PARTITIONING OF THE MAIZE EPIGENOME BY THE NUMBER OF METHYL GROUPS ON HISTONE H3 LYSINES 9 AND 27 & MAPPING MAIZE CENTROMERES AND CENTROMERE EVOLUTION

by

JINGHUA SHI

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Taking advantage of 3D light microscopy and the fine cytological resolution of maize pachytene chromosomes, we compared the distribution of individual methylation events to each other and to DNA staining intensity. The data reveal that three marks (H3K9me1, H3K27me1, and H3K27me2) are associated with heterochromatin. Contrary to expectations, we found that in maize H3K9me2 is not localized to heterochromatin, but mostly distributed in euchromatic regions along with H3K4me2. Other data demonstrate that centromere chromatin contains H3K9me2 and H3K9me3, that H3K27me3 occurs at several focused euchromatic domains, and that H4K20 methylation is rare or absent.

We describe a novel ChIP-display method that maps kinetochore footprints over high-resolution recombination maps. The centromeric retroelement CRM2 was used as the basis for an AFLP strategy that was combined with chromatin immunoprecipitation to yield a display of markers associated with CENH3 (a kinetochore-specific histone H3

variant). Each of the ten centromeres was mapped precisely using a combined set of 264 CRM markers, 57.2% of which interact with CENH3. The novel set of markers provides genetic anchor points throughout centromere cores. We can estimate that within-centromere gene conversion is roughly 10^{-3}. Our data suggest that frequent conversion is an important mechanism for spreading sequence variants among homologous centromeres. We also provide a first measure of linkage disequilibrium (LD) within maize centromere 2. The data reveal that the tandem repeat CentC has been under selection in recent history. FISH analyses show the abundance of CentC in the closest relatives of maize further suggest that CentC has been the major DNA sequences selected by meiotic drive in the genus *Zea*.

INDEX WORDS: Histone methylation, maize, heterochromatin, genetic mapping, centromere, kinetochore, CENH3, ChIP-display, recombination, gene conversion, evolution

PARTITIONING OF THE MAIZE EPIGENOME BY THE NUMBER OF METHYL GROUPS ON HISTONE H3 LYSINES 9 AND 27 & MAPPING MAIZE CENTROMERES AND CENTROMERE EVOLUTION

by

JINGHUA SHI

B.S. Beijing University, People's Republic of China, 2000

M.S. Institute of Botany, Chinese Academy of Sciences, People's Republic of China, 2003

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

PARTITIONING OF THE MAIZE EPIGENOME BY THE NUMBER OF METHYL GROUPS ON HISTONE H3 LYSINES 9 AND 27 & MAPPING MAIZE CENTROMERES AND CENTROMERE EVOLUTION

by

JINGHUA SHI

Major Professor: R. Kelly Dawe

Committee: Jeff Bennetzen
Katrien Devos
Russell Malmberg
Sue Wessler

Electronic Version Approved:

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

DEDICATION

To my husband, my parents and my coming son for their love and support.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my major advisor, Dr. R. Kelly Dawe for his consistent guidance and support during the course of my Ph.D. program. His enthusiasm and foresight in the research have made a deep impression on me and will be the precious asset for my future career. I am deeply indebted to my committee members Drs. Jeff Bennetzen, Katrien Devos, Russell Malmberg and Sue Wessler for their invaluable suggestions and guidance during my studies.

I would like to thank my past and current colleagues in the Dawe lab: Brunie Burgos, Yaqing Du, Lisa Kanizay, Xuexian Li, Yun Li, Amy Luce, Sarah Rushing, Chris Topp and Han Zhang for their help, support and valuable suggestions. I would like to give my special thank to Sarah Wolf who greatly helped me with cloning, sequencing and genotyping.

I would like to thank every member in Plant Functional Genomics Center. I thank the Wessler lab for hosting me when the Dawe lab was under renovation. I thank Dr. John Burke for his help on linkage disequilibrium. I thank the Birchler lab in the University of Missouri for teaching me the FISH technique. I thank the Gernot lab in the University of Hawaii for helping me with sequencing annotation. I would like to thank the plant biology greenhouse staff, Andy Tull and Michael Boyd in particular for taking care of my plants.

I thank my parents who have been always supportive and caring. They traveled from China to US to help me during the past half year, which has been the most difficult period during my life. I owe a special thank to my husband, Jianing Wang for his unconditional love and

support. Finally, I would like to thank my coming son, who keeps growing perfectly and brings me so much happiness.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

CHROMATIN AND CHROMATIN DOMAINS

In eukaryotic cells, DNA and its associated protein complex form chromatin. The nucleosome, which is the fundamental unit of chromatin, is made up of a histone octamer, composing of two of each H2A, H2B, H3 and H4, with ~146 bp DNA wrapping around (Thomas and Kornberg, 1975; Luger et al., 1997). These highly basic histones all contain a well folded C-terminal core domain and relaxed N-terminal tail. Unlike the core domains, the N-terminal tails are not involved in nucleosome formation, but are targets for a series of post-translational modifications, including acetylation, methylation, phosphorylation and ubiquitination (Strahl and Allis, 2000a; Jenuwein and Allis, 2001; Turner, 2002). These modifications, together with other chromatin-remodeling factors, regulate chromatin packaging, and organize genomes into distinct chromatin domains.

In most literature, chromatin is described as being of two types: loosely packed euchromatin, which allows active transcription; and tightly packed heterochromatin, which heavily represses transcription. This characterization was first defined by Heitz (Heitz, 1934) according to the staining pattern during the cell cycle. The originally defined heterochromatin is the chromosomal domain that remains highly condensed throughout the cell cycle. These regions

are considered as “constitutive” heterochromatin, and are found at pericentromere regions and knobs in maize (Dawe and Hiatt, 2004). More recent studies updated the concept of heterochromatin (Bennetzen, 2000). Chromosomal regions that are not condensed at all stages but can change chromatin state and exert developmental silencing, are termed as “facultative” heterochromatin (Wegel and Shaw, 2005).

Centromeres are often characterized as heterochromatin because they contain large blocks of repetitive sequences and repress recombination (Reinhart and Bartel, 2002). However, recent studies showed that centromeric chromatin differs from pericentromeric heterochromatin in several aspects, including: histone components, cytological appearance, chromatin environment and transcription level (Jiang et al., 2003; Black et al., 2004; Sullivan and Karpen, 2004; Topp et al., 2004; Shi and Dawe, 2006; Dalal et al., 2007b; Zhang et al., 2008). In this study, chromatin is characterized into three classes: euchromatin, heterochromatin and centromeric chromatin.

HISTONE VARIANTS

As the major protein components of chromatin, histones are among the most highly conserved proteins in eukaryotes (Thatcher and Gorovsky, 1994). Despite the very slowly evolving feature, all core histones but H4 have multiple nonallelic variants that have acquired diverse functions (Malik and Henikoff, 2003; Kamakaka and Biggins, 2005; Pusarla and Bhargava, 2005). Although often highly similar at sequence level, different histone variants normally associate with different chromatin domains or different development stages. The

histone H3 family contains several variants and illustrates how divergent forms can construct different chromatin domains (Hake and Allis, 2006).

Unlike canonical H3, which is expressed in a replication-dependent (RD) pattern and is deposited over newly synthesized DNA during replication, other H3 variants may undergo replication-independent (RI) assembly (Isenberg, 1979; Doenecke et al., 1997). H3.3 is expressed in a RI fashion and is often enriched in transcriptional active regions (Ahmad and Henikoff, 2002a; Johnson et al., 2004; McKittrick et al., 2004; Chow et al., 2005). H3.1 and H3.2 are both expressed in a RD fashion. H3.1 is only found in animals and is associated with gene activation (Hake et al., 2006). H3.2, present in both plants and animals, is associated with gene silencing (Johnson et al., 2004; McKittrick et al., 2004).

Centromere-specific H3s are functionally conserved in eukaryotes, but they are named differently in different organisms, CENH3 is the general term (Dawe and Henikoff, 2006). CENH3 undergoes RI pathway and replaces H3 in centromeric nucleosomes (Shelby et al., 2000; Ahmad and Henikoff, 2001; Sullivan and Karpen, 2001). CENH3 deposition is an early step in centromere assembly and serves as the base of the kinetochore, which is essential for mediating faithful chromosome segregation during mitosis (Stoler et al., 1995; Buchwitz et al., 1999; Howman et al., 2000; Takahashi et al., 2000; Blower and Karpen, 2001; Van Hooser et al., 2001; Talbert et al., 2002).

Whereas other H3 variants are very similar in structure, CENH3s are highly divergent, with an N-terminal tail that is highly variable in length and sequence (Malik and Henikoff, 2003; Cooper and Henikoff, 2004). Such divergence normally suggests that the protein is under positive selection. Surprisingly, the highly divergent N-terminal tails neither function in binding centromeric DNA, nor evolve adaptively (Talbert et al., 2004). Black et al. showed that CENP-A

(human CENH3) and histone H4 form more compact and rigid sub-nucleosomal tetramers than the corresponding canonical tetramers (Black et al., 2004). Their data also suggested that the Loop1 region of the histone fold domain is responsible for this conformational difference. The Loop1 region was also shown to be necessary and sufficient for CENP-A targeting at centromeres. Pairwise comparisons of synonymous and nonsynonymous substitutions further confirmed that this functionally important Loop1 region is adaptively evolving (Malik and Henikoff, 2001; Cooper and Henikoff, 2004).

HISTONE MODIFICATIONS

The flexible N-terminal tails of histones are targets to a series of post-translational modifications, including lysine acetylation and ubiquitination, lysine and arginine methylations, serine phosphorylation and several other less known modifications (Berger, 2002; Iizuka and Smith, 2003; Fuchs et al., 2006). These modifications either directly regulate chromatin structure, or serve as binding platforms to interact with other effectors (Strahl and Allis, 2000b; Jenuwein and Allis, 2001). Multiple modifications in various combinations may form a ‘‘histone code’’ that extends the information capacity of the associated DNA (Strahl and Allis, 2000b).

Histone acetylation and methylation represent the most common modifications of the histone tails, and are by far the most extensively studied. Acetylation adds a negative charge to the modified lysine residue, causing a weaker association between the histone and DNA and making the chromatin more accessible to effectors that activate transcription (Turner, 2000). Methylation does not affect the charge on the modified lysine and arginine, and it does not greatly alter chromatin structure (Strahl and Allis, 2000b). However, methylated lysines create

binding sites for other protein complex, which may further regulate chromatin structure and gene expression. Depending on the function of the recruited protein complex, histone lysine methylation can be either associated with "on" (transcriptional potent) or "off" (repressive) chromatin.

There are five potential lysine residues for methylation on histone N-terminal tails (K4, 9, 27, 36 of H3 and K20 of H4). Moreover, lysine can be monomethylated (me1), dimethylated (me2), or trimethylated (me3), and each methylation state may have unique biological functions, increasing the complexity of the histone code (Dutnall, 2003; Lachner et al., 2004). In general, methylated H3K4 and H3K36 are associated with "on" chromatin, while H3K9, H3K27 and H4K20 are associated with "off" chromatin. Despite the conservation of most methylated histone isoforms, they may mark different chromatin domains in different organisms.

Whereas histone methylation on H3K4 and K36 is associated with genes and thus marks the gene-rich euchromatin regions in plants and animals (Lippman and Martienssen, 2004; Sims et al., 2004), heterochromatin marks are more variable from organism to organism. In mouse, H3K9me3, H3K27me1 and H4K20me3 mark the most condensed constitutive heterochromatin, while H3K9me2, H3k27me3 and H4K20me1 are associated with the facultative heterochromatin (Peters et al., 2003; Plath et al., 2003; Rice et al., 2003; Silva et al., 2003; Okamoto et al., 2004; Schotta et al., 2004; Heard, 2005). In *Arabidopsis*, H3K9me1, 2 and H3K27me1, 2 and H4K20me1 were shown to be heterochromatin marks (Soppe et al., 2002; Lindroth et al., 2004; Naumann et al., 2005). Although associated with the silenced state, H3K9me3 and H3K27me3 were shown to be excluded from repetitive sequences and only localized to genes, and thus mark euchromatin (Turck et al., 2007; Zhang et al., 2007). These patterns are not conserved among plants. In species with small genome size (<500 Mb), H3K9me2 marks heterochromatin domains,

while in species with larger genome size, it has an even distribution all over the chromosomes (Jasencakova et al., 2003). In barley, H3K27me2 has been found exclusively in euchromatin domains (Fuchs et al., 2006). Taken together, the plant data suggest that H3K9me1 and H3K27me1 are the only two conserved marks for heterochromatin domains, H3K9me3 and H3K27me3 generally mark euchromatin domains, and H3K9me2 and H3K27me2 have variable distributions among different species.

The fact that the trimethylated status of H3K9 and K27 is mostly associated with genes and under developmental regulation was unexpected when it was originally discovered, since methylation was considered to be irreversible and more suitable for long-term memory. Not until 2004, when the identification of the first histone lysine-specific demethylase 1 (LSD1, renamed as KDM1 - lysine demethylase 1) (Allis et al., 2007), did we appreciate that lysine methylation can also be under a versatile regulation (Shi et al., 2004). Later, several JmjC domain-containing proteins were characterized as histone demethylases, many of which can catalyze demethylation of trimethylated substrates (Cloos et al., 2006; Klose et al., 2006; Tsukada et al., 2006; Whetstine et al., 2006; Yamane et al., 2006).

With increasing knowledge on histone demethylases, discovering the interplay among different histone modifications have become an exciting new direction of research. In fission yeast *Schizosaccharomyces pombe*, H3K4 demethylase Lid2 and H3K9 methyltransferase Clr4 were found in the same heterochromatin-promoting complex (Li et al., 2008). In mouse embryonic stem (ES) cells, the H3K4 demethylase JARID1A is associated with PRC2 complex, which is essential to establish H3K27me3 (Pasini et al., 2008). The presence of both histone methyltransferase and demethylase in the same complex allows the corresponding chromatin

domain to undergo transition between "on" and "off" status, and makes it possible to respond to developmental regulation.

CENTROMERES

1) Overview

The centromere, which is the DNA located at the primary constriction on the metaphase chromosome, is the assembly site for the kinetochore – the protein complex that interacts with spindle microtubules. In all eukaryotes, centromeres are essential for ensuring proper segregation of chromosomes during mitosis and meiosis.

Despite the conserved function, centromeres of different organisms are highly variable in size, structure and chromosomal distribution patterns. Whereas in budding yeast *Saccharomyces cerevisiae*, the "point" centromere, which is composed of a defined 125 bp DNA sequence, contains enough information for all centromere functions (Fitzgerald-Hayes et al., 1982). Most other eukaryotes, including fission yeast, animals and plants, have "regional" centromeres which are not defined genetically, but epigenetically, and can extend as long as several megabases (Nakaseko et al., 1986; Chikashige et al., 1989; Sun et al., 1997; Copenhaver et al., 1999; Schueler et al., 2001; Nagaki et al., 2004). Some other organisms, such as *Caenorhabditis elegans*, have a type of specialized centromere -- holokinetic centromeres, in which kinetochores are dispersed along entire chromosome (Maddox et al., 2004).

2) Genetic composition of centromere DNA

The centromere paradox is based on the fact that the centromeric DNA sequences are evolving rapidly while centromere function is conserved among all eukaryotes (Henikoff et al., 2001). In budding yeast, the simple *CEN* sequence is only 125 bp, containing three conserved domains: the imperfect palindromes CDEI and CDEIII and the ~ 85 bp highly AT-rich CDEII (Fitzgerald-Hayes et al., 1982). However, centromeres in most of eukaryotes are much larger. For instance, in fission yeast, centromeres are 40~100 kb, containing an ~15 kb central region flanked by 20~100 kb repetitive outer surrounding sequences (*otr*) (Polizzi and Clarke, 1991; Takahashi et al., 1992). In *Drosophila*, the centromere of a minichromosome spans ~420 kb, primarily composed of 5 bp short repeats (Sun et al., 1997). In human, centromeres contain mainly satellite repeats and can extend to 3~5 Mb (Schueler and Sullivan, 2006).

Although centromere sequences may share very limited homology even in closely related species, most of them have one feature in common -- tandem repeat arrays. For instance, the 171-bp α-satellite is the primary component of human centromeres (Schueler et al., 2001). Satellite repeats are also abundant in most plant species. In *Arabidopsis*, centromeres are mainly composed of 180-bp tandem repeat arrays (Copenhaver et al., 1999). In rice and maize, the 155-bp CentO and the 156-bp CentC repeats were found to be associated with CENH3, respectively (Zhong et al., 2002; Nagaki et al., 2004).

In addition to satellite repeats, transposable elements also reside in centromeres of many species. In human centromeres, the clustered α-satellite arrays are frequently interrupted by interspersed transposable elements, including LINEs (long interspersed elements), SINEs (short interspersed elements) and LTR (long terminal repeat) retrotransposons (Schueler et al., 2001). In the grass family, centromeric retrotransposons (CR elements), a class of LTR retrotransposons

derived from *Ty3/gypsy* family, preferentially target centromere regions (Miller et al., 1998; Presting et al., 1998). Compared with other LTR retrotransposons, CR elements diverge very slowly, suggesting that CR elements may be under selective pressure (Nagaki et al., 2003b). CR elements are also transcribed such that the RNAs remain tightly bound to centromeric nucleosomes, further suggesting a functional role in kinetochore assembly (Jiang et al., 2003; Topp et al., 2004).

3) Functional centromeres are defined epigenetically

Surprisingly, the satellite repeats are neither necessary nor sufficient for the formation of functional centromeres (Sullivan et al., 2001; Cheeseman and Desai, 2008). Studies from different organisms all suggest that functional centromere can be established without any centromeric repeats. The most extreme cases are known as 'neocentromeres', which are ectopic centromeres that form in noncentromeric regions of chromosomes. In human, several cases showed that neocentromeres can form at new chromosomal position where no α-satellite repeats are detected (Karpen and Allshire, 1997; Choo, 2001; Lo et al., 2001; Amor and Choo, 2002). In plants, the first functional neocentromere was reported in barley. Barley telosome 7H, which has completely lost all known barley centromere repeats, is stably transmitted during mitosis and meiosis (Nasuda et al., 2005). In *Drosophila*, overexpression of CENH3 (CID) resulted in formation of ectopic centromeres randomly and throughout chromosome arms (Heun et al., 2006).

Meanwhile, the presence of satellite repeats alone cannot automatically initiate centromere formation. In human dicentric chromosomes that have two domains with α-satellite arrays, only one has centromere function and the other is inactivated (Sullivan and Willard,

1998). In maize line knobless tama flint (KTF), chromosome 8 has two CentC-containing regions, apparently due to an inversion with one breakpoint in the original cluster of CentC. The original centromere region, still containing the majority of CentC arrays, was inactivated, while the new CentC site generated by the inversion acquired centromere function (Lamb et al., 2007a).

There is no doubt that CENH3 has a key role in specifying centromeres and that CENH3 is essential for recruiting other kinetochore components. However, little is known about how CENH3 nucleosomes are specifically deposited into centromeric chromatin. Evidence from human and *Drosophila* showed that CENH3 synthesis is independent from DNA replication. While centromere replication happens in S phase, CENH3 is not deposited until G2 (Shelby et al., 2000; Ahmad and Henikoff, 2001; Sullivan and Karpen, 2001). One model suggested that during each round of replication the old CENH3 nucleosomes are distributed equally between the two daughter strands to properly mark the site for kinetochore assembly. The newly synthesized CENH3 nucleosomes then come in later, and fill the empty spots close to the old CENH3 nucleosomes (Carroll and Straight, 2006). However, other factors may also play important roles in centromere formation. Whereas in *Drosophila*, overexpression of CENH3 resulted in ectopic centromere formation; in human cells, the same experiment did not lead to ectopic kinetochore formation (although other kinetochore proteins, such as CENP-C, were recruited) (Van Hooser et al., 2001). Mis18 and its homologues seem to be good candidates that regulate CENH3 deposition. In fission yeast, CENH3 fail to incorporate to centromere chromatin when Mis18 is depleted (Hayashi et al., 2004). In human cells, the protein complex of hMis18-α, hMis18-β, and hM18BP1is essential for the recruitment of de novo-synthesized CENP-A (Fujita et al., 2007). A recent study from *Drosophila* identified CENPC as another essential factor for regulating CENH3 deposition (Erhardt et al., 2008). These data seem to suggest a self-determining

(epigenetic) process, whereby CENH3 is required to recruit CENPC, but CENPC is required to recruit CENH3.

CENH3 replacement of H3 is critical for establishing the special epigenetic environment of centromere chromatin. The recent discovery of "hemisomes" in centromeres might shed light on what role CENH3 nucleosomes play in centromere identity and function. In this study, Henikoff and colleagues showed that instead of forming an octamer, CENH3 nucleosomes are present as heterotypic tetramers, containing one copy each of H2A, H2B, CENH3 and H4. The authors propose that the asymmetric structure of the hemisomes keep centromeres less condensed during mitosis, thereby helping maintain centromere identity (Dalal et al., 2007b; Dalal et al., 2007a).

MAPPING CENTROMERES

In most eukaryotes, satellite repeats and transposable elements are the major components of centromeric DNA sequences (Copenhaver et al., 1999; Schueler et al., 2001; Cheng et al., 2002b; Jin et al., 2004). This highly repetitive feature makes it difficult to find single-copy markers that are required for genetic mapping. Delimiting the precise boundaries of centromeres is another challenge for centromere mapping: the centromere and the flanking pericentromeric regions are not clearly differentiated by sequence (Jin et al., 2004; May et al., 2005; Luce et al., 2006). In many organisms, heterochromatic pericentromeric regions are very similar to centromeric regions in DNA composition, even though they have entirely different chromatin environments (Palmer et al., 1987; Choo, 2000; Zhong et al., 2002). The only reliable marker of functional centromeric DNA is the CENH3 histone variant.

Traditionally, centromere positions can be roughly placed using trisomic mapping (Frary et al., 1996; Singh et al., 1996; Lin, 2001). It takes generations to get the trisomic lines and requires different lines for each chromosome. Tetrad analysis is another powerful approach for centromere mapping, but it requires that all four products of meiosis are available to score (Fogel and Hurst, 1967; Copenhaver et al., 1999; Copenhaver et al., 2000). Application of tetrad mapping was originally limited to fungal organisms. The *Arabidopsis* quartet1 (qrt1) mutant, which causes pollen walls to remain fused after cytokinesis, made it possible to use tetrad analysis to map all five centromeres in *Arabidopsis* (Preuss et al., 1994; Copenhaver et al., 1999). Unfortunately a functionally equivalent quartet-like mutant has yet to be identified in other plants or animals.

In the grass family, CR elements are confirmed to be enriched in centromeres by CENH3 Chromatin Immunoprecipitation (ChIP) assays (Zhong et al., 2002; Nagaki et al., 2003a; Nagaki et al., 2004). Luce et al. (2006) published a new approach to mapping centromeres in maize using CRM-junction fragments (Luce et al., 2006). Assuming the typical random insertion of retroelements, there are thousands of different ways that a CRM retroelement can insert into another CRM or other retrotransposons. Primers that flank the insertion points are often unique. Amplified CRM junction-junction fragments were shown to be single-copy and polymorphic between the two parents B73 and Mo17, and were mapped to existing maize genetic maps using the IBM recombinant inbred lines. CENH3 chromatin immunoprecipitation (ChIP) was performed to characterize which markers were associated with functional centromeres. The junction markers can directly serve as Real-Time PCR markers for assaying the ChIPed DNA. Maize centromere 8 was mapped with confidence by showing that one of the single copy markers was enriched in CENH3 precipitates (Luce et al., 2006).

CENTROMERE EVOLUTION

1) Origin of centromeres and its evolutionary importance

A major difference between eukaryotes and prokaryotes is the means of segregating genetic information to progeny. Centromeres are critical structure for mitosis as they are the sites for kinetochore assembly and spindle attachment. However, the evolutionary origin of centromeres remains an enigma. Although centromere sequences are highly variable in different organisms, the components of kinetochore proteins are very similar among eukaryotes. Phylogenetic analysis of kinetochore proteins indicated that centromeres in all eukaryotes are derived from a common ancestor that had complex centromeres containing repetitive DNA (Meraldi et al., 2006). Villasante et al. proposed that the first centromere was derived from a telomere, based on their similar repetitive features (Villasante et al., 2007).

Centromeres are highly dynamic and actively participate in genome evolution. Studies of marsupial species suggested that centromeres are hotspots for a variety of genome rearrangements, including fissions, isochromosomes, whole-arm reciprocal translocations and minichromosomes (Bulazel et al., 2007; Metcalfe et al., 2007). Centromere repositioning is another source of chromosome rearrangement. Evolutionary new centromeres (ENC) can appear in novel chromosomal regions, reshape karyotypes and dramatically affect genome evolution (Ferreri et al., 2005; Ventura et al., 2007).

Rapid evolution of centromeric DNA also provides driving force for speciation. Even among closely related species, centromeric sequences can be considerably different (Csink and Henikoff, 1998; Murphy and Karpen, 1998; Lee et al., 2005; Lamb and Birchler, 2006). One of the most striking examples of rapid centromere evolution comes from the rice genus. A

chromatin immunoprecipitation cloning study revealed that *Oryza brachyantha*, which only diverged from other rice species about 7~9 million years ago, has lost nearly all the centromere repetitive sequences from its ancestors (including both CRR and the tandem repeat CentO). In place of CentO, the non-homologous repeat CentO-F was detected in all *O. brachyantha* centromeres (Dawe, 2005; Lee et al., 2005). Key centromere proteins, namely CENH3 and CENPC, also undergo rapid evolution and the two events – centromere and kinetochore evolution – may be related (Talbert et al., 2002; Cooper and Henikoff, 2004; Talbert et al., 2004). The proposed mechanism for change is centromere-based meiotic drive. In plants and animals, only one out of four meiotic products becomes a functional egg during female meiosis. Centromeres that specifically target this singly cell (by any means) can potentially 'cheat' meiosis and increase their representation in a population by meiotic drive. The driven centromere is likely to attract more repeats and more kinetochore proteins, causing a centromeric imbalance that can reduce fertility. In principle, such a genomic conflict could be resolved by mutating the binding domains of inner kinetochore proteins (Henikoff et al., 2001; Dawe and Henikoff, 2006). For any isolated populations without gene flow, the driving process might occur in independent directions. The resulted divergence then leads to reproductive isolation and initiates speciation (Henikoff et al., 2001).

2) Mechanisms for centromere evolution

The meiotic drive explains why the centromeres are evolving so rapidly. However, it does not provide a mechanism for how the sequences evolve. It is noteworthy that saying centromeres are highly dynamic does not necessarily mean centromere repeats are highly polymorphic within a species. On the contrary, the satellite repeats are quickly homogenized

(Henikoff et al., 2001; Henikoff, 2002). Under the meiotic drive, the rapid expansion and contraction select for more copies of the favored repeat variants and remove the less favored repeat variants (Charlesworth et al., 1994; Schueler et al., 2001; Lee et al., 2006).

Classical reciprocal recombination is one major mechanism for sequence evolution. Extensive evidence suggests that DNA sequence polymorphism is positively correlated with recombination rate (Begun and Aquadro, 1992; Dvorak et al., 1998; Nachman et al., 1998; Hamblin and Aquadro, 1999). However, given the extremely low rate of measured recombination in the centromere regions, it would be an unlikely source of centromere divergence (Mahtani and Willard, 1998; Copenhaver et al., 1999; Chen et al., 2002); (Hall et al., 2005; Ma and Bennetzen, 2006).

Another way for sequence evolution is accumulating random mutations. A recent study from the wild yeast species revealed that the centromere was the fastest-evolving part of the chromosome (Bensasson et al., 2008). In yeast, all four meiosis products survive, therefore, there should not be any driving occurring in yeast centromeres. This unexpected result indicated that mutation rates might be elevated in the centromere regions.

Centromere satellite repeats undergo extensive homogenization by expansions and contractions. In both human and rice, the centromeric satellite repeats are divergent, but local homogenization had occurred within each centromere (Lee et al., 2006; Rudd et al., 2006). For instance, in human, although the overall divergence for monomeric alpha-satellites can be as high as 30%, higher-order alpha-satellites are highly homogeneous (typically 97–100% identical) (Rudd and Willard, 2004). The primary underlying mechanism for this concerted evolution was proposed to be unequal crossover. However, most of the primary data supporting the unequal recombination model comes from a 30 year old computer simulation (Smith, 1976). The

remaining empirical support for unequal recombination is indirect, being based on sample sequencing of existing variations (Wevrick and Willard, 1989; Warburton and Willard, 1992; Schueler et al., 2001). In addition to unequal crossover, the expansion and homogenization of repeat arrays is probably facilitated by gene conversion, sequence amplification, segmental duplication and satellite transposition (Stephan, 1986; Walsh, 1987; Kearney et al., 2001; Schindelhauer and Schwarz, 2002; Alkan et al., 2004; Hall et al., 2004; Ma and Bennetzen, 2006; Ventura et al., 2007). Such events may take place intrachromosomally, between sister chromatids, or between homologous chromosomes.

REFERENCES

Ahmad, K., and Henikoff, S. (2001). Centromeres are specialized replication domains in heterochromatin. J. Cell Biol. **153,** 101-109.

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

Alkan, C., Eichler, E.E., Bailey, J.A., Sahinalp, S.C., and Tuzun, E. (2004). The role of unequal crossover in alpha-satellite DNA evolution: a computational analysis. J Comput Biol **11,** 933-944.

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., and Zhang, Y. (2007). New nomenclature for chromatin-modifying enzymes. Cell **131,** 633-636.

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

Begun, D.J., and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature **356,** 519-520.

Bennetzen, J.L. (2000). The many hues of plant heterochromatin. Genome Biol 1, REVIEWS107.

Bensasson, D., Zarowiecki, M., Burt, A., and Koufopanou, V. (2008). Rapid evolution of yeast centromeres in the absence of drive. Genetics **178,** 2161-2167.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev **12,** 142-148.

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.

Blower, M.D., and Karpen, G.H. (2001). The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat. Cell Biol. **3,** 730-739.

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.

Bulazel, K.V., Ferreri, G.C., Eldridge, M.D., and O'Neill, R.J. (2007). Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. Genome Biol **8,** R170.

Carroll, C.W., and Straight, A.F. (2006). Centromere formation: from epigenetics to self-assembly. Trends Cell Biol **16,** 70-78.

Charlesworth, B., Sneglowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371,** 215-220.

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

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., N.K., S., Tyagi, A.K., Soderlund, C., Dean, R.A., and Wing, R.A. (2002). An integrated physical and genetic map of the rice genome. Plant Cell **14,** 1-10.

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. (2002b). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14,** 1691-1704.

Chikashige, Y., Kinoshita, N., Nakaseko, Y., Matsumoto, T., Murakami, S., Niwa, O., and Yanagida, M. (1989). Composite motifs and repeat symmetry in S. pombe centromeres: direct analysis by integration of NotI restriction sites. Cell**,** 739-751.

Choo, K.H. (2000). Centromerization. Trends Cell Biol **10,** 182-188.

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

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO Rep **6,** 354-360.

Cloos, P.A., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K.H., and Helin, K. (2006). The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature **442,** 307-311.

Cooper, J.L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol **21,** 1712-1718.

Copenhaver, G.P., Keith, K.C., and Preuss, D. (2000). Tetrad analysis in higher plants. A budding technology. Plant Physiol **124,** 7-16.

Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., McCombie, W.R., Martienssen, R.A., Marra, M., and Preuss, D. (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286,** 2468-2474.

Csink, A.K., and Henikoff, S. (1998). Something from nothing: the evolution and utility of satellite repeats. Trends Genet. **15,** 200-204.

Dalal, Y., Furuyama, T., Vermaak, D., and Henikoff, S. (2007a). Structure, dynamics, and evolution of centromeric nucleosomes. Proc Natl Acad Sci U S A **104,** 15974-15981.

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

Dawe, R.K. (2005). Centromere renewal and replacement in the plant kingdom. Proc Natl Acad Sci U S A **102,** 11573-11574.

Dawe, R.K., and Hiatt, E.N. (2004). Plant neocentromeres: fast, focused, and driven. Chromosome Res **12,** 655-669.

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

Doenecke, D., Albig, W., Bode, C., Drabent, B., Franke, K., Gavenis, K., and Witt, O. (1997). Histones: genetic diversity and tissue-specific gene expression. Histochem Cell Biol **107,** 1-10.

Dutnall, R.N. (2003). Cracking the histone code: one, two, three methyls, you're out! Mol Cell **12,** 3-4.

Dvorak, J., Luo, M.C., and Yang, Z.L. (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species. Genetics **148,** 423-434.

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

Ferreri, G.C., Liscinsky, D.M., Mack, J.A., Eldridge, M.D., and O'Neill, R.J. (2005). Retention of latent centromeres in the Mammalian genome. J Hered **96,** 217-224.

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

Fogel, S., and Hurst, D.D. (1967). Meiotic gene conversion in yeast tetrads and the theory of recombination. Genetics **57,** 455-481.

Frary, A., Presting, G.G., and Tanksley, S.D. (1996). Molecular mapping of the centromeres of tomato chromosomes 7 and 9. Mol Gen Genet **250,** 295-304.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci **11,** 199-208.

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

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci U S A **103,** 6428-6435.

Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem **281,** 559-568.

Hall, A.E., Keith, K.C., Hall, S.E., Copenhaver, G.P., and Preuss, D. (2004). The rapidly evolving field of plant centromeres. Curr Opin Plant Biol **7,** 108-114.

Hall, S.E., Luo, S., Hall, A.E., and Preuss, D. (2005). Differential rates of local and global homogenization in centromere satellites from Arabidopsis relatives. Genetics **170,** 1913-1927.

Hamblin, M.T., and Aquadro, C.F. (1999). DNA sequence variation and the recombinational landscape in Drosophila pseudoobscura: a study of the second chromosome. Genetics **153,** 859-869.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell **118,** 715-729.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. Curr Opin Genet Dev **15,** 482-489.

Heitz. (1934). Die somatische Heteropyknose bei Drosophila melanogaster und ihre genetische Bedeutung.

Henikoff, S. (2002). Near the edge of a chromosome's 'black hole'. Trends Genet. **18,** 165-167.

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

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

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

Iizuka, M., and Smith, M.M. (2003). Functional consequences of histone modifications. Curr Opin Genet Dev **13,** 154-160.

Isenberg, I. (1979). Histones. Annu Rev Biochem **48,** 159-191.

Jasencakova, Z., Soppe, W.J., Meister, A., Gernand, D., Turner, B.M., and Schubert, I. (2003). Histone modifications in Arabidopsis- high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. Plant J **33,** 471-480.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science **293,** 1074-1080.

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.

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

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res **32,** 6511-6518.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev **19,** 295-310.

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

Kearney, H.M., Kirkpatrick, D.T., Gerton, J.L., and Petes, T.D. (2001). Meiotic recombination involving heterozygous large insertions in Saccharomyces cerevisiae: formation and repair of large, unpaired DNA loops. Genetics **158,** 1457-1476.

Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature **442,** 312-316.

Lachner, M., Sengupta, R., Schotta, G., and Jenuwein, T. (2004). Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. Cold Spring Harb Symp Quant Biol **69,** 209-218.

Lamb, J.C., and Birchler, J.A. (2006). Retroelement genome painting: cytological visualization of retroelement expansions in the genera Zea and Tripsacum. Genetics **173,** 1007-1021.

Lamb, J.C., Meyer, J.M., and Birchler, J.A. (2007). A hemicentric inversion in the maize line knobless Tama flint created two sites of centromeric elements and moved the kinetochore-forming region. Chromosoma **116,** 237-247.

Lee, H.R., Neumann, P., Macas, J., and Jiang, J. (2006). Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol **23,** 2505-2520.

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.

Li, F., Huarte, M., Zaratiegui, M., Vaughn, M.W., Shi, Y., Martienssen, R., and Cande, W.Z. (2008). Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. Cell **135,** 272-283.

Lin, B.Y., S.J. Chang, H.M. Lin. (2001). RFLP mapping of the centromere of chromosomes 1, 6 and 9 by B-A translocations inmaize. Bot. Bull. Acad. Sinica **42,** 273-279.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., and Jacobsen, S.E. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J **23,** 4286-4296.

Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature **431,** 364-370.

Lo, A.W., Magliano, D.J., Sibson, M.C., Kalitsis, P., Craig, J.M., and Choo, K.H. (2001). A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res **11,** 448-457.

Luce, A.C., Sharma, A., Mollere, O.S., Wolfgruber, T.K., Nagaki, K., Jiang, J., Presting, G.G., and Dawe, R.K. (2006). Precise centromere mapping using a combination of repeat junction markers and chromatin immunoprecipitation-polymerase chain reaction. Genetics **174,** 1057-1061.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389,** 251-260.

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

Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). "Holo"er than thou: chromosome segregation and kinetochore function in C. elegans. Chromosome Res **12,** 641-653.

Mahtani, M.M., and Willard, H.F. (1998). Physical and genetic mapping of the human X chromosome centromere: repression of recombination. Genome Res **8,** 100-110.

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

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

May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. PLoS Genet **1,** e79.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc. Natl. Acad. Sci. USA **101,** 1525-1530.

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.

Metcalfe, C.J., Bulazel, K.V., Ferreri, G.C., Schroeder-Reiter, E., Wanner, G., Rens, W., Obergfell, C., Eldridge, M.D., and O'Neill, R.J. (2007). Genomic instability within centromeres of interspecific marsupial hybrids. Genetics **177,** 2507-2517.

Miller, J.T., Dong, F., Jackson, S.A., Song, J., and Jiang, J. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150,** 1615-1623.

Murphy, T.D., and Karpen, G.H. (1998). Centromeres take flight: Alpha satellite and the quest for the human centromere. Cell **93,** 317-320.

Nachman, M.W., Bauer, V.L., Crowell, S.L., and Aquadro, C.F. (1998). DNA variability and recombination rates at X-linked loci in humans. Genetics **150,** 1133-1141.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003a). Chromatin Immunoprecipitation reveals that the 180-bp satellite repeat Is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics **163,** 1221-1225.

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 reveals active genes. Nature Genet. **36,** 138-145.

Nagaki, K., Song, J., Stupar, R., Parokonny, A.S., Yuan, Q., Ouyang, S., Liu, J., Dawe, R.K., Buell, C.R., and Jiang, J. (2003b). Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics **163,** 759-770.

Nakaseko, Y., Adachi, Y., Funahashi, S.-i., Niwa, O., and Yanagida, M. (1986). Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J **5,** 1011-1021.

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.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T., and Reuter, G. (2005). Pivotal role of

AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. Embo J **24,** 1418-1429.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science **303,** 644-649.

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.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev **22,** 1345-1355.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H., and Jenuwein, T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell **12,** 1577-1589.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science **300,** 131-135.

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

Presting, G., Malysheva, L., Fuchs, J., and Schubert, I. (1998). A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16,** 721-728.

Preuss, D., Rhee, S.Y., and Davis, R.W. (1994). Tetrad analysis possible in Arabidopsis with mutation of the QUARTET (QRT) genes. Science **264,** 1458-1460.

Pusarla, R.H., and Bhargava, P. (2005). Histones in functional diversification. Core histone variants. FEBS J **272,** 5149-5168.

Reinhart, R.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science **297,** 1831.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Mol Cell **12,** 1591-1598.

Rudd, M.K., and Willard, H.F. (2004). Analysis of the centromeric regions of the human genome assembly. Trends Genet **20,** 529-533.

Rudd, M.K., Wray, G.A., and Willard, H.F. (2006). The evolutionary dynamics of alpha-satellite. Genome Res **16,** 88-96.

Schindelhauer, D., and Schwarz, T. (2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha-satellite DNA array. Genome Res **12,** 1815-1826.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev **18,** 1251-1262.

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

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science **294,** 109-114.

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

Shi, J., and Dawe, K. (2006). Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. Genetics**,** in press.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell **119,** 941-953.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell **4,** 481-495.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev **18,** 2437-2468.

Singh, K., Ishii, T., Parco, A., Huang, N., Brar, D.S., and Khush, G.S. (1996). Centromere mapping and orientation of the molecular linkage map of rice (Oryza sativa L.). Proc. Natl. Acad. Sci. USA **93,** 6163-6168.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science **191,** 528-535.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsi. EMBO J. **21,** 6549-6559.

Stephan, W. (1986). Recombination and the evolution of satellite DNA. Genet. Res. Camb. **47,** 167-174.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgeral-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 in mitosis. Genes. Dev. **9,** 573-586.

Strahl, B.D., and Allis, C.D. (2000a). The language of covalent histone modifications. Nature **403,** 41-45.

Strahl, B.D., and Allis, C.D. (2000b). The language of covalent histone modifications. Nature **403,** 41-45.

Sullivan, B., and Karpen, G. (2001). Centromere identity in Drosophila is not determined in vivo by replication timing. J Cell Biol **154,** 683-690.

Sullivan, B., Blower, M., and Karpen, G. (2001). Determining centromere identity: cyclical stories and forking paths. Nat. Rev. Genet. **2,** 584-596.

Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. Nat Genet **20,** 227-228.

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

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

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

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol. Biol. Cell **3,** 819-835.

Talbert, P., R Masuelli, Tyagi, A., Comai, L., and 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., and Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. J Biol **3,** 18.

Thatcher, T.H., and Gorovsky, M.A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res **22,** 174-179.

Thomas, J.O., and Kornberg, R.D. (1975). An octamer of histones in chromatin and free in solution. Proc Natl Acad Sci U S A **72,** 2626-2630.

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

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature **439,** 811-816.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet **3,** e86.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays **22,** 836-845.

Turner, B.M. (2002). Cellular memory and the histone code. Cell **111,** 285-291.

Van Hooser, A.A., Ouspenski, II, 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.

Ventura, M., Antonacci, F., Cardone, M.F., Stanyon, R., D'Addabbo, P., Cellamare, A., Sprague, L.J., Eichler, E.E., Archidiacono, N., and Rocchi, M. (2007). Evolutionary formation of new centromeres in macaque. Science **316,** 243-246.

Villasante, A., Abad, J.P., and Mendez-Lago, M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci U S A **104,** 10542-10547.

Walsh, J. (1987). Peristence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics **115,** 553-567.

Warburton, P.E., and Willard, H.F. (1992). PCR amplification of tandemly repeated DNA: analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human alpha satellite DNA. Nucleic Acids Res **20,** 6033-6042.

Wegel, E., and Shaw, P. (2005). Gene activation and deactivation related changes in the three-dimensional structure of chromatin. Chromosoma **114,** 331-337.

Wevrick, R., and Willard, H.F. (1989). Long-range organization of tandem arrays of a satellite DNA at the centromeres of human chromosomes: High-frequency array-length polymorphism and meiotic stability. Proc. Natl. Acad. Sci. USA **86,** 9394-9398.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell **125,** 467-481.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell **125,** 483-495.

Zhang, W., Lee, H.R., Koo, D.H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell **20,** 25-34.

Zhang, X.Y., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. Plos Biology **5,** 1026-1035.

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.

CHAPTER 2

PARTITIONING OF THE MAIZE EPIGENOME BY THE NUMBER OF METHYL GROUPS ON HISTONE H3 LYSINES 9 AND 27

Jinghua Shi and R. Kelly Dawe. (2006) Genetics 173, 1571-1583

Reprinted here with permission of publisher.

ABSTRACT

We report a detailed analysis of maize chromosome structure with respect to seven histone H3 methylation states (dimethylation at lysine 4 and mono-, di-, and trimethylation at lysines 9 and 27). Threedimensional light microscopy and the fine cytological resolution of maize pachytene chromosomes made it possible to compare the distribution of individual histone methylation events to each other and to DNA staining intensity. Major conclusions are that (1) H3K27me2 marks classical heterochromatin; (2) H3K4me2 is limited to areas between and around H3K27me2-marked chromomeres, clearly demarcating the euchromatic gene space; (3) H3K9me2 is restricted to the euchromatic gene space; (4) H3K27me3 occurs in a few (roughly seven) focused euchromatic domains; (5) centromeres and CENP-C are closely associated with H3K9me2 and H3K9me3; and (6) histone H4K20 di- and trimethylation are nearly or completely absent in maize. Each methylation state identifies different regions of the epigenome. We discuss the evolutionary liability of histone methylation profiles and draw a distinction between H3K9me2-mediated gene silencing and heterochromatin formation.

INTRODUCTION

In eukaryotes, the fundamental unit for DNA packing is the nucleosome, a protein octamer/DNA complex composed of four core histones, H2A, H2B, H3 and H4, that are wrapped twice with ~146 bp of DNA (Luger et al., 1997). The histone amino-termini are targets for a series of post-translational modifications, including acetylation, phosphorylation, and methylation. These modifications regulate chromatin structure and gene expression (Strahl and Allis, 2000a; Jenuwein and Allis, 2001; Turner, 2002). Multiple modifications in various

combinations are thought to form a histone 'code' that extends the information capacity of the associated DNA (Strahl and Allis, 2000a). For example, histone H3 and H4 acetylation is consistently associated with transcriptionally active euchromatin, while methylation can be associated with either inactive or active chromatin depending on the residue. Methylation at H3K4, H3K36 and H3K79 are hallmarks for active transcription, whereas methylation at H3K9, H3K27 and H4K20 are correlated with transcriptionally inert heterochromatin (Fischle et al., 2003; Lachner et al., 2003; Margueron et al., 2005; Peters and Schubeler, 2005). Lysine can be monomethylated (me1), dimethylated (me2) or trimethylated (me3), and each methylation state may have unique biological functions, increasing the potential complexity of the histone code (Dutnall, 2003).

At the chromosomal level, transcriptionally inactive regions tend to be associated with heterochromatin (Brown, 1966) and the brightest immunostaining for 'off' marks such as methylated H3K9, H3K27 and H4K20 (Peters et al., 2003; Schotta et al., 2004). In animal interphase cells, H3K9me3, H3K27me1 and H4K20me3 mark the most deeply stained regions while less condensed heterochromatin contains H3K9me1, H3K9me2 and H3K27me3 (Peters et al., 2003; Plath et al., 2003; Rice et al., 2003; Silva et al., 2003; Okamoto et al., 2004; Schotta et al., 2004). Available data suggest that at least part of this description applies to plants as well. Several reports indicate that the mono- and dimethylated forms of H3K9 and H3K27 are enriched in heterochromatin (Jackson et al., 2004; Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005)**,** though there is variation among small and large genome plants (Houben et al., 2003). Unlike in animals, *Arabidopsis* H3K27me3 is associated with euchromatin and active transcription, while H3K9me3 is extremely rare (Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005). There is also evidence that H4K20 methylation is present in

Arabidopsis (Naumann et al., 2005; Ng et al., 2006). It is not clear, however, whether there is a direct relationship between heterochromatin and histone methylation in any species since quantitative comparisons are not yet available. Further, only a brief description is available for the staining patterns in large genome plants (Houben et al., 2003), which make up the bulk of the angiosperms (Arumuganthan and earle, 1991).

Although most of the genome interacts with histone H3, centromeric DNA also interacts with Centromeric Histone H3 (CENH3) (Henikoff et al., 2001). CENH3 is an important variant of H3 that recruits inner kinetochore proteins such as CENP-C (Centromere Protein C) and mediates kinetochore formation (Van Hooser et al., 2001). Arrays of CENH3-containing nucleosomes are not continuous, but intervened by blocks of histone H3-containing nucleosomes (Blower et al., 2002; Sullivan and Karpen, 2004; Chueh et al., 2005; Yan et al., 2005b). Centromeres are also transcribed to produce long stable RNAs – an observation which sets them apart from other highly repetitive regions of the genome (Saffery et al., 2003; Nagaki et al., 2004; Topp et al., 2004; May et al., 2005). The available data suggest that the chromatin structure of centromeres is quite different from flanking heterochromatin, but the extent and importance of the differences are poorly understood.

Here we provide a quantitative whole-genome view of the distribution of mono-, di-, and trimethylation at H3K9 and H3K27 in maize, a model species with a large genome. The excellent cytological resolution of maize pachytene chromosomes, 3D light microscopy, and the ability to quantify staining patterns relative to cytological features provide an unparalleled view of histone methylation. The data reveal that three marks (H3K9me1, H3K27me1, and H3K27me2) correlate with DAPI (DNA) staining, but that only H3K27me2 is specifically enriched in condensed areas. Contrary to expectations, we found that in maize H3K9me2 is not

abundant in heterochromatin, but accumulates along with H3K4me2 in the between-chromomere euchromatic gene space. Other data demonstrate that centromere cores contain H3K9me2 and H3K9me3, that H3K27me3 occurs at several brightly focused euchromatic domains, and that H4K20 methylation is rare or absent.

MATERIALS AND METHODS

Antibodies

Chicken anti-maize CENPC (1:50 for immunofluorescence) was described in a previous study (Zhong et al., 2002). Antibodies (Table 2.1) were purchased from Upstate (Lake Placid, NY) and Abcam (Cambridge, MA).

Cytological preparation and indirect immunofluorescence

Anthers from the inbred lines KYS, W23, and B73 were fixed and processed as described previously (Dawe et al., 1994), except that 2% paraformaldehyde and 0.1% Triton X-100 were added during fixation. All data described here were confirmed in at least two inbreds (KYS and either W23 or B73). Cells were adhered to polylysine-coated cover slips by spinning at 100g. The coverslips were then incubated overnight with primary antibodies, rinsed in PBS three times, and blocked with 10% goat serum for 2 hrs. Secondary antibodies (FITC-conjugated goat anti-rabbit, rhodamine-conjugated donkey anti-chicken, or rhodamine conjugated donkey anti-rat, Jackson ImmunoResearch, West Grove, PA) were diluted 1:100 and applied for 2 hours. These preparations contain both meiotic cells and anther wall cells known as tapetal cells; the tapetal

cells were used as material for the analysis of interphase (Fig. 2.5). Cultured mouse neocortical neuron cells were obtained from the laboratory of Dr. Thomas F. Murray (University of Georgia) and processed for immunocytochemistry as previously reported (Dravid et al., 2005). Mouse cells were blocked with 5% BSA in PBS for 30 min, primary antibodies were applied for 1hr at 1:200 dilution in PBS, and secondary antibodies were applied for 30 min. All cells were stained with 0.01 mg/ml DAPI (4, 6-diamidino-2-phenylindole) for 10 min, mounted in Mowiol mounting medium (Harlow and Lane, 1988) and sealed with nail polish.

Protein blot analysis

Maize protein was extracted from young ears about 10 cm in length (Zhang et al., 2005), and from cultured mouse neocortical neuron cells (Dravid et al., 2005). Protein extraction and protein blot analysis was carried out as previously described (Zhang et al., 2005). A single nitrocellulose membrane was cut into four pieces and incubated with different primary antibodies. After detection, the membrane was stripped and reused for other antibodies.

Microscopy and image analysis

All data except those shown in Figures 2.1, 2.2, 2.3C, 2.6A and 2.7A were captured and processed using a Zeiss Axio Imager and Slide Book 4.0 software (Intelligent Imaging Innovations, Inc., Denver, CO). Figures 2.1, 2.2, 2.3C, 2.6A and 2.7A were captured and analyzed using a DeltaVision 3D light microscopy system and associated software (Applied Precision, Inc., Issaquah, WA). The steps of chromosome straightening were described previously (Dawe et al., 1994). In Figure 2.2, data from two straightened chromosomes 9 were

normalized (to the highest value on the histogram) and averaged. Statistical smoothing by the Lowess method (Cleveland, 1979), using GraphPad Software, Inc., San Diego, CA) was applied to the resulting histograms.

Table 2.1. List of the tested antibodies, their localization patterns and dilutions used.

Antibody	**Source/Cat.**	**Host**	**Pachytene immunofluoresence staining**	**WB ***	**Dilution for**	**Dilution for WB**
H3K4me2	Upstate/07-	Rabbit	Between chromomeres	+	1:50	1:2000
H3K9me1	Upstate/07-	Rabbit	Pericentromeres, chromomeres, and between chromomeres	+	1:75	1:10000
H3K9me2	Upstate/05-	Rabbit	Between chromomeres, in kinetochores (weak)	+	1:50	1:5000
H3K9me2	Upstate/07-	Rabbit	Between chromomeres, in kinetochores (strong)	+	1:50	1:2000
H3K9me2	Upstate/07-	Rabbit	Between chromomeres, in kinetochores (medium)	/	1:50	/
H3K9me3	Abcam/	Rabbit	Between chromomeres, in kinetochores	+	1:50	1:2000
H3K27me1	Upstate/07-	Rabbit	Pericentromeres, chromomeres and continuously between	+	1:100	1:12000
H3K27me2	Upstate/07-	Rabbit	Pericentromeres and chromomeres	+	1:50	1:2000
H3K27me2	Abcam/	Rat	Pericentromeres and chromomeres	+	1:50	/
H3K27me3	Upstate/07-	Rabbit	Several discrete domains	+	1:50	1:1000
H3K27me3	Abcam/	Mouse	Same as above (only stained in W23)	/	1:25	/
Acetyl H4	Upstate/06-	Rabbit	/	+	1:50	1:2500
H4K20me1	Upstate/07-	Rabbit	No visible signals	+	1:25-1:200	1:2000
H4K20me2	Upstate/07-	Rabbit	No visible signals	-	1:25-1:200	1:2000
H4K20me3	Upstate/07-	Rabbit	No visible signals	-	1:25-1:200	1:2000
H4K20me3	Abcam/	Rabbit	No visible signals	/	1:25-1:200	/

***WB** - western blot

RESULTS

H3K4me2 identifies euchromatin: the between-chromomere space in distal halves of chromosome arms

The major features of the maize karyotype can be observed on chromosomes stained for DAPI. As shown in Figure 2.1, the ten chromosomes vary in length and arm ratio. Pericentromeres are evident as deeply stained regions flanking centromeres, and the distal portions of chromosome arms contain numerous small regions of heterochromatin known as chromomeres. In addition, maize and most other grasses contain dense heterochromatic regions called knobs (see (Dawe and Hiatt, 2004)). In KYS there are four knobs, on 5L, 6L, 7L and 9S (Fig. 2.1). As in other species the majority of maize genes are known to lie towards the ends of chromosomes, away from the pericentromeres (Anderson et al., 2004).

Based on prior data we would expect antisera against H3K4me2 to stain brightly at the ends of chromosomes, with the intensity of the stain roughly indicating the number or activity of the genes present (Figs. 2.1B, 2.2B). Likewise, antisera against H3K27me2 should label the heterochromatic domains of the chromosome, such as pericentromere and knobs (Figs. 2.1B, 2.2D, 2.3E). Analysis of computationally straightened chromosomes confirmed this expectation, showing that H3K4me2-stained regions are located in the distal halves of chromosome arms (Fig. 2.1B). No visible labeling was detectable within CENP-C-marked centromeres or knobs.

For quantitative interpretations we focused on chromosome 9 where gene and recombination frequencies have been carefully documented (Anderson et al., 2003; Anderson et al., 2004). Chromosome 9 from cells stained for DAPI and H3K4me2 were straightened, intensity histograms extracted from the linear axes, and statistical smoothing used to reveal

general staining trends (Cleveland, 1979). The data show that DAPI and H3K4me2 staining complement each other, with H3K4me2 being distributed towards chromosome ends away from pericentromeres (Figs. 2.2A, 2.2B). Further conforming to expectations, the data show that H3K4me2 staining is roughly correlated with the gene/recombination map of Anderson and colleagues (Fig. 2.2E; (Anderson et al., 2003)).

High magnification views revealed that H3K4me2 staining is limited to the between-chromomere spaces where DAPI staining was weak or absent (Fig. 2.4E). When combined with the trend analysis showing a general correlation with genes (Fig. 2.2E), these data strongly suggest that the between-chromomere space represents euchromatin in maize.

H3K9me2 marks euchromatin

Localization with three different anti-H3K9me2 antisera on maize pachytene chromosomes failed to support the general consensus that H3K9me2 is a heterochromatic marker in plants (Houben et al., 2003; Jackson et al., 2004; Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005). Analysis of over 50 pachytene nuclei revealed consistently bright signal along chromosome arms and little staining in knobs and pericentromeres (Fig. 2.3B). Similarly, no significant staining of knobs was observed in mitotic interphase cells (Fig. 2.5). H3K9me2 is abundant at the end of chromosome 9L (Fig. 2.2C) and between chromomeres (Fig. 2.4D), mirroring the staining patterns for H3K4me2. It is important to note, however, that H3K9me2 staining did not match H3K4me2 precisely. For instance H3K9me2 showed light staining in knobs whereas H3K4me2 did not (Figs. 2.2C, 2.3B, 2.4D).

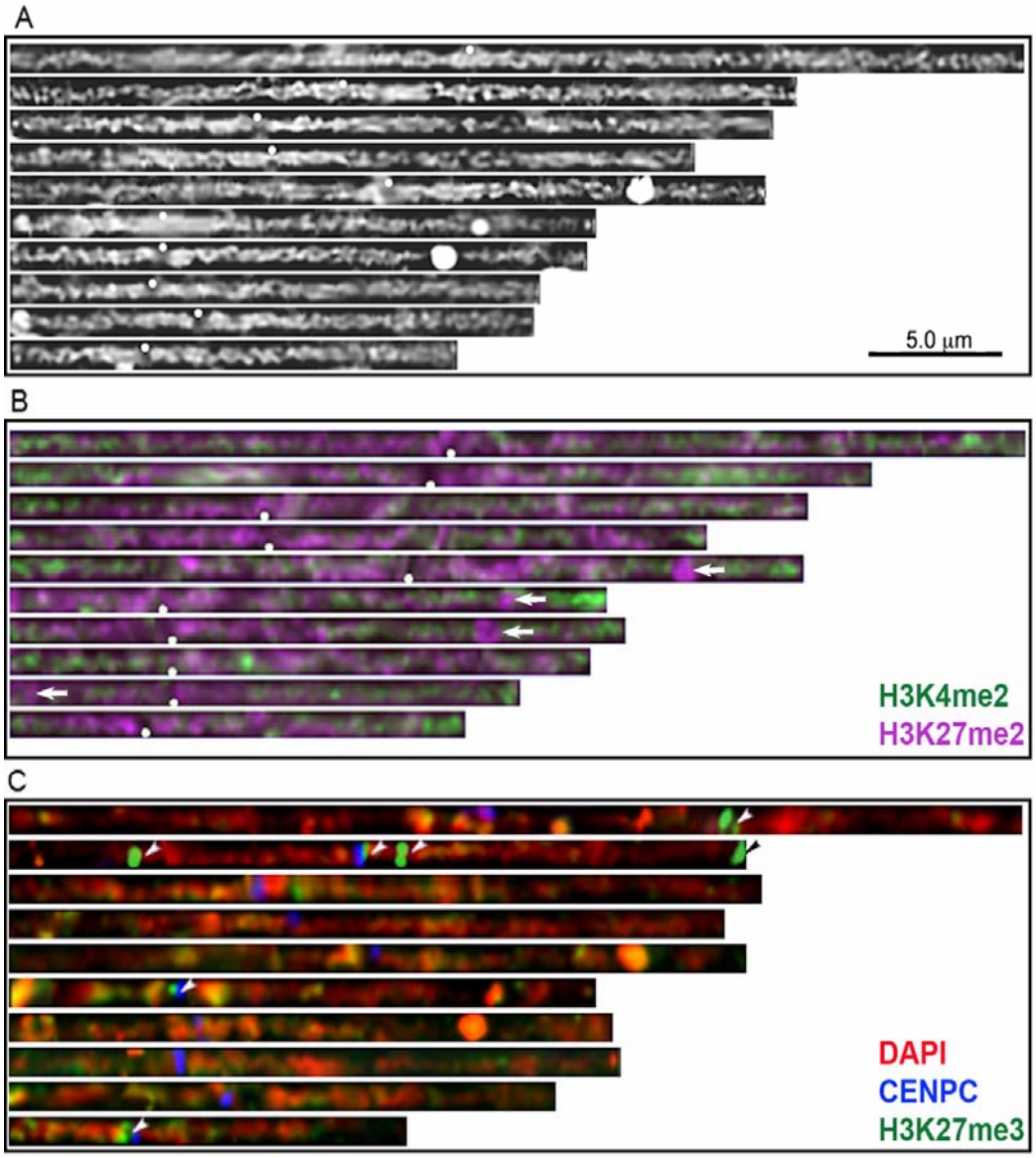

Figure 2.1. **Complete KYS karyotypes showing DAPI, H3K4me2, H3K27me2 and H3K27me3.** **A**) The KYS karyotype as it appears after staining with DAPI. **B**) The KYS karyotype after staining with anti-H3K4me2 (green) and anti-H3K27me2 (magenta). **C**) The KYS karyotype showing DAPI (red), anti-H3K27me3 (green) and CENPC (blue). The white dots in A and B show the position of centromeres, arrows point to the positions of knobs and the arrowheads point to H3K27me3-rich domains.

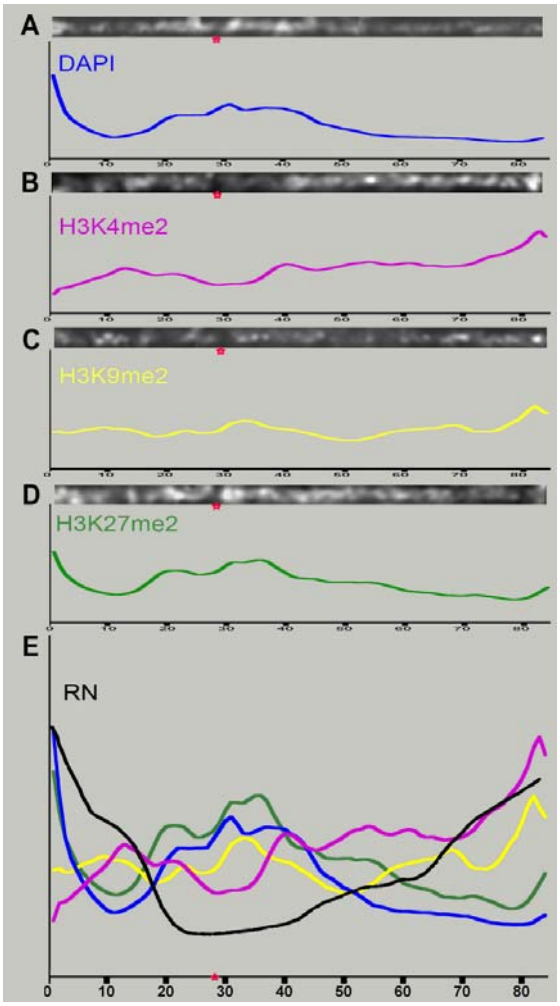

Figure 2.2. **Staining patterns on chromosome 9. A-D**) Lowess-smoothed curves showing the general trends for several staining patterns. **A**) DAPI staining, showing the distribution of heterochromatin. **B**) H3K4me2 staining, showing a distribution that is skewed up towards chromosome ends and negatively correlated with heterochromatin. **C**) H3K9me2 staining, showing a flat distribution that trends upwards similar to H3K4me2. **D**) H3K27me2 staining, showing a distribution that closely follows heterochromatin. **E**) A combination plot of A-D and the recombination nodules map for chromosome 9 in black (adapted from (Anderson et al., 2003)). Red stars show the positions of centromeres.

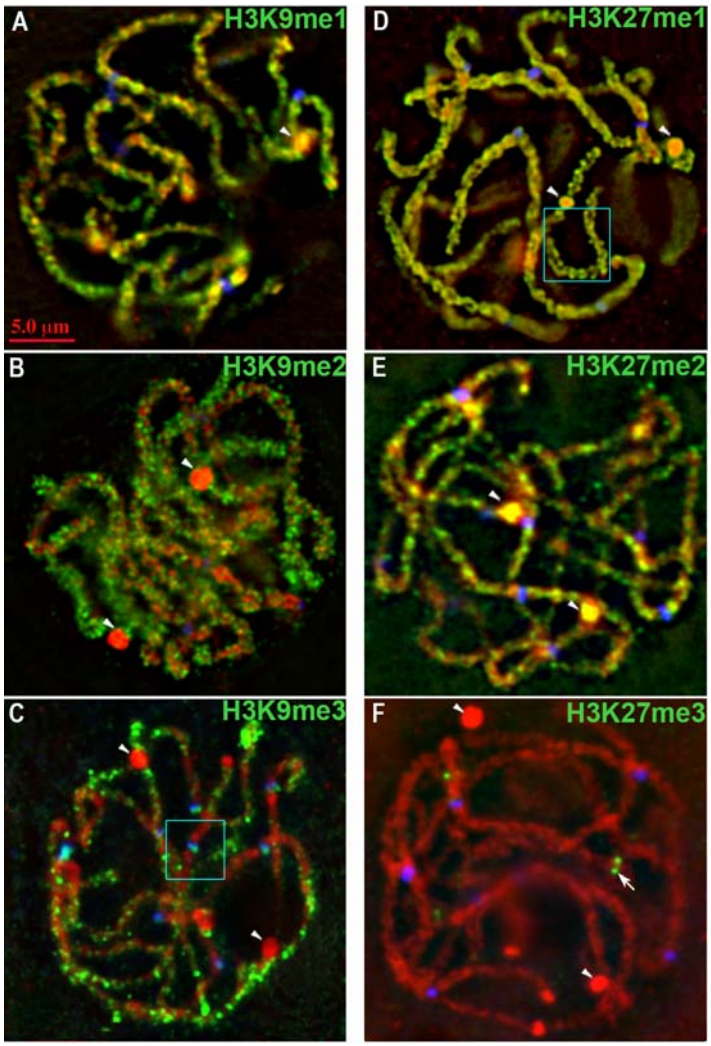

Figure 2.3. **Whole-nucleus distribution of histone H3K9 and K27 mono, di, and tri-methylation at pachytene.** Antibody staining is shown in green, DAPI in red, and CENP-C in blue. **A**) H3K9me1 stains heterochromatin in a spotty pattern. **B**) H3K9me2 stains euchromatin. **C**) H3K9me3 stains euchromatin weakly. A 2x magnification of centromere staining with this antiserum (boxed region) is shown in Figure 7C. **D**) H3K27me1 stains chromosomes uniformly. The boxed region is shown in a 2x magnification in Figure 4A. **E**) H3K27me2 stains heterochromatin. **F**) H3K27me3 staining is enriched in several focused domains (one is shown in the center, arrow). Arrowheads indicate knobs.

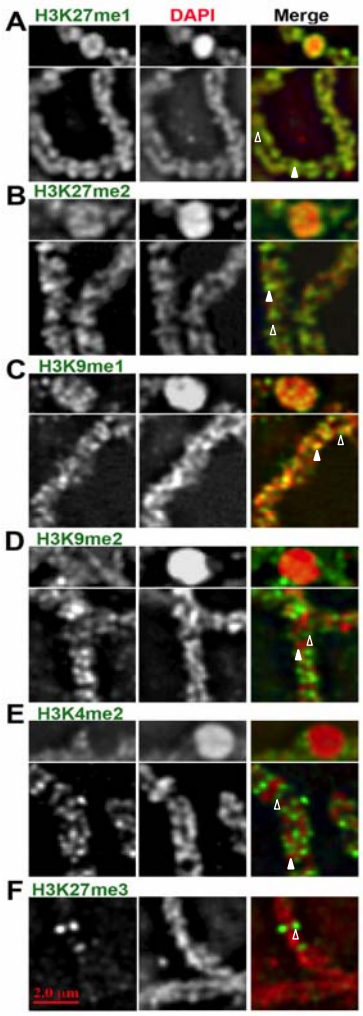

Figure 2.4. Histone methylation staining relative to chromomeres. **A**) H3K27me1 uniformly stains chromomeres and between-chromomere regions, but only stains the exposed surfaces of knobs. **B**) H3K27me2 is enriched in chromomeres. Staining occurs unevenly throughout knobs. **C**) H3K9me1 is enriched in chromomeres as well occasional between-chromomere regions. H3K9me1 staining is pronounced in the middle of knobs. **D**) H3K9me2 stains between-chromomere spaces and rarely within knobs. **E**) H3K4me2 only stains between and around chromomeres. **F**) H3K27me3-rich domains do not overlap with DAPI-rich chromomeres. Upper panels in A-E show the knob staining for each methylation. Closed arrowheads point to chromomeres, while open arrowheads point to between-chromomere spaces.

H3K9me1, H3K27me1 and H3K27me2 stain heterochromatin, but only H3K27me2 is enriched there

Pachytene staining patterns for the mono, di, and trimethylated forms of H3K9 and H3K27 are shown in Fig. 2.3. Anti-H3K27me1 and DAPI co-localized almost perfectly, both within and between chromomeres (Figs. 2.3D, 2.4A). The only exception was in knobs where staining was limited to exposed surfaces (at both pachytene and interphase; Figs. 2.4A, 2.5). Similarly, although H3K9me1 was abundant in pericentromeric heterochromatin and knobs (Figs. 2.3A, 2.5), there was consistent evidence of localized accumulation between chromomeres (Fig. 2.4C). These data suggest that H3K27me1 and H3K9me1 do not fall neatly into either the heterochromatin or euchromatin categories, since they are represented within both domains.

In contrast, chromosomes labeled for H3K27me2 revealed a pattern that closely matches expectations for a marker of classical heterochromatin. In whole pachytene nuclei (Fig. 2.3E) and straightened chromosomes (Figs. 2.1B, 2.2D), staining was not uniform but enriched in pericentromeres and knobs. H3K27me2 was also the only modification that strongly stained knobs in interphase (Fig. 2.5). Fine-scale analyses of chromomeres provided the expected co-localization of H3K27me2 with chromomeres, but unlike H3K27me1, it was not uniformly distributed between chromomeres. With few exceptions the between-chromomere space was either devoid of staining or stained very weakly (Fig. 2.4B).

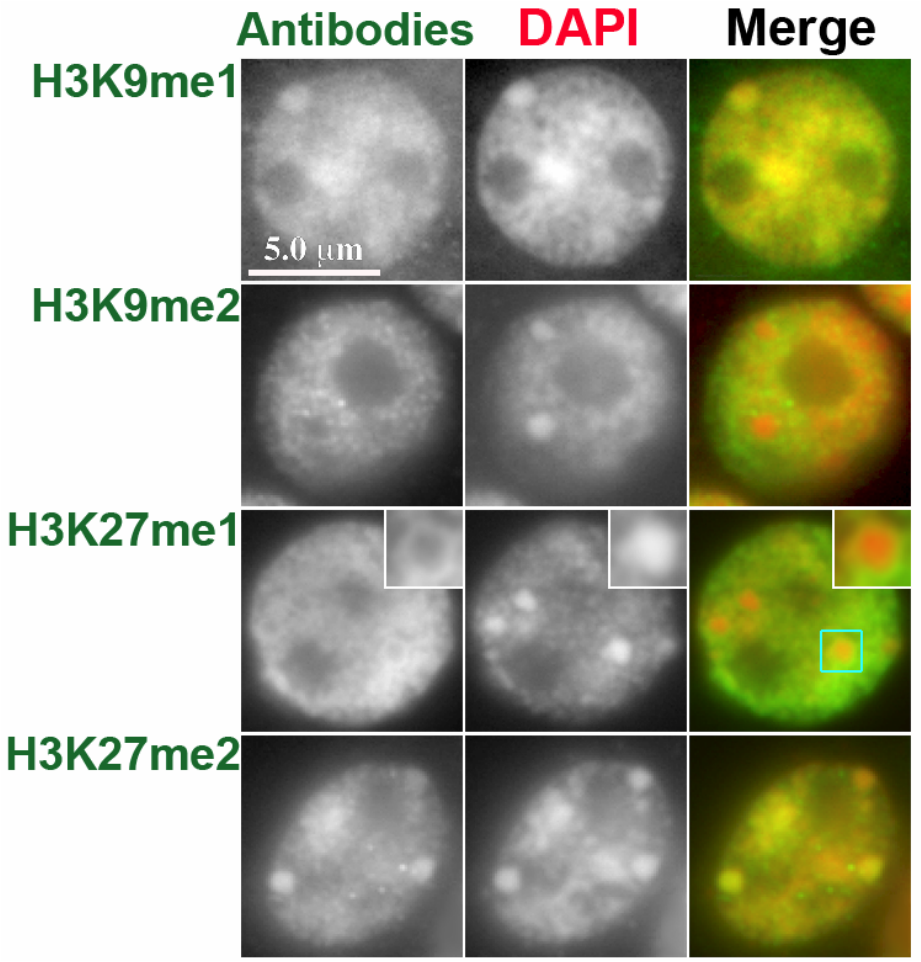

Figure 2.5. **Histone H3K9 and H3K27 mono and di-methylation in interphase cells. A**) H3K9me1 stains knobs weakly. **B**) H3K9me2 does not stain knobs. **C**) H3K27me1 stains the outer surface of knobs (inset is a 2x magnification of the knob). **D**) H3K27me2 stains knobs brightly.

As an independent test of both the heterochromatic distribution of H3K27me2 and the euchromatic distribution of H3K9me2, we doubly labeled cells for both markers (Fig. 2.6B). The data revealed very little overlap between the antisera: chromosomes appeared as a collage of red and green, with virtually none of the yellow color that indicates staining overlap. We can rule out the possibility that the results are confounded by antibody competition of some form (i.e. that the presence of one antibody excludes the binding of a second in the same vicinity). Control localizations using two different primary antibodies to the same epitope (H3K27me2) showed nearly perfect overlap (Fig. 2.6A).

K9-methylated H3 is intermingled with CENP-C in primary constrictions

In several species histone H3 is interspersed with CENH3 in alternating blocks, such that an extended array of CENH3-containing nucleosomes is followed by an array of histone H3-containing blocks, and so on (Blower et al., 2002; Sullivan and Karpen, 2004; Chueh et al., 2005; Yan et al., 2005b). Although H3K4me2 has been documented in both *Drosophila* and rice centromeres (Sullivan and Karpen, 2004; Yan et al., 2005b), we found no evidence by immunolocalization that H3K4me2 is present in maize centromeres. By our assays the major H3 modifications in centromeres are H3K9me2 and H3K9me3 (Figs. 2.3C, 2.7). Both marks are localized in the vicinity of the kinetochores, but surprisingly, we detected very little co-localization between CENP-C and either H3K9me2 or H3K9me3 (Fig. 2.7). Most of the K9-methylated histone H3 was detected just outside of the main concentrations of CENP-C (Fig. 2.7). These data suggest that the specialized histone H3 arrays around (and perhaps within) the kinetochore may not directly facilitate kinetochore assembly.

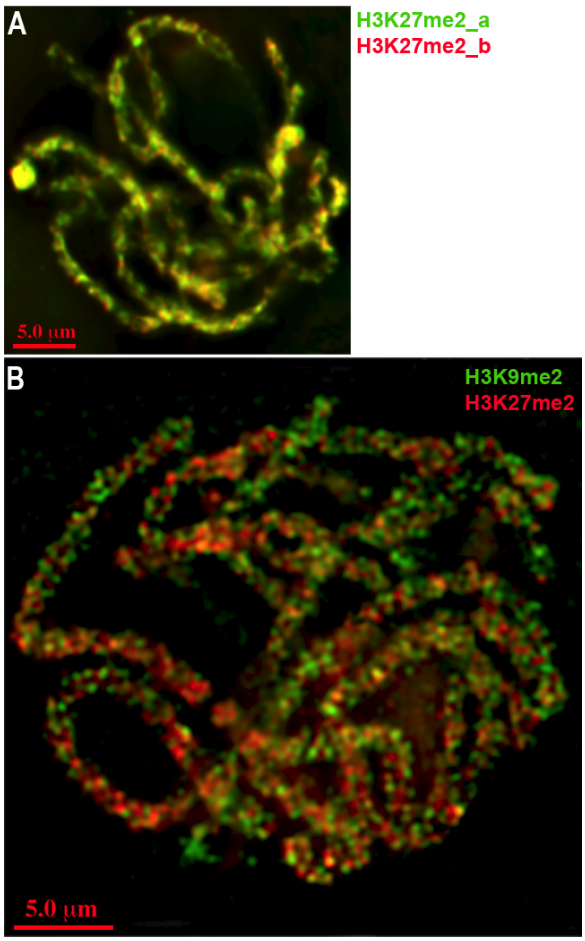

Figure 2.6. **H3K9me2 and H3K27me2 staining do not overlap. A**) Co-labeling control showing that two different antibodies against the same modification H3K27me2 (Upstate 07-452 from rabbit and Abcam 14222 from rat) when applied together, label the same regions of pachytene chromosomes. Minor non-concordance can be attributed to chromatic aberration. **B**) Co-labeling with anti-H3K9me2 and anti-H3K27me2 reveals no significant overlap.

We observed distinct variability among H3K9me2 antisera with respect to centromere staining patterns. One anti-H3K9me2 antiserum rarely stained kinetochores (upstate 05-768, Figs. 2.3B, 2.7B), one stained kinetochores more consistently (upstate 07-441), and one was almost entirely limited to kinetochores (upstate 07-212, Fig. 2.7A). Different labeling affinities could be caused by different chromatin accessibility/condensation features, or by other post-translational modifications that interfere with (or promote) the binding of the antibodies.

H3K9me3 and H3K27me3 are specialized euchromatic marks

Although the trimethylated form of lysine 27 marks inactive chromatin in animals (Plath et al., 2003; Cao and Zhang, 2004; Okamoto et al., 2004), in *Arabidopsis* it is localized to euchromatin (Mathieu et al., 2005). In maize, two anti-H3K27me3 antisera stained chromosomes weakly except for several very bright focused domains. These H3K27me3-rich domains did not overlap DAPI-rich chromomeres (Fig. 2.4F) and mapped cytologically to chromosomes 1, 2, 6 and 10 in the KYS inbred (Figs. 2.1C, 2.3F). Areas of rich H3K27me3 staining were found flanking centromeres, in mid-chromosome arm, and close to a telomere. Likewise in the W23 inbred (where chromosomes are not easily identified) there were at least three spots near centromeres and one at the end of a chromosome.

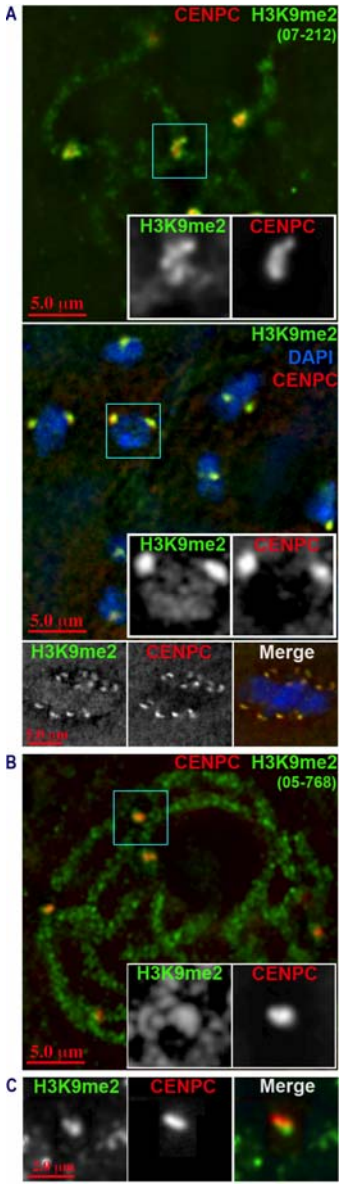

Figure 2.7. K9-methylated H3 is intermingled in primary constrictions but does not perfectly overlap with CENP-C. A) Anti-H3K9me2 (07-212) stains kinetochores very brightly in all stages of meiosis (pachytene, diakinesis and metaphase II are shown; blue represents DAPI). B) Anti-H3K9me2 (05-768) stains chromosome arms brightly but kinetochores weakly. C) Kinetochore staining with anti-H3K9me3 antisera. Insets in A and B are shown at higher magnification to reveal the imperfect co-localization with CENP-C.

In *Arabidopsis* the trimethylated state of lysine 9 (H3K9me3) has been reported to be absent (Jackson et al., 2004) or extremely rare (Johnson et al., 2004). However, we found clear evidence of H3K9me3 in pachytene cells (Fig. 2.3C), root tips cells (not shown) and western blots (Fig. 2.8B). The staining patterns resembled what was observed with H3K9me2 antisera, but with much weaker signal. H3K9me3 was undetectable in pericentromeres and knobs, and enriched in the distal portions of chromosome arms (Fig. 2.3C).

H4K20 di- and tri-methylation is undetectable in maize

Methylation of histone H4 lysine 20 (H4K20) is one of the most important markers of inactive chromatin in animals (Sims et al., 2003). However, we failed to detect any specific immunolocalization using antibodies to the mono, di or trimethylated forms of H4K20 (Fig. 2.8A). It is unlikely that the absence of staining is a consequence of improperly handled or inactive antisera. When the same antibodies to H4K20me3 were incubated with mouse neuronal cells (Fig. 2.8C), strong punctate staining coinciding with DAPI-rich pericentromeres was observed, consistent with prior results in the same species (Sarg et al., 2004).

For the di- and trimethylated forms of H4K20, the immunolocalization data were confirmed by western analysis of maize protein samples. Bands of the expected size were observed for nine of the H3 antibodies used in this study, and for an antibody to acetylated H4 (Fig. 2.8B). However, among the three anti-H4K20 antibodies, only those to anti-H4K20me1 revealed an expected 14 kD band (Fig. 2.8B). H4K20me2 and H4K20me3 were consistently undetectable in maize, although in mouse extracts the appropriate bands were clearly visible (Fig. 2.8B). The very poor representation of H4K20 suggests that it is not a major repressive mark in maize.

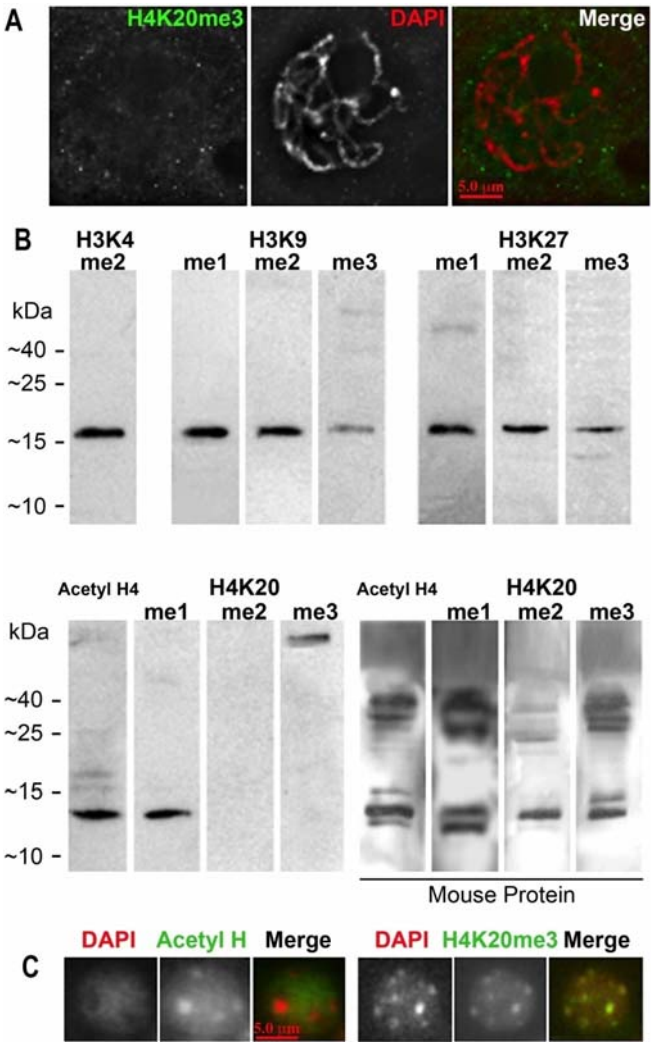

Figure 2.8. Methylation at H4K20 is rare in maize. **A**) No H4K20me3 signal was detected in maize by immunolocalization. **B**) Western analysis indicates that dimethylated H3K4, mono, di and trimethylated H3K9, mono, di and trimethylated H3K27 and monomethylated H4K20 exist in maize. Di- and trimethylated H4K20 are not detectable by protein blotting in maize, but are detectable in mouse extracts. **C**) In mouse cells, anti-acetyl H4 antibodies stain euchromatin (left), while anti-H4K20me3 antibodies (same as used in A) stained heterochromatic regions (right).

The primary K20 methylases in animals are SET8/PR-Set7 and Suv-20 (Schotta et al., 2004; Couture et al., 2005; Xiao et al., 2005). We failed to identify any significant sequence homology to SET8/PR-Set7 or Suv-20 in either the complete *Arabidopsis* or near-complete rice genomes. However, in *Arabidopsis*, SUVH2 can catalyze monomethylation at H4K20 (Naumann et al., 2005) and this may be responsible for the monomethylation we observed on western blots.

DISCUSSION

The term heterochromatin was first used by Heitz in 1928 to describe segments of chromosomes that stain deeply with DNA stains and which do not fully decondense in interphase (Zacharias, 1995). Using *Drosophila* as a model, Heitz went on to conclude that "..the density of genes in a chromosome is related to the longitudinal differentiation in euchromatin and heterochromatin. Euchromatic pieces are rich, whereas heterochromatic ones are at least poor in genes" (Heitz, 1934)as translated by Zacharias,(Zacharias, 1995). Maintaining this integrated view of genetics and chromosome structure has become increasingly difficult as cyto-based genetics has transitioned to DNA-based genomics. However, recent results showing that heterochromatin is marked by specific histone methylation events have the potential to bridge the divide between genomics and chromosome structure. In principle, assays for histone methylation at lysines 9 and 27 can provide high-resolution cytological markers, and add new clarity to the relationship between chromosome and gene.

Although both classical heterochromatin and histone modification are generally assayed by their staining intensities, heterochromatin and histone modification have not been quantified and carefully compared. As a result it is not clear how well histone methylation marks heterochromatic regions, which methylation states provide the best markers, or whether the modifications are conserved among organisms. Here we address these questions in maize, an important model for large-genome cereal grains and one of a handful of species with well-developed genetics and cytogenetics. Transposon-rich intergenic spaces in maize can extend for hundreds of kilobases (Chan et al., 2006), and it this context heterochromatin and euchromatin have clear foundations in differential condensation. Our data show that heterochromatin contains a relatively simple set of histone modifications that is distinct from the more complex mixture of 'on' and 'off' histone modifications that make up the gene-rich euchromatin space.

H3K27me2 marks classical heterochromatin

Mass spectrometry of *Arabidopsis* H3K9 and H3K27 revealed that the monomethylated forms were predominant, the dimethyl forms were less abundant, and that the trimethyl forms were present at levels several-fold lower (Johnson et al., 2004). H3K27me1 was present on more than 60% of all canonical histone H3 in inflorescence tissue (Johnson et al., 2004). Our observations using multiple antisera (Table 2.1) in several inbred lines generally confirm these conclusions. The intensity signals from anti-H3K27me1 antibodies were as bright or brighter than DAPI staining, while the H3K27me2 signals were less intense, and H3K27me3 showed limited localized staining and was difficult to detect on westerns (Fig. 2.8B). A similar trend was observed with the antibodies against the H3K9 mono, di and trimethylated epitopes.

One of the goals of our study was to determine which, if any, of the six methylation states at H3K9 and H3K27 accurately mark heterochromatin. Although three appear to mark heterochromatin in some capacity (Fig. 2.9: H3K9me1, H3K27me1, and H3K27me2) our analysis suggests that H3K27me2 is the only modification that marks heterochromatin specifically (Fig. 2.9). At pachytene, H3K27me2 is enriched in pericentromeres (Figs. 2.1B, 2.2, 2.3E), chromomeres (Fig. 2.4B), and knobs (Fig. 2.3E). It is also the only modification we tested that thoroughly stains knobs at interphase (Fig. 2.5) and the only marker that shows a clear reduction in staining between chromomeres at pachytene (Fig. 2.4B). At present H3K9me1 and H3K27me1 are difficult to interpret since very little is known about them from ChIP studies and they appear to stain cytologically condensed regions as well as non-condensed regions.

A cytological definition of euchromatin in maize

As a highly conserved marker of transcribed or 'poised' genes (Schneider et al., 2004; Alvarez-Venegas and Avramova, 2005), H3K4me2 provides an excellent molecular marker of the active gene space. Our data show that H3K4me2 staining complements the staining for H3K27me2; both in terms of trends along the linear axis of chromosome 9 (Fig. 2.2) and among the chromomeres that dot distal regions of chromosome arms (Fig. 2.4E). H3K4me2-stained chromatin lies not just between but also around H3K27me2-stained chromomeres, with some of the brightest H3K4me2 labeling over areas with the weakest DAPI staining (Fig. 2.4E). These data appear to confirm the interpretation first made by McClintock (McClintock, 1944) that the between-chromomere space is where the majority of genes are located in maize.

The distribution of H3K9 di and trimethylation within euchromatin

Prior data from *Arabidopsis* (Houben et al., 2003; Jackson et al., 2004; Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005) indicate that H3K9me2 is a heterochromatin mark. Houben et al. (2003) showed that in faba bean cells and other large-genome plants, H3K9me2 was also distributed towards the ends of chromosomes. The authors suggested that in large-genome plants, with many more retroelements, that the heterochromatin is more widely distributed along chromosomes. We show that the unexpected distribution of H3K9me2 in maize is not because heterochromatin is broadly distributed but because H3K9me2 is a euchromatic mark.

We make this conclusion based on three forms of data. General observations and trend analyses indicate that H3K9me2 is preferentially distributed towards chromosome ends where it is correlated with H3K4me2 and estimated gene frequency (Fig. 2.2). Also like H3K4me2, H3K9me2 staining is most intense in the between-chromomere gene space where DAPI staining is weakest (Fig. 2.4D). Finally and perhaps most convincingly, when H3K9me2 and the bona-fide heterochromatin mark H3K27me2 are labeled simultaneously there is very little overlap (Fig. 2.6B). Staining patterns were essentially identical with either of two independently generated anti-H3K9me2 antisera (Figs. 3B, 6B, 7). The fact that H3K9me2 can be observed at low frequency in pericentromeres and knobs (Figs. 4D, 9) is not contrary to the interpretation that it is a euchromatic mark. Small genic (or at least low copy sequences) are distributed at a low frequency in pericentromere and knobs, as suggested by the presence recombination proteins (visualized as recombination nodules at pachytene; (Anderson et al., 2003); summarized in Fig. 2E). The idea that H3K9me2 functions in euchromatin is also supported by several *Arabidopsis* studies: mutations of *kryptonite* (a Su(var)3-9 family histone methyltransferase) and *dnmt1*

(DNA methyltransferase 1) show significant reductions in H3K9me2 but no changes in the structure of pericentromeric heterochromatin (Jasencakova et al., 2003; Tariq et al., 2003; Jackson et al., 2004; Johnson et al., 2004).

The presence of both H3K4me2 and H3K9me2 in euchromatin can be interpreted as supporting the generally-held views that H3K4me2 marks active or poised genes (Santos-Rosa et al., 2002; Schneider et al., 2004), while H3K9me2 marks those that are temporally or spatially inactive (Jenuwein and Allis, 2001). The data further suggest that in maize most genes are at least poised for transcription, and that the 'on' and 'off' marks are frequently in close proximity. This interpretation is supported by recent ChIP data which suggest that active genes may have substantial amounts of H3K9me2, and vice versa, that inactive genes often contain H3K4me2 (Santos-Rosa et al., 2002; Bastow et al., 2004; Alvarez-Venegas and Avramova, 2005). Alvarez-Venegas et al. (Alvarez-Venegas and Avramova, 2005) showed that H3K4me2 and H3K9me2 coexisted at nearly equal levels in several *Arabidopsis* genes. This was a consistent result for inactive or moderately expressed sequences; only when gene expression was very high was there evidence that H3K4me2 dominated over H3K9me2.

H3K27me3 is limited to a small number of bright, focused euchromatic regions

In animals H3K27me3 is a marker of inactive chromatin (e.g. Plath et al., 2003; (Plath et al., 2003; Cao and Zhang, 2004; Okamoto et al., 2004) in *Arabidopsis* it marks euchromatin (Mathieu et al., 2005), and in maize H3K27me3 accumulates in a small number of highly focused domains (Fig. 2.3F). The H3K27me3-rich domains lie between chromomeres (Fig. 2.4F) and map to disparate regions, including adjacent to centromeres, mid-chromosome arm, and

telomere. It is not clear what these regions of the genome have in common, and comparisons to the genetic map provide no obvious clues. What sequences underlie the H3K27me3-rich spots remain a mystery, though we assume they are either clusters of specific repeats or clusters of co-regulated genes as in *Drosophila* (Ringrose et al., 2004).

Histone H3 in centromeres is di- and trimethylated at K9

Centromeres can be defined by their interactions with the specialized histone variant CENH3 and associated proteins such CENP-C (Dawe et al., 1999; Henikoff et al., 2001; Zhong et al., 2002). Prior data suggest that CENH3-containing nucleosomes are closely associated with intervening arrays of nucleosomes containing histone H3 (Blower et al., 2002; Chueh et al., 2005; Yan et al., 2005b). The centromere-embedded H3 may be specially modified to facilitate centromere/kinetochore function (Sullivan and Karpen, 2004) or may function as an extension of pericentromeric heterochromatin (Yan et al., 2005b); perhaps mediating chromatid cohesion (Bernard et al., 2001). Our data from maize show that the H3 modifications in centromeres are distinctly different from those in flanking pericentromeric heterochromatin. We observed clear centromere labeling with antisera to H3K9me2 and H3K9me3, but found no evidence of H3K9me1, H3K27me1, and H3K27me2, which are enriched in pericentromeres. Although resolution is limited at the sub-micron level of the centromere, our data further suggest that histone H3 may be only partially associated with the CENP-C-marked kinetochore, at least at pachytene (Fig. 2.7).

Differences in histone modifications between centromeres and pericentromeres are likely to reflect the transcriptional activity of the two domains (Jiang et al., 2003). While

pericentromeres are heterochromatic by our assays, maize centromeres are known to be transcribed (Topp et al., 2004), and as shown here, contain euchromatic histone modifications (H3K9me2 and H3K9me3). Available data from other species are consistent with this view. H3K4me2 is abundant in *Drosophila* centromeres ((Sullivan and Karpen, 2004) where it is presumed to mark a transcriptionally poised state. Rice and *Arabidopsis* centromeric DNAs are transcribed and associated with both H3K9me2 and H3K4me2 (May et al., 2005; Yan et al., 2005b). The relative abundance of centromeric H3K4me2 and H3K9me2 appears to vary among species, but it is not yet clear if these are biological differences or simply differences in the methods used (i.e. ChIP versus immunolocalization). An initial study in rice detected only H3K9me2 within centromere cores (Nagaki et al., 2004), but when more sensitive methods were used, H3K4me2 was also detected (Yan et al., 2005b).

Evolutionary liability of histone methylation patterns

The proposition that histone modification patterns might provide a universal code for interpreting eukaryotic gene activity (Strahl and Allis, 2000a; Jenuwein and Allis, 2001) has in many ways been borne out in recent years, at least with respect to the acetylation events that activate gene expression (e.g. (Ng et al., 2006). However, it was already apparent at the outset of our study that plants and animals differed with respect to the patterns of histone methylation in heterochromatin (Houben et al., 2003; Jackson et al., 2004; Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005). Because we focused our efforts on a large-genome plant and used pachytene chromosomes as our subject material, we are now able to confirm and extend these

comparisons to animals as well as within plants. The data suggest that the genome-wide localization of histone methylation patterns is much more variable than previously recognized.

A tabular comparison of the pericentromeric staining patterns maize, *Arabidopsis*, and human cells is shown in Table 2.2. Strikingly, of the three known histone methylation states in human pericentromeric heterochromatin, only one, H3K27me1, is also found in the corresponding regions of *Arabidopsis* and maize chromosomes (Jackson et al., 2004; Lindroth et al., 2004; Naumann et al., 2005). There is also an unexpected degree of variability within the angiosperms: two of the five marks present in *Arabidopsis* heterochromatin are rare or absent in the corresponding regions of maize. H3K9me2 is readily detectable in *Arabidopsis* heterochromatin (Houben et al., 2003; Jackson et al., 2004; Lindroth et al., 2004; Mathieu et al., 2005; Naumann et al., 2005) but is limited to euchromatin in maize. We also observed a near-complete absence of H4K20 methylation in maize, which is apparently prevalent in *Arabidopsis* (Naumann et al., 2005). H4K20me2 and H4K20me3 were undetectable by immunostaining or western blot in our experiments. H4K20me1 could be detected on western blots (Fig. 2.8B), but it is either insufficiently abundant or the epitope is insufficiently exposed to be detected immunocytochemically.

The data demonstrate that over evolutionary time there have been important shifts in the distribution of the mono, di, and trimethylated forms of histone H3 lysines 9 and 27. As yet there is no evidence that these shifts affect the basic function of H3K9 and H3K27 in gene silencing. However, the results do suggest that the presumed epigenetic code has a capacity to evolve along with changes in genome architecture. In maize where retroelements dominate as the most abundant form of repeat (SanMiguel and Bennetzen, 1998; Chan et al., 2006), histone marks that are specialized for retroelement inactivation (e.g. H3K27me2) may have played a

larger role in differentiating chromosomes. The distribution of marks that function primarily in gene silencing (e.g. H3K9me2) may have a more striking euchromatic distribution in maize only because there is a sharper differentiation of chromosome structure in this species. In the larger context our data suggest that genetic inactivity is not always manifested as heterochromatin, and that the repeat structure of an organism is likely to have had a major impact on the distribution and prevalence of histone methylation.

Table 2.2. Comparison of pericentromeric staining patterns in maize, *Arabidopsis*, and human cells.

Methylation State	Maize	*Arabidopsis*	Human
H3K9me1	+	+	
H3K9me2		+	
H3K9me3			+
H3K27me1	+	+	+
H3K27me2	+	+	
H3K27me3			
H4K20me1	**?***	+	
H4K20me2			
H4K20me3			+

*Detectable by western blot but not by immunofluorescence

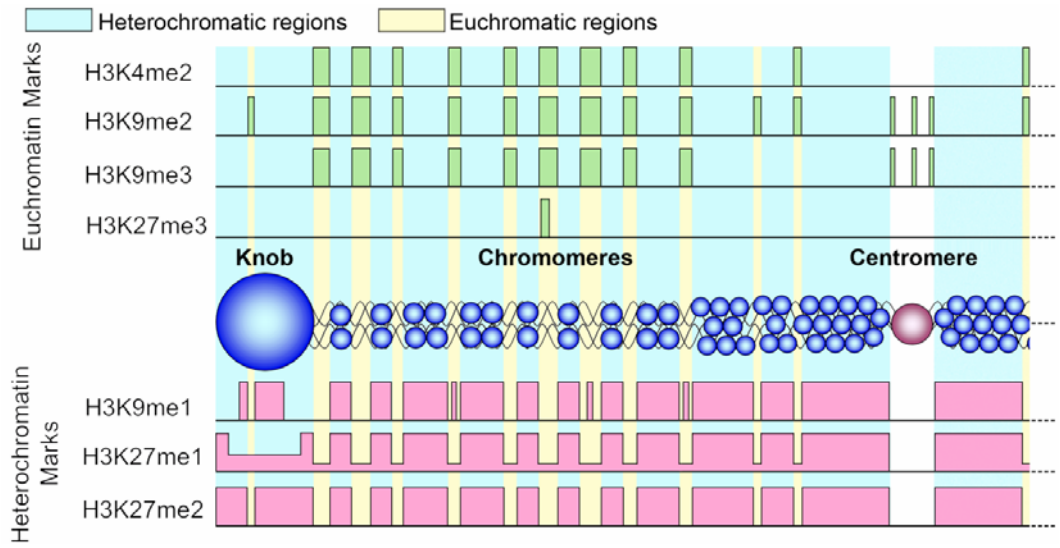

Figure 2.9. **Summary of histone methylation patterns in maize.** H3K4me2, H3K9me2 and H3K9me3 are distributed between chromomeres in the presumed euchromatin, with H3K9me2 also showing limited staining in knobs. H3K27me3 is enriched in specific euchromatic domains; one is shown here as an example, though in reality few chromosomes contain such domains. H3K9me1 is distributed in heterochromatic regions as well as a subset of the between-chromomere spaces. H3K27me1 is abundant in heterochromatin but is also uniformly present in between-chromomere spaces. H3K27me2 is the only marker that is found specifically heterochromatin.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation (0421619). We thank Dr. Thomas F. Murray and Dr. Zhengyu Cao for supplying materials and expertise during the analysis of mouse neuronal cells.

REFERENCES

Ahmad, K., and Henikoff, S. (2001). Centromeres are specialized replication domains in heterochromatin. J. Cell Biol. **153,** 101-109.

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

Alkan, C., Eichler, E.E., Bailey, J.A., Sahinalp, S.C., and Tuzun, E. (2004). The role of unequal crossover in alpha-satellite DNA evolution: a computational analysis. J Comput Biol **11,** 933-944.

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., and Zhang, Y. (2007). New nomenclature for chromatin-modifying enzymes. Cell **131,** 633-636.

Alvarez-Venegas, R., and Avramova, Z. (2005). Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants. Nucleic Acids Res **33,** 5199-5207.

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

Anderson, L.K., Salameh, N., Bass, H.W., Harper, L.C., Cande, W.Z., Weber, G., and Stack, S.M. (2004). Integrating genetic linkage maps with pachytene chromosome structure in maize. Genetics **166,** 1923-1933.

Anderson, L.K., Doyle, G.G., Brigham, B., Carter, J., Hooker, K.D., Lai, A., Rice, M., and Stack, S.M. (2003). High-resolution crossover maps for each bivalent of Zea mays using recombination nodules. Genetics **165,** 849-865.

Andolfatto, P., and Przeworski, M. (2001). Regions of lower crossing over harbor more rare variants in African populations of Drosophila melanogaster. Genetics **158,** 657-665.

Ardlie, K.G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. Nat Rev Genet **3,** 299-309.

Arumuganthan, K., and earle, E. (1991). Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9,** 208-218.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature **427,** 164-167.

Beadle, G.W. (1932). A Possible Influence of the Spindle Fibre on Crossing-Over in Drosophila. Proc Natl Acad Sci U S A **18,** 160-165.

Begun, D.J., and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature **356,** 519-520.

Bensasson, D., Zarowiecki, M., Burt, A., and Koufopanou, V. (2008). Rapid evolution of yeast centromeres in the absence of drive. Genetics **178,** 2161-2167.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev **12,** 142-148.

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

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.

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

Blower, M.D., and Karpen, G.H. (2001). The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat. Cell Biol. **3,** 730-739.

Brown, S.W. (1966). Heterochromatin. Science **151,** 417-425.

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.

Bulazel, K.V., Ferreri, G.C., Eldridge, M.D., and O'Neill, R.J. (2007). Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. Genome Biol **8,** R170.

Cambareri, E.B., Aisner, R., and Carbon, J. (1998). Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol. Cell Biol. **18,** 5465-5477.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr Opin Genet Dev **14,** 155-164.

Carroll, C.W., and Straight, A.F. (2006). Centromere formation: from epigenetics to self-assembly. Trends Cell Biol **16,** 70-78.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., and Wessler, S.R. (2000). Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci U S A **97,** 10083-10089.

Chan, A.P., Pertea, G., Cheung, F., Lee, D., Zheng, L., Whitelaw, C., Pontaroli, A.C., SanMiguel, P., Yuan, Y., Bennetzen, J., Barbazuk, W.B., Quackenbush, J., and Rabinowicz, P.D. (2006). The TIGR Maize Database. Nucleic Acids Res **34,** D771-776.

Charlesworth, B., Sneglowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371,** 215-220.

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

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., N.K., S., Tyagi, A.K., Soderlund, C., Dean, R.A., and Wing, R.A. (2002). An integrated physical and genetic map of the rice genome. Plant Cell **14,** 1-10.

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. (2002b). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14,** 1691-1704.

Chikashige, Y., Kinoshita, N., Nakaseko, Y., Matsumoto, T., Murakami, S., Niwa, O., and Yanagida, M. (1989). Composite motifs and repeat symmetry in S. pombe centromeres: direct analysis by integration of NotI restriction sites. Cell**,** 739-751.

Choo, K.H. (2000). Centromerization. Trends Cell Biol **10,** 182-188.

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

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks

promoters of transcriptionally active genes during mammalian cell division. EMBO Rep **6,** 354-360.

Chueh, A.C., Wong, L.H., Wong, N., and Choo, K.H. (2005). Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. Hum Mol Genet **14,** 85-93.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc.**,** 829-836.

Cloos, P.A., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K.H., and Helin, K. (2006). The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature **442,** 307-311.

Cooper, J.L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol **21,** 1712-1718.

Copenhaver, G.P., Browne, W.E., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. Proc Natl Acad Sci U S A **95,** 247-252.

Copenhaver, G.P., Keith, K.C., and Preuss, D. (2000). Tetrad analysis in higher plants. A budding technology. Plant Physiol **124,** 7-16.

Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., McCombie, W.R., Martienssen, R.A., Marra, M., and Preuss, D. (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286,** 2468-2474.

Couture, J.F., Collazo, E., Brunzelle, J.S., and Trievel, R.C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes Dev **19,** 1455-1465.

Csink, A.K., and Henikoff, S. (1998). Something from nothing: the evolution and utility of satellite repeats. Trends Genet. **15,** 200-204.

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

Dalal, Y., Furuyama, T., Vermaak, D., and Henikoff, S. (2007b). Structure, dynamics, and evolution of centromeric nucleosomes. Proc Natl Acad Sci U S A **104,** 15974-15981.

Davis, C.R., Kempainen, R.R., Srodes, M.S., and McClung, C.R. (1994). Correlation of the physical and genetic maps of the centromeric region of the right arm of linkage group III of Neurospora crassa. Genetics **136,** 1297-1306.

Dawe, R.K. (2005). Centromere renewal and replacement in the plant kingdom. Proc Natl Acad Sci U S A **102,** 11573-11574.

Dawe, R.K., and Hiatt, E.N. (2004). Plant neocentromeres: fast, focused, and driven. Chromosome Res **12,** 655-669.

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

Dawe, R.K., Sedat, J.W., Agard, D.A., and Cande, W.Z. (1994). Meiotic chromosome pairing in maize is associated with a novel chromatin organization. Cell **76,** 901-912.

Dawe, R.K., Reed, L., 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.

Devos, K.M., Ma, J., Pontaroli, A.C., Pratt, L.H., and Bennetzen, J.L. (2005). Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci U S A **102,** 19243-19248.

Doebley, J.F. (1990a). Molecular systematics of *Zea* (Gramineae). Maydica **35,** 143–150.

Doebley, J.F. (1990b). Molecular evidence and the evolution of maize. Econ. Bot. **44 (3 Suppl.),** 6-27.

Doenecke, D., Albig, W., Bode, C., Drabent, B., Franke, K., Gavenis, K., and Witt, O. (1997). Histones: genetic diversity and tissue-specific gene expression. Histochem Cell Biol **107,** 1-10.

Dooner, H.K., and He, L. (2008). Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. Plant Cell **20,** 249-258.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin **19,** 11-15.

Dravid, S.M., Baden, D.G., and Murray, T.F. (2005). Brevetoxin augments NMDA receptor signaling in murine neocortical neurons. Brain Res **1031,** 30-38.

Dutnall, R.N. (2003). Cracking the histone code: one, two, three methyls, you're out! Mol Cell **12,** 3-4.

Dvorak, J., Luo, M.C., and Yang, Z.L. (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species. Genetics **148,** 423-434.

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

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., and Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci U S A **95,** 4441-4446.

Ferreri, G.C., Liscinsky, D.M., Mack, J.A., Eldridge, M.D., and O'Neill, R.J. (2005). Retention of latent centromeres in the Mammalian genome. J Hered **96,** 217-224.

Fischle, W., Wang, Y., and Allis, C.D. (2003). Histone and chromatin cross-talk. Curr Opin Cell Biol **15,** 172-183.

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

Fogel, S., and Hurst, D.D. (1967). Meiotic gene conversion in yeast tetrads and the theory of recombination. Genetics **57,** 455-481.

Frary, A., Presting, G.G., and Tanksley, S.D. (1996). Molecular mapping of the centromeres of tomato chromosomes 7 and 9. Mol Gen Genet **250,** 295-304.

Fu, Y., Wen, T.J., Ronin, Y.I., Chen, H.D., Guo, L., Mester, D.I., Yang, Y., Lee, M., Korol, A.B., Ashlock, D.A., and Schnable, P.S. (2006). Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics **174,** 1671-1683.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci **11,** 199-208.

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

Fukui, K.N., Suzuki, G., Lagudah, E.S., Rahman, S., Appels, R., Yamamoto, M., and Mukai, Y. (2001). Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. Plant Cell Physiol **42,** 189-196.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci U S A **103,** 6428-6435.

Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem **281,** 559-568.

Hall, A.E., Keith, K.C., Hall, S.E., Copenhaver, G.P., and Preuss, D. (2004). The rapidly evolving field of plant centromeres. Curr Opin Plant Biol **7,** 108-114.

Hall, S.E., Luo, S., Hall, A.E., and Preuss, D. (2005). Differential rates of local and global homogenization in centromere satellites from Arabidopsis relatives. Genetics **170,** 1913-1927.

Hamblin, M.T., and Aquadro, C.F. (1999). DNA sequence variation and the recombinational landscape in Drosophila pseudoobscura: a study of the second chromosome. Genetics **153,** 859-869.

Harlow, E., and Lane, D. (1988). Antibodies, a laboratory manual. (Cold Spring Harbor Laboratory Press).

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell **118,** 715-729.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. Curr Opin Genet Dev **15,** 482-489.

Heitz. (1934). Die somatische Heteropyknose bei Drosophila melanogaster und ihre genetische Bedeutung.

Henikoff, S. (2002). Near the edge of a chromosome's 'black hole'. Trends Genet. **18,** 165-167.

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

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

Hosouchi, T., Kumekawa, N., Tsuruoka, H., and Kotani, H. (2002). Physical map-based sizes of the centromeric regions of Arabidopsis thaliana chromosomes 1, 2, and 3. DNA Res. **9,** 117-121.

Houben, A., Demidov, D., Gernand, D., Meister, A., Leach, C.R., and Schubert, I. (2003). Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA conten. Plant J. **33,** 967-973.

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

Hudson, R.R., and Kaplan, N.L. (1995). Deleterious background selection with recombination. Genetics **141,** 1605-1617.

Iizuka, M., and Smith, M.M. (2003). Functional consequences of histone modifications. Curr Opin Genet Dev **13,** 154-160.

Isenberg, I. (1979). Histones. Annu Rev Biochem **48,** 159-191.

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

Jackson, M.S., See, C.G., Mulligan, L.M., and Lauffart, B.F. (1996). A 9.75-Mb map across the centromere of human chromosome 10. Genomics **33,** 258-270.

Jaco, I., Canela, A., Vera, E., and Blasco, M.A. (2008). Centromere mitotic recombination in mammalian cells. J Cell Biol **181,** 885-892.

Jasencakova, Z., Soppe, W.J., Meister, A., Gernand, D., Turner, B.M., and Schubert, I. (2003). Histone modifications in Arabidopsis- high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. Plant J **33,** 471-480.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science **293,** 1074-1080.

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.

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

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res **32,** 6511-6518.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev **19,** 295-310.

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

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

Kearney, H.M., Kirkpatrick, D.T., Gerton, J.L., and Petes, T.D. (2001). Meiotic recombination involving heterozygous large insertions in Saccharomyces cerevisiae: formation and repair of large, unpaired DNA loops. Genetics **158,** 1457-1476.

Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature **442,** 312-316.

Kraft, T., Sall, T., Magnusson-Rading, I., Nilsson, N.O., and Hallden, C. (1998). Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (Beta vulgaris subsp. maritima). Genetics **150,** 1239-1244.

Lachner, M., O'Sullivan, R.J., and Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. J Cell Sci **116,** 2117-2124.

Lachner, M., Sengupta, R., Schotta, G., and Jenuwein, T. (2004). Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. Cold Spring Harb Symp Quant Biol **69,** 209-218.

Lamb, J.C., and Birchler, J.A. (2006). Retroelement genome painting: cytological visualization of retroelement expansions in the genera Zea and Tripsacum. Genetics **173,** 1007-1021.

Lamb, J.C., Meyer, J.M., and Birchler, J.A. (2007). A hemicentric inversion in the maize line knobless Tama flint created two sites of centromeric elements and moved the kinetochore-forming region. Chromosoma **116,** 237-247.

Lambie, E.J., and Roeder, G.S. (1986). Repression of meiotic crossing over by a centromere (CEN3) in Saccharomyces cerevisiae. Genetics **114,** 769-789.

Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J., Jones, R., and Jenkins, G. (2000). Retrotransposon evolution in diverse plant genomes. Genetics **156,** 313-325.

Lee, H.R., Neumann, P., Macas, J., and Jiang, J. (2006). Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol **23,** 2505-2520.

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.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D., and Hallauer, A. (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol **48,** 453-461.

Li, F., Huarte, M., Zaratiegui, M., Vaughn, M.W., Shi, Y., Martienssen, R., and Cande, W.Z. (2008). Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. Cell **135,** 272-283.

Liebman, S.W., Symington, L.S., and Petes, T.D. (1988). Mitotic recombination within the centromere of a yeast chromosome. Science **241,** 1074-1077.

Lin, B.Y., S.J. Chang, H.M. Lin. (2001). RFLP mapping of the centromere of chromosomes 1, 6 and 9 by B-A translocations inmaize. Bot. Bull. Acad. Sinica **42,** 273-279.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., and Jacobsen, S.E. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J **23,** 4286-4296.

Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature **431,** 364-370.

Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, E., and Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165,** 2117-2128.

Lo, A.W., Magliano, D.J., Sibson, M.C., Kalitsis, P., Craig, J.M., and Choo, K.H. (2001). A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res **11,** 448-457.

Luce, A.C., Sharma, A., Mollere, O.S., Wolfgruber, T.K., Nagaki, K., Jiang, J., Presting, G.G., and Dawe, R.K. (2006). Precise centromere mapping using a combination of repeat junction markers and chromatin immunoprecipitation-polymerase chain reaction. Genetics **174,** 1057-1061.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389,** 251-260.

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

Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). "Holo"er than thou: chromosome segregation and kinetochore function in C. elegans. Chromosome Res **12,** 641-653.

Mahtani, M.M., and Willard, H.F. (1998). Physical and genetic mapping of the human X chromosome centromere: repression of recombination. Genome Res **8,** 100-110.

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

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

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr Opin Genet Dev **15,** 163-176.

Mather, K. (1939). Crossing over and Heterochromatin in the X Chromosome of Drosophila Melanogaster. Genetics **24,** 413-435.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. Embo J **24,** 2783-2791.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A **99,** 6080-6084.

May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential Regulation of Strand-Specific Transcripts from Arabidopsis Centromeric Satellite Repeats. PLoS Genet **1,** e79.

McClintock, B. (1944). The relation of homozygous deficiencies to mutations and allelic series in maize. Genetics **29,** 478-502.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc. Natl. Acad. Sci. USA **101,** 1525-1530.

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.

Metcalfe, C.J., Bulazel, K.V., Ferreri, G.C., Schroeder-Reiter, E., Wanner, G., Rens, W., Obergfell, C., Eldridge, M.D., and O'Neill, R.J. (2007). Genomic instability within centromeres of interspecific marsupial hybrids. Genetics **177,** 2507-2517.

Miller, J.T., Dong, F., Jackson, S.A., Song, J., and Jiang, J. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150,** 1615-1623.

Mroczek, R.J., Melo, J.R., Luce, A.C., Hiatt, E.N., and Dawe, R.K. (2006). The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. Genetics **174,** 145-154.

Murphy, T.D., and Karpen, G.H. (1998). Centromeres take flight: Alpha satellite and the quest for the human centromere. Cell **93,** 317-320.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S.N., Milan, D., Ostrander, E.A., Pape, G., Parker, H.G., Raudsepp, T., Rogatcheva, M.B., Schook, L.B., Skow, L.C., Welge, M., Womack, J.E., O'Brien S, J., Pevzner, P.A., and Lewin, H.A. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science **309,** 613-617.

Nachman, M.W., Bauer, V.L., Crowell, S.L., and Aquadro, C.F. (1998). DNA variability and recombination rates at X-linked loci in humans. Genetics **150,** 1133-1141.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003a). Chromatin Immunoprecipitation reveals that the 180-bp satellite repeat Is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics **163,** 1221-1225.

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 reveals active genes. Nature Genet. **36,** 138-145.

Nagaki, K., Song, J., Stupar, R., Parokonny, A.S., Yuan, Q., Ouyang, S., Liu, J., Dawe, R.K., Buell, C.R., and Jiang, J. (2003b). Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics **163,** 759-770.

Nakaseko, Y., Adachi, Y., Funahashi, S.-i., Niwa, O., and Yanagida, M. (1986). Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J **5,** 1011-1021.

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.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T., and Reuter, G. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. Embo J **24,** 1418-1429.

Ng, D.W., Chandrasekharan, M.B., and Hall, T.C. (2006). Ordered Histone Modifications Are Associated with Transcriptional Poising and Activation of the phaseolin Promoter. Plant Cell **18,** 119-132.

O'Neill, R.J., Eldridge, M.D., and Metcalfe, C.J. (2004). Centromere dynamics and chromosome evolution in marsupials. J Hered **95,** 375-381.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science **303,** 644-649.

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.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev **22,** 1345-1355.

Peters, A.H., and Schubeler, D. (2005). Methylation of histones: playing memory with DNA. Curr Opin Cell Biol **17,** 230-238.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H., and Jenuwein, T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell **12,** 1577-1589.

Piperno, D.R., and Flannery, K.V. (2001). The earliest archaeological maize (Zea mays L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc Natl Acad Sci U S A **98,** 2101-2103.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science **300,** 131-135.

Pohl, M.E., Piperno, D.R., Pope, K.O., and Jones, J.G. (2007). Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres, Tabasco, Mexico. Proc Natl Acad Sci U S A **104,** 6870-6875.

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

Presting, G., Malysheva, L., Fuchs, J., and Schubert, I. (1998). A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16,** 721-728.

Preuss, D., Rhee, S.Y., and Davis, R.W. (1994). Tetrad analysis possible in Arabidopsis with mutation of the QUARTET (QRT) genes. Science **264,** 1458-1460.

Pusarla, R.H., and Bhargava, P. (2005). Histones in functional diversification. Core histone variants. FEBS J **272,** 5149-5168.

Rafalski, A., and Morgante, M. (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet **20,** 103-111.

Reinhart, R.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science **297,** 1831.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S.t. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A **98,** 11479-11484.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Mol Cell **12,** 1591-1598.

Ringrose, L., Ehret, H., and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol Cell **16,** 641-653.

Roizes, G. (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Res **34,** 1912-1924.

Rudd, M.K., and Willard, H.F. (2004). Analysis of the centromeric regions of the human genome assembly. Trends Genet **20,** 529-533.

Rudd, M.K., Wray, G.A., and Willard, H.F. (2006). The evolutionary dynamics of alpha-satellite. Genome Res **16,** 88-96.

Saffery, R., Sumer, H., Hassan, S., Wong, L.H., Craig, J.M., Todokoro, K., Anderson, M., Stafford, A., and Choo, K.H. (2003). Transcription within a functional human centromere. Mol. Cell **12,** 509-516.

SanMiguel, P., and Bennetzen, J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Annals Bot. **82,** 37-44.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science **274,** 765-768.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature **419,** 407-411.

Sarg, B., Helliger, W., Talasz, H., Koutzamani, E., and Lindner, H.H. (2004). Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. J Biol Chem **279,** 53458-53464.

Schindelhauer, D., and Schwarz, T. (2002). Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous alpha-satellite DNA array. Genome Res **12,** 1815-1826.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol **6,** 73-77.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev **18,** 1251-1262.

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

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science **294,** 109-114.

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

Shi, J., and Dawe, K. (2006). Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. Genetics**,** in press.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell **119,** 941-953.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell **4,** 481-495.

Sims, R.J., 3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet **19,** 629-639.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev **18,** 2437-2468.

Singh, K., Ishii, T., Parco, A., Huang, N., Brar, D.S., and Khush, G.S. (1996). Centromere mapping and orientation of the molecular linkage map of rice (Oryza sativa L.). Proc. Natl. Acad. Sci. USA **93,** 6163-6168.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science **191,** 528-535.

Song, R., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci U S A **100,** 9055-9060.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsi. EMBO J. **21,** 6549-6559.

Stephan, W. (1986). Recombination and the evolution of satellite DNA. Genet. Res. Camb. **47,** 167-174.

Stephan, W., and Langley, C.H. (1998). DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics **150,** 1585-1593.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgeral-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 in mitosis. Genes. Dev. **9,** 573-586.

Strahl, B.D., and Allis, C.D. (2000a). The language of covalent histone modifications. Nature **403,** 41-45.

Strahl, B.D., and Allis, C.D. (2000b). The language of covalent histone modifications. Nature **403,** 41-45.

Sullivan, B., and Karpen, G. (2001). Centromere identity in Drosophila is not determined in vivo by replication timing. J Cell Biol **154,** 683-690.

Sullivan, B., Blower, M., and Karpen, G. (2001). Determining centromere identity: cyclical stories and forking paths. Nat. Rev. Genet. **2,** 584-596.

Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. Nat Genet **20,** 227-228.

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

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

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

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol. Biol. Cell **3,** 819-835.

Talbert, P., R Masuelli, Tyagi, A., Comai, L., and 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., and Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. J Biol **3,** 18.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and J., J.P. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. USA **100,** 8823-8827.

Tenaillon, M.I., U'Ren, J., Tenaillon, O., and Gaut, B.S. (2004). Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol **21,** 1214-1225.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., and Gaut, B.S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci U S A **98,** 9161-9166.

Thatcher, T.H., and Gorovsky, M.A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res **22,** 174-179.

Thomas, J.O., and Kornberg, R.D. (1975). An octamer of histones in chromatin and free in solution. Proc Natl Acad Sci U S A **72,** 2626-2630.

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

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature **439,** 811-816.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet **3,** e86.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays **22,** 836-845.

Turner, B.M. (2002). Cellular memory and the histone code. Cell **111,** 285-291.

Van Hooser, A.A., Ouspenski, II, 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.

Ventura, M., Antonacci, F., Cardone, M.F., Stanyon, R., D'Addabbo, P., Cellamare, A., Sprague, L.J., Eichler, E.E., Archidiacono, N., and Rocchi, M. (2007). Evolutionary formation of new centromeres in macaque. Science **316,** 243-246.

Villasante, A., Abad, J.P., and Mendez-Lago, M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci U S A **104,** 10542-10547.

Walsh, J. (1987). Peristence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics **115,** 553-567.

Wang, Q., and Dooner, H.K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci U S A **103,** 17644-17649.

Warburton, P.E., and Willard, H.F. (1992). PCR amplification of tandemly repeated DNA: analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human alpha satellite DNA. Nucleic Acids Res **20,** 6033-6042.

Warburton, P.E., Haaf, T., Gosden, J., Lawson, D., and Willard, H.F. (1996). Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. Genomics **33,** 220-228.

Wegel, E., and Shaw, P. (2005). Gene activation and deactivation related changes in the three-dimensional structure of chromatin. Chromosoma **114,** 331-337.

Wevrick, R., and Willard, H.F. (1989). Long-range organization of tandem arrays of a satellite DNA at the centromeres of human chromosomes: High-frequency array-length polymorphism and meiotic stability. Proc. Natl. Acad. Sci. USA **86,** 9394-9398.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell **125,** 467-481.

Xiao, B., Jing, C., Kelly, G., Walker, P.A., Muskett, F.W., Frenkiel, T.A., Martin, S.R., Sarma, K., Reinberg, D., Gamblin, S.J., and Wilson, J.R. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes Dev **19,** 1444-1454.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell **125,** 483-495.

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

Yan, H., Jin, W., Nagaki, K., Tian, S., Ouyang, S., Buell, C.R., Talbert, P.B., Henikoff, S., and Jiang, J. (2005a). Transcription and Histone Modifications in the Recombination-Free Region Spanning a Rice Centromere. Plant Cell.

Yan, H., Jin, W., Nagaki, K., Tian, S., Ouyang, S., Buell, C.R., Talbert, P.B., Henikoff, S., and Jiang, J. (2005b). Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell **17,** 3227-3238.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178,** 539-551.

Zacharias, H. (1995). Emil Heitz (1892-1965): chloroplasts, heterochromatin, and polytene chromosomes. Genetics **141,** 7-14.

Zhang, W., Lee, H.R., Koo, D.H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell **20,** 25-34.

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.

Zhang, X.Y., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. Plos Biology **5,** 1026-1035.

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.

CHAPTER 3

RECOMBINATION WITHIN CENTROMERE CORES

Jinghua Shi and R. Kelly Dawe. To be submitted to Nature Genetics.

ABSTRACT

Centromeres are constant features of all eukaryotic genomes, yet they are the most dynamic at the genomic level. Here we describe a novel CENH3 ChIP-display method (based on CRM2) that maps kinetochore footprints over high-resolution recombination maps. Each of the ten centromeres was genetically mapped using a set of 264 CRM2 markers, 57.2% of which interact with CENH3. Multiple, sequenced, polymorphic markers span each centromere. Segregation mapping revealed that at least two within-centromere gene conversion events occurred in a maize mapping population (in an 11 generation period). The complete physical map of centromere 2 allowed us to provide a first measure of linkage disequilibrium (LD) within a centromere. The data show an absence of LD over TE-rich regions, and suggest that the primary mode of LD decay is short conversion tracks. A region of the centromere containing the major tandem repeat in maize centromeres, CentC, shows a different pattern of evolution and LD extending to 300-400 kb. Our data suggest that frequent conversion is an important mechanism for spreading sequence variants among homologous centromeres. The unusually high degree of LD in proximity to the primary CentC, and the abundance of CentC in the closest relatives of maize further suggest that CentC has been under selection in recent history.

INTRODUCTION

Centromere is the site for kinetochore assembly and spindle attachment. The presence of a centromere-specific histone H3 (CENH3) demarcates CEN chromatin from the flanking heterochromatin. Although functionally conserved, centromeres are the most dynamic regions of

complex genomes. Even among closely related species, the centromeric DNA components, copy number and structural organization are highly variable (Warburton et al., 1996; O'Neill et al., 2004; Lee et al., 2005; Murphy et al., 2005). The apparent conflict between essentiality and sequence dispensability remains one of the major unresolved paradoxes in genetics (Henikoff et al., 2001).

Satellite repeats and transposable elements are the major components of centromeric DNA sequences in most eukaryotes. For instance, human centromeres are composed primarily of 171-bp alpha satellite (Schueler et al., 2001), and *Arabidopsis* centromeres are rich in180-bp satellite repeats (Copenhaver et al., 1999; Hosouchi et al., 2002). However, in other species, the simple satellite is not the major repeat, and the bulk of the centromere is composed of transposon-derived sequences (Cambareri et al., 1998; Fukui et al., 2001). An example is maize. A 156-bp tandem repeat CentC is present in all centromeres but a *Ty3/gypsy* retrotransposon known as CRM is more abundant and consistent in copy number among centromeres (Presting et al., 1998; Nagaki et al., 2003b).

Centromeres are specified epigenetically by the presence of centromeric specific histone H3 variant, CENH3 (Palmer et al., 1987; Karpen and Allshire, 1997). In both animals and plants, CENH3 nucleosomes are not distributed continuously, but interspersed with H3 nucleosomes (Blower et al., 2002; Yan et al., 2008). Delimiting the precise boundaries of centromeres is another challenge for centromere mapping: the same DNA components constitute both the functional centromeres and the flanking regions as well (Schueler et al., 2001; Zhong et al., 2002; Yan et al., 2005a). CENH3 is the only reliable marker of functional centromeric DNA, and antisera to CENH3 have been used to identify centromere repeats and single-copy markers in maize (Zhong et al., 2002; Luce et al., 2006).

Centromeres were found to severely repress meiotic recombination in the 1930's (Beadle, 1932; Mather, 1939). Later evidence demonstrated that the suppression of recombination is a common feature of centromere in all eukaryotes (Lambie and Roeder, 1986; Nakaseko et al., 1986; Davis et al., 1994; Jackson et al., 1996; Mahtani and Willard, 1998; Copenhaver et al., 1999; Anderson et al., 2003). In genetic maps, centromeres are defined as the regions where frequency of crossovers approaches zero (Copenhaver et al., 1998; Copenhaver et al., 1999; Yan et al., 2005a; Fu et al., 2006).

Although classical crossing over is limited, other mechanisms of genetic exchange are presumed to occur (Smith, 1976; Charlesworth et al., 1994). Genetic exchanges within and between sister chromatids have been identified experimentally and by sequence analysis. For instance, a high frequency of mitotic recombination in centromeres has been documented in several species (Liebman et al., 1988; Jaco et al., 2008); large intrachromosomal segmental duplication was detected in rice centromeres (Ma and Bennetzen, 2006); and sequence comparison among human centromeres supported the hypothesis of unequal crossover between sister chromatids (Roizes, 2006). Although such exchanges are likely to account for much of the amplification and or removal of local repeats, interchromosomal exchanges are also essential for allowing sequence variants to spread between homologs. There have been no interchromosomal genetic exchanges detected within centromeres, but this observation is not necessarily meaningful because in most species there are very few markers within centromeres that could be used to assay for exchange.

The highly repetitive nature of centromeres makes them difficult to study by modern sequencing or mapping methods. This is a particularly acute problem within tandem repeat arrays but can be overcome in species such as maize that have centromeres with densely

populated retrotransposons. Over time the CR elements cluster around and within each other to produce a layered nested arrangement of TEs (SanMiguel et al., 1996; Miller et al., 1998; Presting et al., 1998; Nagaki et al., 2003b; Ma and Bennetzen, 2006). Each insertion site has a high probability of being unique (Devos et al., 2005; Luce et al., 2006). Primers can be designed to flank the insertion points and amplify single copy markers. The insertion points are highly polymorphic among inbreds and can be used for centromere mapping (Luce et al., 2006).

Maize (*Zea mays ssp. mays*) has long been known for dramatic genetic diversity. Archaeological and molecular evidence indicates that modern maize was originated from a single domestication event from Balsas teosinte, *Zea mays ssp. parviglumis* (parviglumis hereafter) (Piperno and Flannery, 2001; Matsuoka et al., 2002). Although cultivated maize underwent a domestication bottleneck, maize inbreds still retain up to 80% diversity from the open-pollinated landraces and ~60% diversity from parviglumis (Tenaillon et al., 2001; Liu et al., 2003). Extensive polymorphisms among maize inbreds are also accumulated by rearrangement and transposon insertions (Song and Messing, 2003; Wang and Dooner, 2006; Dooner and He, 2008). The high level of polymorphisms facilitates generating centromeric markers and detecting genetic exchanges within centromere.

In this study, we used CRM2 transposon display to generate high density of centromeric markers, and further overlaid functional centromeres on the genetic map using CENH3 ChIP. The abundance of our centromere markers led us to the discovery of interchromosomal conversion-type genetic exchanges within centromeres. Then we took advantage of the completed physical map of centromere 2 to measure linkage disequilibrium (LD, a measure of historic genetic exchange) within the centromere core. The results revealed that in proximity to the major tandem repeat array CentC, LD correlates with physical distance. Outside of the simple

repeat array domain, LD dissolves due to frequent short gene conversion-like events. Presumably as a consequence of rapid intra-centromere exchange, the number of different centromere haplotypes is higher than expected, with many centromeres having as little as 10-20% homology overall. The data suggest that frequent conversion is an important mechanism for spreading sequence variants among homologous centromeres. The unusually high degree of LD in proximity to the primary CentC arrays suggests that CentC has been under selection in recent history.

MATERIALS AND METHODS

Genetic stocks

A ninety-four line IBM DNA Kit, provided by the Maize Genetics Cooperation Stock Center (http://www.maizemap.org/94_ibm.htm), was used for CRM2 display. MO005 and MO007 were excluded from the analysis due to high levesl of heterozygosity (four and six centromeres were heterozygous, respectively). Additional accessions of IBM lines used for confirmation and further ChIP and FISH analysis were obtained from the MaizeGDB stock center (http://www.maizegdb.org/stock.php).

A set of 53 maize inbred lines, including the 50-line core set (Liu et al., 2003) with additional lines within the NAM (nested association mapping) founder lines (Yu et al., 2008), was chosen to represent the genetic diversity for LD analysis. A272, CML14, CML91, CML281, CML349, CO159, NC350 and NC364 were not included due to seed unavailability or poor germination. All lines including B73 and Mo17 were obtained from USDA. DNA was extracted from 3-week-old seedlings using a modified CTAB protocol (Doyle and Doyle, 1987).

Transposon Display

Transposon display was carried out as described in Casa et al. (Casa et al., 2000) with following modifications. The full-length sequence of CRM2 was obtained from NCBI (GenBank AY129008). Primers were designed to specifically amplify the flanking sequences of CRM2 but not other CRM families. Genomic DNA was digested using BfaI. The primers for primary amplification were CRM2_R1 (5'- GAGGTGGTGTATCGGTTGCT) and BfaI + 0 (5'- GACGATGAGTCCTGAGTAG), and for selective amplification were P^{33} or FAM labeled CRM2_R2 (5'- CTACAGCCTTCCAAAGACGC) and BfaI + 3 selective bases (5'- GACGATGAGTCCTGAGTAG + ACC/AGC/TAT/TTC/TCG/TGC/GAC/GCA). The final annealing temperature for selective amplification was 58 °C. The PCR products were electrophoresed on 6% polyacrylamide gels. For further validating the bands that had aberrant scores, an extra selective base was added for selective amplification.

Mapping of CRM2 markers

All polymorphic bands that showed a Mendelian segregation ratio were scored. The data were initially sent to a community IBM mapping service (CIMDE) which constructed a linkage map using two-point mapping method ("build" and "place" MapMaker commands) from a framework of 580 loci. After getting the rough positions in this way, we further constructed a finer "multi-point" map (hereafter referred to as "centromere map") for each chromosome using MapMaker ver. 3.0. In each centromere map, mapping scores for 20 flanking markers from the IBM2 2008 Neighbors linkage interpretation (www.maizegdb.org) were added to the file of centromere markers scores. The closest IBM2 core bin markers were added as the first and last marker for each centromere map. In addition, we included as many "skeleton" markers (ISU

map4, (Fu et al., 2006)) as possible. The CRM2 markers were then placed into centromere framework using a multi-point method (the "try" MapMaker command). In each centromere map, the IRIL lines that showed heterozygosity or contamination for that centromere were excluded from the mapping data set (Table 3.1). Heterozygosity is 1.74% and contamination is 0.65% in the IBM population.

Table 3.1. Removed IBM lines for each centromere.

Centromere	Heterozygosity	Contamination
1	None	None
2	15	58
3	12, 20, 30	11
4	47, 85	11, 27
5	86	35
6	34, 36, 44	None
7	61	None
8	58	None
9	22, 66, 85	None
10	None	35

Identifying CENH3-associated markers by ChIP-display

Native chromatin immunoprecipitation (ChIP) was carried out as described in Topp et al. (Topp et al., 2004) with minor modifications. Chromatin was extracted from young kernels (8~15 cm) or young roots (~ 1 week after germination). RNase-free DNase I (Promega) was utilized for chromatin digestion. Chromatin was digested to ~300-3000 bp fragments. To reduce nonspecific binding, the digested chromatin was precleared with 50% protein A Sepharose (Amersham Pharmacia Biotech) for 3 hours. After immunoprecipitation with anti-CENH3 antisera (Zhong et al., 2002), the supernatant (unbound) and IP (bound) fractions were purified with a PCR purification Kit (invitrogen). The purified DNAs were then used for CRM2 transposon display. For both B73 and Mo17, ChIP-display was replicated three times. Bands

that were amplified in the IPed DNAs from all three experiments were considered to be associated with centromere cores.

Recovery and sequencing of CRM2 markers

Sixty-four CRM2 bands were excised from TD gels and re-amplified with primer set BfaI + 0 and CRM2_R2. The PCR products were purified using QIAGEN Gel Purification kit and were either directly sequenced or cloned into a TOPO TA vector (Invitrogen, Carlsbad, CA) and then sequenced. As controls for the ChIP-display method, 33 of the same bands, except from the ChIP-display (IP) lanes, were also cloned and sequenced. Among the 33 sequenced IP bands, 31 exactly matched their counterparts from the corresponding inbred, and were considered to be associated with CENH3.

FISH

FISH on mitotic cells was performed as described in Kato et al. (Kato et al., 2004). The following four repetitive DNA sequences were included in the probe cocktail: subtelomeric 4-12-1 (labeled with fluorescein-12-dCTP or FITC), CRM2 LTR (labeled with FITC), CentC (labeled with Texas red-5-dCTP) and knob180 (labeled with Texas red). The clones of 4-12-1, CentC and knob180 were generously provided by Dr. Birchler (University of Missouri). The CRM2 LTR was PCR amplified from genomic DNA using the following primer set: forward, 5'-TCGTCAACTCAACCATCAGGT and reverse, GCAAGTAGCGAGAGCTAAACTTGA. All images were captured and processed using Zeiss Axio Imager and SlideBook 4.0 software (Intelligent Imaging Innovations, Denver).

RESULTS

Generating unique centromeric markers using CRM2-display

In maize centromeres there are thousands of TEs of the CRM family that fall into four major subclasses. The oldest is CRM4, which dates to the earliest detectable stages of centromere evolution, and the youngest is CRM1, which is probably still active on a limited basis. In the intervening period CRM2 proliferated, beginning a dramatic expansion roughly 2 mya and continuing to dominate for another million years (the last known transposition was roughly 1 mya). TEs in this age bracket retain >90% of their initial sequence identity but are often interrupted by insertions and deletions (often caused by more recently active TEs). Thus CRM2 has the features of an excellent polymorphism tag, being conserved enough to easily identify but still lying within a diverse area of the genome.

Transposon display (known as TD; see (Casa et al., 2000)) makes it possible to capture such transposon-induced polymorphisms. In this method a transposon-specific primer is paired with a restriction site adapter, and the resulting PCR products resolved on a polyacrylamide gel. When we used TD to display all the CRM2 elements, we found the number of products exceeded what could be displayed. Therefore to make the data manageable, three selective bases were added to the adapter primer such that only 1/64 of the total number of bands was amplified in any given experiment. The data suggest that there are roughly 1500 CRM2 elements in the maize genome and that there is extreme level of polymorphism (~ 80%) between inbreds B73 and Mo17 (Fig. 3.1).

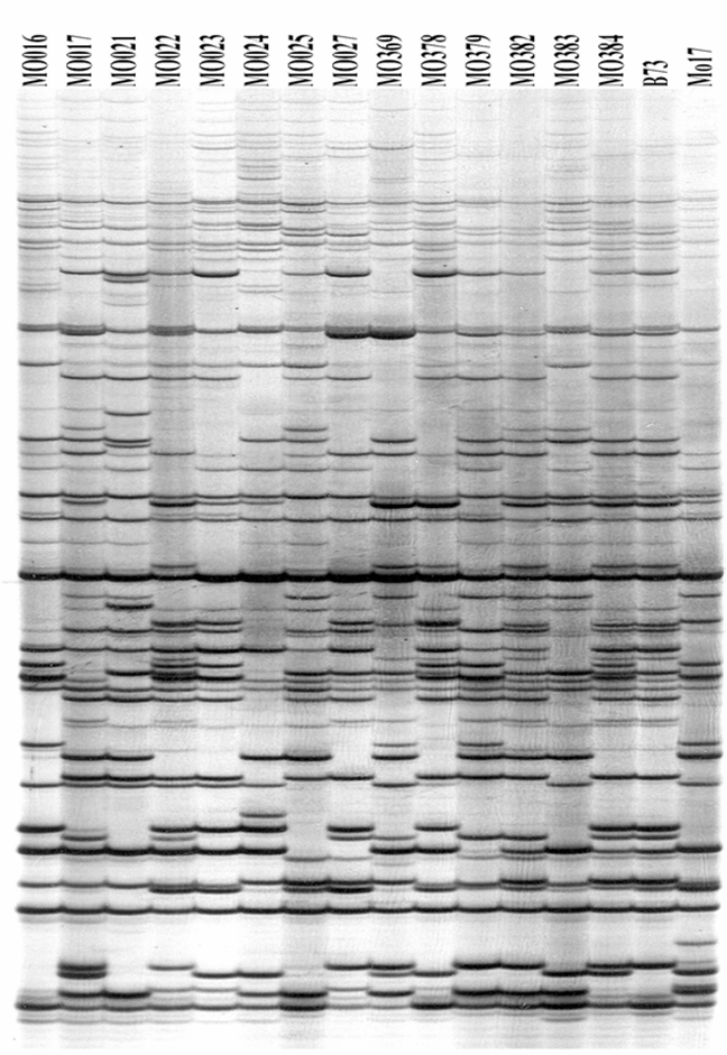

Figure 3.1. Autoradiographs of CRM2 display. CRM2 display in the two inbreds B73 and Mo17 and IBM mapping population, using BfaI + AGC and CRM2_R2 for selective amplification. This pair of primers generated 55 polymorphic bands, among which 41 were converted to mappable markers.

A genetic map of maize centromeres

The IBM (intermated B73 x Mo17) population was built by Lee and coworkers to facilitate the genetic mapping in maize (Lee et al., 2002). It is similar to a standard recombinant inbred population, except that intercrossing was used to increase the total number of recombinants to roughly four times higher than would be observed in a single generation (Lee et al., 2002). A total of 283 CRM2 markers generated from eight combinations of primers were scored in the 94 IBM RILs. All markers were first placed on a standard backbone map by a 2-point strategy. This first-pass mapping step revealed that nine of the markers mapped onto chromosome arms in gene-rich regions of the genome. The remaining 274 markers fell into ten unlinked clusters that identified each maize centromere. We then compiled high-density maps of known markers in the vicinity of each centromere, and fine-mapped the 274 new markers using a multipoint mapping algorithm (Fig. 3.2). Ten additional markers were shown to map outside of the main marker clusters by this method (by 1-5 cM). These 10 markers were classified as pericentromeric.

Based on FISH analysis, we expected the number of CRM2 markers to differ slightly among centromeres and among homologous centromeres from different inbreds (James Birchler, unpublished correspondence). As can be seen in Table 3.2, these expectations were borne out in the mapping data. Although all centromeres have markers derived from both parents, chromosome 1 has 2 Mo17-type markers but 11 B73-type markers. The opposite situation occurs in ch9, which has 18 Mo17-type markers but only 2 B73-type markers. In general the number of polymorphic markers correlates with the intensity of the FISH signal. For instance, centromeres 2 and 8, with the most B73-type makers (34 and 26, respectively) have the brightest FISH signal using a CRM2 probe.

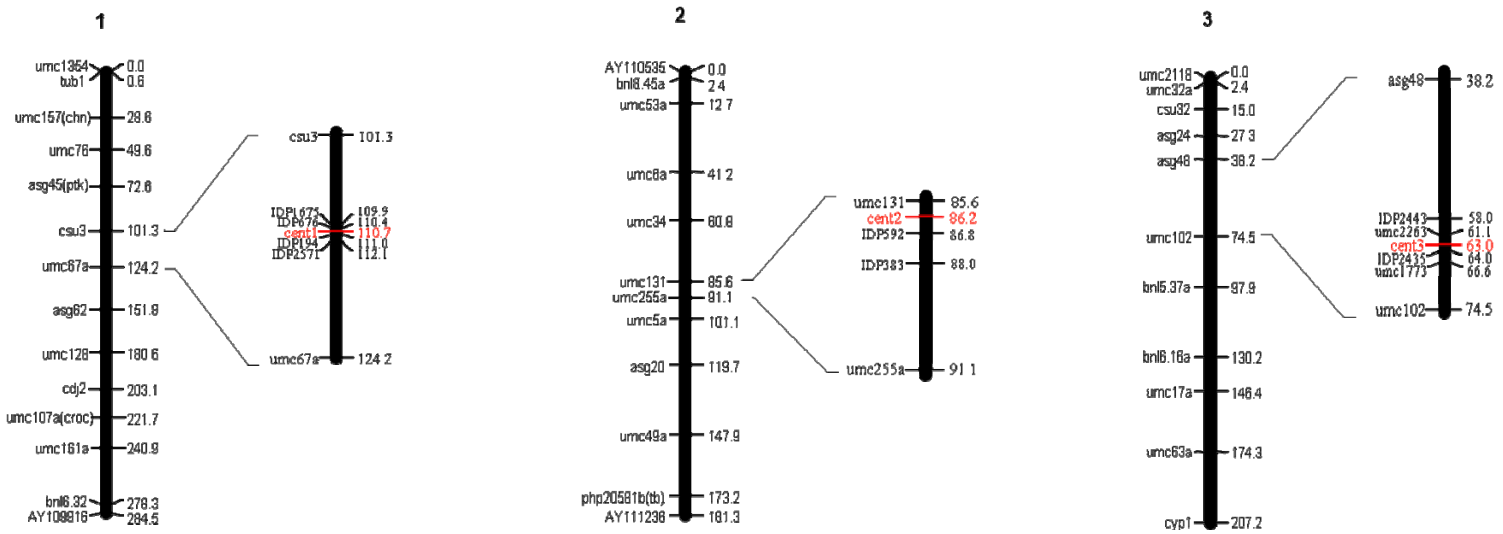

Figure 3.2. The map position of CRM2 marker clusters on the IBM2 neighbor Map. (*To be continued***)**

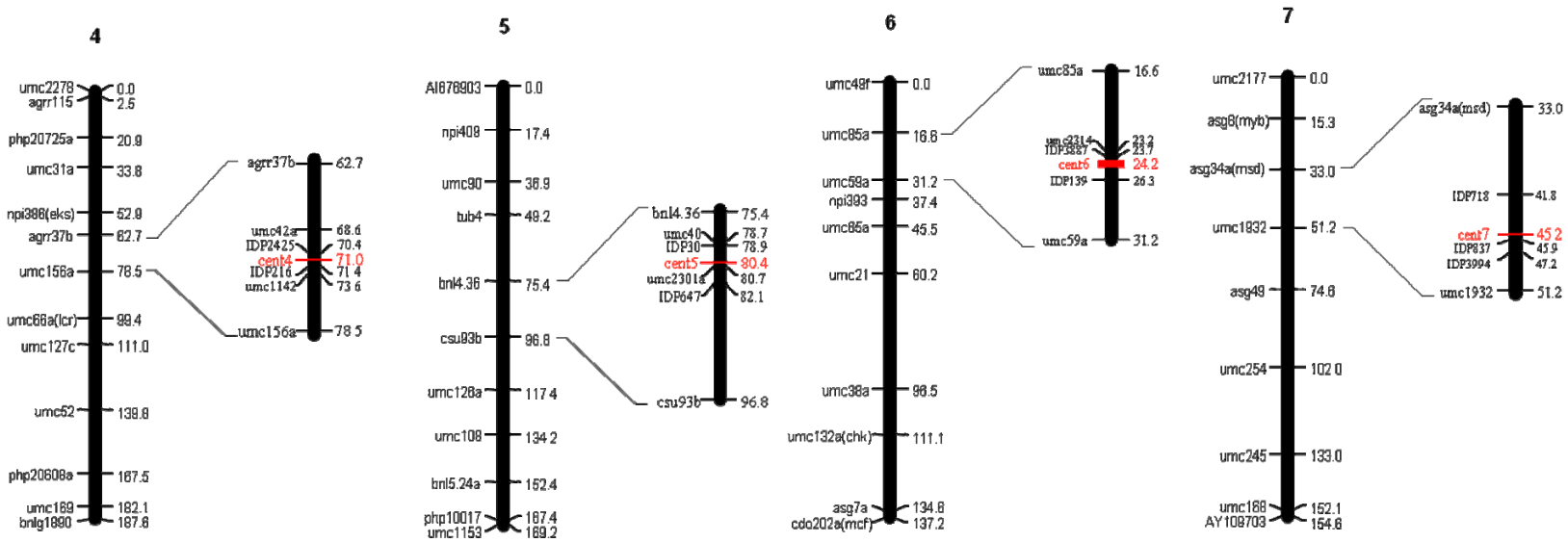

Figure 3.2. The map position of CRM2 marker clusters on the IBM2 neighbor Map. (*To be continued***)**

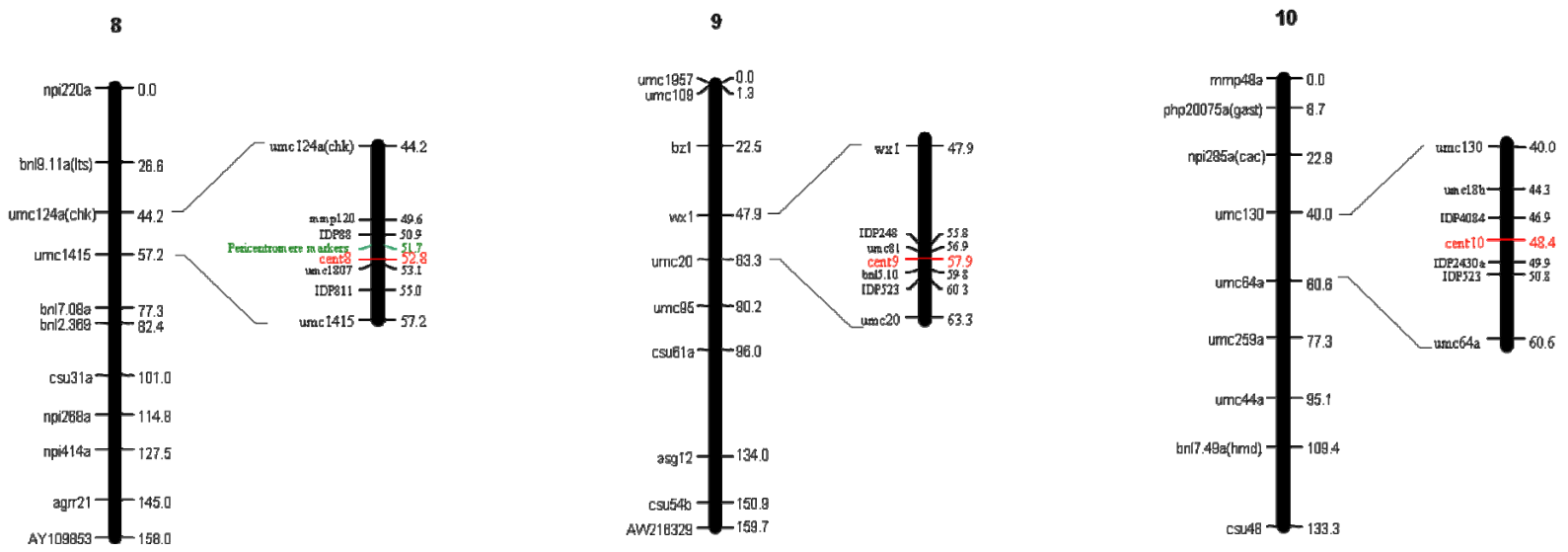

Figure 3.2. The map position of CRM2 marker clusters on the IBM2 neighbor Map.

Table 3.2. Number of markers for each centromere.

Centromere	B73-type	Mo17-type
1	11	2
2	34	10
3	30	6
4	17	19
5	12	18
6	9	6
7	9	13
8	26	11
9	2	18
10	8	3

Excluding heterozygosity and contamination

We detected several lines that contained markers from either both or neither of the parental centromeres. For those centromeres that contain markers from both parents (14 cases, see Table 3.1), heterozygosity is a likely explanation. In centromere 4, B73 and Mo17 contain distinguishable amount of CentC arrays. FISH was applied to RIL IBM85, which contained markers from both parents, to detect heterozygosity. The results showed that the CentC amount of two centromere 4s in IBM85 were clearly different, and each represented the centromere 4 from B73 or Mo17 (Fig. 3.3). We further genotyped the apparent heterozygous regions using primers that could detect flanking insertion deletion polymorphisms (known as IDPs; (Fu et al., 2006)). In all lines that appeared to be heterozygous for a centromere, the IDPs directly flanking the centromere were also heterozygous (Fig. 3.3). These data suggest that the level of heterozygosity for centromeres is 1.52%, which is in line with predicted levels from a 6X self-crossed inbred population.

Less predictable are four IBM lines with centromeres that lack markers from either parent (Table 3.1). Since our results suggest centromeres from one line are not likely to share markers with the same centromere from another line, the absence of markers could reflect the presence of

contaminant chromosomes. Under this assumption the contaminated centromeres would contribute non-parental CRM bands to TD gels. Indeed there were numerous unexpected bands in three of the aberrant IBM lines (all of the non parental bands in the study were limited to those lines). We also subjected the four lines to FISH analysis. Each of the centromeres scored as 'absent' by B73/Mo17 marker criteria nevertheless contained abundant quantities of CentC and CRM. We also noted that line IBM58 segregates for knobs that are not present in either parent (Fig. 3.3). Taken together the data suggest that IBM lines 11, 27, 35, and 58 are contaminated by centromeres from other inbreds, and were subsequently removed from the mapping population. We note that since only one or two centromeres are affected in each IBM line, the contamination must have occurred early in the crossing scheme.

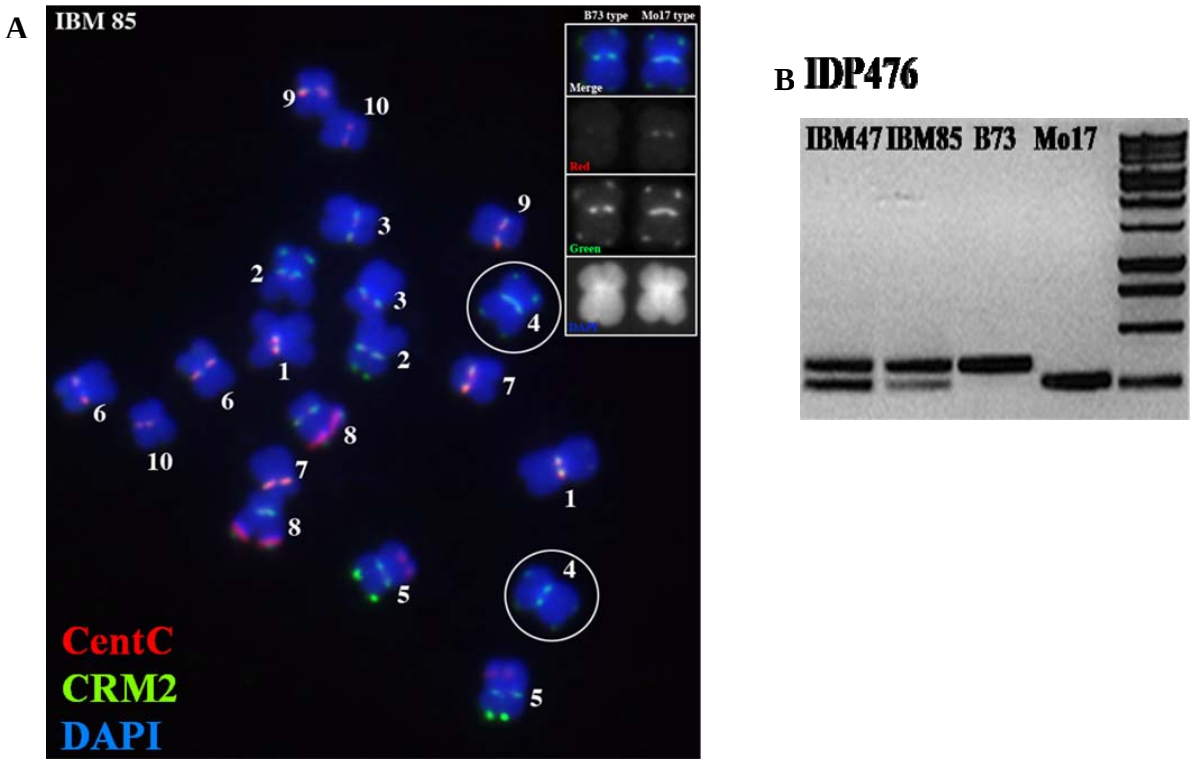

Figure 3.3. Confirmation of heterozygosity and contamination in centromeres. (*To be continued***)**

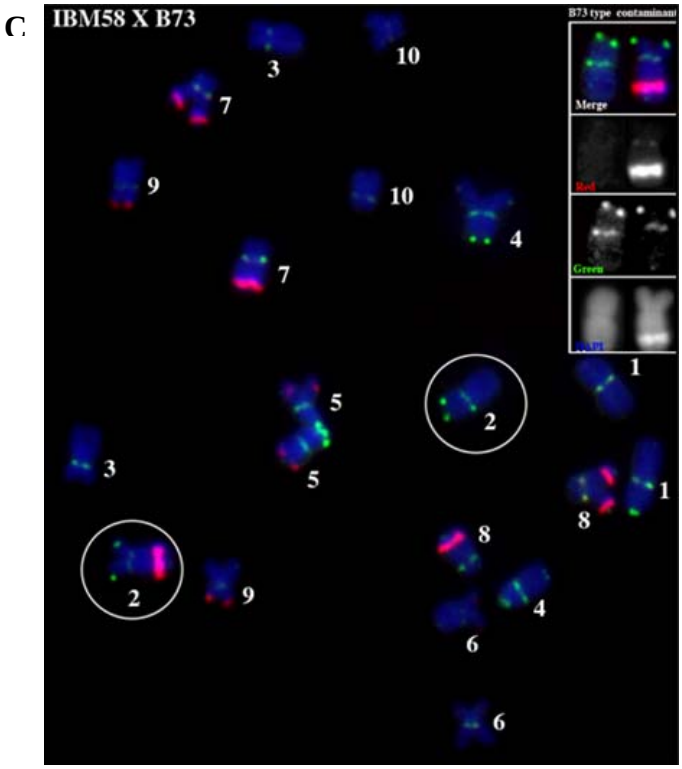

Figure 3.3. Confirmation of heterozygosity and contamination in centromeres. (A) A FISH image of IBM85, showing different CentC amount at two centromere 4s. (B) A photograph of PCR results for marker IDP476, showing that IBM47 and 85 are heterozygous in centromere 4 flanking regions. (C) A FISH image of IBM58 X B73, showing the presence of CRM2 and the aberrant knob of IBM 58 in chromosome 2. Red, CentC and knob; green, CRM; blue, DAPI.

CRM2 markers interact with CENH3

Kinetochores in all species are anchored by a histone H3 variant known as CENH3 (Choo, 2000; Dawe and Henikoff, 2006). However, CENH3 chromatin is not continuously distributed over the regions that contain centromere repeats. Blocks of CENH3 and histone H3 nucleosomes are interspersed in centromeric regions (Blower et al., 2002; Yan et al., 2008). In

addition, the centromere sequences extend to the flanking pericentromere regions as well (Schueler et al., 2001; Zhong et al., 2002; Yan et al., 2005a). As a result, any assay of common centromere repeats provides only a partial view of the functional centromere/kinetochore regions.

To identify CRM2 markers that lie within functional regions, we added a chromatin immunoprecipitation (ChIP) step to the protocol. Centromeric chromatin was precipitated with anti-CENH3 antibodies, the DNA purified from the associated chromatin and the sample further processed for CRM2 display. Input DNA (before adding antibodies) was used as a positive control and a treatment without antibodies (NoIgG) was used as a negative control. As shown in Fig. 3.4, only a fraction of the input bands (marked with "*") were displayed after CENH3 IP treatment. Each of these bands mapped to the major centromere clusters (i.e. none of the known pericentromeric bands were precipitated). Several bands are absent in the CENH3 IP fraction (marked with "-"). AGC157, which maps to the long arm of chromosome 6, is apparently not associated with CENH3. B_2_AGC190 and B_7_AGC229, which map to centromere 2 and 7 respectively, are also absent in IP. It is likely that these two markers are associated with the H3 nucleosomes that are interspersed in the CENH3 nucleosome blocks. A much smaller subset of non-specific background bands can be identified in the negative control.

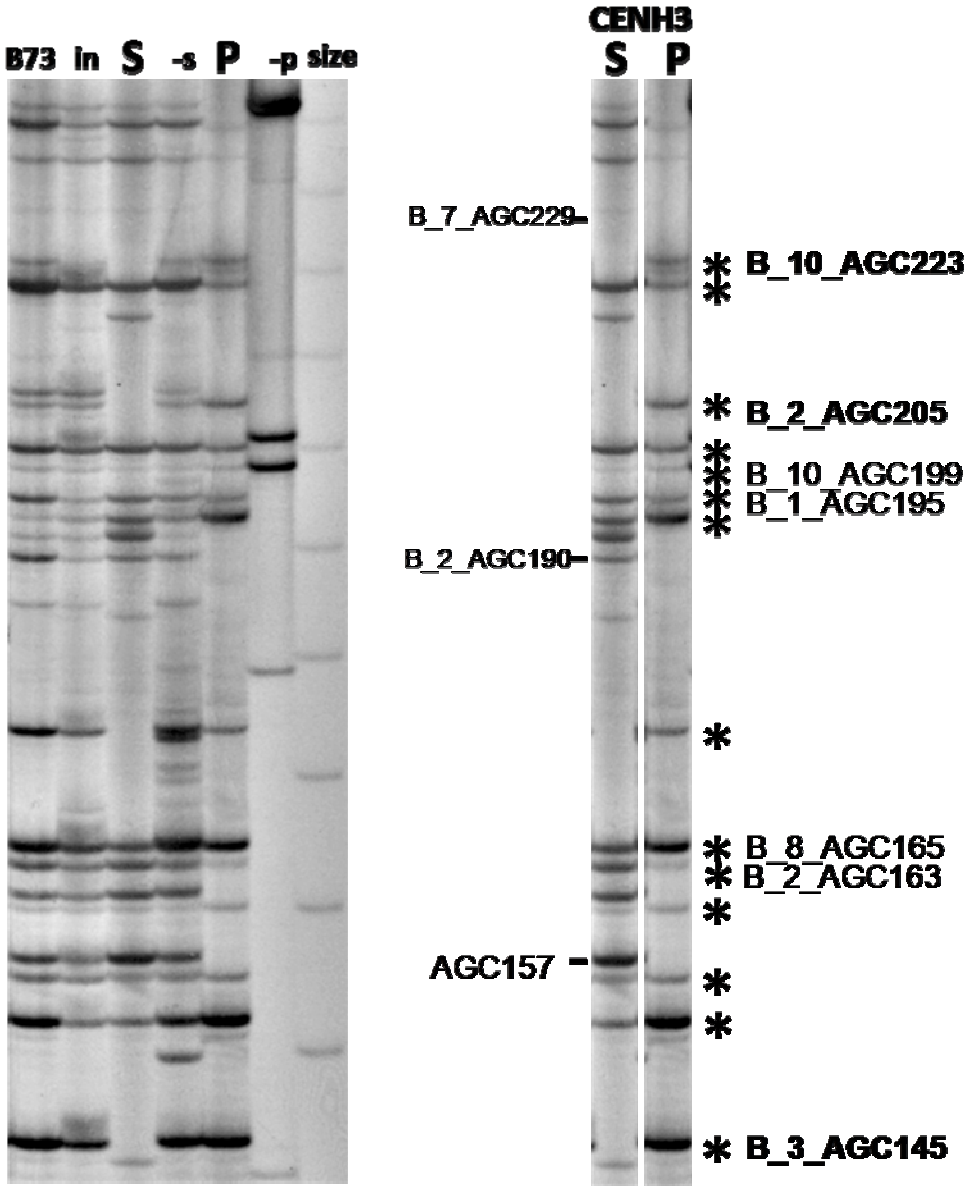

Figure 3.4. ChIP-display. Left panel shows the full suite of controls (in=input, S=supernatant, P=pellet, -p=no Ab). The right panel shows direct comparison between supernatant and IP. * mark the bands that are present in IP. AGC157, which maps to a genic region, is absent in IP. Two other markers, B_2_AGC190 and B_7_AGC229, which map to centromere 2 and 7 respectively, are also absent in IP.

The ChIP-display method is complex and prone to potential errors at several stages. To rule out artifactual bands and confirm the method, we cloned and sequenced 33 displayed bands from both genomic DNA and CENH3 IPed DNA (GenBank IDs here, all are submitted and have BankIt numbers). All but two of the IPed bands matched their genomic counterparts. The exceptions were high molecular weight, weak bands, which were difficult to accurately recover from the TD gel. The selectivity of the assay was further confirmed by scoring bands that mapped to pericentromeric regions: all of the 19 markers that mapped outside the centromere clusters were not ChIPed by CENH3. Among our centromere clustered markers, 57.2% were associated with CENH3, indicating that they are distributed in the CENH3 nucleosome blocks, while the remaining 42.8% markers are likely distributed in the H3 nucleosome blocks that are interspersed in the functional centromeres. We note that only 208 of the 264 markers were testable using ChIP (the rest were either of high molecular weight or very weak bands).

Sequence conversion events within centromeres

The IBM population presents a unique opportunity for identifying rare genetic exchanges within centromere cores. Since recombination is suppressed in centromeres, the markers from a single centromere haplotype should always be inherited as a unit. While this is true for the great majority of centromeres, we also detected aberrant inheritance patterns. These fell into to two categories: loss of a marker from known centromere haplotype and gain of a marker. The gain of marker category is of particular interest. In these cases a full centromere haplotype is present from one parent but the line also contains a single marker from the homologous centromere in the other parent.

Since marker loss is a negative result it is difficult to confirm. In principle these may represent deletion events, but they are just as likely to represent PCR errors and were not pursued further. However there are several definitive ways to confirm the gain of a marker in our scoring system. We first cloned and sequenced each affected band from its parental line. With sequence in hand, new primers with four selective bases were used for screening. In eight such cases, the gained bands were not observed using the 4-sel-bp primer, indicating the bands were not the same sequence as the original bands from the parents. However two bands, B73_8_ACC165 and Mo17_5_TCG264, showed the same aberrant segregation pattern with a 4-sel-bp primer. These were further confirmed with additional tests.

For marker B73_8_ACC165 a specific 11-sel-bp primer was designed for confirmation. With this primer the segregation was identical to the original observation, such that RIL IBM10 (MO0017), which contains the complete Mo17 centromere 8 haplotype, also contains marker B73_8_ACC165 from B73 centromere 8 (Fig. 3.5). The chance for a false positive amplification using 11 selective bases is $1/4^{11}$ (effectively zero). For Mo17_5_TCG264, we directly sequenced the aberrantly scored bands in the affected RILs IBM24 (MO0035) and IBM54 (MO0269). Both lines contain the complete B73 centromere 5 haplotype as well as the Mo17_5_TCG264 marker from Mo17 centromere 5 (Fig. 3.5). These data established in two cases that single marker exchange occurred within centromere cores at some time during the 11-generation time span required to prepare the IBM mapping population.

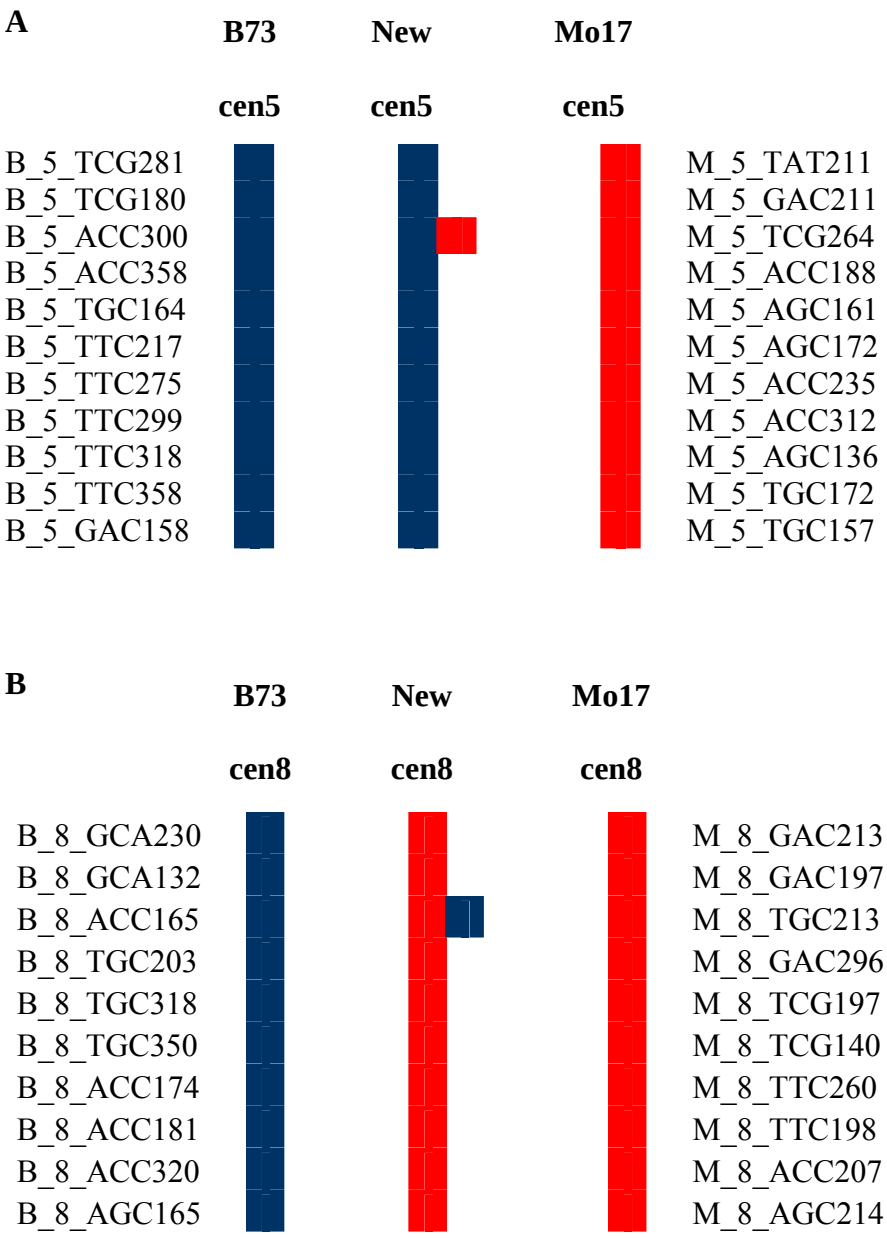

Figure 3.5. Conversion events in centromeres. (A) B73 and Mo17 centromere 5 haplotypes and new centromere 5 haplotype in IBM 24 (MO0035) and 54 (MO0269). (B) B73 and Mo17 centromere 8 haplotypes and new centromere 8 haplotype in IBM 10 (MO0017).

Linkage disequilibrium in centromere 2

LD is often used as a complementary analysis to linkage analysis. As an out crossing species, LD decays rapidly over regions longer than 2000 bp (Remington et al., 2001). LD mainly represents the history of recombination. In regions where crossing over is completely shut down, such as centromeres, we should expect complete LD (Rafalski and Morgante, 2004). A fully contiguous sequence of centromere 2 has recently been completed (Gernot Presting lab, including Shi and Dawe, in preparation). This physical map, in conjunction with our marker system, provides an opportunity to directly test within-centromere LD. For this purpose fourteen continuously distributed markers were chosen (including both PCR markers and TD markers). A set of 53 maize inbred lines, including the 50-line core set (Liu et al., 2003) with additional lines within NAM (nested association mapping) founder lines (Yu et al., 2008), was chosen for LD analysis. This set of inbred lines was estimated to capture ~70% of the total genetic diversity in maize (Liu et al., 2003).

LD between pairs of markers within centromere 2 is summarized in Fig. 3.6. The data revealed that LD does not persist along the entire centromere. Many single-marker disruptions in the LD map are detected. This phenomenon resembles the conversion events we detected in IBM population, and supports the view that a high frequency of conversion-like genetic exchanges is an important resource for centromere evolution. The data also provide evidence of linkage disequilibrium over distances as long as 3 BACs (where LD declines to <0.1). Since the average length of the 16 centromere 2 BACs is ~160 kb, we can estimate that LD in centromere 2 is ~480 kb. However, this measure of LD should be taken in context. Within B73 centromere 2 there are several clustered, short CentC tracts (totaling ~50kb) imbedded within the larger CRM-rich centromere. A closer look at the LD distribution revealed that the region most

affected by LD falls precisely over the region that includes known CentC arrays. As discussed below, these data suggest that CentC regions show a distinct pattern of evolution.

A

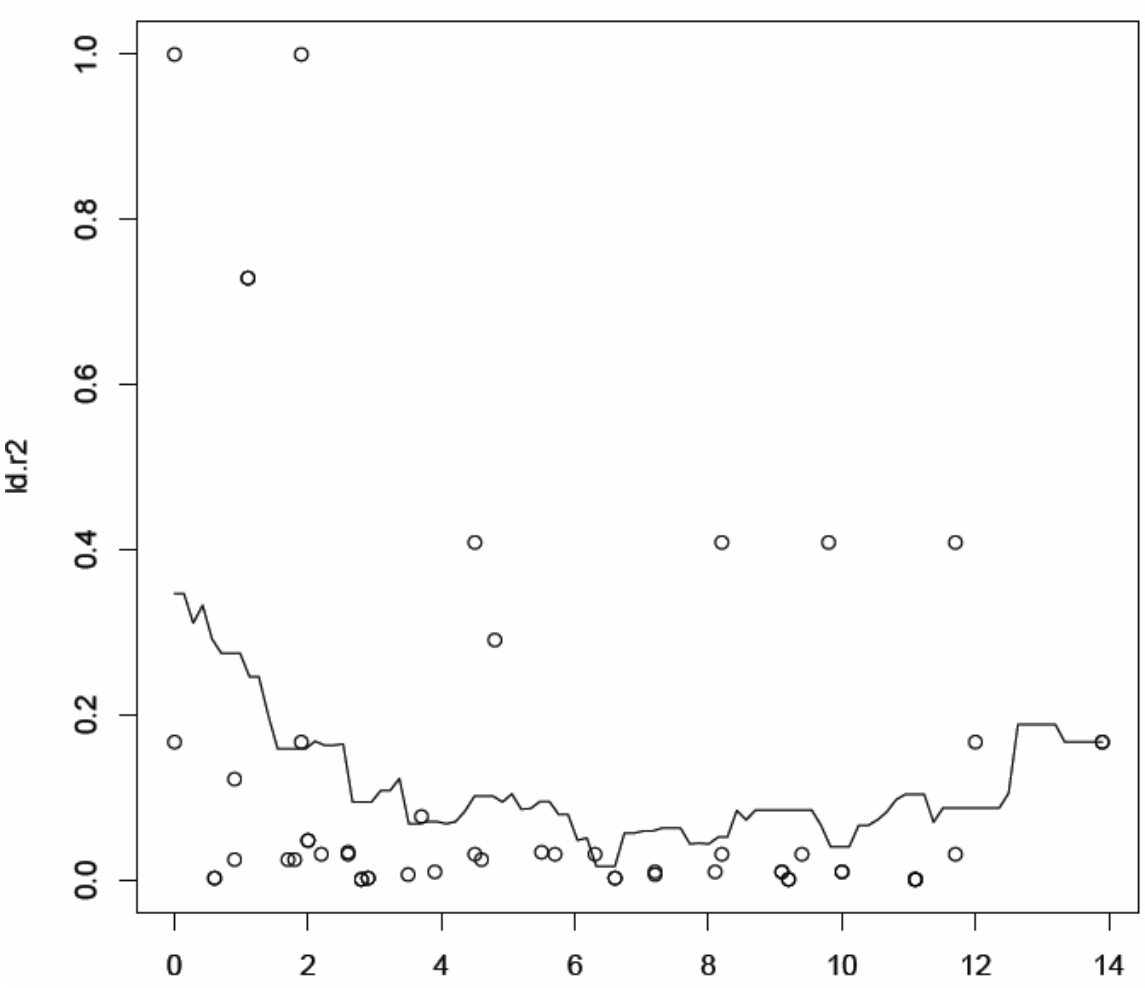

Figure 3.6. LD in centromere 2. (*to be continued*)

B

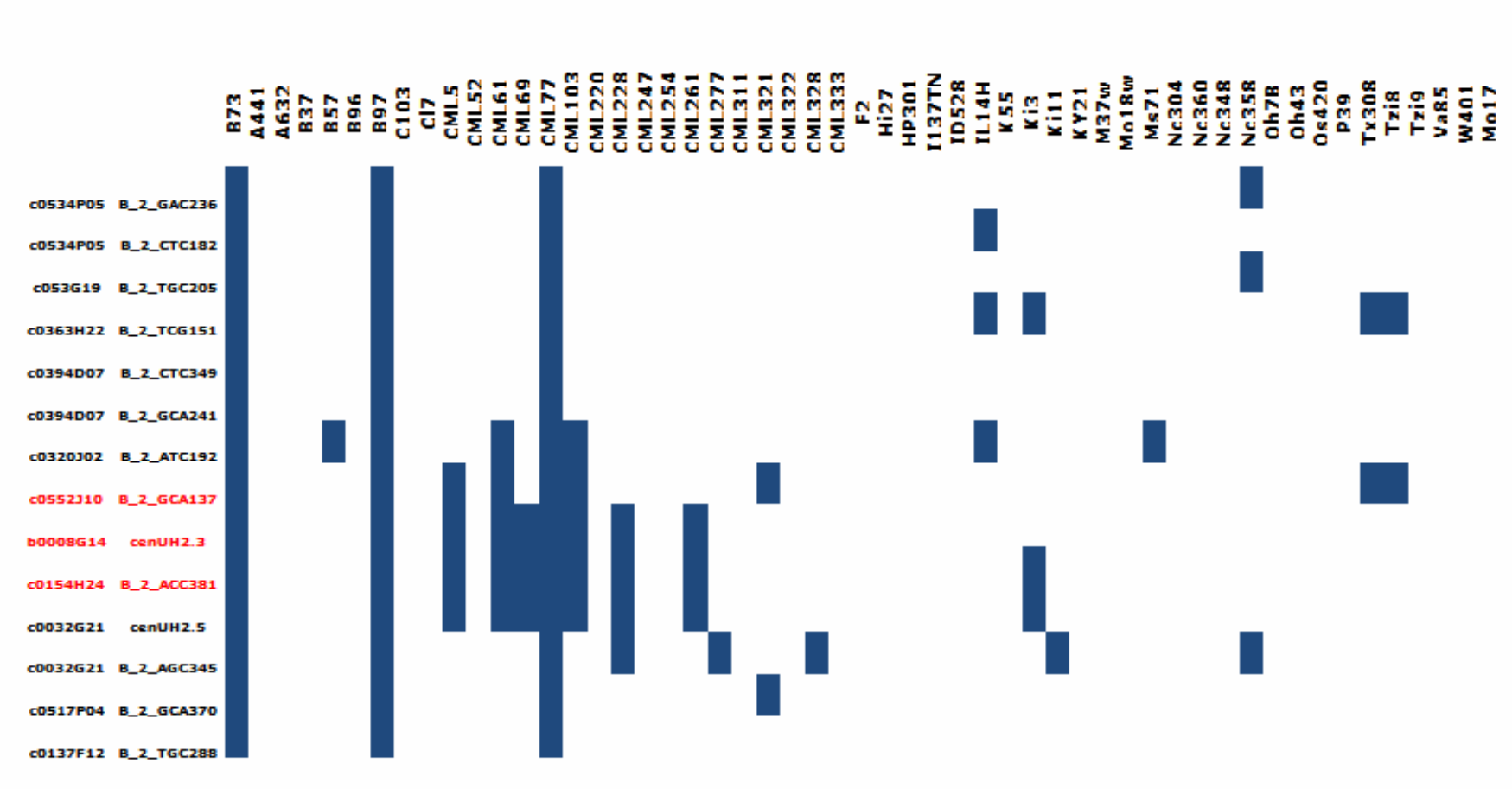

Figure 3.6. LD in centromere 2. (A) Plots of squared correlations of allele frequencies (r^2) against estimated distance between polymorphic sites in centromere 2. (B) Distribution of markers for LD analysis in centromere 2. BACs/Markers in red: CentC containing BACs.

Maize centromere haplotype estimates

Comparison between B73 and Mo17 revealed that the homologous centromeres lack overall similarity. Only <20% of CRM2 markers are non-polymorphic. This suggests that centromeres in different inbreds can be so different as to be accurately described as independent haplotypes. Since centromeres are effectively blocked from crossing over, we can interpret centromeres as single long alleles, most of which appear to have diverged from each other hundreds of thousands of years ago and show genetic exchange primarily by gene conversion.

To estimate the frequency of centromere haplotypes in maize, we scored the 53 inbreds for extended or complete homology to either B73 or Mo17. Table 3.3 shows which centromere were analyzed and their frequency of B73 or Mo17-homologous centromeres in the 53 inbreds (centromeres with less than five markers were not considered). We find that each of the centromeres is represented at an average frequency of 5.4 times among the 53 inbreds. The distribution shows little or no correlation with known inbred phylogeny (Liu et al., 2003). This distribution suggests that maize as a subspecies contains a relatively small number of founder centromeres.

Perhaps more noteworthy is the fact that major haplotypes represent only a fraction of the total variation present. Features of the B73 haplotype are spread widely throughout the maize inbred sample, most often as single marker additions to other haplotypes. These short exchange events increase the measurable haplotype diversity considerably. We presume that the measured diversity would be much higher if the marker set were increased, or if full sequence were available.

Table 3.3. Centromere haplotype frequency.

Centromere	B73-haplotype	Mo17-haplotype
1	8	N/A
2	3	N/A
3	4	N/A
4	N/A	5
5	5	N/A
6	10	N/A
7	N/A	4
8	4	7
9	N/A	4
10	N/A	N/A

DISCUSSION

Here we have described a novel approach to centromere mapping and show that gene conversion is common within centromeres. In addition, the large collection of markers allowed us to address other major questions in centromere evolution. These include the mode and tempo of change, the role of tandem repeat arrays, and centromere haplotype frequency.

Recombination was empirically measured within a standard maize mapping population to show that within an 11-generation period, at least two gene conversion-like events occurred. A total of 264 markers were assayed in each of 92 inbred lines for aberrant segregation. These data reveal that each small region has a high frequency of being involved in a conversion event. Linkage disequilibrium (LD) analysis supports the inference that a high level of conversion-like exchange is a major cause of centromere diversity, showing that small segments of the B73 haplotype are widely scattered in multiple inbreds.

Evidence for neutral evolution

We observed that centromeres lie within regions that are almost completely divergent and are isolated from normal levels of exchange. Thus centromeres can be classified as haplotypes with unique evolutionary histories. Extensive evidence suggests that DNA sequence polymorphism is positively related with recombination rate. Originally discovered in *Drosophila* (Begun and Aquadro, 1992), a significant association between genetic diversity and crossing over was observed in many other plants and animals (Dvorak et al., 1998; Kraft et al., 1998; Nachman et al., 1998; Stephan and Langley, 1998; Hamblin and Aquadro, 1999). In the regions near centromeres where crossing over is severely suppressed, lower levels of variation were observed (Hudson and Kaplan, 1995; Stephan and Langley, 1998). However, in theory the positive relationship between recombination and polymorphism can be a consequence of 'background selection' for removal of deleterious mutations or fixation of advantageous mutations (Charlesworth et al., 1994; Andolfatto and Przeworski, 2001). In centromeres, where genetic determination is limited and genes are few, we might expect the trend to reverse: crossing over is reduced but polymorphism is high.

Maize was domesticated from parviglumis between 6,250 and 10,000 B.P. (before present) (Piperno and Flannery, 2001; Tenaillon et al., 2004). Although the domestication bottleneck led to some reduction of genetic diversity, maize is known for its dramatic allelic divergence among inbreds. A previous study of 2039 alleles revealed that maize inbreds capture ~80% of the diversity in open-pollinated landraces, which retain the ~80% diversity from parviglumis (Tenaillon et al., 2001; Liu et al., 2003). The high level of genetic divergence in maize indicates a relatively large founder population. However, Eyre-walker et al. (Eyre-Walker et al., 1998) constructed a computer simulated model to demonstrate that the founding

population could be very small when the individuals are highly diverse – a few hundred parviglumis individuals were sufficient to capture the diversity in maize. Based on this assumption, we propose that the founder parviglumis population may have contained a relatively small number of centromere haplotypes. In our data set, these are now visible as major 'founder' centromere haplotypes. Subsequently during domestication, an open-pollination plant such as maize would have maintained most centromeres in a heterozygous state and made possible the genetic exchanges we have detected as single marker exchange events, which accordingly expanded the total haplotype diversity dramatically.

Evidence for centromere drive

Several authors have emphasized the potential role of meiotic drive in centromere evolution (Henikoff et al., 2001). In plants and animals, only one out of four meiotic products becomes a functional egg during female meiosis. Centromeres that specifically target this single cell (by any means) can potentially 'cheat' meiosis and increase their representation in a population by meiotic drive. The driven centromere is likely to attract more repeats and more kinetochore proteins, causing a centromeric imbalance that can reduce fertility. In principle, such a genomic conflict could be resolved by mutating the binding domains of inner kinetochore proteins (Henikoff et al., 2001; Dawe and Henikoff, 2006).

According to this hypothesis, centromere evolution should have several identifiable stages: expansion of selfish repeats that leads to deleterious effects on the genomes, loss of drive by kinetochore protein mutation, accumulation of new repeats that gradually take the place of selectively neutral sequences, then over longer periods the acquisition of new repeats that show drive characteristics (Dawe and Henikoff, 2006). Our data analysis reveals evidence supporting

this view. In modern maize, distribution of CentC is highly polymorphic (Kato et al., 2004) and has little apparent function in centromere function. Only about 50% of the CentC monomers in maize interact with CENH3 (Zhong et al., 2002; Jin et al., 2004). In B73 centromere 2 there is only ~50 kb of CentC, while a functional maize centromere 2 is close to 1000 kb. Other evidence that CentC is not required for centromere function comes from centromere 5, which is fully sequenced and contains a small CentC array that is localized in pericentromeric regions that are not associated with CENH3 (G. Presting, personal communication).

Among the centromere 2 haplotypes assayed, there was a single "recombination free' zone flanking the single CentC containing region. One possible explanation for the localized LD pattern in the vicinity of CentC is crossing over. Although there is no crossing over in this central core domain in modern maize, it is possible that deeper in evolutionary history CentC arrays were under selection and consequently dominated centromere cores. Under this view, CentC array in maize progenitors were much longer and traversed the bulk of centromere cores. In support of this view, we have shown that *Zea diploperennis* and *Zea luxurians*, which are close maize relatives, contain vast expanses of CentC on every centromere (see Chapter 4).

In the absence of selection, we expect CentC arrays to rapidly contract by intrachromosomal and between-sister exchange events (Smith, 1976; Liebman et al., 1988; Charlesworth et al., 1994; Roizes, 2006; Jaco et al., 2008). In maize and all other grasses, CR elements are enriched within and around centromeres. During the process contracting repeat arrays and forming a new balance between centromere sequences and proteins, new genetic materials from pericentromere regions are presumably recruited into centromere cores (Langdon et al., 2000). The fact that we can still detect evidence of selection around CentC suggests that the transition from long CentC arrays to short arrays occurred relatively recently. Immediately

adjacent to the CentC array, LD drops off over a distance of roughly 200 kb. The rapid decline of LD is likely to be a result of frequent conversion events, in which a small piece of a chromosome is transferred to another chromosome. Recent studies showed that LD could be broken down between tightly linked markers when there was high rate of conversion events (Ardlie et al., 2002).

From a mechanistic perspective our data show that there is genetic mixing among maize centromeres, and that it occurs at a frequency that could contribute to centromere evolution over the long term. The frequent conversion events provide a way to rapidly spread different sequence variants between chromosomes. Although all of our markers are based on retrotransposon sequences, the mechanism for genetic exchanges that we observed do not rely on retrotransposon features, and are likely represent a universal mechanism for centromere evolution. In human, higher order alpha satellites (HOR) are located in the functional centromeres, while the monomeric alpha satellites are present in the pericentromere (Schueler et al., 2001; Rudd and Willard, 2004; Schueler et al., 2005). Unequal crossover is widely proposed to be the mechanism for generating and spreading HORs. However, unequal crossover is probably the major mechanism for homogenization; gene conversion is more likely to be the mechanism that spreads the novel satellite monomers in the genome.

REFERENCES

Anderson, L.K., Doyle, G.G., Brigham, B., Carter, J., Hooker, K.D., Lai, A., Rice, M., and Stack, S.M. (2003). High-resolution crossover maps for each bivalent of Zea mays using recombination nodules. Genetics **165,** 849-865.

Andolfatto, P., and Przeworski, M. (2001). Regions of lower crossing over harbor more rare variants in African populations of Drosophila melanogaster. Genetics **158,** 657-665.

Ardlie, K.G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. Nat Rev Genet **3,** 299-309.

Beadle, G.W. (1932). A Possible Influence of the Spindle Fibre on Crossing-Over in Drosophila. Proc Natl Acad Sci U S A **18,** 160-165.

Begun, D.J., and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature **356,** 519-520.

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

Cambareri, E.B., Aisner, R., and Carbon, J. (1998). Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol. Cell Biol. **18,** 5465-5477.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., and Wessler, S.R. (2000). Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci U S A **97,** 10083-10089.

Charlesworth, B., Sneglowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371,** 215-220.

Choo, K.H. (2000). Centromerization. Trends Cell Biol **10,** 182-188.

Copenhaver, G.P., Browne, W.E., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. Proc Natl Acad Sci U S A **95,** 247-252.

Copenhaver, G.P., Nickel, K., Kuromori, T., Benito, M., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L.D., McCombie, W.R., Martienssen, R.A., Marra, M., and Preuss, D. (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286,** 2468-2474.

Davis, C.R., Kempainen, R.R., Srodes, M.S., and McClung, C.R. (1994). Correlation of the physical and genetic maps of the centromeric region of the right arm of linkage group III of Neurospora crassa. Genetics **136,** 1297-1306.

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

Devos, K.M., Ma, J., Pontaroli, A.C., Pratt, L.H., and Bennetzen, J.L. (2005). Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci U S A **102,** 19243-19248.

Dooner, H.K., and He, L. (2008). Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. Plant Cell **20,** 249-258.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin **19,** 11-15.

Dvorak, J., Luo, M.C., and Yang, Z.L. (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species. Genetics **148,** 423-434.

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., and Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci U S A **95,** 4441-4446.

Fu, Y., Wen, T.J., Ronin, Y.I., Chen, H.D., Guo, L., Mester, D.I., Yang, Y., Lee, M., Korol, A.B., Ashlock, D.A., and Schnable, P.S. (2006). Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics **174,** 1671-1683.

Fukui, K.N., Suzuki, G., Lagudah, E.S., Rahman, S., Appels, R., Yamamoto, M., and Mukai, Y. (2001). Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. Plant Cell Physiol **42,** 189-196.

Hamblin, M.T., and Aquadro, C.F. (1999). DNA sequence variation and the recombinational landscape in Drosophila pseudoobscura: a study of the second chromosome. Genetics **153,** 859-869.

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

Hosouchi, T., Kumekawa, N., Tsuruoka, H., and Kotani, H. (2002). Physical map-based sizes of the centromeric regions of Arabidopsis thaliana chromosomes 1, 2, and 3. DNA Res. **9,** 117-121.

Hudson, R.R., and Kaplan, N.L. (1995). Deleterious background selection with recombination. Genetics **141,** 1605-1617.

Jackson, M.S., See, C.G., Mulligan, L.M., and Lauffart, B.F. (1996). A 9.75-Mb map across the centromere of human chromosome 10. Genomics **33,** 258-270.

Jaco, I., Canela, A., Vera, E., and Blasco, M.A. (2008). Centromere mitotic recombination in mammalian cells. J Cell Biol **181,** 885-892.

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

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

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

Kraft, T., Sall, T., Magnusson-Rading, I., Nilsson, N.O., and Hallden, C. (1998). Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (Beta vulgaris subsp. maritima). Genetics **150,** 1239-1244.

Lambie, E.J., and Roeder, G.S. (1986). Repression of meiotic crossing over by a centromere (CEN3) in Saccharomyces cerevisiae. Genetics **114,** 769-789.

Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J., Jones, R., and Jenkins, G. (2000). Retrotransposon evolution in diverse plant genomes. Genetics **156,** 313-325.

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.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D., and Hallauer, A. (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol **48,** 453-461.

Liebman, S.W., Symington, L.S., and Petes, T.D. (1988). Mitotic recombination within the centromere of a yeast chromosome. Science **241,** 1074-1077.

Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, E., and Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165,** 2117-2128.

Luce, A.C., Sharma, A., Mollere, O.S., Wolfgruber, T.K., Nagaki, K., Jiang, J., Presting, G.G., and Dawe, R.K. (2006). Precise centromere mapping using a combination of repeat junction markers and chromatin immunoprecipitation-polymerase chain reaction. Genetics **174,** 1057-1061.

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

Mahtani, M.M., and Willard, H.F. (1998). Physical and genetic mapping of the human X chromosome centromere: repression of recombination. Genome Res **8,** 100-110.

Mather, K. (1939). Crossing over and Heterochromatin in the X Chromosome of Drosophila Melanogaster. Genetics **24,** 413-435.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A **99,** 6080-6084.

Miller, J.T., Dong, F., Jackson, S.A., Song, J., and Jiang, J. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150,** 1615-1623.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S.N., Milan, D., Ostrander, E.A., Pape, G., Parker, H.G., Raudsepp, T., Rogatcheva, M.B., Schook, L.B., Skow, L.C., Welge, M., Womack, J.E., O'Brien S, J., Pevzner, P.A.,

and Lewin, H.A. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science **309,** 613-617.

Nachman, M.W., Bauer, V.L., Crowell, S.L., and Aquadro, C.F. (1998). DNA variability and recombination rates at X-linked loci in humans. Genetics **150,** 1133-1141.

Nagaki, K., Song, J., Stupar, R., Parokonny, A.S., Yuan, Q., Ouyang, S., Liu, J., Dawe, R.K., Buell, C.R., and Jiang, J. (2003). Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics **163,** 759-770.

Nakaseko, Y., Adachi, Y., Funahashi, S.-i., Niwa, O., and Yanagida, M. (1986). Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J **5,** 1011-1021.

O'Neill, R.J., Eldridge, M.D., and Metcalfe, C.J. (2004). Centromere dynamics and chromosome evolution in marsupials. J Hered **95,** 375-381.

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.

Piperno, D.R., and Flannery, K.V. (2001). The earliest archaeological maize (Zea mays L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc Natl Acad Sci U S A **98,** 2101-2103.

Presting, G., Malysheva, L., Fuchs, J., and Schubert, I. (1998). A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16,** 721-728.

Rafalski, A., and Morgante, M. (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet **20,** 103-111.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S.t. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A **98,** 11479-11484.

Roizes, G. (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Res **34,** 1912-1924.

Rudd, M.K., and Willard, H.F. (2004). Analysis of the centromeric regions of the human genome assembly. Trends Genet **20,** 529-533.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science **274,** 765-768.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science **294,** 109-114.

Schueler, M.G., Dunn, J.M., Bird, C.P., Ross, M.T., Viggiano, L., Rocchi, M., Willard, H.F., and Green, E.D. (2005). Progressive proximal expansion of the primate X chromosome centromere. Proc Natl Acad Sci U S A.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science **191,** 528-535.

Song, R., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci U S A **100,** 9055-9060.

Stephan, W., and Langley, C.H. (1998). DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics **150,** 1585-1593.

Tenaillon, M.I., U'Ren, J., Tenaillon, O., and Gaut, B.S. (2004). Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol **21,** 1214-1225.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., and Gaut, B.S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci U S A **98,** 9161-9166.

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

Wang, Q., and Dooner, H.K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc Natl Acad Sci U S A **103,** 17644-17649.

Warburton, P.E., Haaf, T., Gosden, J., Lawson, D., and Willard, H.F. (1996). Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. Genomics **33,** 220-228.

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

Yan, H., Jin, W., Nagaki, K., Tian, S., Ouyang, S., Buell, C.R., Talbert, P.B., Henikoff, S., and Jiang, J. (2005). Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell **17,** 3227-3238.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178,** 539-551.

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.

CHAPTER 4

CROSSING OVER IN PERICENTROMERE REGIONS PLAYS A ROLE IN CENTROMERE EVOLUTION

Jinghua Shi and R. Kelly Dawe. To be submitted to Genetics.

INTRODUCTION

There are four species in the genus *Zea*, and they are divided into two sections. Section *Luxuriantes* consists of the annual *Z. luxurians*, the protected perennial taxa *Z. diploperennis* and its autotetraploid derivative *Z. perennis*. The only species in Section *Zea* is *Z. mays*, which is further divided into four subspecies: ssp. *huehuetenangensis*, ssp. *mexicana*, ssp. *parviglumis* and ssp. *mays* (Doebley, 1990a). *Z. mays* ssp. *mays* is maize, the domesticated species, and the other three are its wild relatives. *Z. mays* ssp. *parviglumis* is widely considered to be the direct progenitor of ssp. *mays* (Doebley, 2004). Archaeological data suggest that maize domestication from ssp. *parviglumis* occurred at least 7,000 years before present (BP) (Pohl et al., 2007), and current phylogenies support this view, indicating that maize arose from a single domestication event about 9,000 years ago (Matsuoka et al., 2002).

In chapter 3, we provided evidence for conversion-like genetic exchanges in centromere cores and proposed that it is a universal mechanism for rapidly spreading sequence variants between chromosomes. The conversion mechanism efficiently explains the phenomenon that small tracts of B73 centromere are wildly distributed in many other inbreds and that LD declines rapidly in centromere. However, we also showed that CentC arrays were under selection in recent evolutionary history, and presented the hypothesis that meiotic drive was the selective force. These observations raise the question of whether the pre-existing centromere diversity in maize was inherited from its progenitors or arose during (or after) domestication, and whether the meiotic drive interpretation can be supported by broader comparisons. In this chapter, preliminary data provide some evidence of CentC selection in the genus *Zea*. We speculate that crossing over occurring in the pericentromere regions may create new combinations of

pericentromere DNA sequences which might be recruited to the centromere cores during evolution.

PRELIMINARY RESULTS

Centromeres in maize relatives

Work described in chapter 3 indicates that maize centromeres can be distinguished as divergent haplotypes, and that many different haplotypes exist for each centromere. Prior cytological data show that CentC abundance also differs widely among inbreds. For instance, centromeres 8 and 9 from B73 appear to have at least three times less CentC than centromeres 8 and 9 from Mo17 (Kato et al., 2004).

Zea is an outcrossing species and, by molecular estimates, began diverging recently (about 300,000 years ago) (Ross-Ibarra et al., 2009). Close maize relatives may still retain the majority of diversity from their common ancestor. A recent study suggested that CentC is very rich in *Z. diploperennis* (Lamb et al., 2007b). In an effort to better address this question, we assayed the distribution of CentC in *Z. mays* ssp. *parviglumis* and *Z. luxurians*. Assays were made in F1 individuals that were heterozygous for species of interest and maize (providing a B73 internal control for CentC labeling). We 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 (Fig. 4.1). In addition, we noticed that the distribution patterns of the two major centromeric DNA components (CRM and CentC) complement each other (Fig. 4.1). In ssp. *parviglumis*, each centromere contains a similarly high amount of CentC

arrays and relatively low amount of CRMs. Within maize, a similar relationship exists: when CentC is abundant CRM staining is weak, and when CRM is abundant CentC is not.

Under the meiotic drive interpretation, we anticipate that when CentC is under selection, CentC arrays will be more abundant and consistent among chromosomes. Cytological assays suggest that CentC staining pattern fits the meiotic drive interpretation in *Z. diploperennis* (Lamb et al., 2007b), *Z. luxurians*, and strikingly, in *Z. may* ssp. *parviglumis*, the immediate ancestor of maize (Fig. 4.1).

These results are surprising, since maize was only domesticated from ssp. *parviglumis* about 10,000 BP. There are three possible explanations: (1) Prior evidence suggests that maize originated from a single domestication event (Doebley, 1990b; Matsuoka et al., 2002). It is possible that great diversity in centromeres exists among different ssp. *parviglumis* populations, and the line we used in FISH does not represent the diversity in ssp. *parviglumis*. This seems unlikely, however, given that maize is an outcrossing species. (2) Although much of the maize genome was inherited from ssp. *parviglumis*, it is known that gene flow occurred between early maize landraces and other maize relatives such as ssp. *mexicana*, which was not included in our tests. It is possible that these untested teosintes have highly diverged centromeres and strongly contributed to the centromere diversity in maize. (3) During and/or after the domestication process, a key kinetochore protein sustained a mutation that changed its CentC binding characteristics. In this event, the most recent meiotic drive interpretation predicts that CentC arrays will rapidly contract (Dawe and Henikoff, 2006). The process of CentC removal must have been associated with acquisition of other DNA components into functional centromeres. This may have subsequently led to the rapid diversification of maize centromeres.

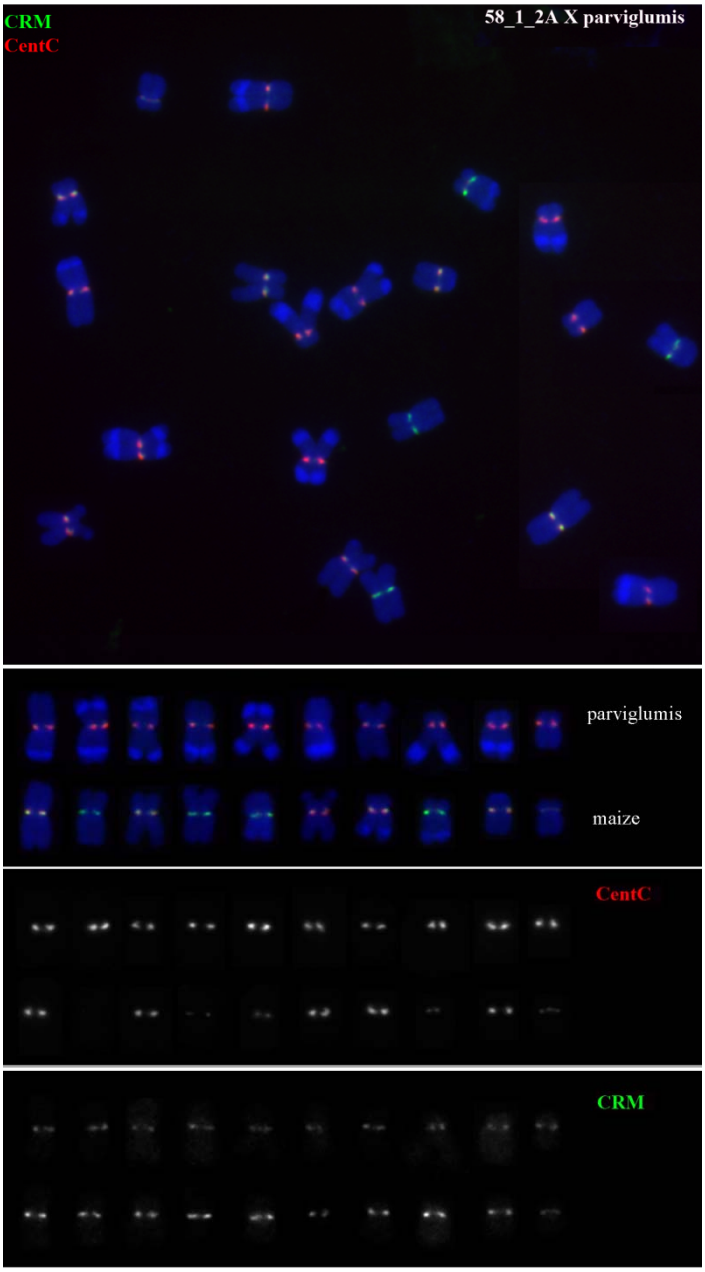

Figure 4.1. **Comparison of CentC and CRM distribution in maize and ssp. *parviglumis*.** FISH image from a hybrid between maize (in a HiII/B73 background; Dawe lab plant 58_1_2A) and ssp. *parviglumis*. The upper row shows the karyotype for ssp. *parviglumis* and the lower row shows the maize karyotype.

Crossing over among CRM2 clusters in pericentromeric regions

Large changes in CRM content may occur in pericentromeric regions during periods when centromeres are largely occupied by CentC arrays. In our genetic map, we detected 10 CRM2 markers that are located in the pericentromeric regions. Although many are interspersed in different chromosomes, six are located on cen8 and form a cluster that is genetically separate from the CENH3-defined cen8. None of the six markers are precipitated by CENH3 antibodies, indicating that they lie outside the functional centromere. Although we cannot yet interpret the physical distance between this pericentromeric cluster and the functional centromere, the genetic distance suggests it is very near the centromere.

We wondered if the genetic separation of cen8 pericentromere could be identified in our diversity assays from 53 inbreds. The data confirm that the pericentromeric cluster is segregating in an independent manner (Fig. 4.2). These results showed that CRM2 clusters are not always located within centromere cores, and when distributed in the pericentromeric regions, they can undergo classical crossing over with the centromere core.

A model for large-scale changes in centromere makeup

The observations that CentC shows specialized LD patterns, that CentC is highly polymorphic in *Zea* relatives, and that large blocks of CRM may exchange among centromeres by crossing over suggest a general model for centromere diversification at the structural level. This is shown in Figure 4.3. When CentC arrays are long enough to occupy the bulk of a centromere, much of the CRM will be localized far from the centromere cores where crossing over occurs. However, when CentC loses its dominant position, such flanking CRM clusters could be recruited to the centromere cores. Crossing over within the flanking regions of different

centromere haplotypes can create novel combinations of CRM clusters, which may then be recruited into cores independently. Such a process could be recurrent, leading to increasing diversification over time (Fig. 4.3).

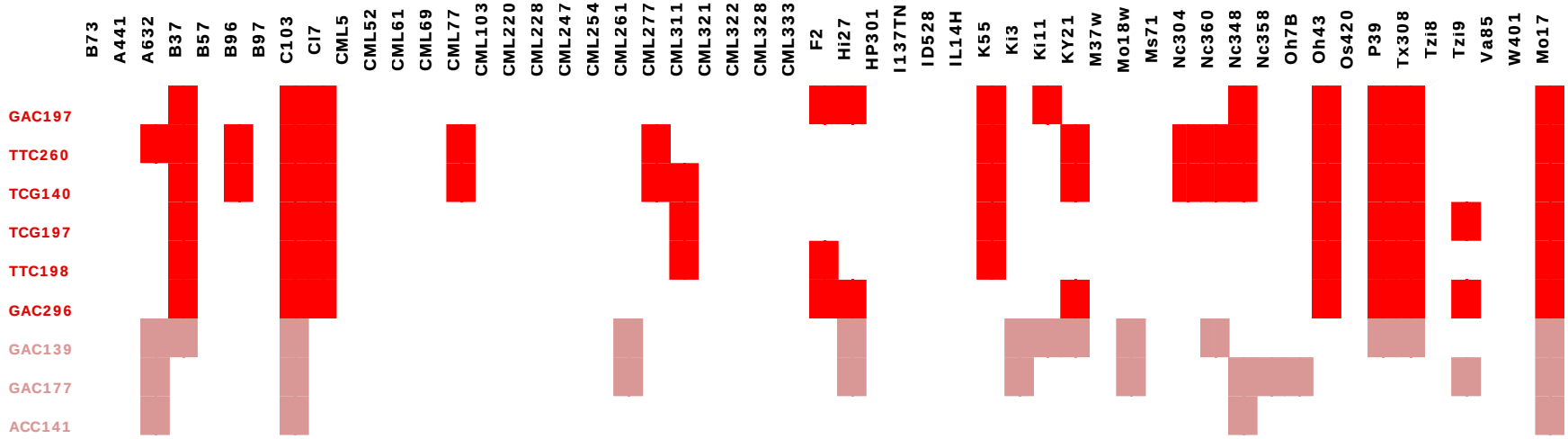

Figure 4.2. **LD in centromere 8.** Distribution of Mo17-derived markers for LD analysis in centromere 8. Red: centromere markers; Pink: pericentromere markers.

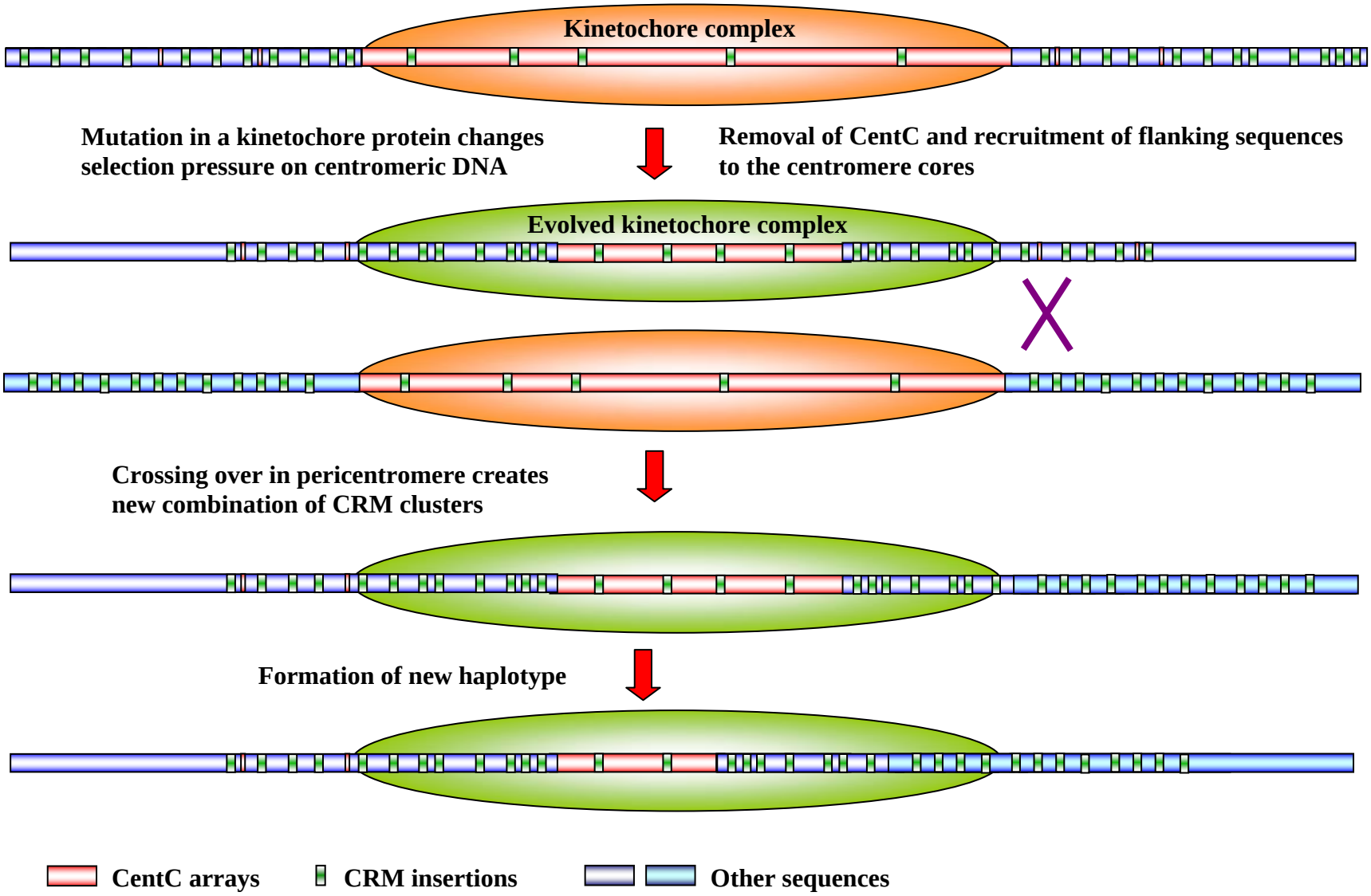

Figure 4.3. **A mechanism that may generate centromere haplotypes.**

REFERENCES

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

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

Doebley, J.F. (1990a). Molecular systematics of *Zea* (Gramineae). Maydica **35,** 143–150.

Doebley, J.F. (1990b). Molecular evidence and the evolution of maize. Econ. Bot. **44 (3 Suppl.),** 6-27.

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

Lamb, J.C., Meyer, J.M., Corcoran, B., Kato, A., Han, F., and Birchler, J.A. (2007). Distinct chromosomal distributions of highly repetitive sequences in maize. Chromosome Res **15,** 33-49.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A **99,** 6080-6084.

Pohl, M.E., Piperno, D.R., Pope, K.O., and Jones, J.G. (2007). Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres, Tabasco, Mexico. Proc Natl Acad Sci U S A **104,** 6870-6875.

Ross-Ibarra, J., Tenaillon, M., and Gaut, B.S. (2009). Historical divergence and gene flow in the genus *Zea*. Genetics, **181**(4), 1399-1413.

CHAPTER 5

CONCLUSIONS

CHARACTERIZING THE MAIZE EPIGENOME

It is widely recognized that the state of chromatin regulates gene activity. The association of H3K4 and K36 methylation and gene activity has been well established and known to be conserved among eukaryotes (Lippman and Martienssen, 2004; Sims et al., 2004). However, the epigenetic markers that are responsible for silencing vary among different species. In plants, it has been shown that differences in epigenetic regulation tend to be associated with genome size.

While most research in epigenetics focuses on regulation of a single gene, characterizing the epigenome provides a global view of epigenetic changes across the entire genome. As a large genome species, maize has distinctive cytological features, which provide an excellent system to characterize chromatin states and their associated histone methylation patterns. Chapter 2 described a whole-genome view of the distribution of each histone methylation that is associated with transcriptional repression.

The high resolution of maize pachytene chromosomes makes it possible to clearly distinguish gene-rich euchromatin and transposon-rich heterochromatin. My results show that within the heterochromatin marks in *Arabidopsis*, H3K9me1, H3 K27me1 and 2 are associated with heterochromatin in maize. In contrast, H3K9me2 are mostly associated with euchromatic

regions. My data also suggest that both H3K9me3 and H3K27me3 are distributed exclusively in euchromatin regions. Whole-genome ChIP-chip studies in *Arabidopsis* confirmed that H3K9me3 and H3K27me3 are excluded from repetitive sequences and only localize to genes (Turck et al., 2007; Zhang et al., 2007). The discovery of localized H3K27me3 distribution also indicates that it may play a role in regulation of important developmental genes.

The centromere is a special chromatin domain that is neither euchromatic nor heterochromatic. Studies have shown that in centromere chromatin, H3-containing nucleosomes are interspersed with CENH3-containing nucleosomes. My results demonstrate that the centromere-embedded H3s are modified significantly different from those in flanking pericentromeric regions. While the pericentromeric heterochromatin is marked by H3K9me1, H3K27me1 and 2, the centromere domain is associated with H3K9me2 and 3.

Taken together, my results support that epigenetic marks evolve along with genome evolution. In a large genome such as maize, the abundantly distributed transposable elements make up a substantial portion of the genome. The difference between small genome and large genome plants in their transposon enrichment is likely to be the reason for the differentiation of the epigenetic code.

MAPPING MAIZE KINETOCHORES

In a highly repetitive region such as the centromere, it is difficult to find the single-copy markers that are required for genetic mapping. In addition, delimiting the functional centromeres

from its flanking pericentromeric regions is another challenge for centromere mapping. Chapter 3 described a novel method to generate genetic markers in centromeres. By combining ChIP with transposon display, I was able to identify which of the mapped markers are located in the functional centromeres. This method is very powerful for differentiating markers within kinetochores from those outside kinetochores.

The CR element is a universal component of centromeres in the grass family (Miller et al., 1998; Presting et al., 1998; Mroczek et al., 2006). In addition, although lacking the CR element, a lot of other species have abundant transposable elements in centromeres. Therefore, the ChIP-display method can be easily applied in many species that have CENH3 antibodies available.

EVOLUTION OF MAIZE CENTROMERES

Despite of their conserved function, centromeres sequences are evolving rapidly. Several mechanisms have been proposed to drive centromeres evolution, but none has sufficient experimental support. With the large number of centromere markers generated in our system, it is possible to detect genetic exchanges occurring within functional centromeres and further address questions in centromere evolution.

In chapter 3, two conversion-like events were detected in the IBM population. Considering the short history of the population, the frequency of conversion is very high, and should make substantial impact in centromere evolution. The LD analysis further supports the discovery of high frequency conversion events. The results suggest conversion is an important mechanism to create centromere haplotypes.

Meiotic drive hypothesis is widely accepted as a major force to drive centromere evolution. In chapter 3, we found clues that CentC arrays used to be under selection. According

to the meiotic drive hypothesis, these results indicate that CentC arrays have been historically selected by the driving system. In chapter 4, we show that maize wild relatives all contain large amounts of CentC on every centromere. Given all of above evidence, our data support the meiotic drive model, and suggest that CentC was the primary centromeric component of ancient *Zea* species.

We also observed that centromeres in different inbreds are highly diverged and can be viewed as haplotypes. Previous studies estimated that the founder population of domesticated maize was relatively small, containing as few as several hundred individuals (Eyre-Walker et al., 1998). In such a small population, it is unlikely that too many centromere haplotypes existed. Why there are numerous centromere haplotypes in modern maize populations? Based on the results in chapter 3 and 4, I propose two major mechanisms for creating centromere haplotypes:

(1) Crossing over in pericentromere creates novel combinations of CRM clusters, which later could be recruited into the centromere cores. These lead to the formation of major centromere haplotypes;

(2) Frequent conversion-like events in centromere cores could rapidly spread sequence variants to other already existing centromere haplotypes, and lead to the formation of minor centromere haplotypes.

Taken together, my work provides evidence for genetic exchanges in maize centromere, and we propose the mechanisms for how centromeres evolve.

REFERENCES

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., and Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci U S A **95,** 4441-4446.

Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature **431,** 364-370.

Miller, J.T., Dong, F., Jackson, S.A., Song, J., and Jiang, J. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150,** 1615-1623.

Mroczek, R.J., Melo, J.R., Luce, A.C., Hiatt, E.N., and Dawe, R.K. (2006). The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. Genetics **174,** 145-154.

Presting, G., Malysheva, L., Fuchs, J., and Schubert, I. (1998). A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16,** 721-728.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev **18,** 2437-2468.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet **3,** e86.

Zhang, X.Y., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. Plos Biology **5,** 1026-1035.