation between genome size and centromere size. (B) Correlation between chromosomes size and centromere size.

A





B

Generation	Genome from Oaxaca (%)	Genome size*	SD
B73	0	1.00	0.00
Oaxaca	100%	1.30	0.02
F2	50%	1.15	0.05
BC1F2	75%	1.20	0.01

Generation	Genome from luxurians (%)	Genome size*	SD
B73	0	1.00	0.00
luxurians	100%	1.61	0.02
F1	50%	1.31	0.01
F2	50%	1.31	0.04
BC1F2	75%	1.47	0.03

* measured as a ratio over the B73 maize reference genome.

Figure 2.5 The workflow for generating maize lines with different genome sizes.

(A) Crossing schemes for integrating B73 centromere into Oaxaca and *Zea. luxurians*. (B) Genome size of B73, Oaxaca, *Zea. luxurians* and their hybrids. GS, average genome size among 3-11 plants per family, SD indicates standard deviation.

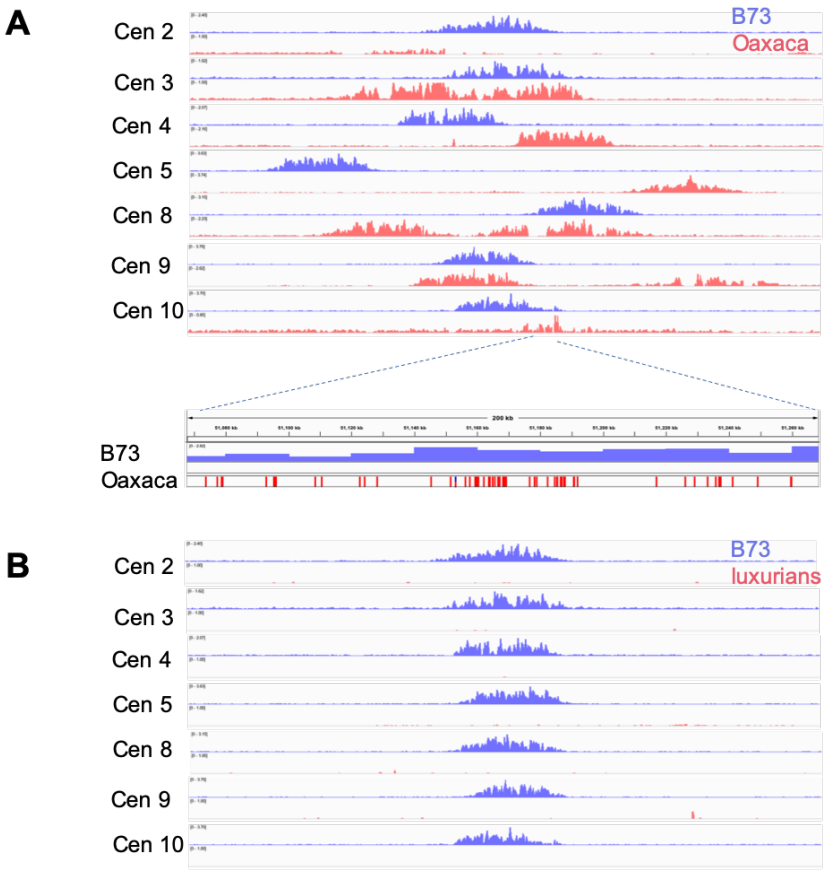


Figure 2.6 Comparison of B73 centromeres to those in Oaxaca and *Zea luxurians*.

(A) CENH3 ChIP profiles for seven centromeres in B73 and Oaxaca. Both B73 and Oaxaca reads were mapped to the B73 genome. Tick marks below the plot in the inset show SNPs on the Oaxaca ChIP reads. (B) CENH3 ChIP profiles for seven centromeres B73 and *Zea luxurians*. Window size: 10Mb.

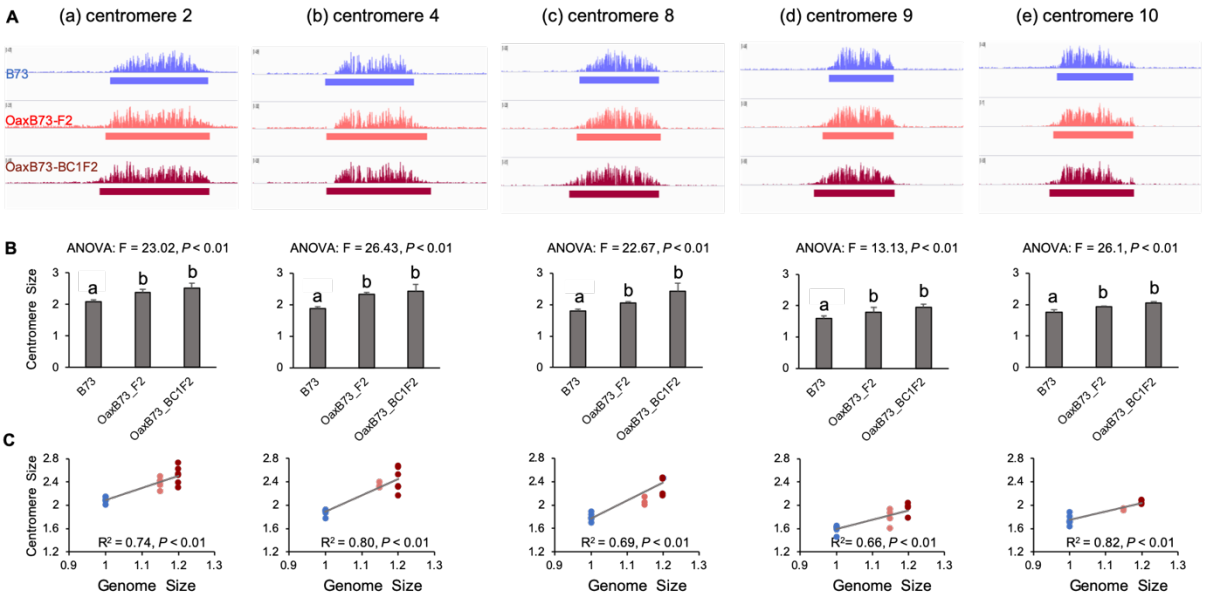


Figure 2.7 Centromeres are expanded in B73-Oaxa hybrids.

(A) CENH3-ChIP profiles of B73, OaxB73-F2 and OaxB73-BC1F2 for five centromeres. Window size: 5Mb. (B) ANOVA analysis of centromere sizes across different lines. Bar graphs show mean centromere size comparison among different lines. Letters represent different groups that are statistically different ($P < 0.05$). The sizes of all five centromeres in both OaxB73-F2 and OaxB73-BC1F2 progeny are significantly larger than in B73. (C) Linear regression analyses of centromere sizes across different lines. Circles represent different individuals. Blue: B73, red: F2, wine: BC1F2. Genome sizes are averages based on 3-11 individuals. The unit of centromere size is Mb.

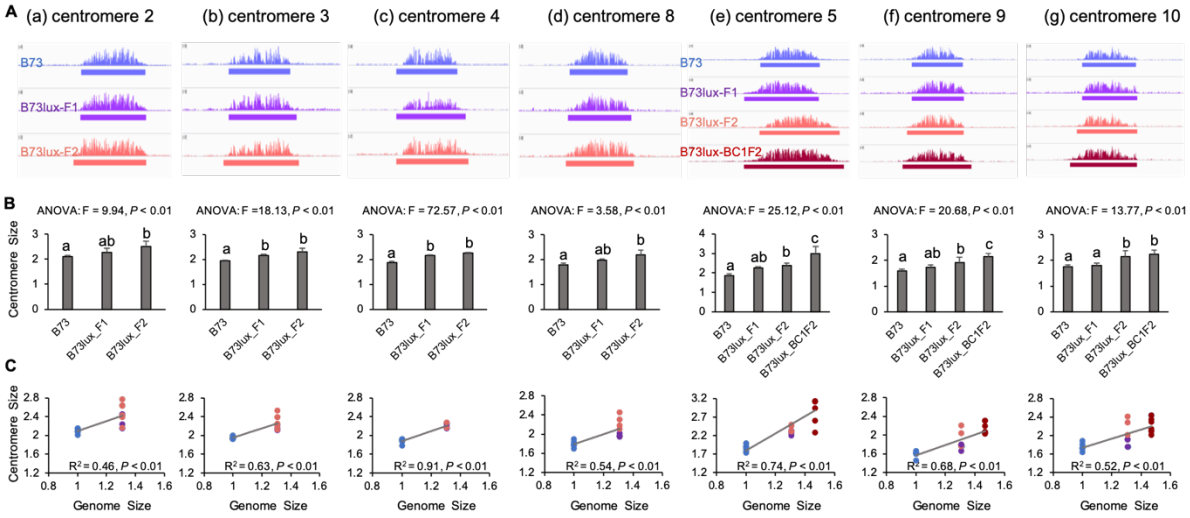


Figure 2.8 Centromeres are expanded in the B73-luxurians F2, BC1F2 progeny.

(A) CENH3-ChIP profiles of B73, B73lux-F1, B73lux-F2 and B73lux-BC1F2 on centromere 2, 3, 4, 5, 8, 9 and 10. Window size: 5Mb. Blue: B73, purple: F1, red: F2, wine: BC1F2. (B) ANOVA analysis of centromere size across B73, F1, F2 and BC1F2 progeny. Bar graph showing mean centromere size comparison between different lines. Letters represent significant differences among different subgroups and similar letters indicate no significant differences ($P < 0.05$). The sizes of Cen3 and Cen4 in B73lux-F1 are significantly bigger than B73, while the size of Cen2, Cen5, Cen8, Cen9 and Cen10 in B73lux-F1 are not significantly bigger than B73. The sizes of Cen5, Cen9 and Cen10 in B73lux-BC1F2 are significantly bigger than B73, and the sizes of Cen2, Cen3, Cen4, Cen5, Cen8, Cen9 and Cen10 in B73lux-F2 are significantly bigger than B73. (C) linear regression analysis on centromere size across F2, BC1F2 progeny. In the boxplots, the circles represent individual data points. Blue: B73, purple: F1, red: F2, wine: BC1F2. The genome size is the average genome size based on 3-10 individuals. The unit of centromere size is Mb.

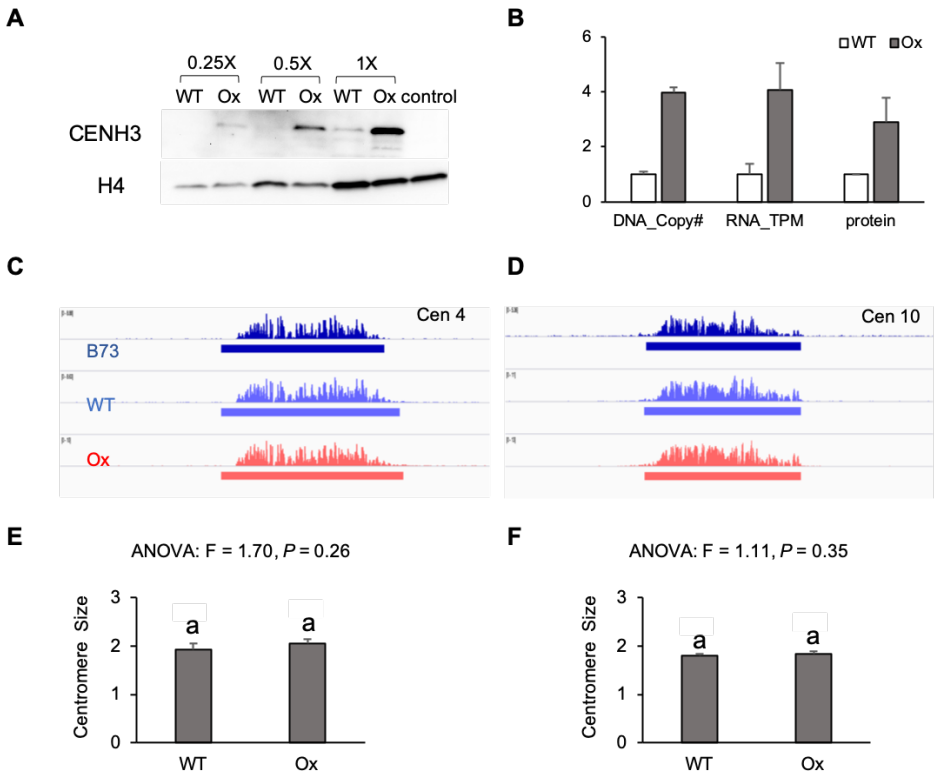


Figure 2.9 Centromere size is stable in the CENH3-overexpression lines.

(A) Protein blot analysis of maize CENH3 expression levels in roots of wild type (WT) and overexpression (Ox) lines. Nuclear proteins were diluted to 0.25X, 0.5X and 1X. Rice nuclear protein was used as a control. The same blot was incubated with antibodies to histone H4 as a loading control. (B) Quantification of *CENH3* gene copy number, mRNA expression levels based on RNA-seq, and protein levels from at least three independent biological replicates. Wild type expression was set to one in each experiment. CENH3 is a single copy gene in wild type lines. CENH3 protein expression levels were normalized to H4 levels. (C) and (D) CENH3-ChIP profiles of B73, WT and Ox for centromeres 4 and 10. Window size: 5Mb. (E) and (F) ANOVA analysis of centromere size across different lines. Bar graph showing mean centromere size comparison between different lines. There were no significant differences in the sizes of Cen4 and Cen10 between WT and Ox ($P < 0.05$). The unit of centromere size is Mb.

REFERENCES

Albert, P.S., Gao, Z., Danilova, T.V., and Birchler, J.A. (2010). Diversity of chromosomal karyotypes in maize and its relatives. Cytogenet. Genome Res. **129**: 6–16.

Aldrup-MacDonald, M.E., Kuo, M.E., Sullivan, L.L., Chew, K., and Sullivan, B.A. (2016). Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. Genome Res. **26**: 1301–1311.

Allshire, R.C. and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat. Rev. Genet. **9**: 923–937.

Amodeo, A.A. and Skotheim, J.M. (2016). Cell-Size Control. Cold Spring Harb. Perspect. Biol. **8**: a019083.

Bennett, M.D., Smith, J.B., Ward, J., and Jenkins, G. (1981). The relationship between nuclear DNA content and centromere volume in higher plants. J. Cell Sci. **47**: 91–115.

Bilinski, P., Albert, P.S., Berg, J.J., Birchler, J.A., Grote, M.N., Lorant, A., Quezada, J., Swarts, K., Yang, J., and Ross-Ibarra, J. (2018). Parallel altitudinal clines reveal trends in adaptive evolution of genome size in Zea mays. PLoS Genet. **14**: e1007162.

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

Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T., Cleveland, D.W., Black, B.E., Shah, J.V., and Jansen, L.E. (2014). The quantitative architecture of centromeric chromatin. Elife **3**: e02137.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120.

Boudichevskaia, A., Houben, A., Fiebig, A., Prochazkova, K., Pecinka, A., and Lermontova, I. (2019). Depletion of KNL2 Results in Altered Expression of Genes Involved in Regulation of the Cell Cycle, Transcription, and Development in Arabidopsis. Int. J. Mol. Sci. **20**: 5726.

Cavalier-Smith, T. (2005). Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann. Bot. **95**: 147–175.

Chen, C.-C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., Luger, K., and Mellone, B.G. (2014). CAL1 is the Drosophila CENP-A assembly factor. J. Cell Biol. **204**: 313–329.

Chia, J.-M. et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. **44**: 803–807.

Clarke, J.D. (2009). Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. Cold Spring Harbor Protocols **2009**: db.prot5177–pdb.prot5177.

Dawe, R.K. et al. (2018). A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. Cell **173**: 839–850.e18.

Dawe, R.K. and Henikoff, S. (2006). Centromeres put epigenetics in the driver's seat. Trends Biochem. Sci. **31**: 662–669.

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

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

Feng, C., Yuan, J., Bai, H., Liu, Y., Su, H., Liu, Y., Shi, L., Gao, Z., Birchler, J.A., and Han, F. (2019). The deposition of CENH3 in maize is stringently regulated. Plant J.

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

French, B.T., Westhorpe, F.G., Limouse, C., and Straight, A.F. (2017). Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. Dev. Cell **42**: 190–199.e10.

Gent, J.I., Dong, Y., Jiang, J., and Dawe, R.K. (2012). Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. Nucleic Acids Res. **40**: 1550–1560.

Gent, J.I., Wang, K., Jiang, J., and Dawe, R.K. (2015). Stable Patterns of CENH3 Occupancy Through Maize Lineages Containing Genetically Similar Centromeres. Genetics **200**: 1105–1116.

Gent, J.I., Wang, N., and Dawe, R.K. (2017). Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. Genome Biol. **18**: 121.

Gillooly, J.F., Hein, A., and Damiani, R. (2015). Nuclear DNA Content Varies with Cell Size across Human Cell Types. Cold Spring Harb. Perspect. Biol. **7**: a019091.

Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol. Rev. Camb. Philos. Soc. **76**: 65–101.

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

Iwata-Otsubo, A., Dawicki-McKenna, J.M., Akera, T., Falk, S.J., Chmátal, L., Yang, K., Sullivan, B.A., Schultz, R.M., Lampson, M.A., and Black, B.E. (2017). Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes

that Drive in Female Meiosis. Curr. Biol. **27**: 2365–2373.e8.

Jupe, F., Rivkin, A.C., Michael, T.P., Zander, M., Motley, T.S., Sandoval, J.P., Slotkin, K.R., Chen, H., Castagnon, R., Nery, J.R., and Ecker, J.R. (2018). The complex architecture of plant transgene insertions. bioRxiv: 282772.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**: 907–915.

Kynast, R.G. et al. (2001). A complete set of maize individual chromosome additions to the oat genome. Plant Physiol. **125**: 1216–1227.

Le Goff, S. et al. (2020). The H3 histone chaperone NASPSIM3 escorts CenH3 in Arabidopsis. Plant J. **101**: 71–86.

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., Demidov, D., Schubert, V., and Schubert, I. (2013). Arabidopsis KINETOCHORE NULL2 Is an Upstream Component for Centromeric Histone H3 Variant cenH3 Deposition at Centromeres. Plant Cell **25**: 3389–3404.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079.

Liu, J. et al. (2020). Gapless assembly of maize chromosomes using long-read technologies. Genome Biol. **21**: 121.

Marshall, W.F. (2016). Cell Geometry: How Cells Count and Measure Size. Annu. Rev. Biophys. **45**: 49–64.

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

McMullen, M.D. et al. (2009). Genetic properties of the maize nested association mapping population. Science **325**: 737–740.

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

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

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

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell,

C.R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. Nat. Genet. **36**: 138–145.

O'Connell, C.B., Khodjakov, A., and McEwen, B.F. (2012). Kinetochore flexibility: creating a dynamic chromosome-spindle interface. Curr. Opin. Cell Biol. **24**: 40–47.

Osterman, J.C. and Dennis, E.S. (1989). Molecular analysis of the ADH1-Cm allele of maize. Plant Mol. Biol. **13**: 203–212.

Pellicer, J. and Leitch, I.J. (2014). The application of flow cytometry for estimating genome size and ploidy level in plants. Methods Mol. Biol. **1115**: 279–307.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**: 290–295.

Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples.

Price, H.J., Sparrow, A.H., and Nauman, A.F. (1973). Correlations between nuclear volume, cell volume and DNA content in meristematic cells of herbaceous angiosperms. Experientia **29**: 1028–1029.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics **47**: 11.12.1–34.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. **42**: W187–91.

Robinson, D.O., Coate, J.E., Singh, A., Hong, L., Bush, M., Doyle, J.J., and Roeder, A.H.K. (2018). Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. Plant Cell **30**: 2308–2329.

Rosner, B. (2015). Fundamentals of Biostatistics (Cengage Learning).

Roy, B. and Sanyal, K. (2011). Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. Eukaryot. Cell **10**: 1384–1395.

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

Sandmann, M., Talbert, P., Demidov, D., Kuhlmann, M., Rutten, T., Conrad, U., and Lermontova, I. (2017). Targeting of Arabidopsis KNL2 to Centromeres Depends on the Conserved CENPC-k Motif in Its C Terminus. Plant Cell **29**: 144–155.

Schmoller, K.M. and Skotheim, J.M. (2015). The Biosynthetic Basis of Cell Size Control. Trends Cell Biol. **25**: 793–802.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**: 671–675.

Schneider, K.L., Xie, Z., Wolfgruber, T.K., and Presting, G.G. (2016). Inbreeding drives maize centromere evolution. Proc. Natl. Acad. Sci. U. S. A. **113**: E987–96.

Schrader, F. (1936). THE KINETOCHORE OR SPINDLE FIBRE LOCUS IN AMPHIUMA TRIDACTYLUM. Biol. Bull. **70**: 484–498.

Shou, H., Frame, B.R., Whitham, S.A., and Wang, K. (2004). Assessment of transgenic maize events produced by particle bombardment or Agrobacterium-mediated transformation. Mol. Breed. **13**: 201–208.

Shrestha, R.L., Ahn, G.S., Staples, M.I., Sathyan, K.M., Karpova, T.S., Foltz, D.R., and Basrai, M.A. (2017). Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. Oncotarget **8**: 46781–46800.

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

Tenaillon, M.I., Hufford, M.B., Gaut, B.S., and Ross-Ibarra, J. (2011). Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians. Genome Biol. Evol. **3**: 219–229.

Thiel, T., Kota, R., Grosse, I., Stein, N., and Graner, A. (2004). SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res. **32**: e5.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. **14**: 178–192.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J. Cell Sci. **114**: 3529–3542.

Wang, K., Wu, Y., Zhang, W., Dawe, R.K., and Jiang, J. (2014). Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res. **24**: 107–116.

Wang, N., Gent, J.I., and Dawe R.K. (2020). Haploid induction by a maize cenh3 null mutant. Manuscript submitted for publication.

Wolfgruber, T.K. et al. (2009). Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet. **5**: e1000743.

Zhang, H. and Dawe, R.K. (2012). Total centromere size and genome size are strongly correlated in ten grass species. Chromosome Res. **20**: 403–412.

Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K. (2002). Centromeric retroelements and satellites interact with

maize kinetochore protein CENH3. Plant Cell **14**: 2825–2836.

Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J., and Nurse, P. (2010). A coordinated global control over cellular transcription. Curr. Biol. **20**: 2010–2015.

CHAPTER 3

HAPLOID INDUCTION BY A MAZIE CENH3 NULL MUTANT [2]

[2] Na Wang, Jonathan I. Gent, R. Kelly Dawe. Manuscript submitted to *Science*.

Abstract

The production of haploids is an important first step in creating many new plant varieties. One approach used in *Arabidopsis* involves crossing plants expressing different forms of Centromeric Histone H3 (CENH3) and subsequent loss of the genome with weaker centromeres. However, the method has been ineffective in crop plants. Here we describe a greatly simplified method based on crossing maize lines that are heterozygous for a *cenh3* null mutation. Crossing +/*cenh3* to wild type plants in both directions yielded haploid progeny. Genome elimination was determined by the *cenh3* genotype of the gametophyte, suggesting that centromere failure is caused by CENH3 dilution during the post-meiotic cell divisions that precede gamete formation. The *cenh3* haploid inducer works as a vigorous hybrid and can be transferred to other lines in a single cross, making it versatile for a variety of applications.

Introduction

Tens of thousands of maize haploid lines are generated by breeding companies around the world each year as a prerequisite for creating new inbreds, which are ultimately used to produce hybrids for sale. The induced haploids are doubled by colchicine and immediately tested for agronomic performance. The traditional technology is based on an inbred called Stock 6 that induces haploids when crossed as a male (Coe, 1959) and in modern lines has been reported to induce haploids at frequencies as high as ~15% (Uliana Trentin et al., 2020). The key underlying gene, called *Matrilineal* (MATL/ZMPLA1/NLD), is a patatin-like phospholipase expressed primarily in pollen (Kelliher et al., 2017; Liu et al., 2017; Gilles et al., 2017). Its mechanism of action is not understood, but may involve a change in membrane properties during fertilization that leads to a loss of the paternal chromosomes. Mutations in *matrilineal* also induce haploids in

rice and wheat (Yao et al., 2018; Liu et al., 2020b, 2020a). The *Matrilineal* gene is not conserved in dicotyledonous plants, although recent work has revealed a conserved enhancer of *matrilineal* that induces haploids at low levels in *Arabidopsis* (Zhong et al., 2020).

A potentially superior and broadly useful method of inducing haploids was pioneered by Simon Chan and colleagues, who showed that crossing *Arabidopsis* lines with a structurally altered Centromeric Histone H3 (CENH3) protein yielded haploids and aneuploids at frequencies as high as 25-45% (Ravi and Chan, 2010). CENH3 is a histone variant that defines centromere location and recruits overlying kinetochore proteins (Zhong et al., 2002; Cheeseman and Desai, 2008). The original study involved a construct called *GFP-tailswap* where the N-terminal tail of CENH3 was replaced with sequence from another H3 variant and modified with a GFP tag. Recent data demonstrate that point mutations and small deletions of CENH3 can also induce haploids at similar frequencies (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015, 2020). However, outside of *Arabidopsis*, centromere-mediated haploid induction has proven to be less effective, generally producing <1% haploids (Kalinowska et al., 2019).

In an earlier report we proposed that the mechanism of centromere-mediated haploid induction is based on differences in effective centromere size between haploid inducers and their wild type crossing partners (Wang and Kelly Dawe, 2018). *CENH3* point mutations reduce CENH3 loading in somatic cells (Karimi-Ashtiyani et al., 2015) and *GFP-tailswap* lines show impaired CENH3 loading in meiosis (Ravi et al., 2011), suggesting that haploid inducers transmit small or weak centromeres to gametophytes. When crossed to wild type plants, the progeny have a centromere size imbalance, which we argued leads to targeted destruction of the smaller centromeres by natural clearing mechanisms that remove misplaced CENH3 and spurious small centromeres (Wang and Kelly Dawe, 2018). The centromere size model predicts

that *cenh3* mutants with the most severe loss of function will be the best haploid inducers. However, severe *cenh3* mutants also compromise plant growth, making it difficult to breed a healthy haploid inducer line. This contradiction in goals likely explains the poor success record of centromere-mediated haploid induction in species other than *Arabidopsis*.

Results and Discussion

The current study was designed to test the centromere size model in maize, initially using the *GFP-tailswap* method. However, this approach is complicated by the fact that it requires both a mutant of native *cenh3* and a functional *GFP-tailswap* transgene that complements the mutant. Another group had already shown some success using an existing maize mutant (*cenh3-mu1015598*) caused by a *Robertsons Mutator* (*Mu*) insertion in the 5' UTR of the gene (Kelliher et al., 2016; Feng et al., 2019). They crossed *GFP-tailswap* into the *cenh3-mu1015598* background and observed an average of 0.86% haploids when crossed as a male and no haploids when crossed as a female (Kelliher et al., 2016). We also obtained *cenh3-mu1015598* and self-crossed three heterozygous plants. Genotyping revealed that two ears segregated a low frequency of homozygous mutants that grew to various states of maturity (Table 3.1). The recovery of homozygous mutants indicates that *cenh3-mu1015598* is not a null, and that the prior results may have been confounded by a low level of wild type *CENH3* expression. The variable penetrance of the *cenh3-mu1015598* allele can be explained by the fact that *Mu* elements can promote low levels of expression when inserted into 5' UTR regions (Barkan and Martienssen, 1991).

To overcome the selection against true null *cenh3* alleles, we opted to create a *cenh3* null using a two-construct CRISPR/Cas9 approach. One line was transformed with a simple construct expressing Cas9 driven by a ubiquitin promoter. A second was transformed with a construct

expressing a gRNA targeting the fourth exon of the native *CENH3* gene and an "*ImmuneCENH3*" gene that contains a full-length native *CENH3* gene with five silent nucleotide changes in the gRNA target area (Figure 3.1A, 3.1B). After we crossed the two lines together, Cas9 generated mutations in the native *CENH3* gene but left the transgene unaffected. We chose a *cenh3* allele with a single nucleotide deletion that causes an immediate stop codon in the N-terminal tail of CENH3 (Figure 3.1C). In the presence of *ImmuneCENH3,* the *cenh3* mutation segregates as a simple Mendelian recessive trait (Table 3.2). We then created transgenics with *TailswapCENH3,* a close replica of the *Arabidopsis GFP-tailswap* construct, and crossed it to the *cenh3* mutation. We were unable to obtain any plants that contained *TailswapCENH3* and were homozygous for *cenh3* (Table 3.2), suggesting that the transgene does not complement a true null. This is likely because the large GFP tag interferes with the binding of other kinetochore proteins that interact with CENH3 (Cheeseman and Desai, 2008). *Arabidopsis GFP-tailswap* plants are severely stunted (Maheshwari et al., 2015) and nearly sterile (Ravi et al., 2011), indicating a similar negative impact on kinetochore function.

Except for tailswap-CENH3, we also crossed *cenh3* null mutants with other *CENH3* transgenics to test whether they can complement *cenh3* null mutants. In Arabidopsis, the *cenh3* null mutant can also be complemented by CENH3 from a wide variety of angiosperms, including monocot *Zea mays* (Ravi et al., 2011). Our lab has created a large number of transgenics for a prior aim of tethering CENH3 to an array of synthetic repeat arrays (Zhang et al., 2012). In the present study, we crossed the *cenh3* heterozygous mutant with *CENH3* promoter-LexA-CENH3, where LexA is DNA binding module fused to the N-terminus of the CENH3 protein. This transgene induces obvious meiotic abnormalities which are very similar to the tailswap-CENH3 meiotic phenotype in Arabidopsis (Ravi et al., 2011; Zhang, 2011). The maize *cenh3*

heterozygous mutants were also crossed with other variants of CENH3 transgenic lines.These include *MaizeCENH3* (the cDNA of maize CENH3 gene driven by maize CENH3 promoter), *RiceCENH3* (the cDNA of rice CENH3 gene driven by maize CENH3 promoter) and *SorghumCENH3* (the cDNA of sorghum CENH3 gene driven by maize CENH3 promoter) (Figure 3.4). Our results showed that only in the presence of *MaizeCENH3*, we were able to get three *cenh3* homozygous mutants out of 73 samples (Table 3.3), which was much less than the ImmuneCENH3 complementation ratio (43/172, Table 3.2). However, the other CENH3 variants are not able to rescue the null mutant (Table 3.3). These results suggested that the introns of CENH3 also play a role in the maintenance of CENH3 function and maize CENH3 protein is very stringent across different species.

During the course of these studies, we noted that *cenh3* was occasionally transmitted in the absence of *ImmuneCENH3.* By crossing to wild type lines, we were able to obtain a simple segregating *cenh3* line that lacked both of the original transgenes. Among selfed progeny from a +/*cenh3* line there were 163 +/+ wild type individuals, 55 +/*cenh3* heterozygotes, and zero *cenh3*/*cenh3* homozygotes, indicating that the mutant is homozygous lethal and poorly transmitted through gametophytes. We also carried out reciprocal crosses between +/*cenh3* heterozygotes and wild type plants. A Mendelian trait is normally transmitted to 50% of testcross progeny, however we observed that only 12.1% of the progeny received *cenh3* when crossed through the male and 25% when crossed through the female (Table 3.4). The reduction in transmission is expected because sperm and eggs are carried within multicellular haploid gametophytes. Two haploid cell divisions precede the formation of sperm and three haploid cell divisions precede the formation of an egg. Those gametophytes with the *cenh3* allele must use CENH3 carried over from the sporophytic phase while it is naturally diluted at each cell cycle

(Lermontova et al., 2006). A general expectation is that *cenh3* sperm would have about ¼ of the normal amount of CENH3 and an egg carrying *cenh3* would have about ⅛ relative to the *cenh3* heterozygous parent. Assuming no dosage compensation, those values would be reduced by an additional ½ relative to a normal homozygous wild type parent. As a result, sperm and eggs carrying *cenh3* presumably have smaller centromeres.

To test whether +/*cenh3* heterozygous mutants are able to induce haploids, we crossed *cenh3* heterozygotes with tester lines in both directions. In the first test we crossed wild type and +/*cenh3* plants to a line that is homozygous for a recessive *glossy8* (*gl8*) mutation on chromosome 5 that causes seedling leaves to have a shiny appearance. We observed that 0.5% of the progeny were glossy when +/*cenh3* heterozygotes was crossed as male, and 5.0% of the progeny were glossy when +/*cenh3* plants were crossed as female (Table 3.5). Flow cytometry analysis revealed that all 45 of the glossy plants were haploids, an interpretation that was confirmed by counting chromosomes in root tip cells of three plants (Figure 3.2A, 3.2B). When grown to maturity the haploid plants were short and sterile as expected (Chase, 1969) (Figure 3.2C, 3.2D). We also observed two non-glossy plants with stunted phenotypes that we hypothesized might be aneuploids. These two plants were skim sequenced along with six haploids. While the haploids showed uniform sequence coverage, the stunted plants did not; one was trisomic for chromosome 3, and the other was monosomic for chromosome 2 and 4 and trisomic for chromosome 10 (Figure 3.3).

We then carried out a second set of tests using *glossy1* (*gl1)*, which has a similar phenotype but the mutation is on chromosome 7. In these crosses we also scored the germination rate, which is an indirect measure of karyotypic abnormality commonly used to estimate the efficacy of *Arabidopsis* haploid inducers (Ravi et al., 2014; Kuppu et al., 2015). In crosses where

+/*cenh3* heterozygotes were the female, 5.2% of the progeny were glossy and haploid by flow cytometry measurements. Another 3.3% of the progeny showed the glossy phenotype but had a higher DNA content than expected for haploids, which were scored as aneuploids (Table 3.6). Different crosses differed considerably in the germination rate (65-91%), frequencies of haploids (1.2-8.9%) and aneuploids (2.1-5.1%) (Table 3.6). Sequence data from five aneuploid plants confirmed that all except one were missing chromosome 7, sometimes in conjunction with the loss of other chromosomes. One glossy plant that appeared to have two complete copies of chromosome 7 may have had a small interstitial deletion that was not detectable by skim sequencing (segmental aneuploids are common in *Arabidopsis GFP-tailswap* crosses (Tan et al., 2015)). The results from the *gl1* tests are more in line with what has been observed in *Arabidopsis*, where any given cross with GFP-*tailswap* generally yields haploids and aneuploids in similar proportions (Ravi and Chan, 2010; Ravi et al., 2014; Maheshwari et al., 2015).

If CENH3 dilution is the underlying mechanism for haploid induction, then only gametes carrying the *cenh3* mutation from the +/*cenh3* parent should induce haploids. Unfortunately it is not possible to score seedlings for the presence of the *cenh3* allele because the genome of the haploid inducer is lost. However, data from *Arabidopsis GFP-tailswap* crosses show that endosperm rarely displays complete uniparental genome elimination when the seedling is haploid (Ravi et al., 2014). If true in maize as well, the genotype of the endosperm could be used to determine the original genotype of the seedling. We genotyped the remnant endosperm from a set of eleven haploid plants produced from a +/*cenh3* X *gl8* cross. All eleven were heterozygous for the *cenh3* allele, indicating that haploid induction is caused by gametes carrying *cenh3*. Since only 25% of the progeny from a female cross receive *cenh3* (Table 3.4), yet our haploid frequencies are calculated based on total seed counts, the effective haploid induction frequency is

on the order of 20%. Bringing this level of haploid induction into practice will require screening for kernels that received *cenh3*. There are multiple publicly available endosperm-expressed GFP transgenes within 100 Kb of the *CENH3* gene that could be linked to the null for this purpose (Li et al., 2013) (http://www.acdsinsertions.org/).

One of the striking elements of centromere-mediated haploid induction is that it is effective in only a subset of progeny. In some individuals, all the chromosomes from the haploid inducer parent are lost, and in another much larger subset, no chromosome loss occurs. The relatively small aneuploid class represents "partial haploid induction" events, where some chromosomes were lost but others survived. The fact that the *gl8* crosses yielded more true haploids than the *gl1* crosses may be related to the fact that the former were carried out in the summer while the latter were carried out in winter. It is also possible that the selection scheme played a role. Studies using the maize *r-X1* deletion line, which generates monosomics at high frequency, have demonstrated that some chromosomes are recovered as monosomics at higher frequency than others (Weber, 1994). Monosomics for chromosome 5 (with *gl8*) are rarely recovered whereas monosomics for chromosome 7 (with *gl1*) are far more common (17 times more common (Weber, 1994)). Indeed, two of five sequenced aneuploids from *gl1* crosses were monosomic for chromosome 7 only (Table 3.6). These data suggest that the *gl8* tester favors the recovery of haploids while the *gl1* tester recovers both haploids and aneuploids.

All prior literature on centromere mediated haploid induction describes the complementation of a null allele with structural variants of CENH3 or alleles that produce altered or partially deleted forms of CENH3 (Ravi and Chan, 2010; Maheshwari et al., 2015; Karimi-Ashtiyani et al., 2015; Ishii et al., 2016; Kuppu et al., 2020). These data have served to sustain the original interpretation that haploid induction is caused by a competition between two

structurally different forms of CENH3, and ultimate rejection of the altered centromeres by a surveillance mechanism for improper assembly (Ravi and Chan, 2010; Britt and Kuppu, 2016; Kalinowska et al., 2019; Kuppu et al., 2020). In contrast, we have achieved high levels of haploid induction using a *cenh3* mutation in the N-terminal tail that removes all sequence that interacts with DNA or other histones. Therefore, quantitative reductions in CENH3 alone can induce centromere-mediated haploid induction, as predicted by the centromere size model (Wang and Kelly Dawe, 2018). As null alleles are readily obtained using CRISPR technology, it should be possible to rapidly develop centromere-mediated haploid inducers in other species as well.

In maize it is likely that the frequency of haploid induction can be improved under standard breeding practices, similar to how the original ~3% haploid frequency observed with Stock 6 (Coe, 1959) was improved to ~15% in multiple breeding programs around the world (Uliana Trentin et al., 2020). The major advantages of the *cenh3* approach are that it can be used to create either paternal or maternal haploids, that it does not require a transgene, and that the plants are phenotypically wild type and can be used as vigorous hybrids (Figure 3.5). Since the inducer works as a heterozygote, *cenh3* can be crossed to any line and the F1 will become a haploid inducer. This feature should make it particularly useful when combined with other technologies that are built upon haploids, such as genotype-independent gene editing (Kelliher et al., 2019), synthetic apomixis (Wang et al., 2019), and passing engineered chromosomes from one line to another (Birchler, 2014).

MATERIALS AND METHODS

Plant materials

The *gl1*, *gl8*, and *cenh3-mu1015598* transposon insertion lines were obtained from the Maize Genetics Cooperation Stock Center, Urbana, Illinois. The *cenh3-mu1015598* allele is one of several mutations in the UFMu-01386 stock line. All plants were grown in the University of Georgia Plant Biology greenhouses. The *gl8* crosses were carried out in August of 2019 and the *gl1* crosses were carried out in December of 2019.

Construct preparation and transformation

The Ubi-Cas9 construct contains 1991 bp of the maize polyubiquitin promoter (GenBank: S94464.1) driving a maize codon-optimized version of Cas9 terminated by the Nos terminator.

The *gRNA-ImmuneCENH3* construct contains two components, a guide RNA module and the *ImmuneCENH3* gene. The guide RNA portion contains the maize U6 promoter (Svitashev et al., 2015) driving a guide RNA (TCCCGCAGCGCTACAGTCCC) terminated by the PolIII terminator TTTTTTTT. The *ImmuneCENH3* portion contains 6455 bp of the native *CENH3* gene (coordinates Chr6:166705239-166711693 on Zm-B73-REFERENCE-NAM-5.0) but has five silent codon changes in the gRNA target area (CCAGGTACGGTCGCCCTGCGCGA). The promoter includes 2184 bp of sequence upstream of the ATG.

To create the *gRNA-TailswapCENH3* construct, the natural 5' UTR of *CENH3* was retained and a codon-optimize GFP sequence was inserted at the ATG of *ImmuneCENH3*. This was followed by a linker sequence ATGGATGAACTATACAAGGGCGGAGGCGGTGGAGGCGTCGAC and the tail sequence of

the maize H3.3 gene (Genbank NM_001294303.2) including its intron, fused to the native *CENH3* gene 3 bp upstream of the guide RNA target area. The Arabidopsis *GFP-tailswap* transgene also includes the H3.3 portion. Our construct was based on the sequence of Arabidopsis *GFP-tailswap* obtained from the Comai laboratory.

The three constructs were synthesized by GenScript (www.genscript.com) and cloned into the binary vector pTF101.1 (Paz et al., 2004). The constructs were transformed into the maize HiII line at the Iowa State University Plant Transformation Facility (Ames, IA) and grown in University of Georgia greenhouses. To generate the *cenh3* mutation, transgenic lines carrying *Ubi-Cas9* were crossed with lines carrying *gRNA-ImmuneCENH3*.

The *LexA-CENH3* was from a previous study in our lab (Zhang et al. 2012), and the three transgenic lines *MaizeCENH3*, *RiceCENH3* and *SorghumCENH3* were generated by Jeffrey Ross-Ibarra's lab.

DNA Extraction, genotyping and sequence analysis

For standard leaf genotyping, genomic DNA was prepared using a CTAB protocol (Clarke, 2009). Endosperm tissue was collected after the kernels had germinated and the glossy phenotype could be distinguished. Embryos and pericarps were removed with forceps, and the endosperm ground to a powder with a mortar and pestle. The endosperm DNA was extracted with the IBI Plant Genomic DNA Mini Kit (IBI Scientific IB47231).

To identify the presence of *ImmuneCENH3* and Cas9 in transgenic lines, primers CENH3-F2 and CENH3-R3 were used to amplify *ImmuneCENH3*, and primers Cas9-F1 and Cas9-R1 were used to amplify Cas9 (Table 3.7). To identify the original *cenh3* mutation in Cas9 plants, PCR was carried out using the Phusion High-Fidelity PCR Kit (New England Biolabs,

Ipswich, MA) with primers CENH3-F1 and CENH3-R1 in Table 3.7. The PCR products were either directly Sanger sequenced or cloned using a TOPO TA cloning kit (Thermo Fisher #K457501) and then Sanger sequenced.

In lines that lack *ImmuneCENH3*, the *cenh3* null allele was differentiated from the native *CENH3* allele by PCR and restriction enzyme digestion. PCR amplifies a 496 bp PCR product using primers CENH3-F2 and CENH3-R2. When this product is digested with the restriction endonuclease AlwNI (New England Biolabs), the wild type allele is cleaved into two pieces of size 284 bp and 212 bp while the mutant *cenh3* allele is not cleaved.

The *cenh3-mu1015598* allele was scored using the primers CENH3-F4, CENH3-R4 and Mumix (a 1:1 mix of the two primers Mu1 and Mu2 in Table 3.7). The wild type allele is amplified with CENH3-F4 and CENH3-R4 while the *Mu* allele is amplified with CENH3-F4 and Mumix.

To identify the presence of *MaizeCENH3*, *RiceCENH3*, *SorghumCENH3* and *LexA-CENH3* in the transgenic lines, primers maize-cenh3-F1 and maize-cenh3-R1 were used to amplify *MaizeCENH3*, primers rice-cenh3-F1 and rice-cenh3-R1 were used to amplify *RiceCENH3*, primers sorghum-cenh3-F1 and sorghum-cenh3-R1 were used to amplify *SorghumCENH3*, and primers LexA-F and LexA-R were used to amplify *LexA-CENH3* (Table 3.7).

Ploidy Evaluation

Progeny from +/*cenh3* crosses were grown indoors under grow lights for 10-13 days and water sprayed on the seedlings to identify the glossy phenotype. All glossy plants were subsequently assayed by flow cytometry. For each individual, about 1 g of flash-frozen leaves or

roots were collected and chopped into 1.5 ml of pre-chilled nuclei extraction buffer (2 mM EDTA, 15 mM Tris-HCl pH 7.5, 20 mM NaCl, 80 mM KCl, 0.5 mM spermine, 15 mM 2-mercaptoethanol, 0.1 mM PMSF, 0.1% Triton X-100). After chopping, the mixture was filtered through a 40 µm cell strainer twice. The nuclei were stained with 4,6-diamidino-2-phenylindole and loaded into flow cytometers hosted by the CTEGD Cytometry Shared Resource Lab at the University of Georgia.

Chromosome spreads

Chromosome analysis was carried out as described in (Dawe et al., 2018). Briefly, root tips were collected from the haploid and diploid plants, incubated in a chamber with nitrous oxide for three hours, and fixed with 90% acetic acid. Root tips were cut with a razor blade and digested in an enzyme solution (1% pectolyase Y-23, 2% cellulase Onozuka R-10) at 37°C for 50 minutes. The root section was washed in ethanol then immersed in 90% acetic acid. A metal pick was used to crush the roots tips and 10 µl of the cell suspension was dropped onto microscope slides. Slides were dried and mounted with a glass coverslip using ProLong Gold with DAPI (Thermo Fisher Cat# P36931). Slides were imaged on a Zeiss Axio Imager.M1 fluorescence microscope with a 63X Plan-APO Chromat oil objective, and slidebook software (Intelligent Imaging Innovations, Denver, CO, USA) used to analyze the data.

Skim sequencing of haploids and aneuploids

For each sample, 12 ng/µl DNA was sonicated in a 100 µl volume with a Diagenode Bioruptor for seven minutes on high setting with 30-second on-off intervals, yielding fragments averaging about 500 bp in length. DNA sequencing libraries were prepared using the KAPA

Hyperprep Kit (KK8502) with KAPA single-indexed adapters (KK8700). 600 ng of sonicated DNA was used as input for each sample, and 3 cycles of PCR were used to amplify libraries. 150-nt Illumina sequencing reads were adapter-trimmed and quality-filtered using Cutadapt version 1.9.1 (Martin, 2011) with parameters as follows: "-q 20 -a AGATCGGAAGAGC -e .05 -O 1 -m 50". Reads were aligned to Zm-B73-REFERENCE-NAM-5.0 (https://nam-genomes.org/) using BWA-mem version 0.7.15 in single-end mode with default parameters (Li and Durbin, 2009). Read coverage was visualized using IGVTools version 2.3.98 (Thorvaldsdóttir et al., 2013) with coverage calculated on 25Mb intervals.

Acknowledgements

We thank Mary Tindall Smith, Fangfang Fu, and Sydney Greenlaw for help with genotyping, and William A. Ricci for help with the flow cytometry analysis, and Jianing Liu for critically reading the manuscript. We thank Jeffrey Ross-Ibarra for generously sharing the transgenic lines. We appreciate the support of the Georgia Genomics and Bioinformatics Core facility and the Cytometry Shared Resource Laboratories at the University of Georgia. This work was funded by grant 1444514 from the National Science Foundation.

Table 3.1 The *cenh3-mu1015598* mutant is not a null [1].

Cross	# seedlings	expected WT:het:hom	observed WT:het:hom
+/*cenh3-mu1015598*-1 ⊗	76	19:38:19	23:44:**8**
+/*cenh3-mu1015598*-2 ⊗	37	9:19:9	6:25:**6**
+/*cenh3-mu1015598*-3 ⊗	25	6:13:6	8:18:0

[1] WT indicates wild type, het indicates +/*cenh3-mu1015598* heterozygote, and hom indicates *cenh3-mu1015598*/*cenh3-mu1015598* homozygote. Note that the first two plants yielded homozygous *cenh3-mu1015598* progeny.

Table 3.2 Segregation of *cenh3* in *ImmuneCENH3* and *TailswapCENH3* backgrounds [1].

Cross	transgene genotypes	*cenh3* genotypes
+/ImmuneCENH3, +/cenh3 ⊗	*+/ImmuneCENH3* or *ImmuneCENH3/ImmuneCENH3* (172)	+/+ (54) *+/cenh3* (75) *cenh3/cenh3* (43)
	+/+ (33)	
+/TailswapCENH3, +/cenh3 ⊗	*+/TailswapCENH3* or *TailswapCENH3/TailswapCENH3* (56)	+/+ (47) *+/cenh3* (6) *cenh3/cenh3* (0)
	+/+ (16)	

[1] Numbers in parentheses show the number of plants of each genotype.

Table 3.3 *cenh3* null mutant is able to be complemented by intronless maize CENH3 but not other CENH3 variants [1].

Cross	transgene genotypes	*cenh3* genotypes
+/MaizeCENH3, +/cenh3 ⊗	+/*MaizeCENH3* or *MaizeCENH3*/*MaizeCENH3* (73)	+/+ (34) +/*cenh3* (36) *cenh3*/*cenh3* (**3**)
	+/+ (13)	
+/SorghumCENH3, +/cenh3 ⊗	+/*SorghumCENH3* or *SorghumCENH3*/*SorghumCENH3* (63)	+/+ (60) +/*cenh3* (3) *cenh3*/*cenh3* (0)
	+/+ (22)	
+/RiceCENH3, +/cenh3 ⊗	+/*RiceCENH3* or *RiceCENH3*/*RiceCENH3* (36)	+/+ (30) +/*cenh3* (6) *cenh3*/*cenh3* (0)
	+/+ (11)	
+/LexA-CENH3, +/cenh3 ⊗	+/*LexA-CENH3* or *LexA-CENH3*/*LexA-CENH3* (67)	+/+ (33) +/*cenh3* (34) *cenh3*/*cenh3* (0)
	+/+ (11)	

[1] Numbers in parentheses show the number of plants of each genotype.

Table 3.4 Transmission of *cenh3* through male and female crosses [1].

Cross	# seedlings	expected WT:het:hom	expected het frequency	observed WT:het:hom	observed het frequency
+/cenh3 ⊗	218	55:109:55	50%	163:55:0	25.2%
+/cenh3 ♀ X B73 ♂	184	90:90	50%	138:46:0	25.0%
B73 ♀ X *+/cenh3* ♂	140	70:70	50%	123:17:0	12.1%

[1] WT indicates wild type, het indicates +/*cenh3* heterozygote, hom indicates *cenh3*/*cenh3* homozygote.

Table 3.5 Haploid and aneuploid induction by +/*cenh3* heterozygotes.

Cross	# seedlings	# glossy plants	# haploid	haploid ratio	# aneuploid	aneuploid ratio
gl8♀X *+/cenh3* ♂	597	3	3	0.5%	0	0
gl8♀ X WT ♂	826	0	-	-	-	-
+/cenh3♀X *gl8* ♂	838	42	42	5.0%	2 [1]	0.2%
WT♀X *gl8* ♂	1000	0	-	-	-	-
+/cenh3♀X *gl1* ♂	844	75	44	5.2%	28	3.3%
WT♀X *gl1* ♂	1114	0	-	-	-	-

[1] The two aneuploid plants are non-glossy plants with stunted phenotypes.

Table 3.6 Results of individual crosses between +/*cenh3* plants and the *gl1* tester.

cross[1]	seeds	seedlings	germination ratio	# glossy plants	# haploid	haploid ratio	# aneuploid	aneuploid ratio
NW222	192	147	76.6%	13	7	4.8%	6	4.1%
NW223	192	174	90.6%	7	2	1.2%	5	2.9%
NW224	192	157	81.8%	19	14	8.9%	5	3.2%
NW225	192	140	72.9%	14	11	7.9%	3	2.1%
NW227	180	117	65.0%	10 [2]	3	2.6%	6	5.1%
NW228	165	109	66.1%	12 [2]	7	6.4%	3	2.8%
Total	1113	844	75.8%	75 [2]	44	5.2%	28	3.3%
WT X *gl1*	1125	1114	99.0%	0	0	0	0	0

[1] NW222, NW223, NW224, NW225, NW227, and NW228 are different ears from the cross +/*cenh3* ♀ X *gl1* ♂.

[2] We were unable to interpret the ploidy level in three glossy plants.

Table 3.7 Primers used in this study.

Primer name	Primer sequence	Purpose
CENH3-F1	TGCAAGATGAGGGCGAATGTG	Genotyping on native*CENH3* and *ImmuneCENH3*
CENH3-R1	TACTTCCTGATCTCCCGCAGC	Genotyping on native *CENH3* but not *ImmuneCENH3*
CENH3-F2	GGCTGCTCTTACTTGCTTGC	Genotyping on native *CENH3* and *ImmuneCENH3*
CENH3-R2	CGCTCTACTTTCCCGTTTGTTAC	Genotyping on native *CENH3* and *ImmuneCENH3*
CENH3-R3	TACTTCCTGATCTCGCGCAGGGCG	Genotyping on *ImmuneCENH3* but not native *CENH3*
Cas9-F1	ACGAGAAGTACCCGACAATCTACC	Genotyping on *Cas9*
Cas9-R1	TGATTTGAAGTTCGGCGTCAGG	Genotyping on *Cas9*
CENH3-F4	CGCTCGAACTGGAGCTTCTT	Genotyping on cenh3::mu
CENH3-R4	AGGTTGGCAGGTAGCCGTTA	Genotyping on cenh3::mu
Mu1	GCCTCTATTTCGTCGAATCCG	Genotyping on cenh3::mu
Mu2	GCCTCCATTTCGTCGAATCCC	Genotyping on cenh3::mu
maize-cenh3-F1	AGGTGACTCAGTTAAGAAGACGAAACC	Genotyping on *MaizeCENH3*
maize-cenh3-R1	GTCCTTTTGCATGATTGTGACACG	Genotyping on *MaizeCENH3*
rice-cenh3-F1	TGGAACATCGGCTTCAGCAG	Genotyping on *RiceCENH3*
rice-cenh3-R1	AGGGTCCACCGTGACACAT	Genotyping on *RiceCENH3*
sorghum-cenh3-F1	AGGAAGCTGCCGCAGAA	Genotyping on *SorghumCENH3*
sorghum-cenh3-R1	CATGACTGTTACGCGCTTGG	Genotyping on *SorghumCENH3*
LexA-F	AGCAAGAGGTGTTCGACCTG	Genotyping on *LexA-CENH3*
LexA-R	CCGAACTTGAGCTGCTTCTT	Genotyping on *LexA-CENH3*

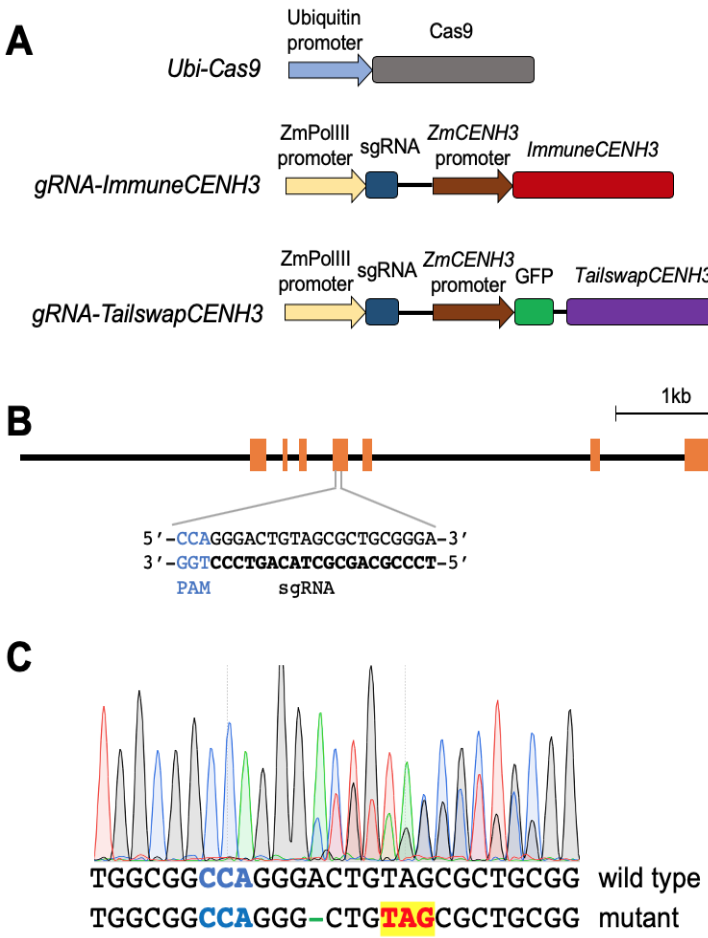


Figure 3.1 Generation of a maize *cenh3* null mutation by CRISPR/Cas9.

(A) Constructs used. Ubi-Cas9 includes a codon-optimized Cas9 driven by the maize polyubiquitin promoter. gRNA-*ImmuneCENH3* includes a gRNA targeting the fourth exon of *Cenh3* and an uncleavable *ImmuneCenH3* gene driven by 2.1 kb of the *Cenh3* native promoter. *TailswapCENH3* is based on *ImmuneCENH3* but includes a modified N-terminal tail and a GFP tag. (B) The maize *Cenh3* gene showing the sequence targeted for gene editing. Exons are shown as orange boxes. The protospacer adjacent motif (PAM) and 20 bp target sequence of the sgRNA are shown. (C) Chromatogram of the sequence from a heterozygous line showing the frameshift in the *cenh3* null mutation. The PAM is in blue, the deletion is in green, and the stop codon is in red.

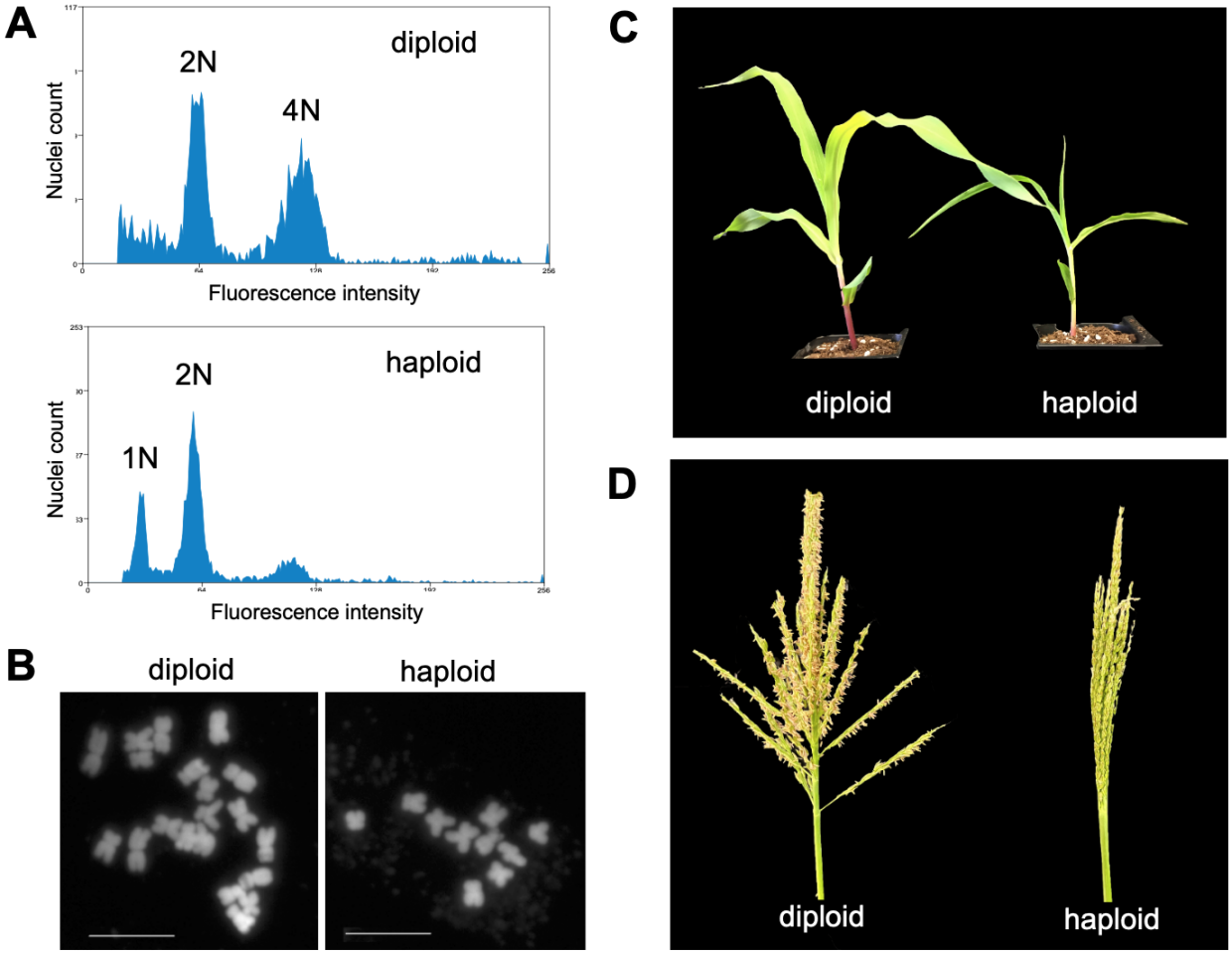


Figure 3.2 Confirmation that plants are haploid.

(A) Flow cytometric analysis of haploids. Diploid plants show peaks at 2N and 4N, where 4N is the result of endoreduplication in differentiated tissues. Haploid plants have 1N and 2N peaks. (B) Chromosome spreads. Maize diploids have 20 chromosomes, whereas haploids have 10. (C) Haploids plants have a shorter stature. (D) Haploid plants are sterile without exerted anthers.

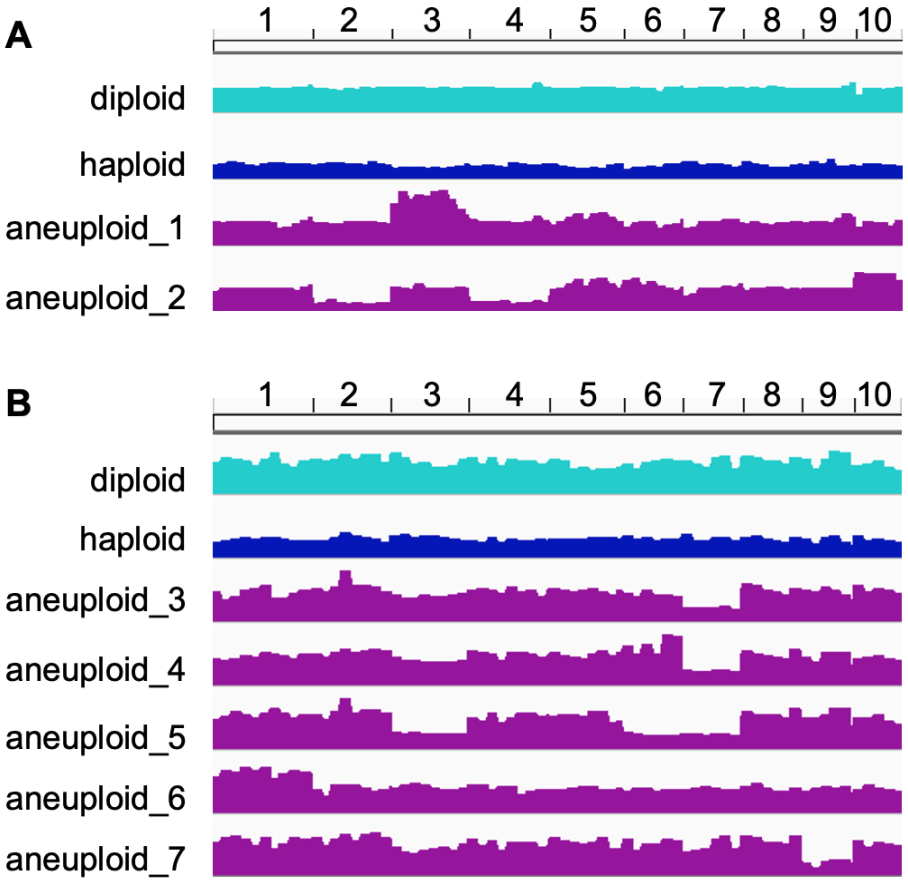


Figure 3.3 Molecular karyotypes of aneuploids.

For both panels, the chromosomes are shown end to end across the top. (A) Aneuploids produced from *gl8* crosses. Aneuploid_1 is trisomic for chromosome 3, and aneuploid_2 is monosomic for chromosome 2 and 4 and trisomic for chromosome 10. (B) Aneuploids produced from *gl1* crosses. Aneuploid_3 and aneuploid_4 are monosomic for chromosome 7, aneuploid_5 is monosomic for chromosome 3, 6 and 7, aneuploid_6 is monosomic for all the chromosomes except chromosome 1, and aneuploid_7 is monosomic for chromosome 9. The coverage in each sample was normalized to the coverage in the relevant diploid from each cross.

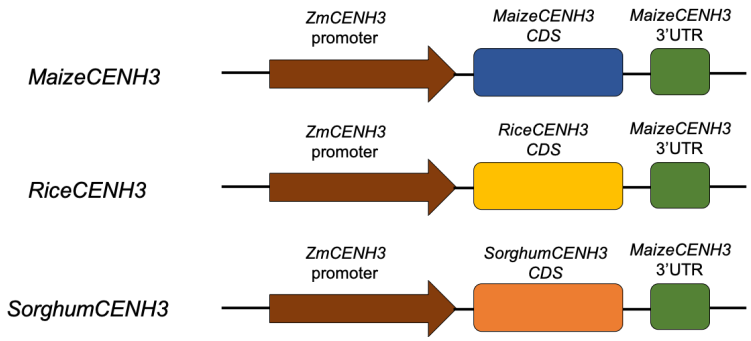


Figure 3.4 Constructs of *CENH3* transgenes.

MaizeCENH3 includes the CDS of Maize *CENH3* gene driven by 3 kb of the *Cenh3* native promoter. *RiceCENH3* includes the CDS of Rice *CENH3* gene driven by 3 kb of the *Cenh3* native promoter. *SorghumCENH3* includes the CDS of Sorghum *CENH3* gene driven by 3 kb of the *Cenh3* native promoter.

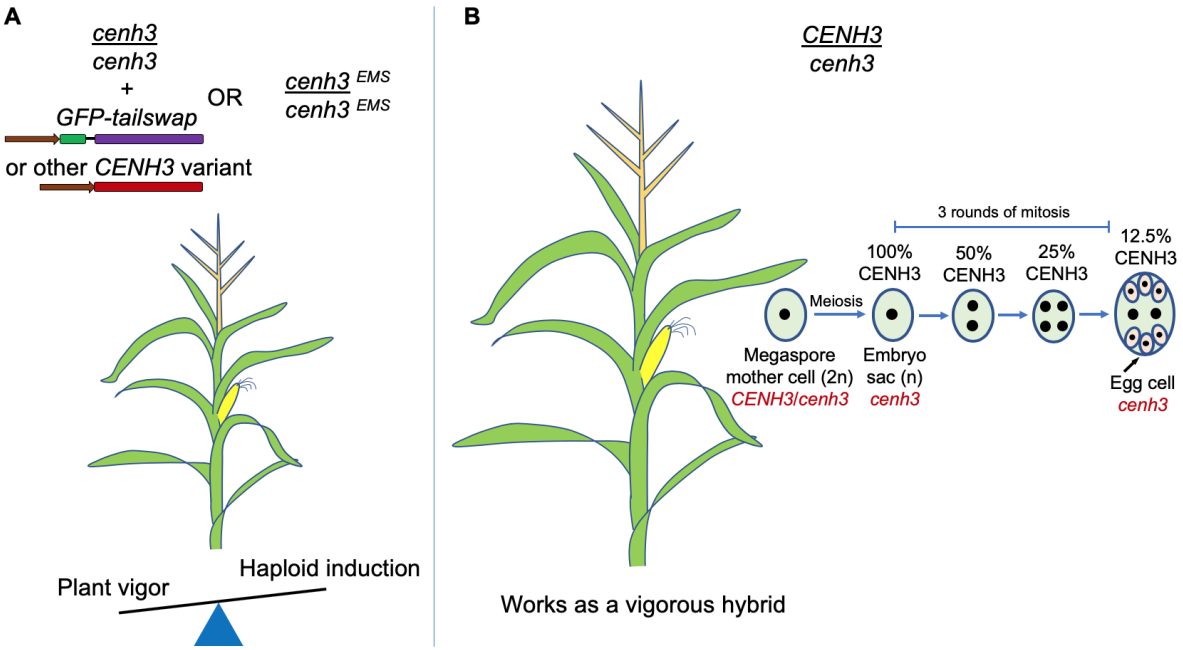


Figure 3.5 Comparison of the GFP-*tailswap*-based method to the *cenh3* null method. (A) The GFP-*tailswap* method and its improved forms. In most applications, a transgene expressing a structurally altered *CENH3*, or other mutant form, is used to complement a loss of function mutation (Ravi and Chan, 2010; Kuppu et al., 2020; Karimi-Ashtiyani et al., 2015; Maheshwari et al., 2015). An EMS-induced point mutation of native CENH3 has also been used (Kuppu et al., 2015). In all cases the plant must grow to maturity with a partially disfunctional *CENH3* gene, which affects plant performance. The most effective haploid inducers are weak plants with poor fertility. (B) The *cenh3* null method. The plant is heterozygous and can be used as a vigorous hybrid. Haploid induction occurs at the gametophyte level. The female gametophyte is shown, where three mitotic cell divisions dilute CENH3 to low levels in the egg cell.

REFERENCES

Barkan, A. and Martienssen, R.A. (1991). Inactivation of maize transposon Mu suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. Proc. Natl. Acad. Sci. U. S. A. **88**: 3502–3506.

Birchler, J.A. (2014). Engineered minichromosomes in plants. Curr. Opin. Plant Biol. **19**: 76–80.

Britt, A.B. and Kuppu, S. (2016). Cenh3: An Emerging Player in Haploid Induction Technology. Front. Plant Sci. **7**: 357.

Chase, S.S. (1969). Monoploids and monoploid-derivatives of maize (Zea mays L.). Bot. Rev. **35**: 117–168.

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

Clarke, J.D. (2009). Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. Cold Spring Harbor Protocols **2009**: db.prot5177–pdb.prot5177.

Coe, E.H. (1959). A Line of Maize with High Haploid Frequency. Am. Nat. **93**: 381–382.

Dawe, R.K. et al. (2018). A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. Cell **173**: 839–850.e18.

Feng, C., Yuan, J., Bai, H., Liu, Y., Su, H., Liu, Y., Shi, L., Gao, Z., Birchler, J.A., and Han, F. (2019). The deposition of CENH3 in maize is stringently regulated. Plant J.

Gilles, L.M., Khaled, A., Laffaire, J.-B., Chaignon, S., Gendrot, G., Laplaige, J., Bergès, H., Beydon, G., Bayle, V., Barret, P., and Others (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J. **36**: 707–717.

Ishii, T., Karimi-Ashtiyani, R., and Houben, A. (2016). Haploidization via Chromosome Elimination: Means and Mechanisms. Annu. Rev. Plant Biol. **67**: 421–438.

Kalinowska, K., Chamas, S., Unkel, K., Demidov, D., Lermontova, I., Dresselhaus, T., Kumlehn, J., Dunemann, F., and Houben, A. (2019). State-of-the-art and novel developments of in vivo haploid technologies. Theor. Appl. Genet. **132**: 593–605.

Karimi-Ashtiyani, R. et al. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proc. Natl. Acad. Sci. U. S. A. **112**: 11211–11216.

Kelliher, T. et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature **542**: 105–109.

Kelliher, T. et al. (2019). One-step genome editing of elite crop germplasm during haploid induction. Nat. Biotechnol. **37**: 287–292.

Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M.L., and Martin, B. (2016). Maternal Haploids Are Preferentially Induced by CENH3-tailswap Transgenic Complementation in Maize. Front. Plant Sci. **7**: 414.

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., and Britt, A.B. (2020). A variety of changes, including CRISPR/Cas9 mediated deletions, in CENH3 lead to haploid induction on outcrossing. Plant Biotechnol. J.

Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B. (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. PLoS Genet. **11**: e1005494.

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

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754–1760.

Liu, C. et al. (2017). A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. Mol. Plant **10**: 520–522.

Liu, C. et al. (2020a). Extension of the in vivo haploid induction system from diploid maize to hexaploid wheat. Plant Biotechnol. J. **18**: 316–318.

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z., Du, L., Pei, X., and Ye, X. (2020b). Efficient induction of haploid plants in wheat by editing of TaMTL using an optimized Agrobacterium-mediated CRISPR system. J. Exp. Bot. **71**: 1337–1349.

Li, Y., Segal, G., Wang, Q., and Dooner, H.K. (2013). Gene tagging with engineered Ds elements in maize. Methods Mol. Biol. **1057**: 83–99.

Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L. (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet. **11**: e1004970.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**: 10.

Paz, M.M., Shou, H., Guo, Z., Zhang, Z., Banerjee, A.K., and Wang, K. (2004). Assessment of conditions affecting Agrobacterium-mediated soybean transformation using the cotyledonary node explant. Euphytica **136**: 167–179.

Ravi, M. et al. (2014). A haploid genetics toolbox for Arabidopsis thaliana. Nature Communications **5**.

Ravi, M. and Chan, S.W.L. (2010). Haploid plants produced by centromere-mediated genome

elimination. Nature **464**: 615–618.

Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M., and Chan, S.W.L. (2011). Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana. PLoS Genet. **7**: e1002121.

Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C., and Cigan, A.M. (2015). Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. Plant Physiol. **169**: 931–945.

Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W. (2015). Catastrophic chromosomal restructuring during genome elimination in plants. Elife **4**.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. **14**: 178–192.

Uliana Trentin, H., Frei, U.K., and Lübberstedt, T. (2020). Breeding Maize Maternal Haploid Inducers. Plants **9**.

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., Wu, M., Sun, T., Cheng, Z., Mercier, R., and Wang, K. (2019). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. Nat. Biotechnol. **37**: 283–286.

Wang, N. and Kelly Dawe, R. (2018). Centromere Size and Its Relationship to Haploid Formation in Plants. Molecular Plant **11**: 398–406.

Weber, D.F. (1994). Use of Maize Monosomics for Gene Localization and Dosage Studies. In The Maize Handbook, M. Freeling and V. Walbot, eds (Springer New York: New York, NY), pp. 350–358.

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., and Kelliher, T. (2018). OsMATL mutation induces haploid seed formation in indica rice. Nat Plants **4**: 530–533.

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

Zhang, H., Phan, B.H., Wang, K., Artelt, B.J., Jiang, J., Parrott, W.A., and Dawe, R.K. (2012). Stable integration of an engineered megabase repeat array into the maize genome. Plant J. **70**: 357–365.

Zhang, H. (2011). Production of a synthetic repeat array system for artificial centromere formation in maize & epigenetic regulation of kinetochore size in plants. Diss. University of Georgia

Zhong, Y. et al. (2020). A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous Arabidopsis. Nat Plants **6**: 466–472.

CHAPTER 4

CONCLUSIONS AND FUTURE STUDY

The primary goal of this study is to explore the centromere size scaling factors and detect the impacts of centromere size on haploid induction. In **chapter 2**, we compared the centromere size across a wide variety of inbred lines, and showed that centromere size is not impacted by centromere sequence but positively correlated with genome size and chromosome size. The relationship of genome size and centromere size was also confirmed by the hybrids between B73 and two larger genome relatives Oaxaca and *Zea luxurians*. Although overexpression of CENH3 in *Drosophila* caused ectopic formation of functional centromere, we found threefold of CENH3 overexpression in maize has no measurable effect on centromere size, indicating CENH3 itself probably is not enough to regulate centromere size. CENH3 may need to incorporate with a chaperone, other kinetochore proteins KNL2 and CENPC to form a complex to control the effective centromere size.

Haploid plants are of great value to crop breeders because they can rapidly accelerate the production of new inbred lines. In **chapter 3**, to explore whether manipulating the centromeric protein CENH3 can induce haploids in maize, I created a *cenh3* null mutant using CRISPR/Cas9, which resulted in a frameshift and premature mutation in *CENH3*. As the homozygous *cenh3* mutant is inviable, I used *cenh3* heterozygotes in this study. Reciprocal crosses between *cenh3* heterozygotes and wild-type plants showed that the *cenh3* null mutation transmits poorly both maternally and paternally. However, haploids were obtained at high efficiency when homozygous wild type plants were crossed with *cenh3* heterozygotes either as male (haploid

induction ratio 0.5%) or female (5.0%). The conclusion of this study is that a simple *cenh3* null mutant is enough to induce haploid.

In this dissertation, I found the centromere size scaling factors and successfully induced haploid in maize. However, there are several future directions that can help us better understand the centromere size regulation network and improve the haploid induction in maize. These future directions are discussed in the following sections.

Explore centromere regulation network in maize

Multiple kinetochore proteins are required for maintaining the function of centromere. Centromere network is well established in *Arabidopsis*, CENH3 localizes within the nucleosome and binds to CENP-C and KNL2 to maintain its accurate recruitment, and $NASP^{SIM}$ escorts non-nucleosomal CENH3 to regulate the level of soluble CENH3 for its deposition (Hori et al., 2017; Sandmann et al., 2017; Le Goff et al., 2020). However, in maize and other crop species, neither KNL2 homolog nor any chaperone has been identified yet. Compared with *Arabidopsis* whose centromeres are all satellite centromere and hard to quantify the size (Hall et al., 2003), maize has seven fully assembled complex centromeres which make the centromere measurement easier than *Arabidopsis* (Liu et al., 2020). Possible directions for the future study are to identify and characterize the KNL2 homolog and CENH3 chaperones in maize and figure out what are the key components in regulating centromere size. Although we overexpressed CENH3 by threefold in this study, the successful incorporation of these excess CENH3 are limited, to identify the limiting factors is of significant importance. Are CENH3 chaperones enough to incorporate new CENH3 or does it require all the inner kinetochore proteins together to work as a complex? We

can overexpress these proteins including KNL2, CENP-C and NASP^{SIM}, to test the key limiting factors for centromere size.

Enhance the haploid induction rates in maize

We proposed a CENH3 dilution model to explain the haploid induction in the *cenh3* null mutant. CENH3 is naturally divided equally between the replicated DNA strands at S phase and replenished later in G2 (Lermontova et al., 2006). As we discussed in **Chapter 3**, although the actual amount of protein will depend on the stability of the mRNA and protein, a reasonable expectation is that *cenh3* sperm would have 25% of the normal amount of CENH3 and an egg carrying the *cenh3* would have no more than 12.5%. Sperm and eggs carrying *cenh3* should have smaller centromeres. These smaller centromeres are easier to degrade and ultimately fail to align on the mitotic spindle, causing chromosome loss and haploid formation (Sanei et al., 2011; Wang and Kelly Dawe, 2018).

The haploid induction capacity of maize Stock6 is associated with eight quantitative trait loci indicating that a potential way to enhance haploid induction ratio is to combine multiple haploid induction-related genes. It is reported that haploid induction is enhanced in maize when combined *ZmDMP* and *ZmMTL* mutation together (Zhong et al., 2019). The haploid induction ratio of the modern Stock6 line is ~10%. Two major quantitative trait loci on *qhir1* and *qhir8* involved in the haploid induction, the single mutant *mtl* in *qhir1* can only induce ~1% haploids, the single mutant in *ZmDMP* in *qhir8* can only induce ~0.15% haploids, however, the haploid induction ratio of the double mutant *mtl*/*Zmdmp* is distinctly increased to ~7% (Zhong et al., 2019). In our study, *cenh3* heterozygous mutants are able to induce haploid at a ratio ~5%. It has been reported that a mutation causing a substitution within the C-terminal region of the

centromeric protein KNL2 is able to induce haploid in *Arabidopsis*, although this mutant is not a null mutant, the haploid induction ratio is about 10% (Lermontova 2019). There are also other centromeric proteins in plants, such as CENPC, KNL2 and NASP^{SIM}, which can bind with CENH3 to maintain its accurate deposition. Once the KNL2 homolog and CENH3 chaperon are characterized in maize, a promising future study is to increase the haploid induction ratio by creating double or triple mutants on CENH3 and its interactive factors. Although the replication of these kinetochore proteins is not as clear as CENH3, we expected the egg and sperm carrying *cenh3/knl2/nasp^{sim}* would have a smaller and weaker centromere than the single mutant cenh3, and the chance to be degraded should be higher. Since kinetochore proteins are essential for plant development, the homozygous *cenh3* mutant is inviable, we expected the homozygous *cenpc*, *knl2* or *nasp^{sim}* mutant might also be lethal. We can utilize the CRISPR-Cas9 system to create the CENP-C, KNL2 and NASP^{SIM} knockout lines. As these mutants may not be able to be transmitted through gametophytes, when we create the null mutant, we would also introduce an *immuneCENPC*, *immuneKNL2* or *immuneNASP^{SIM}* to the embryogenic calli to rescue the null mutant. Since homozygous *knl2* mutant is viable in Arabidopsis (Lermontova et al. 2013), we expect to obtain homozygous *knl2* mutant in maize. However, we expect to obtain the heterozygous *cenpc* mutant or *NASP^{SIM}* mutant, then we would cross these mutants with *cenh3* heterozygous mutants to test the impact on haploid induction and plant development.

Several other factors may also affect the rate of haploid induction. The source germplasm and season have been shown to have a significant effect on haploid induction ratio for tropical maize (Kebede et al., 2011). Besides, the genotype of the donor may also affect the haploid induction ratio. For *indeterminate gametophyte 1*, which is a paternal haploid inducer, the frequency of haploids was affected by the genotype background of the pollinator (Lashermes and

Beckert, 1988; Pollacsek, 1992). All the above factors may also affect the rate of centromere-mediated haploid induction. It's interesting to explore what kind of conditions (including season, source germplasm and genotype) would improve the centromere-mediated haploid induction ratio.

Detecting factors that affect haploid induction

Different mutation types may affect the haploid induction ratio. Liu et al (2017) created a *ZmPLA1* mutant using CRISPR-Cas9 system and targeting the first exon of the gene which lead to an average of ~ 2% haploid induction, while Kelliher et al. (2017) utilize transcription-activator-like effector nuclease (TALEN)-mediated targeted mutagenesis to generate a small deletions nearby the site of the 4-bp insertion in Stock6-derived lines showed an average of ~6.65% haploid induction. The haploid induction ratio of the latter one is much higher than the former one. The reason might lie in the mutated location and mutation types of the gene. In this study, we created two different *CENH3* mutations by CRISPR-Cas9 (Figure 4.1), the first one is a single nucleotide deletion that causes an immediate stop codon in the N-terminal tail of CENH3, the other one is a single nucleotide insertion that cause frame shift and stop codon in the histone fold domain. In **chapter 3**, we only tested the haploid induction in the first mutant, which mutated in the N-terminal tail and removed all the histone fold domain that can interact with DNA and other histones. Further, there are also several different types of point mutations on *CENH3* in the maize EMS mutation database, e.g. mutations cause premature stop codon, mutations in the splice site region and mutations cause nonsynonymous coding changes at different locations (Lu et al., 2018). These different mutation types may induce different ratios of haploid. Therefore, one possible direction of future study is to explore the different haploid

induction ratios among these different mutation types and test the impact of the mutation locations on the haploid induction, which perhaps to better understand the structural function of CENH3.

Applying the CENH3-based haploid induction to gene editing

Genome editing through CRISPR/Cas9 technology is a powerful tool for crop breeding, but because the plant transformation techniques is highly genotype dependent, the application of this technology to the vast majority of crop varieties is limited. The *matl* has been reported to deliver a CRISPR/Cas9 cassette into any inbred background and edit genes in more than 3% of the haploid progeny (Kelliher et al., 2019). As the edited haploids lack both haploid inducer parent genome and the editing cassette, it can be doubled to generate inbred with homozygous mutations in an efficient and non-transgenic way. However, this approach is hard to handle since it requires generating a homozygous *matl* mutant before starting incorporation with gene editing in another genetic background. Further, it can only induce haploid when crossed as a male which only generates haploids with homologous cytoplasmic and nuclear genomes. More importantly, the mechanism how the *matl* mutant causes haploid induction is not clear. The CENH3-based haploid inducer is vigorous and genetically dominant, which will be more efficient than the Stock6 derived lines. Additionally, CENH3-based inducer lines are used as female in the cross that can transfer the nuclear genome of a male parent to a heterologous cytoplasm, which is very useful for creating cytoplasmic male sterile lines. Lastly, the mechanism of centromere-mediated haploid induction is known and the genome elimination happens after zygote fertilization, which ensures the genes, e.g. CRISPR/Cas9 and guide RNA cassette from the haploid inducer parent

can be present in the fertilized zygote. Therefore, applying CENH3-based haploid induction to genotype independent gene editing is very promising.

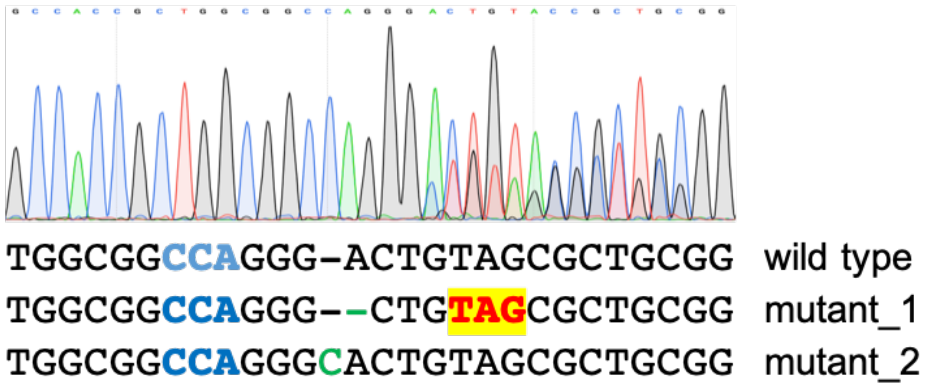


Figure 4.1 *cenh3* mutants generated by the CRISPR/Cas9 system. Chromatogram of the sequence from two heterozygous lines showing the frameshift in the *cenh3* null mutation. Mutant_1 has a single nucleotide deletion mutation causing immediate premature stop codon in the N-terminal tail part of CENH3, and the mutant_2 has a single nucleotide insertion mutation causing stop codon in the histone fold domain. The PAM is in blue, the deletion and insertion is in green, and the stop codon is in red.

REFERENCES

Hall, S.E., Kettler, G., and Preuss, D. (2003). Centromere satellites from Arabidopsis populations: maintenance of conserved and variable domains. Genome Res. **13**: 195–205.

Hori, T., Shang, W.-H., Hara, M., Ariyoshi, M., Arimura, Y., Fujita, R., Kurumizaka, H., and Fukagawa, T. (2017). Association of M18BP1/KNL2 with CENP-A Nucleosome Is Essential for Centromere Formation in Non-mammalian Vertebrates. Dev. Cell **42**: 181–189.e3.

Kebede, A.Z., Dhillon, B.S., Schipprack, W., Araus, J.L., Bänziger, M., Semagn, K., Alvarado, G., and Melchinger, A.E. (2011). Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. Euphytica **180**: 219–226.

Kelliher, T. et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature **542**: 105–109.

Kelliher, T. et al. (2019). One-step genome editing of elite crop germplasm during haploid induction. Nat. Biotechnol. **37**: 287–292.

Lashermes, P. and Beckert, M. (1988). Genetic control of maternal haploidy in maize (Zea mays L.) and selection of haploid inducing lines. Theor. Appl. Genet. **76**: 405–410.

Le Goff, S. et al. (2020). The H3 histone chaperone NASP escorts CenH3 in Arabidopsis. Plant J. **101**: 71–86.

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

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., Demidov, D., Schubert, V., and Schubert, I. (2013). Arabidopsis KINETOCHORE NULL2 Is an Upstream Component for Centromeric Histone H3 Variant cenH3 Deposition at Centromeres. Plant Cell **25**: 3389–3404.

Lermontova, Inna. (2019). Generation of haploid plants based on knl2. U.S. Patent Application No. 15/770,049.

Liu, J. et al. (2020). Gapless assembly of maize chromosomes using long-read technologies. Genome Biol. **21**: 121.

Lu, X. et al. (2018). Gene-Indexed Mutations in Maize. Mol. Plant **11**: 496–504.

Pollacsek, M. (1992). Management of the ig gene for haploid induction in maize. Agronomie **12**: 247–251.

Sandmann, M., Talbert, P., Demidov, D., Kuhlmann, M., Rutten, T., Conrad, U., and Lermontova, I. (2017). Targeting of Arabidopsis KNL2 to Centromeres Depends on the Conserved CENPC-k Motif in Its C Terminus. Plant Cell **29**: 144–155.

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

Wang, N. and Kelly Dawe, R. (2018). Centromere Size and Its Relationship to Haploid Formation in Plants. Molecular Plant **11**: 398–406.

Zhong, Y. et al. (2019). Mutation of ZmDMP enhances haploid induction in maize. Nat Plants **5**: 575–580.