CHARACTERIZING MAJOR KINETOCHORE PROTEINS IN GENUS ZEA SUPPORTS MEIOTIC DRIVE MODEL OF CENTROMERE EVOLUTION

by

YUN LI

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Centromeric "meiotic drive" is a widely discussed model for explaining centromeric evolution. It posits that long arrays of centromeric DNA acquire protein-binding preferences that allow them to segregate more efficiently to progeny. The major centromeric binding proteins are Centromeric Histone H3 (CENH3) and Centromere Protein C (CENPC). We used maize (*Zea mays*) to test whether there is a correlation between lines with very long arrays of centromeric repeats (CentC) and unique CENH3 or CENPC sequence polymorphisms. We were surprised to find no discernable differences in CENH3 or CENPC protein sequence within the Section *Zea*. However, sequencing and expression data demonstrate clear evidence of polymorphism within the promoter region of a key *Cenpc* gene, and that this polymorphism might possibly affect the relative quantities of two CENPC isoforms. Our data seems to indicate that expression polymorphism can change the profile of centromere binding proteins and influence the abundance of centromeric repeats.

INDEX WORDS: Centromere, Meiotic drive, Kinetochore protein, Centromeric DNA, CENH3, CENPC, epigenetic inheritance, Genus Zea

CHARACTERIZING MAJOR KINETOCHORE PROTEINS IN GENUS ZEA SUPPORTS MEIOTIC DRIVE MODEL OF CENTROMERE EVOLUTION

by

YUN LI

B.S, Nanjing University, P.R.China, 2000

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2009

© 2009

Yun Li

All Rights Reserved

CHARACTERIZING MAJOR KINETOCHORE PROTEINS IN GENUS ZEA SUPPORTS MEIOTIC DRIVE MODEL OF CENTROMERE EVOLUTION

by

YUN LI

Major Professor:

Kelly Dawe

Committee:

John Burke Katrien Devos

Electronic Version Approved:

Maureen Grasso Dean of the Graduate School The University of Georgia August 2009

DEDICATION

To my parents, my husband, and my son

ACKNOWLEDGEMENTS

I am indebted to my major professor, Dr. R. Kelly Dawe, for the support, guidance, ideas and encouragements during the entire period of my graduate study for my M.S. at the University of Georgia. To me, Kelly has not only nurtured my growth in science with great enthusiasm and patience, but has also become a big reason of who I am today, with his perspectives and beliefs deeply influenced mine. I would also like to thank my committee members, Dr. John Burke and Dr. Katrien Devos for their valuable suggestions and advice.

I thank the current and previous Dawe lab members: Chris Topp, Yaqing Du, Xuexian Li, Jinghua Shi, Han Zhang, Sarah Rushing, Lisa Kanizay, Brunillis Burgos, Amy Luce, Tom Stahl, Sarah Wolf and Nate Ellis for their selfless support and help over the last four years. I would like to thank Sarah Wolf particularly for her considerable help with verifying PCR results in the promoter regions and RNA. I thank Michael Boyd and the Botany green house staff for taking care of my teosinte plants. The USDA generously provided all the teosinte seeds.

I thank my parents for their endless love and support. I thank my husband for his accompany. I thank all my friends for their friendship during my graduate study in the University of Georgia.

TABLE OF CONTENTS

		Page
ACKNOW	WLEDGEMENTS	V
LIST OF	TABLES	viii
LIST OF	FIGURES	ix
CHAPTE	R	
1	INTRODUCTION AND LITERATURE REVIEW	1
	Kinetochore and Centromere	1
	Purpose of the Study	4
2	MATERIALS AND METHODS	6
	Seed Origins	6
	Polymerase Chain Reactions	7
	Sample Collection and Preparation	7
	Detecting the Ratio Between Cenpc1/3 and Cenpc2 Expression: qRT-PCR	e e
	Analysis	7
	CAPS Assay (Cleaved Amplified Polymorphic Sequence)	8
	Sequencing of PCR Products	9
3	RESULTS	10
	Characterization of CENH3 gene in genus Zea	10
	Characterization of three CENPC homologues in the genome of Zea	11
	1) Discovery of CENPC3 in Zea mays	11

	2) Confirmation of the CENPC3 cDNA Sequence	12
	3) Test the Conservation of three CENPC Homologues in Genus Zea	12
	4) Study of the Pomoter Regions in Three CENPC Homologues Unrav	els Possible
	difference of Expression Regulation of All Three Genes	13
	5) Real-time PCR assay uncovers the relative expression ratio between	n cenpc1/3
	and <i>cenpc2</i>	14
	6) PCR results flanking a region near the 3' end including the 16-nt po	lymorphic
	region and Restriction enzyme digests discovered that Cenpc3 is ra	arely
	expressed in Zea mays B73	15
4	DISCUSSION AND CONCLUSIONS	16
REFERE	NCES	19
APPEND	ICES	23
А	PRIMERS USED IN THIS STUDY	23
В	REALTIME PCR DATA	26
С	PCR RESULTS OF PROMOTER STUDY	

LIST OF TABLES

	Page
Table 1: Seed Origins Used in the Study	6
Table 2: All Available CENPC Homologues from EST Libraries	41
Table 3: Real-time PCR Results	42
Table 4: Efficiencies of Primers Used in Real-time PCRs	43
Table 5: The Information of Primers Used in the Real-time PCRs	

LIST OF FIGURES

	Page
Figure 1: CENH3 have Very Dissimilar N-terminal Tails	28
Figure 2: CENH3 are Generally Conserved in Genus Zea	29
Figure 3: Three CENPC Homologues	30
Figure 4: PCR results Using Primers Specific to Each of Three Homologues	31
Figure 5: The Sequences of CENPC Isoforms	32
Figure 6: Explanation of the 16-nt Polymorphism	
Figure 7: Summary of Promoter Study	34
Figure 8: <i>Cenpc3</i> is rarely expressed in <i>Zea mays</i>	35
Figure 9: NspI restriction enzyme digests confirmed that Cenpc3 is rarely expressed in	Zea mays36
Figure 10: The Model Based on Our Study	
Figure 11: Amplification Curves of Real-time PCR	
Figure 12: Melting Curves of Real-time PCR	
Figure 13: PCR Results of Promoter Study	40

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Kinetochore and Centromere

Proper chromosome segregation during cell division ensures the accurate inheritance of genetic material. The centromeres are loci that allow chromosomes to associate with spindle microtubules and segregate chromosomes to daughter nuclei during cell division. Every chromosome requires a kinetochore, a eukaryotic specific, proteinaceous structure that forms the interface between centromeric DNA and the microtubules that pull the chromosomes to the poles at mitosis and meiosis. It is paradoxical that centromere function and kinetochore proteins are essential and well conserved from yeast to plants to humans (Meraldi *et al.* 2006) yet centromeric DNA sequences display dramatic plasticity across species (Henikoff 2001).

In humans and most other animals centromeres are exclusively composed of long stretches of short tandem repetitive 'satellite' DNA sequences that are the most rapidly evolving components of eukaryotic genomes (Csink and Henikoff, 1998). Plant centromeres contain satellite DNAs comparable to those found in animals, as well as very abundant retrotransposons (Jiang *et al.* 2003). Both forms of repeat vary substantially in abundance and arrangement, both within and among species (Copenhaver *et al.* 1999, Cheng *et al.* 2002, Wu *et al.* 2004, Zhang 2004, Kato *et al.* 2004).

In Zea mays centromeres contain long arrays of a 156-bp satellite repeat called CentC that range in size from 50-2000kb on different chromosomes. Similarly, the amount of satellite DNA in rice centromeres (CentO) ranges from 60kb to 1900kb on different chromosomes (Cheng et al. 2002). Besides varying in copy number, the sequence of centromeric satellite repeats differs remarkably among organisms, even in closely related species (Malik and Henikoff 2002, Lamb et al. 2004, Jiang et al. 2003, Henikoff and Dalal 2005). For example, the CentO repeats in rice are absent in Oryza brachyantha, a species of wild rice that last shared a common ancestor with rice only ~7-9 million years ago (Lee et al. 2005, Dawe 2005). It is suggested that these repeats are crucial for centromere function, considering their specificity to the centromere loci. Centromere composition also shows dramatic variance with respect to Centromeric Retrotransposons (CRs), such as CRRs (in rice) (Cheng et al. 2002) and CRMs (in maize) (Zhong et al. 2002). CRs belong to the Ty3/Gypsy long terminal repeat (LTR) retrotransposon family (Nagaki et al. 2004, Ma and Bennetzen 2006). Highly conserved motifs were found in the LTRs of the CR elements from rice (Oryza sativa), maize (Zea mays), and Hordeum vulgare (barley) (Nagaki et al. 2003a), suggesting they may play a critical role in centromere function.

Whereas centromeric DNAs change rapidly, several proteins specific to the centromere/kinetochore complex display high conservation. Among these are two important inner kinetochore "foundation proteins" that have been shown to interact with centromeric DNAs- Centromeric Histone H3 (CENH3) and Centromeric Protein C (CENPC). CENH3 is a centromere-specific histone H3-like protein that replaces the regular histone H3 and interspersed with blocks of histone H3-containing nucleosomes in centromeric chromatin (Yoda *et al.* 2000). First discovered in humans as CENP-A (centromeric protein A) (Palmer *et al.* 1987, 1991), CENH3 has been found in all model eukaryotes (Henikoff *et al.* 2001). CENH3s are similar to

histone H3 in their histone folding domains but differ in their noncanonical NH2-terminal tails even from each other. Compared to histone H3, CENH3 also has a slightly longer 'loop 1' region in the histone fold (Shelby *et al.* 1997, Malik *et al.* 2001, Malik *et al.* 2002). Both diverging regions contact nucleosomal DNA (Vafa and Sullivan 1997), and Malik and Henikoff (2002) provided evidence that the NH2-terminal tail and the loop 1 region were under adaptive evolution and suggested that CENH3 serves as a linker molecule between the rapidly evolving centromeric DNA and the conserved kinetochore proteins (Malik *et al.* 2002, Dawe *et al.* 2002). In plants, an interaction between defined centromeric sequences has been demonstrated for maize (Zhong *et al.* 2002), Arabidopsis (Nagaki *et al.* 2003a), rice (Nagaki *et al.* 2004), and sugarcane (Nagaki and Murata 2005).

CENPC has been shown to be necessary to form a functional centromere in human, mouse and chicken (Sullivan *et al.* 1994, Fukagawa and Brown 1997, Kalitsis *et al.* 1998). It is also conserved across all eukaryotes though the sequence homology is limited to a mere 23 amino acids (Brown, 1995; Henikoff *et al.* 2001). In plants, two CENPC homologs were first identified in maize (Dawe *et al.* 1999). Mammalian and plant CENPC proteins contain adaptively evolving regions that overlap with regions of DNA-binding activity (Talbert *et al.* 2004). Particularly in grasses, exon pairs 9-10 and 11-12 have been subjected to ancient duplications (Talbert *et al.* 2004) and bind to centromeric DNA (Du 2008). CENPC is present at centromeres throughout the cell cycle, and is necessary but not sufficient for kinetochore assembly (Fukagawa *et al.* 1999).

Many observations from animals, plants and yeast indicate that centromeric repeats are dispensable for centromeric function and that epigenetics plays an essential role in mediating centromere function (Henikoff 2001, Choo 2001 and Dawe 2005). The paradox between rapid divergence of centromeric DNA and the conserved kinetochore proteins can be explained by the

"centromeric drive" hypothesis (Malik and Henikoff, 2002). During female meiosis in both plants and animals, centromeres are presumed to compete for access to the single meiotic product that is passed on to the next generation. This process can promote the evolution of selfish repeats that bind tightly to "foundation proteins" and in principle can drive rapid change in the size of repeat arrays (Dawe and Henikoff, 2006). However, in light of the equal likelihood of inheritance of parental chromosomes in Mendelian genetics, centromere drive would be detrimental. A mutation that changes the structure of the foundation protein would disrupt meiotic drive and restore the centromere to a more stable epigenetic state (Dawe and Henikoff, 2006). Therefore, centromeric DNA would be selected for preferential segregation properties while "foundation proteins" would be selected for Mendelian properties. The "centromeric drive" model can explain most of the available data from centromeres. Nevertheless the only strong support to date is the fact finding that both CENH3 and CENPC are both under positive selection. Empirical evidence for preferential segregation of centromeres are lacking and it is not established that either protein binds to specific DNA sequences (Talbert et al. 2002, Malik and Henikoff 2001, Dawe and Henikoff 2006).

Purpose of the Study

Direct tests of the centromere drive hypothesis will require genetic tests in a welldeveloped model system such as maize and its allied species. There are four species in the genus *Zea* divided into two sections. Section *Luxuriantes* consists of the annual *Z. luxurians*, the protected perennial *Z. diploperennis* and its autotetraploid derivative *Z. perennis*. Both *Z. luxurians* and *Z. diploperennis* can be crossed with *Zea mays*. The Section *Zea* is further divided into four subspecies: ssp. *huehuetenangensis*, ssp. *mexicana*, ssp. *parviglumis* and ssp. *mays* (Doebley 1990). *Z. mays* ssp. *mays* is maize, the domesticated species, and the other three are its wild relatives. *Z. mays* ssp. *parviglumis* is generally considered to be the direct progenitor of ssp. *mays* (Doebley 2004).

If CentC is under positive selection, we expect CentC arrays to be more abundant and consistent among chromosomes. Recent studies have suggested that CentC is indeed very rich in some subspecies of Section Zea, including Z. diploperennis (Lamb, Meyer et al. 2007), Z. *luxurians* and Z. mays ssp. parviglumis (Shi 2009), but not in Zea mays. Assays were made in F1 individuals that were heterozygous for species of interest and maize (providing a B73 internal control for CentC labeling). It was found that Z. *luxurians* CentC is strikingly uniform across chromosomes and much more abundant than in B73. Assays of ssp. parviglumis revealed that this subspecies, too, is rich in CentC (Shi 2009). The latter observation was particularly surprising since Z. parviglumis is presumed to be a direct ancestor of maize. These data suggest that centromere drive that promotes rapid change in the size of repeat arrays is actively occurring in some subspecies of Section Zea but not others..

The meiotic drive hypothesis further predicts that variation in centromeric DNA content will be associated with unique inner kinetochore proteins. Here we tested whether maize lines with very long CentC repeat arrays show unique CENH3 or CENPC sequence polymorphisms. We were surprised to find no discernable differences in CENH3 protein sequence within the Section *Zea*. However, we find clear evidence of sequence polymorphism within the promoter region of a key *Cenpc* gene (*Cenpc3*). We argue that the ratio between two different CENPC isoforms, CENPC1/3 and CENPC2 has been altered by promoter mutations, and that the two isoforms have different impacts on the abundance of CentC at centromeres: one isoform promotes centromere drive while the other suppresses it.

CHAPTER 2

MATERIALS AND METHODS

Seed Origins

Plant seeds of maize ancestors were obtained from USDA-ARS (Table 1).

Table 1.	Seed	origins	used	in	the	study.	
		0				2	

Species	Accession No.	Origin
Z. mays. ssp.	PI566685	Mexico, Mexico
mexicana		
Z. mays. ssp.	PI 441934	Guatemala,
huehuetenangensis		Huehuetenango
Z. mays. ssp.	PI 566687	Mexico, Mexico
parviglumis		
Z. luxurians	PI 422162	Mexico
Z. diploperennis	PI 462368	Mexico, Jalisco
Z. perennis	Ames 21869	Mexico

Polymerase Chain Reactions

Polymerase Chain reactions (PCRs) were performed with either Extaq Polymerase (TaKaRa) or AccuTaq (Sigma) in a total of 100uL volume per reaction, which contains Extaq (or Accutaq) 0.5uL, 10Xbuffer 10uL, 2.5uM dNTP 8uL, template (<=500ng/uL) 2uL, 2.5uM forward primer 10uL, 2.5uM reverse primer 10uL and deionized, doubly-distilled water (dd-water) 59.5uL. The reactions were run in the Eppendorf AG 22331 Hamburg machine or Eppendorf Mastercycler personal using the thermocycling conditions below: 95 °C for 3 min for initial penetration and activation of *Taq* polymerase, followed by 40 thermal cycles of 95 °C for 30 s, 55 °C (Annealing Temperature=(T_m of forward primer + T_m of reverse primer)/2 -5 °C) for 30 s, and 72 °C for 1 min per 1000bp of the amplicon length.

Primer sequences for each reaction are listed in Appendix A.

Sample Collection and Preparation

DNAs were prepared using the DNeasy kit (Qiagen) from approximately 4-week-old leaf tissues. Total RNAs were extracted from approximately 4-week-old leaf tissues using Trizol (Invitrogen) method. Total RNAs were reverse transcribed using the Superscript III first-strand synthesis kit (Invitrogen) with Oligo dT to produce cDNA after DNase treatment. cDNAs were treated by RNaseH before being used in PCR reactions.

Detecting the ratio between Cenpc1/3 and Cenpc2 expression: qRT-PCR Analysis

Quantitative RT-PCRs were used to analyze cDNA expression levels using ubiquitin universal primers as the endogenous control. Real-time PCR was performed on a Roche LightCycler 480 Real-Time PCR System using SYBR Green detection chemistry. Experiments are average relative quantities from at least two biological replicates. Primer sequences for realtime reactions are listed in Appendix A.

Real-time PCR and melting analysis were performed on a Roche LightCycler 480 Real-Time PCR System using SYBR Green detection chemistry. PCR was done in Roche Real-time PCR plates that contained 15 uL of reaction mixture. Each 15 uL of reaction mixture contained: 0.45uL of each primer in 2.5uM concentration, 2.25 uL of 25mM MgCl₂, 1.5 uL of 2.5uM dNTP, 1.5uL of 10X PCR Gold Buffer, 1.5 uL of 50% DMSO, 0.15uL AmpliTaq Gold with GeneAmp Taq, 0.113 SYBR Green, 3.0uL template and 4.09uL deionized, doubly-distilled water (dd-water).

The thermocycling conditions for amplification were as follows: 95 °C for 10 min for initial penetration and activation of *Taq* polymerase, followed by 45 thermal cycles of 95 °C for 10 s, 60 °C for10 s, and 72 °C for 20 s. Fluorescence was measured during each 60 °C stage. After amplification, the instrument performed a melting analysis by heating the capillary at 95 °C for 5 s, incubating it at 55 °C for 1 min, and then slowly heating it to 97 °C. Fluorescence was monitored continuously during the melting experiment. To convert melting curves to melting peaks, the LightCycler software (Roche Molecular Biochemicals) calculated the negative derivative of each measured fluorescence with respect to the temperature (dF/d*T*), and then plotted dF/d*T* against temperature for the entire melting experiment. Primer sequences for all reactions are listed in Appendix A.

Restriction Enzyme Digests

Restriction enzyme digests were performed in PCR plates that contained 20 uL of reaction mixture. Each 20 uL of reaction mixture contained: 2ul of buffer, 10uL of PCR product,

1uL of restriction NEB enzyme, 0.2 uL of BSA for NspI and deionized, doubly-distilled water (dd-water) to reach the volume to 20 uL. The mixture was incubated in a PCR machine at 37°C for 1 hour, then 20 min of inactivation at 65°C for NspI.

Sequencing of PCR products

100 uL of PCR products were purified using QIAquick PCR purification kit (QIAGEN), diluted to 20ng/uL and sent to AgenCourt Sequencing facility for sequencing.

CHAPTER 3

RESULTS

Characterization of CENH3 gene in genus Zea

Human CENP-A and its counterparts known as Centromeric H3 (CENH3) in other organisms are known to bind DNA and organize kinetochore assembly (Choo *et al.* 2004). The assembly in most if not all other kinetochore proteins requires the presence of CENH3. For instance CENPC localization depends on the presence of CENH3, but not vice versa (Moore and Roth 2001, Choo *et al.* 2000, Hyman *et al.* 2001, Brinkley *et al.* 2001, Barnes *et al.* 2003). As a direct DNA binding protein CENH3 is a particularly inviting candidate for mediating centromere drive. As first proposed by Henikoff *et al.* (2001) and Malik and Henikoff (2002), mutations in CENH3s could have a profound effect on centromere repeat abundance and centromere drive.

Therefore maize CENH3 was the first candidate for our study. As the most rapidly evolving region of CENH3 is the highly divergent N-terminal tail region (Malik and Henikoff 2002), we chose this region to sequence and analyze (Fig.1). Surprisingly, we found that the N-termini of *Cenh3* genes in maize, *Z. luxurians, Z. diploperennis, Z. perennis, Z. huehuetenangensis, Z. parviglumis, and Z. mexicana* are very similar at the nucleotide and protein levels (Fig 2).

Characterization of three CENPC homologues in the genomes of Zea

1) Discovery of CENPC3 in Zea mays

Like CENH3, CENPC is essential for kinetochore activity. It is also conserved across all eukaryotes though sequence homology is limited to a mere 23 amino acids (Brown 1995, Henikoff *et al.* 2001). Besides CENH3, CENPC is the only other kinetochore protein that has been shown to possess definite DNA binding characteristics, although it binds in a sequence-independent manner (Sugimoto *et al.* 1994, Yang *et al.* 1996, Sugimoto *et al.* 1997, Dawe *et al.* 1999). Two CENPC homologues have been identified in maize (*Cenpc1* and *Cenpc2*) and they share 84% homology, with single-nucleotide changes as well as small insertion/deletion polymorphisms throughout the sequence (Dawe *et al.* 1999).

Using the protein sequence of maize *Cenpc1* as a query, we identified three high throughput genomic sequences (And their GenBank accession numbers are AC212807.4, AC191082.3 and AC190975.3). AC212807.4 and AC191082.3 are separately located on chromosome 3 and chromosome 8, and *Cenpc1* and *Cenpc2*. In addition, a previously unknown CENPC gene located on chromosome 1 was also identified. This gene, later named *Cenpc3*, shares some 97% homology with *Cenpc1* in the genomic DNA, with a major insertion in the intron between exon 11 and 12. A comparison of the predicted protein sequence suggested that CENPC3 is similar if not identical to CENPC1. The predicted cDNAs are also nearly identical, with a 16nt-deleton in the 3' UTR region of *Cenpc1* as the only major difference at the nucleotide level.

In order to find out whether *Cenpc3* also exists in other *Zea* species, we designed several primer pairs that specifically amplify *Cenpc3* genomic DNA and found that *Cenpc3* exists in all *Zea* species, including *Zea luxurians* and *Zea diploperennis*.

To further verify that all three CENPC homologues are present throughout Zea, primers that are specific to the genomic DNAs of *Cenpc1* and *Cenpc3*, and the cDNAs of *Cenpc1* and *Cenpc2* (Fig. 3A) were designed. PCR results verified that all three homologues are present in the genome (Fig. 3B, Fig. 3C and Fig. 4) and that both *Cenpc1/3* and *Cenpc2* are expressed (Fig.4) in each species in genus Zea.

2) Confirmation of the Cenpc3 cDNA sequence

As described above, comparison of the available sequence suggested that cDNA sequences of *Cenpc1* and *Cenpc3* are nearly identical with a 16nt-deletion at the 3' UTR region in *Cenpc1* as the only major difference at the nucleotide level. As these data were derived from predicted exon/intron junctions and draft sequence, we prepared full-length cDNAs using the 16nt deletion. Sequencing of the full-length cDNAs confirmed that the predicted CENPC3 protein sequence is identical to CENPC1.

Additional sequencing from other *Zea* lines further revealed that the defining 16 nt deletion in *Cenpc1* is unique to maize. The sequences from 3' UTRs were obtained from PCR products using *Cenpc1*-specific and *Cenpc3*-specific primers flanking the 3' UTR region (Fig. 3A). In *Z. luxurians, Z. diploperennis, Z. perennis, Z. ssp. huehuetenangensis* and *Z. ssp. parviglumis*, the *Cenpc1* and *Cenpc3* 3' UTR genomic DNAs are 100% identical and indiscernible (Fig.6).

3) Conservation of three CENPC homologues in genus Zea

Under the meiotic drive model, mutation of a major kinetochore protein disrupts sequence-specific interactions between centromeric repetitive DNA and kinetochore foundation

proteins, restoring epigenetic inheritance. As our study of CENH3 seems to exclude the possibility that CENH3 plays a major role in this process, we went on to test whether CENPC protein sequences vary among *Zea* species.

The DNA binding domain of maize CENPC lies between exons 9 and 12 (Du, Topp and Dawe submitted). Primers were designed flanking exon 9-12 of all three homologues and subsequent PCR products from each *Zea* species were sequenced. As with CENH3, we observed that CENPC1/3 and CENPC2 are nearly invariant among *Zea* species (Fig.5).

4) Promoter regions in three CENPC homologues unravels possible difference of expression regulation of all three genes

The fact that maize contains two different isoforms of CENPC raises interesting questions relative to the centromere drive hypothesis. As the two isoforms differ substantially in the exon 9-12 region, it is likely that they differ in their DNA binding characteristics. Furthermore, it is likely that one or the other isoform has a more dominant role in any given tissue, and by extension, it is quite possible that the expression levels vary among *Zea* species.

Therefore we developed the new hypothesis that differential expression among the three *Cenpc* genes has impacted the abundance of CentC in *Zea*. The fact that *Z. mays* ssp. *Parviglumis, Z. luxurians* and *Z. diploperennis* have more abundant CentC may be explained by differences in CENPC isoform abundance in the tissue that produce gametes.

Sequence comparisons reveal that in B73, *Cenpc1* and *Cenpc3* share no homology upstream of position -193 (Fig. 3A). In order to find out whether such a difference also exists in other *Zea* species, especially *Zea luxurians* and *Zea diploperennis*, primers were designed in the promoter region using *Cenpc1* and *Cenpc3* sequences of *Zea mays*. PCR assays were performed

on genomic DNAs of *Z. mays*, *Z. luxurians* and *Z. diploperennis*. Product sizes were compared in the three species as a means to test the sequence conservation in light of major indels (insertions or deletions). As was expected, promoter regions of *Cenpc1* between *Z. mays* and *Z. diploperennis* showed no indel till 2450bp upstream from the start codon (Fig. 7 and Appendix B). However, primers specific to the *Cenpc3* promoter upstream of -193 showed no specific amplification (Appendix B). These data demonstrate a major difference between *Z. mays* B73 and its relatives *Z. luxurians* and *Z. diploperennis* – the *Cenpc3* promoter is not conserved. *Zea mays* B73 has a novel promoter insertion immediately upstream of the start codon that might be expected to dramatically change gene expression.

5) Real-time PCR assay uncovers the relative expression ratio between cenpc1/3 and cenpc2

Our assays of promoter variation prompted us to consider where in the plants the three genes are expressed. Although microarray data are not yet available for maize, we can estimate the expression profile by searching the maize EST and cDNA database (Table 2). The data show that all three homologues are expressed in *Zea mays*. *Cenpc2* seems to be widely distributed throughout all tissues and lines, while *Cenpc1* and *Cenpc3* seem to be more tissue-specific and line-specific.

To investigate the effects of *Cenpc3* polymorphism on gene expression, real-time PCR assays were performed using *Cenpc1/Cenpc3*-cDNA-specific primers and *Cenpc2*-specific primers, using cDNAs from 1 month old leaves (Table 3). Since different primers will have different efficiencies (Table 4) according to their position in the gene, their amplicon length, their T_ms , etc., four sets of primers were designed in the 5' end and 3' end of each gene separately, sharing almost identical T_ms and lengths of amplicons. The average C(t) (Cycle

Threshold) of two primer sets were used in comparison (See Appendix C, Table 5). Results reveal that *Cenpc1/3* is expressed at a higher level than *Cenpc2*, ranging from 6.02 (2 $^{2.59}$) to 56.89 (2 $^{5.83}$) fold more expression using primer pairs near the 5', and from 5.17 (2 $^{2.37}$) to 23.43 (2 $^{4.55}$) fold more expression using primer pairs near the 3'. Real-time PCR results using both sets of primers suggest that the ratio of *Cenpc1/3* to *Cenpc2* is quite different between *Z. mays* and *Z. parviglumis* at 8.07 and 36.64, respectively. Some difference was also observed between *Z. mays* and *Z. luxurians* at 2.37 and 5.64, and between *Z. mays* and *Z. diploperennis* at 6.22 and 1.71. These data suggest that the novel *Cenpc3* promoter insertion observed in maize might possibly have altered relative ratio of CENPC isoforms in maize.

We note that these data were collected in vegetative tissue, which is not necessarily relevant to the germ line lineages that produce gametes. More detailed expression profiling in germline tissue will further enrich our understanding of the impact of the *Cenpc3* promoter insertion on the evolution of maize centromeres.

6) Cenpc3 is rarely expressed in Zea mays B73

Our PCR assays and subsequent restriction enzyme digests flanking a region containing the 16-nt polymorphic region near the 3' end confirmed that *Cenpc3* is barely expressed in *Zea mays* (Fig. 8 and Fig. 9).

The line we used for *Zea mexicana* happened to be a heterozygote between *Zea mays* and *Zea mexicana*, and PCR results verified that it shows both the polymorphism of maize *Cenpc1* and *Cenpc3* (Fig. 8 and Fig. 9). Sequencing result of its genomic DNA in the 3' end confirmed that it also has a *Cenpc1* copy from *Zea mexicana*, which contains the 16-nt polymorphic region (Fig. 6).

CHAPTER 4

DISCUSSION AND CONCLUSIONS

The meiotic drive hypothesis has been proposed to be a major force in centromere evolution (Henikoff *et al.* 2001). Previous studies by Jinghua Shi in the Dawe lab (2009) established that CentC abundance varies dramatically in *Zea*, with maize having the least CentC of all subspecies. It was proposed that CentC was recently under selection, perhaps by meiotic drive, but that in cultivated maize the dynamics have shifted such that CentC is no longer favored. Following the models of Henikoff and colleagues, we pursued the prediction that major kinetochore proteins CENH3 and CENPC would show similar or corresponding levels of polymorphism.

We find that the single gene encoding maize CENH3 shows no significant divergence among *Zea* subspecies, suggesting that CENH3 has not participated in centromere drive. However, the story with CENPC is more complex and interesting. New sequence data revealed that there are three genes encoding CENPC, *Cenpc1*, *Cenpc2* and *Cenpc3*. The two major isoforms of CENPC, CENPC1/3 and CENPC2, sharing 84% homology, differ substantially in the DNA binding domain. Talbert and Henikoff found that CENPC1 and CENPC2 have been subjected to different selective forces since their divergence (Talbert *et al.* 2004). In contrast, CENPC1 and CENPC3 have diverged very recently and are very similar. Although no discernable difference could be observed in the coding sequences of CENPC1 and CENPC3, our data show that *Cenpc3* displays much less conservation in the promoter region in *Zea mays* B73. Strikingly, the *Cenpc3* promoter in maize B73 differs from the *Cenpc3* promoter in all other species.

The above findings prompted us to consider that altering the regulation of CENPC isoforms such that one is expressed at a higher level might have a similar outcome as mutations of protein sequences. If, for instance, CENPC2 is the major isoform in germ lineages of one species, we can imagine that a shift in CENPC1/3 expression could reduce the effective abundance of CENPC2. Although our data do not prove this scenario, they are entirely consistent with it. We show that maize B73 has a novel Cenpc3 promoter that is not present in any other Zea species. In contrast, we found Cenpc1 promoter in maize B73 is much more conserved in Zea. Preliminary expression data from leaves suggest that there is indeed extensive variation in the relative quantities of Cenpc1/3 and Cenpc2 mRNA among Zea allies. PCR assays and subsequent restriction enzyme digests confirmed that Cenpc3 is barely expressed in Zea mays B73. These data supports our model. However, the real-time data remain inconclusive, and as yet do not address the key questions of whether the mRNA levels vary in the tissues that produce gametes. We anticipate that new resources, such as detailed expression profiles in different Zea lines that are likely to be produced in coming years, will help us more thoroughly in addressing this issue.

We also demonstrate that maize B73 *Cenpc3* is unique in lacking a 16nt deletion that is present in *Cenpc1*. The difference between the *Zea mays Cenpc1* and *Cenpc3* cDNAs could possibly affect the translation level of CENPC homologues and also lead to different ratios of CENPC isoforms. There are numerous examples of mutations in the 3' UTR regions of genes

that affect translation (Shanping Wang *et al.* 1997, Bailey-Serres *et al.* 1999, Rita-Ann Monde *et al.* 2000) but more solid evidence is needed to make the statement that the 16-nt difference affects translation in maize.

The data described here suggest that two CENPC isoforms are expressed at different levels in the genus *Zea*, and could possibly indicate that polymorphism in major kinetochore proteins are associated with variation in the abundance of centromeric repeat arrays (Fig. 10). We speculate that by regulating different isoforms of CENPC, meiotic drive has put CentC under selection, reducing the abundance of CentC in cultivated maize, compared to its ancestors.

REFERENCES

Bailey-Serres, J. (1999). "Selective translation of cytoplasmic mRNAs in plants." Trends Plant Sci . 4, 142–148. .

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

Cheng, Z. *et al.*, Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon, Plant Cell 14 (2002), pp. 1691–1704.

Choo, K.H. Domain organization at the centromere and neocentromere, Dev. Cell 1 (2001), pp. 165–177.

Copenhaver, G.P. *et al.*., Genetic definition and sequence analysis of Arabidopsis centromeres, Science 286 (1999), pp. 2468–2474.

Csink, A.K., and S. Henikoff. 1998. Something from nothing: the evolution and utility of satellite repeats. Trends Genet. 14:200–204.

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

Dawe, R.K. (2005). Centromere renewal and replacement in the plant kingdom. Proc. Nat. Acad. Sci. USA 102: 11573-11574.

Dawe, R.K. and Henikoff S. 2006. Centromeres put epigenetics in the drivers seat. Trends Biochem. Sci. 31, 662-669

Doebley, J. (2004). "The genetics of maize evolution." Annu Rev Genet 38: 37-59.

Doebley, J. F. (1990). "Molecular systematics of Zea (Gramineae)." Maydica 35: 143-150.

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

Henikoff, S. *et al.* (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293, 1098–1102

Henikoff, S., Dalal Y (2005) Centromeric heterochromatin: what makes it unique? Curr Opin Genet Dev 15: 177–184

Jiang, J. et al., A molecular view of plant centromeres, Trends Plant Sci. 8 (2003), pp. 570-575.

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

Lamb, J. C., J. M. Meyer, *et al.* (2007). "Distinct chromosomal distributions of highly repetitive sequences in maize." Chromosome Res 15(1): 33-49.

Lee, H.R. *et al.*, Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species, Proc. Natl. Acad. Sci. U. S. A. 102 (2005), pp. 11793–11798.

Ma, J. and Bennetzen, J.L. Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice, Proc. Natl. Acad. Sci. U. S. A. 103 (2006), pp. 383–388.

Malik, H.S., and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157, 1293–1298.

Malik H.S., Henikoff S, Conflict begets complexity: the evolution of centromeres, Current Opinion in Genetics & Development, 12(6), 711-8, December 2002

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

Moore, L. L. and M. B. Roth (2001). "HCP-4, a CENP-C-like protein in Caenorhabditis elegans, is required for resolution of sister centromeres." J Cell Biol 153(6): 1199-208.

Nagaki, K. *et al.*, Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres, Genetics 163 (2003), pp. 759–770.

Nagaki, K. *et al.*, Sequencing of a rice centromere uncovers active genes, Nat. Genet. 36 (2004), pp. 138–145. Paul B Talbert, T. D. B. a. S. H. (2004). "Adaptive evolution of centromere proteins in plants and animals." Journal of Biology 3:18.

Nagaki, K., and Murata, M. (2005). Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. Chromosome Res. 13, 195–203.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. USA 88, 3734–3748.

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

Rita-Ann Monde, Greene J.C."Stern D.B (2000). "The sequence and secondary structure of the 3'-UTR affect 3'-end maturation, RNA accumulation, and translation in tobacco chloroplasts." Journal of Plant Molecular Biology 44(4): 529-542.

Shelby, R.D., Vafa, O., and Sullivan, K.F. (1997). Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J. Cell Biol. 136:501-513

Shi, J. (2009). "PhD thesis: Crossing over in pericentromere regions plays a role in centromere evolution.".

Sugimoto, K., Yata, H., Muro, Y., and Himeno, M. (1994). Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. Journal of Biochemistry 116, 877-881.

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

Sugimoto, K., Kuriyama, K., Shibata, A., and Himeno, M. (1997). Characterization of Internal DNA-binding and C-terminal Dimerization Domains of Human Centromere/Kinetochore Autoantigen CENP-C in vitro: Role of DNA-binding and Self-Associating Activities in Kinetochore Organization. Chromosome Research 5, 132-141.

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

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

Vafa O., Sullivan K.F. (1997) Chromatin containing CENP-A and a-satellite DNA is a major component of the inner kinetochore plate. Curr Biol 7:897–900

Wang S., Browning K.S., Miller W.A. (1997). "A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA " The EMBO Journal 16(13): 4107-4116.

Wu, J. *et al.*, Composition and structure of the centromeric region of rice chromosome 8, Plant Cell 16 (2004), pp. 967–976.

Yang, C.H., Tomkiel, J., Saitoh, H., Johnson, D.H., and Earnshaw, W.C. (1996). Identification of overlapping DNA-binding and centromere-targeting domains in

the human kinetochore protein CENP-C. Mol Cell Biol 16, 3576-3586.

Du, Y. (2008). "PhD thesis: Mechanism of kinetochore assembly in maize.".

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc. Natl. Acad. Sci. USA 97, 7266–7271.

Zhang, Y. *et al.*., Structural features of the rice chromosome 4 centromere, Nucleic Acids Res. 32 (2004), pp. 2023–2030.

Zhong, C.X. *et al.*., Centromeric retroelements and satellites interact with maize kinetochore protein CENH3, Plant Cell 14 (2002), pp. 2825–2836

APPENDICES

A) PRIMERS USED IN THIS STUDY

Primers for CENH3 amplification (See Figure 1)

CENH3_F2 5' GCACC(C/A)GGC(C/G)GTGAGGAA 3' CENH3_F4 5' AGCC(C/G)AAGAAGAAGCTCCAG 3' CENH3_R3 3' TTC CTG ATC TCC CGC AGC GC 5' CENH3_R4 3' TAC AGT CCC TGG CCG CCA GC 5'

Primers for sequencing CENPC1 3' genomic DNAs (See Figure 3A and 3B)

CENPC1_gDNA_specific_F4 5' AACGCTGGTACTTGGACAAGTGGA 3' CENPC_cDNA_R9 3' ATGATACCTTGACGGCATGAGCCA 5'

Primers for sequencing CENPC3 3' genomic DNAs (See Figure 3A and 3C)

CENPC 3_gDNA_F7 specific 5' GTAGCTGCTGGCAATCAGGAGTTT 3' CENPC_cDNA_R9 3' ATGATACCTTGACGGCATGAGCCA 5'

Primers for sequencing whole CENPC3 cDNAs (See Figure 3A and Figure 6) CENPC 3_F1_1 5' ATGGACGCTACCGACCCCCTCT 3' CENPC 3_R2_specific 3' AGTACC<u>ACTATGTATACATGCA</u>AATGTCA 5' (16nt deletion)

Primers to test the conservation of three CENPC homologues in genus *Zea* (See Figure 5) CENPC_F41 5' ACTTCACATGCAGCTGAGGATAGC 3' CENPC_R552 3' TCACCAAGCCAATACTCCAAAGGC 5' Primers for specifically amplifying CENPC1/3 cDNAs (See Figure 4) F1_AcDNA 5' TGGCTGGTGAATCCCTGGAAAT 3' R1_AcDNA 3' AAAGCAACAGGTCACAAGGCGT 5' F2_AcDNA 5' ACGAGTAGCACACTCTCACCAAA 3' R2_AcDNA 3' TGACAAAGCAACAGGTCACAAGGC 5' F3_AcDNA 5' ACCTGGAAGGTTCCTCACTTGGTTT 3' R3_AcDNA

3' TTTGGTGAGAGTGTGCTACTCGT

Primers for specifically amplifying CENPC2 cDNAs (See Figure 4)

F1_BcDNA 5' CAGTTATGAGTAGCGCCAA 3' R1_BcDNA 3' GGCCAGGAGAGAGTATGCTTTGAT 5' F2_BcDNA 5' GATTCTTCCGAGGTTCTGATGACC 3' R2_BcDNA 3' TTGGCGCTACTCATAACTG 5' F3_BcDNA 5' ACCGACAGTTATGAGTAGCGCCAA 3' R3_BcDNA 3' ACAAGTCACAAGGCGTGATCCTCT 5'

Primers for specifically amplifying CENPC1 promoter regions (See Figure 7)

Cenpc1_promoter_F1_1494 5' TTACCTGAATTGCGCCCATCAAGC 3' Cencp1_promoter_R1_2616 3' AGGGTGATGAACAGTGACACGCTA 5' 3' TATCCTGTTTGGCGCGCATGAGGTA 5' Cenpc1_promoter_F3_298 (See Figure 13) 5' TCCCTGGACCGACGAGTAAATTGT 3' Cenpc1_promoter_R3_1517 (See Figure 13) 3' GCTTGATGGGCGCAATTCAGGTAA 5' Cenpc1_promoter_F5_640 3'GTTAGTGCCCTGTGCATGTGATGT 5'

Primers for specifically amplifying CENPC2 promoter regions (See Figure 7) Cenpc2_promoter_F1_428 5' ATACTCAAGCCCTCCCAACCACAT 3' Cenpc2_promoter_R1_1511 AGGATCAGAAGGCTTAAGGGCCAA 5'

3' Cenpc2_promoter_F2_1233 5' ACGTTACTGTTCACGGAGTGGGTT 3' Cenpc2_promoter_R2_2798 3' AGCTCAGACTGTGTGGACTGTGT 5' Cenpc2_promoter_F4_107 5' TGACAATTAAATGCGCGCCAGAGC 3' Cenpc2_promoter_R4_635 3' TGGGAGTCAATACTCCACGCACAA 5'

Primers for specifically amplifying CENPC3 promoter regions (See Figure 7)

Cenpc3_promoter_F2_949 (See Figure 13) 5'ATGCGAAGGTGTGAAGCTCTACCA 3' Cenpc3_promoter_R2_1998 (See Figure 13) 3' TGTTTGAACCTCCGTTCCGGGTAT 5' Cenpc3_promoter_F3_2684 (See Figure 13) 5' TTAGGGCCTGTTTGGTTCGTGACT 3' Cenpc3_promoter_R1_3456 (See Figure 13) 3' CACGGAACGTTGTGAAACGCTGAA 5'

Primers for specifically amplifying CENPC1/3 cDNAs in qRT-PCR (See Figure 11, Figure 12, Table 3, Table 4 and Table 5)

Cenpc1+3 realtime F1 923 specific 5' AGGATGTTATGCATGCTGTTGCGG 3' Cenpc1+3 realtime R1 1042 nonspecific 3'AAGTCAAATCGTCACGGCCATCCT 5' Cenpc1+3 realtime F2 1989 specific 5' AGTGGCTGGTGAATCCCTGGAAAT 3' Cenpc1+3 realtime R2 2114 specific 3' TTGGCACCATTTGGTGAGAGTGTG 5' Cenpc1+3 realtime F3 443 specific 5' TGAAAGGGTCTGAGGAGCTGGTTA 3' Cenpc1+3 realtime R3 596 specific 3' TTACGATCCAGTGCTGGCCTTCTT 5' Cenpc1+3 realtime F4 881 specific 5' CGGCTTCCCAAACAGCAACTATGA 3' Cenpc1+3 realtime R4 1031 nonspecific 3' TCACGGCCATCCTTCTCAGCTAAT 5'

Primers for specifically amplifying CENPC2 cDNAs in qRT-PCR (See Figure 11, Figure 12, Table 3, Table 4 and Table 5)

Cenpc2_realtime_F1_1697 5' GATCACGCCTTGTGACTTGTTGCT 3' Cenpc2_realtime_R1_1880 specific 3' TGTGATACCGCTGCACATACGCTA 5' Cenpc2_realtime_F2_1693 5' AGAGGATCACGCCTTGTGACTTGT 3' Cenpc2_realtime_R2_1882 specific 3' AATGTGATACCGCTGCACATACGC 5' Cenpc2_realtime_F3_1352 5' ACCGACAGTTATGAGTAGCGCCAA 3' Cenpc2_realtime_R3_1504 3' TCCAAAGGCCTCGAGCGTATTCTT 5' Cenpc2_realtime_F4_1150 specific 5' CACCTGTTCTGTGCAATGCACTGT 3' Cenpc2_realtime_R4_1279 specific 3' ATATTTCCAGTGCATGGCTAGATT 5'

Primers for specifically amplifying CENPC1/3 cDNAs in the 3' ends to show that Cenpc3 was rarely expressed in Zea mays (See Figure 8 and Figure 9) Cenpc_cDNA_F4_2456 5' GCCTTGTGACCTGTTGCTTTGTCA 3' Cenpc_cDNA_R9_2860 3' ATGATACCTTGACGGCATGAGCCA 3'

B) REALTIME PCR DATA

For supporting data of the real-time results, please refer to Figure 11, Figure 12 and Table 5.

C) PCR RESULTS OF PROMOTER STUDY

For supporting data of promoter studies, please see Figure 13.



Figure 1. CENH3s align in the histone fold domain (HFD) but have dissimilar tails and loop 1 region (Henikoff and Dalal 2005).

A										1
Zmayshuehue Zluxurians Zmaysparviglumis maize Zmaysmexcicana Zdiploperennis ZPerenis Clustal Consensus		AAGAAG AAGAAG AAGAAG AAGAAG AAGAAG AAGAAG	AAG CAG AAG AAG AAG AAG	AG AG AG AG AG AG AG AG	GAG GAG GAG GAG GAG GAG GAG	G T AG G T AG G T AG G T AG G T AG G T AG	GG GG GG GG GG GG GG GG GG GG GG GG	G GAG G GAG G GAG G GAG G GAG G GAG G GAG		
	50		1 · · · · I 60) l.	80		90		100
Zmayshuehue Zluxurians Zmaysparviglumis maize Zmaysmexcicana Zdiploperennis ZPerenis Clustal Consensus	GA GA GA GA GA GA	G AGG G AGG G CGG G CGG G CGG G CGG	AAAGGG AAAGGG AAAGGG AAAGGG AAAGGG AAAGGG AAAGGG	0 0 0 0 0 0 0 0 0 0 0	566A 566A 566A 566A 566A 566A 566A	GGGGG GGGGG GGGGG GGGGG GGGGG GGGGG GGGG	AAGAG AAGAG AAGAG AAGAG AAGAG AAGAG AAGAG	GG G GG G GG G GG G GG G GG G GG G	GGAG GGAG GGAG GGAG GGAG GGAG GGAG	
	100		110	· · · · ;	20			140		150
Zmayshuehue Zluxurians Zmaysparviglumis maize Zmaysmexcicana Zdiploperennis ZPerenis Clustal Consensus	GA GA GA GA GA GA	AG AG AG AG AG AG AG	TAAGAA TAAGAA TAAGAA TAAGAA TAAGAA TAAGAA TAAGAA	GA GA GA GA GA GA GA GA GA GA GA GA GA GA	AAAA AAAAA AAAAA AAAAA AAAAA	G G G G G G G G G G G G G G G G G G G		GG GG GG GG GG GG GG	AGGGA AGGGA AGGGA AGGGA AGGGA AGGGA AGGGA	GA GA GA GA GA
В										
			10		20		30		40	50
Zmaysmexcicana Perennis Zdiploperennis maize Zmaysparviglumis Zmayshuehuetenange Zluxurians Clustal Consensus	ensis	PKKKI PKKKI PKKKI PKKKI PKKKI	CFERSG CFERSG CFERSG CFERSG CFERSG CFERSG CFERSG	GASTS GASTS GASTS GASTS GASTS GASTS GASTS	ATPERI ATPERI ATPERI ATPERI ATPERI ATCERI	AAGTGG AAGTGG AAGTGG AAGTGG AAGTGG AAGTGG	RAASGO RAASGO RAASGO RAASGO RAASGO RAASGO RAASGO	S SVKR S SVKR S SVKR S SVKR S SVKR S SVKR S SVKR	TKPRHA TKPRHA TKPRHA TKPRHA TKPRHA TKPRHA TKPRHA	WR PG TV WR PG TV WR PG TV WR PG TV WR PG TV WR PG TV WR PG TV

Figure 2. Known maize relatives display very little polymorphism in the N-terminus of their CENH3 genes, both at the nucleotide (A) and the protein levels (B).



Figure 3. (A) Three CENPC homologues are aligned to scale. Blocks in the same color stand for sequences with very high homology. Primers used to verify the presence of CENPC1 and CENPC3 have been separately marked with pink (F1 refers to "CENPC 3_gDNA_F7" and R1 refers to "CENPC_cDNA_R9" in Appendix A) and dark green (F2 refers to "CENPC1_gDNA_specific_F4" and R2 refers to "CENPC_cDNA_R9" in Appendix A). Primers used for sequencing whole CENPC3 cDNAs have been marked with light green (F3 refers to "CENPC 3_F1_1" and R3 refers to "CENPC 3_R2_specific" in Appendix A). (B) PCR products, using genomic DNAs of all *Zea* species and primers F2 ("CENPC1_gDNA_specific_F4" in Appendix A) and R2 ("CENPC_cDNA_R9" in Appendix A), verified *Cenpc1* is present in all genomes in genus *Zea*. (C) PCR products, using genomic DNAs of all *Zea* species and primers F1 ("CENPC 3_gDNA_F7" in Appendix A) and R1 ("CENPC_cDNA_R9" in Appendix A), verified *Cenpc1* is present in all genomes in genus *Zea*. (C) PCR products, using genomic DNAs of all *Zea* species and primers F1 ("CENPC 3_gDNA_F7" in Appendix A) and R1 ("CENPC_cDNA_R9" in Appendix A), verified *Cenpc3* is present in all genomes in genus *Zea*.



Figure 4. (A) Three CENPC homologues are aligned to scale. Blocks in the same color stand for sequences with very high homology. Primers used to verify the presence of CENPC1 and CENPC3 have been separately marked with pink (F1 refers to "F1_AcDNA" and R1 refers to "R1_AcDNA" in Appendix A) and dark green (F2 refers to "F2_BcDNA" and R2 refers to "R2_BcDNA" in Appendix A). (B) PCR products, using genomic DNAs of all *Zea* species and primers F1 and R1, verified *Cenpc1* is present in all genomes in genus *Zea*. (C) PCR products, using genomic DNAs of all *Zea* species and primers F3 and R3, verified *Cenpc3* is present in all genomes in genus *Zea*.



Figure 5. The sequences of CENPC isoforms share very high homology in genus Zea, both in CENPC1 (Fig. 6A), which shows 99% identify, and in CENPC2 (Fig. 6B), which displays 100% identity between Z. *luxurians* and Z. *mays* and 95% identify between Z. *diploperennis* and Z. *mays*.



Figure 6. All Zea species except Zea mays contains the 16nt deletion in the 3' UTR region in their Cenpc1s (Fig. 7A) and all Zea species has the 16nt deletion in 3' UTR region in their Cenpc3s (Fig. 7B). Polymorphism of the 16-nt region has been illustrated in picture (Fig. 7C) and in table (Fig. 7D).



Figure 7. CENPC3 promoter region displays poor conservation between *Z. mays* and *Z. diploperennis* while promoter regions of CENPC1 and CENPC2 are much more conserved.

Α	1 2 3					
500bp		500bp 400bp				
polymorphism	gene	PCR product sizes				
Polymorphism WITHOUT 16nt-deletion in 3' UTR	gene Z.mays Cenpc1	PCR product sizes 390bp				

Figure 8. *Cenpc3* is rarely expressed in *Zea mays*. PCR products were amplified using primer sets Cenpc_cDNA_F4_2456 and Cenpc_cDNA_R9_2860 (see Appendix A), starting from the middle of exon 14 to the end of 3' region in *Cenpc1* and *Cenpc3*. Lane 1 shows maize *Cenpc3*-the product with the 16-nt region barely exists in *Zea mays* while lane 2 and lane 3 serve as controls. Lane 2 shows products of both polymorphisms exist in the *Zea mexicana* we used, which is a heterozygote between *Zea mays* and *Zea mexicana*. Lane3 shows that only the product with the 16-nt region exists in *Zea diploperennis*, in which *Cenpc1* and *Cenpc3* both have the 16-nt region.



Figure 9. NspI restriction enzyme digests confirmed that *Cenpc3* is rarely expressed in *Zea* mays. PCR products were amplified using Cenpc_cDNA_F4_2456 and Cenpc_cDNA_R9_2860 (see Appendix A) and digested using NspI. Lane 1 shows digest products of *Zea* mays cenpc1 was observed while digest products of *Zea* mays cenpc3 was barely observed. Lane 2 and lane 3 serve as controls. Lane 2 shows digest products of both polymorphisms exist in the *Zea* mexicana we used, which is a heterozygote between *Zea* mays and *Zea* mexicana. Lane3 shows that only digest product with the 16-nt region was observed in *Zea* diploperennis, in which *Cenpc1* and *Cenpc3* both have the 16-nt region.



Figure 10. Two CENPC isoforms (CENPC1/3 and CENPC2) are expressed at different levels in the genus *Zea* between *Z. parviglumis* (A) and *Z. mays* (B), and the resulting polymorphism in major kinetochore proteins put the previously abundant CentC repeats under selection, leading to variation in the abundance of centromeric repeat arrays between *Z. mays* and its ancestors.



Figure 11. Amplification curves of real-time PCR amplifying *Cenpc1/3* using four sets of *cenpc1/3*-specific primers (A) and *Cenpc2* using four sets of *Cenpc2*-specific primers (B) are compared.



Figure 12. Melting curves of real-time PCR amplifying *Cenpc1/3* using four sets of *cenpc1/3*-specific primers (A) and *Cenpc2* using four sets of *Cenpc2*-specific primers (B).



Figure 13. PCR results showing different products in Z. mays (1), Z. luxurians (2), and Z. diploperennis (3), with yellow bands showing the predicted size.

Table 2. All available CENPC homologues have been searched for in available EST libraries and each CENPC homologues is categorized by to line and which tissue(s) it comes from.

							Mixed	whole					
		root		shoot			/female	mieotic					
		and		apical		endo-	infl.,	anthers			egg		
	root	shoot	embryo	meristem	leaf	sperm	root	to pollen	tassel	kernel	cell	mixed	all
							tissues	shed					
cenpc2		2	4	4	1	2	2	1				4	20
cenpc1				1	1							5	7
cenpc3	4											2	6
cenpc1or3	3	1	4	16			10		1	2	1	6	44
													77
	B73	inbred B73	F-352	Ohio43	inbred Mo17		W23	unkown					
cenpc2	2	4	1	1	1			11					20
cenpc1	3	2						2					7
cenpc3							2	4					6
cenpc1or3	14	3		1	2		2	22					44
													77
			tł	he biggest n	umbers in each	categor	y are mark	ed red.					

Table 3. Real-time PCR results show that in most Zea species, CENPC1/3 cDNA is more expressed than CENPC2 cDNA, in one-month-old leaves, the ratio of which varies among species in genus Zea.

	(primers ne	ear 5' end)	(primers near 3' end)			
species	ΔCt (Ct of Cenpc2 - Ct of Cenpc1/3)	ratio of Cenpc1/3 to Cenpc2	ΔCt (Ct of Cenpc2 - Ct of Cenpc1/3)	ratio of Cenpc1/3 to Cenpc2		
Zea mays B73	4.34	20.25	3.94	15.35		
Zea mays Mo 17	2.59	6.02	2.37	5.17		
Zea mays B73/Mo 17 X Ki3	3.62	12.3	2.37	5.17		
Z. mays ssp. huehue	3.6	12.13	2.99	7.94		
Z. mays ssp. parviglumis	5.83	56.89	4.55	23.43		
Zea luxurians	4.16	17.88	3.28	9.71		
Zea diploperennis	3.81	14.03	3.77	13.64		

Table 4. Efficiencies of primers used in real-time PCRs, using *Zea mays* ubiquitin primers as the standard. Efficiencies have been tested using diluted templates by 32 fold, and all primers showed a C(t) value increase around 5 and each primers exhibited almost the same efficiency as before the template dilution.

Cenpc1/3 primers	F1R1	F2R2
efficiency (%)	89.29	105.67
Cenpc2 primers	F1R1	F4R4
efficiency (%)	90.21	75.57
ubiquitin		
efficiency (%)	100	

Table 5. The information of primers used in the real-time PCR (A) and the Ct value of each reaction (B).

primers species	Cenpc1/3- F1R1	Cenpc2- F4R4	ubiquitin	Cenpc1/3- F1R1 (normalized)	Cenpc2- F4R4 (normalized)	Δ	2^
Zea mays B73	29.18	33.52	27.22	1.96	6.3	4.34	20.25
Mo 17	27.01	29.6	25.99	1.02	3.61	2.59	6.02
B73/Mo 17 X Ki3	32.86	36.48	30.82	2.04	5.66	3.62	12.3
Zea huehue	28.03	31.63	24.79	3.24	6.84	3.6	12.13
Z parviglumis	31.19	37.02	31.76	-0.57	5.26	5.83	56.89
Zea luxurians	28.79	32.95	27.17	1.62	5.78	4.16	17.88
Zea diploperennis	28.16	31.97	26.78	1.38	5.19	3.81	14.03
efficiency (%)	89.29	75.57	100%				
primers				Cenpc1/3- F2R2	Cenpc2- F1R1		
species	Cenpc1/3- F2R2	Cenpc2- F1R1	ubiquitin	(normalized)	(normalized)	Δ	2 [^]
Zea mays B73	25.34	29.28	27.22	-1.88	2.06	3.94	15.35
Mo 17	22.8	25.17	25.99	-3.19	-0.82	2.37	5.17
B73/Mo 17 X Ki3	30.79	33.16	30.82	-0.03	2.34	2.37	5.17
Zea huehue	24.67	27.66	24.79	-0.12	2.87	2.99	7.94
Z parviglumis	27.9	32.45	31.76	-3.86	0.69	4.55	23.43
Zea luxurians	25.89	29.17	27.17	-1.28	2	3.28	9.71
Zea diploperennis	25.4	29.17	26.78	-1.38	2.39	3.77	13.64
efficiency (%)	105.67	90.21	100%				