Conflict and Coevolution Shape the Primate Kinetochore

Kevin Roach

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Harmit S. Malik, Chair

Willie J. Swanson

Christine M. Disteche

Program Authorized to Offer Degree:

Department of Genome Sciences

University of Washington

Abstract

Conflict and Coevolution Shape the Primate Kinetochore

Kevin Roach

Chair of Supervisory Committee:

Member Harmit S. Malik

Division of Basic Sciences

Fred Hutchinson Cancer Research Center

Centromeres and the kinetochore proteins that bind them are required for chromosome segregation during every eukaryotic cell division. Despite this conserved function, ongoing conflict between selfish centromeric DNA and the proteins of the kinetochore causes both to rapidly evolve. Centromeres (centromeric DNA satellite arrays) are able to gain evolutionary advantages by driving during asymmetric female meiosis, in turn conferring a selective advantage to kinetochore proteins that can suppress centromeric imbalances. For example, the centromeric variant of histone H3 (CenH3), the basis of

kinetochore function and the epigenetic mark of active centromeres, is rapidly evolving across many taxa, including primates. While conflict with driving centromeres can explain the rapid evolution of CenH3, it cannot explain incongruent patterns of functional complementation and localization. I show that that the co-evolutionary constraints of chaperone interaction drive these divergent patterns of CenH3 functional divergence. The presence or absence of a conserved centromeric histone chaperone, SCM3/HJURP, in the genome is able to fully predict the differences in functional complementation and localization of CenH3. I expand previous evolutionary analyses using genomic sequence data and evolutionary analysis to explore how the entire primate kinetochore is shaped by conflict with centromeres. I find that the primate inner kinetochore is defined by rapid evolution, while the outer kinetochore and fibrous corona are undergoing purifying selection. Furthermore, I find that the CenH3 chaperone HJURP is also evolving under positive selection. The extensive adaptive evolution of the primate kinetochore provides new evidence of the breath of the conflict between centromere and kinetochore, as well as new targets to investigate the functional consequences of genetic conflict on cell division and kinetochore localization.

Dedication

To my wife, Amanda Roach, without whom I would not have been able to complete this project.

Acknowledgements

Over the past six years I have received support and encouragement from a great number of individuals. Dr. Harmit S. Malik has been a tireless and dedicated mentor. His guidance has made this a rewarding learning experience. I would like to thank my dissertation committee of Dr. Sue Biggins, Dr. Christine M Disteche, Dr. Mary-Claire King and Dr. Willie J Swanson for their guidance and support over the past six years as I moved from a hypothesis to a completed dissertation. In addition, Dr. King introduced me to laboratory science. Dr. Sarah Zanders and Dr. Biggins provided essential insight and assistance designing and carrying out the fungal experiments. Dr. Swanson provided valuable critiques of the evolutionary data and analysis. I would like to thank Ben Ross and Dr Maulik Patel and Dr. Nitin Phadnis for their helpful suggestions that improved various manuscripts. I would also like to thank all the members, past and present, of the Malik lab who provided a simulating laboratory environment.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

I. Introduction

Accurate chromosome segregation is essential at each eukaryotic mitotic and meiotic cell division. The basic process of segregating chromosomes during cell division has remained virtually unchanged across millions of years of eukaryotic evolution [2,4,5]. This vital function is mediated by the kinetochore complex of proteins, which binds to the centromere and provides the attachment site for spindle microtubules. These microtubules pull apart sister chromatids, ensuring proper chromosome segregation. Incorrectly attached or unattached microtubules can trigger a cascade of signals that halts cell division. Given the high degree of functional constraint and the broad similarities across eukaryotes, the apparatus for chromosome segregation is expected to reflect purifying selection, wherein natural selection acts to conserve the sequence and function of most genes by removing deleterious mutations. Contrary to this expectation, centromeric DNA and a few key proteins required for chromosome segregation evolve rapidly across broad lineages of plants and animals.

A pattern of rapid evolution is often seen in other genes involved in recurrent adaptation or those that participate in genetic conflict. In the latter scenario, classically described as a ‘Red-Queen interaction’, competing entities constantly vie for evolutionary dominance [6]. Such ‘Red-Queen’ scenarios explain rapid evolution in a wide range of biological phenomena including host-pathogen and sperm-egg interactions. Red-Queen interactions may also provide

an explanation for the surprising finding that centromeric DNA and essential genes encoding components of the chromosome segregation apparatus evolve rapidly. Here, I highlight these observations and review a model that posits that competition between homologous chromosomes during female meiosis is the driving force behind this rapid evolution [7]. This model provides broad taxonomic predictions for the evolution of centromeric DNA and proteins and introduces a general explanation for how rapid evolution at the protein-DNA interface might drive incompatibilities and reproductive isolation in animal species.

Centromeres have diverse sequence and structure

Despite carrying out identical function across eukaryotes, centromeric DNA varies widely among species [4]. The first centromeres characterized at the sequence level were the ‘point’ centromeres of *Saccharomyces cerevisiae.* 125 bp of centromeric DNA is necessary and sufficient to recruit and assemble the protein components of the budding yeast kinetochore complex (Figure 1.1). Thus, budding yeast centromeres are genetically defined. However, the simple 'point' centromeres of budding yeasts are exceptions to the rule even amongst fungi, many of which possess larger centromeres. Moreover, in most fungi (e.g., *S. pombe*), centromere identity is not dictated by the sequence of centromeric DNA but by the binding of centromeric proteins like CenH3s.

Centromeres in most multicellular organisms are even more complex, composed of large AT-rich repetitive sequences. These repetitive sequences, also termed “satellite” DNA, were identified through early cloning and sequencing

studies. The repetitive nature of centromeric DNA is challenging for modern sequencing technology and assembly. Current knowledge of metazoan centromeric DNA sequences is therefore based mostly on a few detailed studies of the centromeres of primates, *Drosophila,* and rice that required painstaking assembly and characterization over many years of effort (Figure 1.1).

Primate centromeres are composed of megabases of an AT-rich DNA sequence known as alpha-satellite. Alpha-satellite is a 171-base pair monomeric repetitive sequence that was first identified as human DNA that disrupted chromosome segregation upon introduction into the chromosomes of African green monkey cultured cells [8]. Subsequent analysis of human centromeric sequences revealed that alpha-satellites in humans and primate relatives are arranged in higher-order arrays, where the array size varies from single alpha-satellites in most species [9] to a higher-order array in human centromeres, consisting of multiple tandemly-arranged monomers in repeat units. Conservation between monomers of the same array can be as low as 70-80% identity. In contrast, conservation between multimeric repeats is much higher [10]. The higher order array structure appears to be evolutionarily young, found in only in some great apes. Moreover, some satellite-arrays are both evolutionarily young and chromosome-specific in human centromeres. For instance, the human X-chromosomal centromeric [illegible]lpha-satellite array is a 2-kilobase repeat unit that is composed of 12 monomers of the DXZ1 alpha-satellite, an arrangement that is only found in the closest relatives of humans.

The evolution of centromeric and pericentric DNA sequences is sculpted by recombination (unequal crossing over and gene conversion), which acts to homogenize sequences in the center of centromeric arrays, whereas flanking pericentric sequences accumulate mutations and transpositions [11]. Repetitive monomers of alpha-satellite sequences are therefore not exclusive to primate centromeres; they are also found immediately adjacent to centromeres in pericentric heterochromatin. These pericentric sequences do not recruit centromeric proteins but still function to ensure proper chromosome segregation by recruiting cohesion proteins. Less pair-wise sequence identity is observed among pericentric alpha-satellite monomers than between those found in centromeric arrays, which may reflect relaxed constraint or less efficient homogenization. It has been suggested that pericentric alpha-satellite monomers represent older centromeric satellites that were replaced by newly arisen variants in the middle of the centromere. In the process, the alpha-satellite monomers were gradually displaced to the edges of the centromeric array. Thus, these pericentric sequences serve as fossil records of ancestral centromeric sequences. The best-studied example of this phenomenon is found in the pericentric region of the human X-chromosome, where the oldest alpha-satellite domains are the furthest from the current centromere [10]. Homogenization of alpha-satellites is not always limited to a single chromosomal array. Indeed, a higher-order array can arise at the centromere of one chromosome during recent primate evolution, spread to other chromosomes by transposition, and become fixed [12]. Surprisingly, centromeric satellite sequences are more divergent

between species than are pericentric satellites [10]. The functional centromeric sequences are thus the most rapidly evolving between species, despite being most functionally constrained by their role in chromosome segregation.

In *D. melanogaster*, centromeric DNA from a minichromosome was found to be primarily composed of repetitive pentameric sequences interspersed with transposable elements. 85% of the centromeric sequence was found to be AATAT and AAGAG satellites, with very low sequence variation [13]. While the sequence composition of centromeric satellites seems to be invariant within species, the size of satellite arrays can vary dramatically even within members of the same Drosophila species [14]. However, centromeric satellites differ even more dramatically between species. For example, there is a hundred-fold difference in abundance for the AAGAG satellite between *D. melanogaster* and *D. erecta*, which shared a common ancestor only 5-10 million years ago [15]. Furthermore, some satellites present in the *D. melanogaster* genome are completely absent in the genome of *D. simulans*, suggesting complete turnover of centromeric sequences in less than 2.5 million years [16].

Rapid evolution of centromeric DNA has also been observed in plants. In *Oryza sativa*, centromeric regions are largely composed of two components that are interspersed with each other: a 155 bp centromeric CentO satellite, and a centromeric specific CRR (Centromeric Retrotransposon in Rice) retrotransposon. Using chromatin immunoprecipitation experiments, investigators pulled down DNA associated with rice CenH3 and found a high level of sequence

divergence at the centromeres of closely related species of wild rice. Some species of rice completely lack centromeric CentO. Comparative genomics revealed that the CentO satellites represent evolutionary young inventions that had supplanted ancestral centromeric satellites during recent evolution in rice species [17]. This implies that plants, like primates and Drosophila, have experienced dynamic evolution of their centromeric DNA.

Evolutionary forces shaping centromeric sequences

Observations from primates, *Drosophila,* and plants reveal a dynamic picture of centromeric DNA evolution, despite an essential conserved role in chromosome segregation. How do I reconcile this rapid evolution in the face of extreme functional constraint? I consider three scenarios that may explain this 'centromere paradox'.

In the first scenario, higher mutation rates may introduce rapid changes in centromeric DNA, which then have to be accommodated by changes in centromeric proteins. It is conceivable that centromeric satellite repeats might be subject to a higher mutation rate, perhaps a result of the unique chromatin environment they are in or as a result of their unique AT-rich nucleotide composition. Indeed, recent reports from budding yeast suggest that centromeric DNAs might be subject to elevated mutation rates [18]. Nonetheless, for newly arisen mutant centromeric sequences to survive under this co-evolutionary scenario, they would have to encounter rare compensatory mutations in a centromere binding protein to avoid being eliminated by purifying selection. Thus,

even if centromeric mutation rate were higher, the frequency of compensatory mutations in centromeric proteins would be rare, so this co-evolutionary scenario appears unlikely to account for the centromere paradox.

In a second scenario, biased gene conversion may rapidly alter centromeric DNA satellites, followed by co-evolutionary accommodation by centromeric proteins [19]. A new centromeric allele could arise which was favored by recombination (or by biased gene conversion). This new centromeric variant could then spread throughout the satellite arrays of homologous centromeres in the species. If such biased gene conversion events result in the fixation of a new centromeric DNA array that compromises accurate chromosome segregation, strong selection will act on centromere proteins to restore function. Since centromeric DNA is strongly impacted by recombination, this scenario has some explanatory power. However, the co-evolutionary process should stop once a satellite variant has driven to fixation and centromeric proteins have co-evolved. For it to start anew, new recombinational bias would have to be invoked in which the previously successful centromeres were replaced by newer versions. This process therefore seems unlikely to produce the recurrent patterns of rapid evolution observed across taxa. Furthermore, under this scenario, the new centromere is fixed not because of increased centromeric ability but would have to be fixed in spite of decreased centromeric ability to account for the positive selection of centromeric proteins.

I favor a third possibility in which increased centromeric ability translates directly to *increased transmission* during chromosome segregation [4,7,20]. This opportunity for increased transmission arises from the unique nature of female meiosis in both plants and animals. Unlike mitosis or male meiosis, female meiosis in plants and animals is an asymmetric cell division. While mitosis produces two identical daughter cells and male meiosis results in four viable gametes, only one of the four products of female meiosis can be passed on to the next generation in the oocyte. There is, therefore, an opportunity for competition among loci on homologous chromosomes to compete for positioning and inclusion in the oocyte. Centromeres are ideally positioned to compete during female meiosis. Under this scenario model, centromeres competitively orient towards the 'preferred pole' during meiosis I (Figure 1.2) perhaps by recruiting more microtubules and biasing the acentrosomal spindle in female meiosis, resulting in a transmission advantage in female meiosis. It bears mentioning that unlike post-meiotic dysfunction following male meiosis, centromere drive incurs no fertility cost [21]. I posit that this competition, or centromere-drive is the underlying genetic conflict that explains the rapid evolution of centromeric DNA. If this transmission advantage in female meiosis has a subsequent negative consequence in male meiosis, it would require centromeric proteins to adapt to restore male fertility. The key advantages of the centromere-drive model (over the biased gene conversion scenario) are two-fold. First, the new centromeres have increased rather than compromised centromere function. Second, it has the dynamics of a classical meiotic drive-suppression

system, in which the cheating centromere 'wins' initially, but is 'suppressed' in subsequent steps via adaptation. A key discriminator between these models is that 'centromere drive' should be highly subject to taxonomic differences in meiotic programs, whereas biased gene conversion should not. As I highlight here, the genetic and taxonomic patterns of centromeric protein evolution support this model overwhelmingly.

Evidence of centromeres ‘cheating’ asymmetric meiosis

The first cytological evidence that chromosomes can exploit asymmetries in female meiosis came from studies of B chromosomes in grasshoppers, which exploit this asymmetry to enhance their own transmission [22]. There is now increasing evidence that many animals and plant chromosomes are shaped by similar biases in female meiosis. One of the more dramatic examples is from Robertsonian chromosomes in humans. Robertsonians are formed by the fusion of two acrocentric chromosomes into a single metacentric chromosome. Offspring of heterozygous individuals (carriers) receive either the Robertsonian fusion or the two wild type acrocentric chromosomes. In humans, male Robertsonian carriers transmit the wild-type chromosomes and the Robertsonian fusion to their offspring at equal rates, whereas the Robertsonian fusion chromosome is preferentially transmitted to 58% of the offspring from a female carrier [23]

Further evidence for centromere drive comes from genetic studies in *Mimulus* (monkeyflower) species. In intraspecies crosses of *M. guttatus*, there is

a transmission bias associated with a *D* locus, which is thought to be an expansion/duplication of a centromeric region. Chromosomes bearing the *D* locus were preferentially transmitted at 58% through female meiosis, whereas no distortion was seen in male meiosis [24]. In earlier crosses involving interspecies F1 hybrids between *M. guttatus* and *M. nasutus*, the *D* locus exhibited a 98% transmission bias in female meiosis [25]. In the absence of viability differences, such strong transmission bias could only result from distortion during meiosis I, leaving the centromere as the most probable candidate for the *D* locus [21,25].

Both of these examples highlight a key point about consequences of transmission distortion in female meiosis i.e., an accompanying defect in male meiosis. For instance, human male carriers of Robertsonian fusions suffer a high rate of fertility defects [26]. Similarly, *M. guttatus* males that are homozygous for the *D* locus suffer 20% lower pollen counts [24]. This suggests that either heterozygosity or homozygosity of a driving centromere can be deleterious to male meiosis. One likely explanation is that unequal centromere strengths might result in tension inequity and increased nondisjunction in male meiosis [7,20]. While this could conceivably occur between any pair of chromosomes, the sex chromosomes might be especially susceptible to the deleterious effects of centromere drive. For instance, in XY systems such as mammals, the Y chromosome does not undergo female meiosis and is thus not subjected to centromere drive. Repeated rounds of competition and drive on X chromosomes could result in a ‘super-X’ centromere competing against a much weaker Y centromere. This inequity in centromere affinity can lead to greater rates of

nondisjunction in XY male meiosis and potentially result in male sterility. Thus, selection will favor alleles of centromere binding proteins that alleviate drive and restore meiotic parity by adaptively altering their DNA-binding specificity.

The trilaminar eukaryotic kinetochore

The kinetochore is a complex structure that contains more than 90 proteins (reviewed in [27]). Electron microscopy has revealed that the kinetochore consists of three layers, although this trilaminar appearance is visible only briefly during the cell cycle in most organisms (budding yeasts like *Saccharomyces cerevisiae* are a notable exception). The inner kinetochore associates directly with the DNA of the centromere and is present throughout the cell cycle. The outer kinetochore provides a binding site for microtubules. In most organisms including mammals, this layer is not present during interphase but forms only after the chromosomes condense in preparation for cell division. The outermost layer of the kinetochore is referred to as the fibrous corona and is only visible in the absence of microtubule binding. The fibrous corona consists of proteins involved in microtubule anchoring and in regulating the fidelity of chromosome segregation. This outermost layer is only transiently associated with the kinetochore, forming after DNA condensation and dissociating shortly after microtubule binding is achieved. Despite structural and temporal differences in localization, all three kinetochore layers contain proteins that are essential for chromosome segregation. Intriguingly, the evolutionary patterns of each of these layers also differ widely, reflecting the divergent evolutionary constraints that they are subject to. I begin by focusing on the inner kinetochore proteins that include

those proteins most closely and permanently associated with the centromeric chromatin[27-29].

Eukaryotic genomes are packed into higher order chromatin by wrapping DNA around nucleosomes [30]. Nucleosomes are responsible for the bulk packaging of eukaryotic genomes and consist of an octamer of histone proteins, two copies of each of H2A, H2B, H3, and H4. However, centromeric chromatin is uniquely different, consisting instead of centromeric DNA wrapped around nucleosomes that contain a variant of histone H3. This centromeric variant of H3, hereafter CenH3, was discovered first in biochemical studies in humans and was one of the first kinetochore proteins to be discovered [31,32]. CenH3s have been found in every eukaryotic genome established so far and are a defining characteristic of centromere function [33]. For instance, CenH3s only mark the active centromere of a dicentric chromosome, which contains two potential centromeres. Chromosomes that acquire neocentromeres i.e., are able to undergo chromosome segregation despite losing a centromere, are also marked by CenH3 nucleosomes [34,35]. It is therefore unsurprising that CenH3 nucleosomes are absolutely essential for kinetochore formation and proper chromosome segregation.

CenH3 differs from canonical histone H3 both structurally and evolutionarily [30] (Figure 1.3). Both proteins contain a histone fold domain that binds the other proteins of the nucleosome and DNA, as well as a N-terminal tail, which is bound and modified extensively by other proteins. The Loop1 region of

the histone fold domain of CenH3 is longer than the canonical variant. Loop1 is the region of the histone fold domain that contacts the DNA, suggesting that CenH3 may have increased sequence specificity. In addition, while canonical H3 is highly conserved, the N-terminal tail is so variable in length and sequence it is difficult to align the tails of CenH3 across taxonomic groups. While it is possible that this variability could result from apparent lack of evolutionary constraint, detailed studies of CenH3 evolution from insects, plants and mammals has produced evidence of positive selection (higher than expected numbers of non-synonymous changes) acting on the N-terminal tail [5,36,37]. In contrast, the carboxy-terminal histone fold domains (HFD) of CenH3s are much more constrained by virtue of their interaction with the other canonical histones. However, even the HFD domain of CenH3s evolves much more rapidly than canonical H3s [38,39]. Again, studies in both animals and plants suggest that the HFD domain and the Loop1 region in particular (which is predicted to make specific DNA contacts and is solvent accessible in a recent crystallographic structures [3,40] is subject to positive selection.

Findings of positive selection argue that the rapid evolution of CenH3s in both plants and animals has been driven by selective pressure rather than the lack of selective constraint. This prediction is also borne out by functional experiments using CenH3 chimeras created from closely related species. For instance, swapping of the rapidly evolving Loop1 of *Drosophila melanogaster* for the orthologous domain of *D. bipectinata* CenH3 results in a dramatic loss of localization [38]. This revealed that selection was acting on the centromeric

protein-DNA interface, and that Loop1 was necessary for centromeric localization in *Drosophila*. Likewise, CenH3 from closely related plants proved unable to complement the loss of endogenous CenH3 protein in *Arabidopsis* [41]. The essential and conserved function of CenH3s makes these findings of rapid evolution and functional divergence even more intriguing.

Despite rapid evolution and differences in HFD and N-terminal tails, CenH3 must still substitute for H3 in centromeric nucleosomes. The exact nature of the centromeric nucleosome remains a subject of vigorous, current debate. Some structural studies have proposed a canonical nucleosome-like octameric structure, with 2 copies of CenH3s replacing canonical H3s [42,43]. In contrast, other studies that examined the native confirmation of centromeric nucleosomes have argued both that they represent hemisomes (with only one copy each of H2A, H2B, CenH3 and H4) [44]. Moreover, these latter studies have argued for an unusual left-handedness of the DNA-wrap in centromeric nucleosomes, which would be incompatible with an octameric model. This debate has been summarized elsewhere [45,46], and I include it here only to highlight that there are still important, outstanding questions about the nature of centromeric histones and nucleosomes.

It is imperative that CenH3s be exclusively deposited at the centromere. Manipulations that result in ectopic deposition of CenH3s in euchromatin can result in ectopic kinetochore formation in Drosophila and dicentric chromosomes

[47]. In both fungi and mammalian cells, a domain encompassing Loop1 and the adjacent alpha-helix 2 of the histone fold domain are necessary and sufficient to target CenH3 to the centromere [48,49]. Replacement of the canonical H3 Loop1 and alpha-helix 2 with the corresponding domain from CenH3 results in a change of localization from the chromosome arms to the centromere. Reciprocal experiments replacing CenH3 Loop1 and alpha-helix 2 with canonical variants results in a loss of centromeric localization. Curiously, this region (termed the CATD) does not universally confer this property in all CenH3s. For instance, Loop1 and alpha-helix 2 are not sufficient to confer centromeric targeting in *Arabidopsis thaliana*, although they appear necessary [41]. While the cause of the change in the CATD is not known, both fungi and mammals, but not plants, have a chaperone protein (Scm3 for fungi and HJURP for mammals) that is required for CenH3 localization [50-54]. Since the Scm3/HJURP chaperones bind to the CATD of CenH3 , it is difficult to distinguish between the possibility that the CATD targets the centromere directly versus the possibility that the CATD simply allows the chaperone to bind CenH3 and deposit it at the centromere[54,55]. Recent studies further argue for the possibility of a direct and dominant role of HJURP chaperones in determining sites of CenH3 localization [56]. In either case, it is intriguing that the CenH3 chaperone Scm3/HJURP is at least as old as the common ancestor of fungi and animals, yet has been apparently lost in multiple animal lineages and perhaps replaced by non-orthologous proteins like CAL1 in Drosophila [57,58].

Evolution of the proteins of the inner kinetochore

Kinetochore autoantigens in scleroderma patients lead to the discovery of two additional inner kinetochore proteins, CENP-B and CENP-C. CENP-C shares several characteristics with CenH3; it is present in all eukaryotes, essential for kinetochore function and binds centromeric DNA [59]. CENP-C can be reliably aligned only within a limited phylogenetic group such as mammals, but the variation in sequence and length is so great that it is often not possible to align the protein between more diverged groups. However, all CENP-Cs contain a 24 amino acid region of homology located in the carboxy terminus of the protein, termed the CENP-C domain. This small domain is the only widely shared feature of CENP-C. Studies in mammals have shown that CENP-C recognizes and binds specifically to centromeric DNA [60,61]. Reports from maize suggest that DNA-binding of centromeric DNA by CENP-C requires the stabilizing influence of binding to long single-stranded centromeric RNAs [62]. There are several independent centromere recognition and DNA binding domains present in the middle and carboxy terminus of CENP-C, but CenH3 is still required for localization of CENP-C to the centromere. Dependence on CenH3 is not a universal characteristic of CENP-C, however, as *Drosophila* studies have shown an interdependent relationship between CENP-C and CenH3 [63]. Depletion of CENP-C by RNA interference disrupts the deposition of new CenH3 to the new centromere; likewise, depletion of CenH3 disrupts CENP-C localization. Studies in plants and mammals have found that CENP-C is rapidly evolving in these systems, often much more rapidly than CenH3 [2,5]. In addition to rapid

sequence evolution, exons of CENP-C in plants have been duplicated or deleted along multiple separate lineages (Figure 1.4). Unlike in CenH3, it is not yet clear what the effects of such rapid evolution of CENP-C are on centromere binding or protein interactions. The wide distribution of the rapid evolution throughout the amino acid sequence of CENP-C has confounded the prediction of the effects of the changes to CENP-C. However, given the ubiquity of this protein and its importance for centromeric function, such studies would be important in revealing the functional consequences of such rapid evolution.

CENP-B evolution reveals a markedly different, but no less intriguing, history of genetic innovation. CENP-B is restricted to primates and certain fungi and not present in most eukaryotic genomes [64]. CENP-B recognizes a conserved 17-bp “CENP-B box” present in primate centromeric satellites [65]. Interestingly, despite rapid turnover and divergence of centromeric satellites in primates, CENP-B boxes appear to be retained in most centromeres, suggesting an important centromeric function. CENP-B is not essential for the maintenance or function of active centromeres; indeed, it appears to have been lost in some mammalian lineages and gene knockouts in mice have minimal effects on chromosome segregation. However, when present, it is essential for the formation of neocentromeres if chromosome loses its active centromere [66]. CENP-B will localize to these neocentromeres despite the lack of a canonical CENP-B box. CENP-B shows no evidence of rapid evolution of the protein

coding sequence but is a novel case of domestication of a selfish mobile genetic element [5]. CENP-B is a domesticated remnant of a *pogo-like* transposon. The *pogo-like* transposable elements are present across a wide variety of species but CENP-B was domesticated only in mammals[64]. Even more surprising, is that the domestication of a *pogo-like* transposon has occurred not once but at least twice. *Schizosaccharomyces pombe* has three proteins (Abp1, Cbh1, and Cbh2) that share significant sequence similarity to CENP-B and display centromeric localization [64]. These proteins bind degenerate sequences found at the centromere and are essential but somewhat redundant in function. This is an intriguing example of two very rare forms of evolution, domestication of a selfish element and convergent evolution of non-orthologous but functionally equivalent CENP-B kinetochore genes.

Extensive immunoprecipitation experiments carried out in human cell lines resulted in the discovery of a network of additional inner kinetochore proteins that were closely bound to CenH3 chromatin and CENP-C [28,67]. The 16 proteins of this Constitutive Centromere Associated Network (CCAN) make up most of the mammalian inner kinetochore and as the name suggests are present at the centromere throughout the cell cycle [68]. All the members of the CCAN are essential for proper kinetochore formation and segregation of chromosomes [27]. The CCAN proteins have been discovered only recently and work is just beginning to elucidate their evolutionary history and function but some progress has already been made. Early findings suggest that the CCAN interacts directly with the centromeric nucleosomes and assists CENP-C to create a scaffold for

outer kinetochore proteins to bind. CCAN proteins are dependent on CenH3 and (with the exception of CENP-W and CENP-T) CENP-C for localization [28]. Orthologs of CCAN proteins in yeast show as little as 20% sequence similarity, making the search for orthologs in other taxa very difficult. More research is needed to discover whether CCAN proteins are truly absent from other taxa, as well as whether they are rapidly diverging if present.

Conservation and evolution of transiently associated outer kinetochore proteins

The outer kinetochore and fibrous corona, unlike the inner kinetochore, are only transiently associated with the centromere. There are several essential complexes of outer kinetochore proteins, which are deposited at different points in the cell cycle. This process starts immediately after cell division when the Mis18 complex, which in mammals consists of Mis18a, Mis18b and KNL2 in concert with RbAp48, remodels centromeric chromatin to allow the deposition of new CenH3 [69,70]. Mis18s have been discovered in several metazoans and fungi. However, fungi possess only one Mis18, leading to the possibility that Mis18a and Mis18b are the result of duplication and subfunctionalization. More research is needed to understand the evolutionary forces and consequences of Mis18 duplication.

After CenH3 is deposited, a large group of proteins is recruited to the kinetochore in prophase. These proteins form the physical attachment between the kinetochore and spindle microtubules. The four-subunit Mis12 complex is

thought to bind to CENP-C and provide a scaffold for the rest of the outer kinetochore [28,71]. Each subunit of the Mis12 complex appears to be conserved and widely present in eukaryotes. KNL1 and the NDC80 complex bind to the CCAN and Mis12 to form the attachment sites for microtubules [72-74]. KNL1 is essential for kinetochore formation in *Caenorhabditis elegans* but little is known of its evolution [75]. The NDC80 complex is another protein complex of the outer kinetochore that is thought to be conserved across eukaryotes. However, one Ndc80 protein- *mitch,* or Spc25, was recently shown to be rapidly evolving in the *Drosophila* lineage [76]. Intriguingly, like CenH3 and CENP-C, *mitch* is essential for chromosome congression and segregation. There is no evidence of rapid evolution of *mitch*/Spc25, or any other NDC80 proteins in lineages outside *Drosophila* but this possibility have not been actively pursued. Thus, the evolutionary forces shaping the outer kinetochore proteins remain largely unknown.

Evolutionary dynamics at the outer edge: the fibrous corona

The kinetochore structure most distal to the centromere is referred to as the fibrous corona due to its characteristic appearance by electron microscopy. This structure is transiently formed after nuclear envelope breakdown, and can be visualized only in the absence of microtubules. This evidence argues that microtubule capture by the kinetochore results in alteration of the structure of the fibrous corona. Indeed, the assembly and disassembly of the fibrous corona appears to be dynamically regulated through complex signaling interactions [27]. For instance, proteins that are known to localize to the fibrous corona include

components of the spindle checkpoint apparatus such as Mad1, Mad2, Bub1, and BubR1, which function to monitor spindle tension. Additional proteins that have been found to localize to the fibrous corona in a variety of species include dynein, dynactin, CLIP170, CENP-E, Cdc20, the Rod complex, and LIS1 [77]. The evolution of most of these proteins has not been investigated in the same manner as CenH3 or CENP-C, but most appear present across eukaryotes taxa. However, simple analysis of protein presence and absence in different lineages yields some interesting observations.

One such observation involves the DASH (Dam1) complex in yeast. This complex binds to microtubule plus ends and is essential for chromosome segregation and viability in *S. cerevisiae* [78]. Indeed, the complex is sufficient to segregate DNA *in vitro.* Despite this important function, the DASH complex differs in essentiality between *S. pombe* and *S. cerevisiae*. While mutations in DASH complex members are lethal in *S. cerevisiae*, similar mutations merely result in growth defects in *S. pombe* [79]. Additionally, key residues that have been found to be important in *S. cerevisiae* are not conserved in *S. pombe*. Interestingly, DASH homologs have not been found outside of fungi. Instead, the nonhomologous Ska proteins may play an analogous role in vertebrates and other eukaryotes [80]. A second example concerns the CENP-E kinesin-like protein. Dimeric CENP-E proteins form long fibrils that are transiently associated with the kinetochore. It is essential for chromosome segregation and recent reports show it plays a role in microtubule plus-end elongation [81]. Again, despite this essential function, CENP-E homologs have not been identified in

yeast. These two examples illustrate that components of the kinetochore distal to the centromeric DNA also seem to exhibit features of evolutionary lability.

A model of intragenomic conflict shaping centromeres

Patterns of evolution of kinetochore proteins, and the selective forces that drive them, are poorly understood outside of a handful of basal centromere binding proteins. Both inner and outer kinetochore proteins are essential to segregate chromosomes but potentially have different evolutionary forces acting on them. I proposed a model to explain why some essential proteins of the inner kinetochore rapidly evolving. This model stems from a unique opportunity for intragenomic conflict in female meiosis [4,7,82]. Kinetochore proteins that directly bind the centromere are ideally positioned to adapt and restore fertility [82]. The effect of centromere drive on kinetochore proteins outside the nucleosome has not yet been investigated systematically.

The centromere drive model has all the hallmarks of a classic 'Red Queen' conflict [20]. First, centromeres gain a transmission advantage through innovation, increasing in frequency due to increased fitness in female meiosis. The selection on kinetochore proteins to suppress centromere drive arises only later after the driving centromere begins to disrupt male meiosis. This mirrors the classic genetic conflict wherein one party gains an evolutionary advantage (greater transmission) at the expense of the other (lower male fertility), resulting in repeated rounds of adaptation.

To restore meiotic parity, the selected protein must be able to distinguish and suppress the driving centromere, implying that only proteins closely associated with the centromere, such as CenH3, CENP-C and the CCAN complex will be subject to selection by centromere drive. In this case, rapid evolution will be seen only or primarily in the centromere binding proteins of the kinetochore and not the outer kinetochore proteins that interact with microtubules. I test this hypothesis by leveraging the rapidly expanding availability of primate genomic data to explore the effects of genetic conflict on the kinetochore. First, I performed an unbiased analysis of the evolutionary history of the whole primate kinetochore, cataloging how selection to suppress centromere drive and purifying selection to maintain function have interacted to shape the kinetochore. Next, I show that the requirements of the protein-protein interaction between CenH3 and its unique chaperone, HJURP, have altered the evolutionary and functional dynamics of the centromeric histone. HJURP has locked the interaction domain of CenH3 in a long-term co-evolutionary relationship that has dramatically altered the functional consequences of CenH3's diversification. Finally, the identification of HJURP as an essential chaperone for CenH3, and my findings that it restricts CenH3 adaptive capacity, led me to investigate whether the centromeric chaperone is a potential suppressor of centromere drive.

II. Conflict Shapes the Evolution of Primate Kinetochore

Thus far, studies on the evolution of centromeric DNA evolution have been relatively sparse due to the repetitive nature of centromeric sequences in most species. In addition, evolutionary studies of centromeric proteins have been limited to a few key components of the inner kinetochore. As a result we have a limited understanding of the functional and evolutionary pressures acting on the centromeres or kinetochore proteins. However, we have a clear set of predictions provided by the 'centromere drive' hypothesis; namely, that kinetochore proteins that directly recognize or interact with centromeric DNA are positioned to suppress centromere drive and will be rapidly evolving, whereas structural or mechanical proteins at the kinetochore periphery that do not directly interact with centromeric DNA will be relatively constrained in their evolution.

Thus far, investigation of the evolution of kinetochore proteins has been restricted to CenH3, CENP-B and CENP-C [31,83]. Studies in both *Drosophila* and plants have shown that CenH3, which directly binds centromeric DNA as a part of nucleosomes, is evolving under positive selection. In each case, the evolution is especially rapid in the Loop1 region of the histone fold domain, a region that contacts DNA while bound to the nucleosome and is critical for correct centromeric localization[1,37-39,84]. CENP-B localizes independently of CenH3 and is the only primate kinetochore protein with sequence specific DNA binding, recognizing a 17-bp "CENP-B Box" found in some alpha-satellites [65,85]. CENP-B Boxes are not found at all centromeres and CENP-B is not required for

centromere function or propagation[65,86-89]. Unlike CenH3, CENP-B features high sequence conservation [5]. CENP-C, present in all eukaryotes, was one of the first kinetochore proteins discovered and is believed to be a part of the foundation of the kinetochore [28,31,60,61,63,83,90-97]. While it is dependent on CenH3 for localization to the centromere, CENP-C binds centromeric DNA and is evolving under positive selection in plants and animals[2,5,29,98]. However, until now, there has been no systematic look at how positive selection has shaped the entire kinetochore apparatus, including whether evolutionary pressures differ between inner and outer kinetochore. This is partly because the kinetochore has been well described in only a few key species, most thoroughly in the budding yeast *S. cerevisiae,* and *H. sapiens* [27]. Budding yeast do not undergo asymmetric meiosis; therefore, we do not expect to centromeres to drive selection of their kinetochore. *H. sapiens* undergo asymmetric meiosis, however, and only recently has a rapidly-expanding quantity of sequence data available for *H. sapiens* and related primates become available. I began my analysis of the primate kinetochore by investigating the foundational proteins of the inner kinetochore.

Rapid Evolution of the Inner Kinetochore

A subset of kinetochore proteins must constitutively associate with centromeric DNA to form a base for eventual kinetochore assembly. Recent work suggests that this subset of kinetochore proteins DNA extends well beyond CenH3, CENP-B, and CENP-C [28,29,99-101]. Although not all kinetochore proteins or complexes of proteins have been functionally characterized, recent

sequencing studies of primate genomes allow me to study the extent to which centromere drive shapes the proteins involved in chromosome segregation. The increasing availability of genomic sequence data allows me to use the signals of selection to determine which kinetochore proteins have been shaped by centromere drive. Further, only proteins that directly interact with centromeric DNA are positioned to suppress driving centromeres, allowing selection to uncover potential new centromere binding proteins.

I first obtained sequence for the coding region of each protein from the five assembled primate genomes or alternate high-quality sequence sources. Signals of rapid evolution were numerous in initial sliding window calculations, and for those proteins in which two independent pairwise comparisons had a window with a dN/dS > 1, I also performed a maximum-likelihood analysis to detect positive selection [102] (Table 2.1). The power of the codeml analysis was limited by the low number of primate sequences prompting an additional analysis including sequences from uncurated databases [103]. The additional primate sequences allowed me to increase my ability to detect selection, increasing the number of positively selected proteins from 6 in my initial analysis to 10 in the final analysis.

I began by investigating those proteins most closely associated with the centromeric DNA. It is not known to what extent the proteins of Constitutive Centromere Associated Network (CCAN) recognize or bind alpha-satellite DNA, but they localize constitutively to the centromere throughout the cell cycle

[28,29,67]. The CCAN proteins form part of the inner kinetochore but have begun to be grouped into functional units. CENP-T, CENP-W, CENP-S, CENP-X together form hetrotetrameric nucleosome-like structures at centromeres [68,101,104,105]. These four non-canonical histones have been identified in animal and fungal genomes, indicating this complex of proteins was born in the ancestor of opisthokonts or earlier [105]. However, none of these proteins have been found in the well-annotated *Drosophila* genome, nor any insect genome; thus, these noncanonical nucleosomes can either be lost or have diverged so extensively that homology is undetectable in some lineages. In the T/S/W/X complex, only *CENP-T* showed evidence of positive selection ($p<0.05$) whereas *CENP-S* and *CENP-W* showed signs of strong purifying selection. The CENP-S/X complex is dependent on CenH3, the CENP-H/I/K complex and CENP-T/W for localization [67]. CENP-S/X's downstream role in kinetochore formation, combined with a lack of positive selection suggests that these proteins have limited capacity to suppress centromere drive. In contrast, the CENP-T/W heterodimer utilizes multiple amino acids to bind DNA. CENP-T's K467, R480, and K506 and CENP-W's R19, K23, R24, R34, K66, and R68 are essential for DNA binding. However, none of these amino acids show evidence of rapid evolution[28,105]. Evidence of rapid evolution was spread throughout the protein and in multiple domains of CENP-T. This suggests that there are additional amino acids important for DNA discrimination and targeting, which may not have been recognized via traditional biochemical analyses [28,105]. The CENP-T/W complex provides a platform for kinetochore assembly through a long N-terminal

tail that is bound by outer kinetochore proteins [101,104]. Ectopic kinetochore formation can be induced by targeting CENP-T to chromosomal arms, bypassing the requirement for CenH3 and play a fundamental role in centromere specification[101]. Thus, centromeric DNA binding and discrimination by CENP-T may create an important, alternative path to suppress centromere drive.

Just one other CCAN complex of proteins shows evidence of centromeric DNA interaction. CENP-N specifically recognizes CenH3's CATD and binds CenH3 containing chromatin [106,107]. While CENP-N is dependent on CenH3, CENP-N also facilitates CenH3 deposition and is required for the localization of all other members of the CCAN [99,106,108]. Given this close association with centromeric nucleosomes, it is not surprising that *CENP-N* shows strong evidence ($p<0.01$) of positive selection. CENP-C and CenH3 are also directly bound by CENP-L and CENP-M, two other CCAN proteins that form a complex with CENP-N [68,106,107]. Both CENP-L and CENP-M require CENP-N for localization, but CENP-N is not dependent on either [106,108]. Very little is known about the function of CENP-L or CENP-M, and despite CENP-N's role in CenH3 deposition, it is hypothesized that they provide a binding site for proteins of the outer kinetochore. However despite that proposed structural role, *CENP-L* is rapidly evolving ($p<0.05$), suggesting that it is intimately involved in interactions with driving centromeres.

Table 2.1	**Summary of the rapid evolution of the CCAN**	
	F61	3x4

Protein	Species	P value	% of AA ω >1	ω (for >1)	Sites ω>1 (Pos Prob >95%)	P value	% of AA ω>1	ω (for >1)	Sites ω>1 (Pos Prob >95%)
CENP-S	10	1.00	0.00%	1.00		0.99	0.00001	1.00	
CENP-T	9	0.01	2.13%	4.89	253N	0.04	1.11%	6.37	253N
CENP-W	10	1.00	0.00%	1.00		1.00	0.00%	1.00	
CENP-X	8	0.76	11.39%	1.85		1.00	12.94%	1.51	
CENP-H	9	<0.05	30.73%	2.40	10A, 33A, 121R, 177T, 189S	0.07	34.56%	1.95	
CENP-I	7	0.23	0.28631	1.70		0.48	30.85%	1.45	
CENP-K	11	1.00	0.00%	1.00		0.93	24.02%	1.09	
CENP-L	9	0.02	4.10%	3.13	13A	<0.01	3.85%	3.34	13A, 14S
CENP-M	8	0.93	15.53%	1.00		0.86	14.27%	1.10	
CENP-N	8	<0.01	0.02141	7.30	54H, 60R, 167M	<0.01	0.02346	6.82	54H, 60R, 167M
CENP-O	7	<0.01	35.14%	1.98	28V, 32R, 41V, 66V, 67R, 69R, 93A	0.03	35.57%	1.76	28V, 41V, 66V, 67R
CENP-P	8	<0.01	0.41%	38.81	213R*	<0.01	0.43%	34.85	213R*
CENP-Q	9	0.15	33.90%	1.97		0.29	29.35%	1.81	
CENP-R	9	0.30	7.55%	4.29		0.37	4.68%	4.52	
CENP-U	9	<0.01	10.28%	3.71	111S, 283V, 288Q, 289F	<0,05	7.56%	3.51	283V, 289F

CENP-H, CENP-I, and CENP-K form a complex of proteins, essential for the localization of the outer kinetochore. Depletion of any of these proteins causes cell cycle arrest, kinetochore defects, and chromosome missegregation [73,98,99]. CENP-H/I/K are able to localize only when all other members of the complex are present, in addition to CenH3 [99]. CENP-C is dependent on both CENP-I and CENP-H for centromeric localization, but all three proteins coordinate to form a scaffold for the rest of the kinetochore to bind [73,99,109-114]. Despite this upstream role in kinetochore formation, there is no

evidence of centromere binding by any member of the complex, forming the bridge from centromeric chromatin, defined by CENP-A containing nucleosomes, to the structural and mechanical proteins of the outer kinetochore. CENP-H/I/K's importance in recruiting other inner and outer kinetochore proteins could make them candidates to suppress centromere drive, but only CENP-H shows any evidence of positive selection (Table 2.1). However, this evidence supporting rapid evolution of CENP-H was limited and sensitive to the codon substitution model used (marginally insignificant in the F3x4 model); this conclusion will have to be revisited with additional sequence analyses.

The rest of the members of CCAN are not functionally well characterized but show wide variation in evolutionary history. CENP-O, CENP-P, CENP-Q, CENP-R, and CENP-U depend on the localization of CENP-H/I/K as well as all other upstream proteins [28]. CENP-O, CENP-P, CENP-Q, and CENP-U form a stable complex that transiently associates with CENP-R. All the proteins except for CENP-R were required for kinetochore localization of the complex, and for recovery from spindle damage [67,68,115-120]. There is no evidence of CENP-O/P/Q/R/U recognizing or binding centromeric DNA or chromatin. Despite the proposed lack of interaction with centromeric DNA, *CENP-O* ($p < 0.01$), *CENP-P* ($p<0.01$) and *CENP-U* ($p<0.01$) all show robust evidence of rapid evolution (Table 2.1). Even CENP-Q ($p\sim0.15$) and CENP-R ($p\sim0.30$) show some diversification which does not rise to statistical significance; additional sequencing might provide enough evidence to discriminate whether positive selection has occurred on these two proteins. It is intriguing that so many proteins in this complex

shows evidence of positive selection, despite the downstream role of the complex in the inner kinetochore formation.

Table 2.2		Rapid evolution of the KNL2 in the Mis18 complex							
		F61				3x4			
Protein	# of Primates	P value	% of AA sites ω >1	ω (for >1)	Sites ω>1 (Pos Prob >95%)	P value	% of AA sites ω>1	ω (for >1)	Sites ω>1 (Pos Prob >95%)
Mis18a	10	1.00	0.00%	1.00		1	0.00%	1.00	
Mis18b	10	0.19	3.68%	3.55		0.18	3.32%	3.80	
KNL2	11	1.20478E-07	7.12%	3.74	157K, 865R	5.64549E-07	7.62%	3.47	157K, 490Q, 865R, 1080L

Outside of the CCAN, there are additional proteins that transiently associate with centromeres but are a part of the inner kinetochore. MIS18α, MIS1β, and KNL2, form a complex that localizes to the centromere during CenH3 loading [69,70,121]. While there is no direct interaction between the Mis18 complex and CenH3, depleting any component of the complex abolishes recruitment of new CenH3 to the centromere [69,70,121]. It is hypothesized that the Mis18 complex acts to create a permissive, deacetylated chromatin environment for CenH3 deposition [69,121]. It is unknown how Mis18 recognizes centromeres or modifies centromeric chromatin, making it difficult to predict whether Mis18 proteins will be able to suppress centromere drive. MIS18α showed strong conservation while MIS18β showed signs of diversification, but not enough evidence to support a statistically significant conclusion of adaptive evolution (Table 2.2). *KNL2*, however, showed strong evidence of positive selection ($p<1x10^{-6}$). This evolutionary history indicates that the Mis18 complex

may have a direct role in recognizing centromeric DNA, and specifying kinetochore location. The adaptive evolution of this complex and its role in CenH3 loading makes it an ideal candidates for further characterization of the effects of driving centromeres.

The outer kinetochore localizes to the centromere only after prophase as the chromosomes begin to congress. The first of these outer kinetochore proteins form the Mis12 complex, comprised of MIS12, DSN1, NNF1, and NSF1. The Mis12 complex was thought to help maintain CenH3's centromere localization disrupting ectopically localized CenH3, helping ensure proper localization of the kinetochore [27]. However, more recent evidence shows that the MIS12 complex binds to CENP-C and the CENP-H/I/K complex to join the inner kinetochore and the outer kinetochore. The MIS12 complex universally shows strong purifying selection (Table 2.3). This is the first indication that those kinetochore proteins not directly recognizing centromeric DNA or chromatin are under significantly selective pressure.

The rest of the outer kinetochore is loosely called the NDC80 group of proteins. Though not a complex of directly bound proteins, they coordinate to form a coiled-coiled structure essential for kinetochore-microtubule interactions. This structure links microtubules with the Mis12 complex and CENP-C/H/I/K[68,71,73,90]. The NDC80 proteins localize through to the centromere by binding the CCAN proteins[68,73,90,112,113]. Unsurprisingly, most of these proteins show strongly sequence conservation in primates, with only *CASC5* found to be rapidly evolving (Table

2.3). CASC5 is a large protein with relatively poor sequence coverage. Despite an exhaustive search of sequence databases, CASC5 alignments contained many gaps and unalignable regions. The relative scarcity of CACS5 sequence and evidence for selection could be a result of high rates of mutation and rearrangement at the genomic locus or sequence features. CASC5 localizes to the kinetochore from G2 until late anaphase, directly targeting protein phosphatase 1 (PP1) to the outer kinetochore through direct interaction by a conserved motif[122]. After localizing to the kinetochore, PP1 dephosphorylates Aurora B substrates, stabilizing microtubule attachment. Therefor, CASC5 opposes Aurora B activity while still itself a substrate for phosphorylation by Aurora B [122,123]. PP1 binding of CASC5 stabilizes microtubule–kinetochore attachments; silencing the spindle checkpoint while the phosphorylation dependent recruitment of BUB1 and BUB3 to an N-terminal domain of CASC5 maintains the spindle checkpoint [124-126]. These interactions indicate CASC5 is a hub for regulating microtubule-kinetochore attachments and the spindle checkpoint.

TABLE 2.3	**Conservation of the outer kinetochore**		
Complex	Gene	Species	P Value
Mis12	MIS12	6	1
	DSN1	6	0.905
	PMF1	7	1
	NSL1	5	0.67
	MAD2L1	5	1
	MAD2L2	5	1
NDC80	NDC80	5	0.247
	NUF2	5	1
	SPC24	5	0.161
	SPC25	4	1
	ZWINT	7	0.498
	CASC5	5	<0.001

The final proteins to localize to the centromere comprise the fibrous corona, only arriving as chromosomes prepare to align at metaphase and

carrying out the mechanical action of segregating chromosomes (see Chapter I). The chromosomal passenger complex is critical for proper bi-polar spindle attachment. When sister chromatids are captured by microtubules from the same spindle pole the kinase activity of the chromosomal passenger complex subunit Aurora B is activated [127]. Aurora B phosphorylates the Ndc80 subunit of the Ndc80 complex reducing the complex's microtubule binding affinity, perhaps allowing another chance for bi-polar attachment. The RZZ complex targets the minus-end direct motor dynein and the complex dynactin to the kinetochore. Dynein and dynactin are necessary to provide the poleward directed force for attached microtubules. Like the proteins of the outer kinetochore, none of these outer kinetochore or corona complexes show evidence of positive selection (Table 2.4).

TABLE 2.4	Evolution of fibrous corona						
	Gene	# of Primates	P Value	Complex	Gene	# of Primates	P Value
	APITD1	6	1.0	RZZ	ROD1	7	1.0
	CCDC99	5	0.82		ZW10	7	1.0
	CDCA8	6	0.91		Zwilch	5	1.0
	CENPE	4	0.27				
	CENPF	4	0.91	Mitotic	BUB1	5	1.0
	CLIP1	6	0.74	Checpoint	BUB1B	7	0.27
	D4S234E	7	1.0		BUB3	6	1.0
	DCTN1	7	1.0		MAD1L1	4	0.61
	ERCC6L	3	1.0		MAD2L1	6	1.0
	ITGB3BP	4	0.61		MAD2L2	7	1.0
	KIF18A	6	0.82		TTK	5	1.0
	KIF2B	7	1.0		TAOK1	6	1.0
	KIF2C	7	0.50		CDC20	6	1.0
	MAPRE1	7	1.0				
	MLF1IP	3	0.67	Chromosome	AURKB	7	0.82
	NDE1	7	1.0	Passenger	INCENP	7	0.91
	NDEL1	7	1.0	Complex	BIRC5	6	0.15
	NUDC	7	1.0				
	PAFAH1B1	7	1.0	NUP80-160	NUP37	7	0.74
	PLK1	6	1.0		NUP43	6	1.0
	RANGAP1	7	0.55		NUP85	7	1.0
	RCC2	7	0.14		NUP98	6	0.55
	ROD1	7	1.0		NUP107	7	1.0
	SGOL2	3	0.50		NUP133	6	1.0
					NUP160	6	1.0
					SEC13	7	1.0
					SEH1L	7	1.0

Contrasting Evolution of the Inner Kinetochore and more Distant Proteins

My analyses reveal stark differences in patterns of selection in primates. Each kinetochore protein performs a conserved and essential function, and should experience strong purifying selection; however, my work shows a wide range of selective pressure acting on the kinetochore. This variation in selection clearly falls into two categories; the rapidly evolving inner kinetochore versus the

conserved outer kinetochore and fibrous corona. Previous evolutionary studies had focused exclusively on CenH3 and a handful of foundational kinetochore proteins. I have expanded this work to define how selection has shaped the entire kinetochore.

The outer kinetochore and fibrous corona show remarkably uniform conservation. These proteins form much of the physical structure of the kinetochore and are responsible for the regulation and mechanical motion of the chromosomes. Unsurprisingly, these crucial, conserved functions have preserved the presence of outer kinetochore and corona proteins across many eukaryotic taxa and have allowed for the detailed functional characterization. I show that the conservation has continued at the sequence level in primates, with single exception of CASC5.

The inner kinetochore was defined by the proteins' close association with CenH3 containing nucleosomes and constitutive centromeric localization [28,67,108]. Unlike the outer kinetochore, inner kinetochore proteins are not uniformly distributed across taxa. CenH3 and CENP-C are present in all eukaryotic genomes that have been characterized to date, while other critical components, such as CENP-U, have not been found outside of mammals. Many of the inner kinetochore proteins have just begun to be functionally characterized but the CCAN contains both centromeric chromatin binding proteins and proteins that, thus far, have exclusively structural or regulatory roles. Research has thus far been exclusively focused on rapid evolution small a subset of known centromeric DNA binding proteins, narrowing my understanding of how

centromere drive shapes the genome [2,5,37-39,84,128]. My analysis reveals that many of the proteins of the inner kinetochore are rapidly evolving, greatly expanding the extent of centromere drive's effect on the kinetochore.

Genetic conflict with driving centromeres is hypothesized to primarily drive the evolution of the centromere binding proteins. The CCAN is upstream of the localization of almost all other kinetochore complex, including most of the proteins believed to recognize the centromere. Two complexes of CCAN proteins, CENP-S/T/W/X and CENP-L/M/N, bind centromeric DNA or chromatin and are essential for the localization of the rest of the kinetochore [101,106,107]. Proteins from both these complexes are rapidly evolving, in particular CENP-T with both a known DNA binding domain and the ability to bypass the kinetochore's requirement for CenH3, as well as CENP-N, which is required for efficient localization of CenH3. These rapidly evolving proteins are excellent candidates for characterization of the functional consequences of centromere drive. While ectopic localization of CenH3 causes ectopic localization of most kinetochore proteins, CENP-T is not dependent on CenH3 for localization making characterization of changes in DNA affinity or localization experimentally feasible.

The CCAN also performs essential structural and regulatory roles in the kinetochore. Two CCAN complexes perform have essential structural and regulatory roles, CENP-H/I/K and CENP-O/P/Q/R/U, without any DNA recognition role. There is no previously hypothesized conflict between structural kinetochore proteins and centromere. However, unlike the structural and

regulatory proteins of the outer kinetochore, CENP-O, CENP-P, and CENP-U are rapidly evolving. CENP-H may also be rapidly evolving but this conclusion awaits more sequencing efforts. The CENP-O/P/Q/R/Q complex recruits HEC1 and Mammalian polo-like kinase 1 (PLK1) to the centromere, directing and stabilizing microtubule/kinetochore attachment[108,115-117]. Depletion of the complex results mitotic defects and premature sister chromatid separation[120], suggesting that the CENP-O/P/Q/R/Q plays an important role in regulating microtubule dynamics. The rapid evolution of multiple members of this complex points to either a previously unrecognized centromere-binding role for this complex or the interesting possibility that regulators of microtubule binding or dynamics are shaped by centromere drive. A role for regulators of microtubule binding in suppressing centromere drive is especially interesting given the hypothesized role of meiotic positioning in centromere drive. Perhaps the CENP-O/P/Q/R/U complex is able to suppress centromere drive through regulation of rates of microtubule binding or meiotic positioning rather than modulating the binding of centromeric DNA, suggesting a potentially novel avenue to suppress centromere drive.

Mis18α, Mis18β, KNL2 are unique inner kinetochore proteins, which localize to the kinetochore only transiently, during the G2 phase of the cell cycle and CenH3 loading. Mammals and birds have two Mis18 homologs, MIS18α and MIS18β, while amphibians, fish and fungi have only one [70]. Homologous proteins have not yet been found in *S. cerevisiae*, fly, or nematodes, perhaps due to low sequence conservation. The role that Mis18 plays in centromeric

recognition and kinetochore formation is critical but not yet fully described. All three proteins are essential for CenH3 localization, however, the Mis18 complex requires CENP-C presence, which itself is dependent on CenH3 for localization [70,129-131]. The complex's role ensuring proper CenH3 localization places it directly in the midst of the centromere drive conflict and the rapid evolution of KNL2 provides evidence of its role as a possible suppressor of drive.

Positive selection is almost exclusively focused on proteins of the inner kinetochore and proteins that are required to deposit inner kinetochore proteins. The proteins of the inner kinetochore are significantly more likely to rapidly evolving, 9 of 21 proteins of the inner kinetochore contrasted with only 1 of 59 outer kinetochore and fibrous corona. This discordant pattern is highly significant ($p < 0.001$ by Fisher's exact test) confirming that the inner and outer kinetochore are shaped by very different forces; the inner kinetochore recurrently adapts to suppress centromere drive while the outer kinetochore and fibrous corona face strong selection to maintain their function during cell division.

Materials and methods

Phylogenetic and Evolutionary Analysis of the Kinetochore

I created a list of proteins that comprise and regulate the kinetochore (reviewed in [27]). Nucleotide sequence for each gene from *Homo sapiens, Pan troglodytes, Pongo pygmaeus, Macaca mulatta,* and *Callithrix jacchus,* were obtained from the UCSC genome browser [132,133]. The coding sequence sequences for each gene *from H. sapiens, M. mulatta* (Rhesus macaque), and *C.*

jacchus (Marmoset) were used as a query to interrogate the nr, wgs, and hgts databases to obtain a more diverse group of primate sequences [134]. Translational alignments of primate sequences were created using Geneious with manual curation of gaps [135]. Trees conforming to the accepted primate phylogeny were created using neighbor joining (clustalX). Nucleotide sequences were aligned using Geneious 5.3.6 with manual refinement of gap positions [135].

Sequencing of CenH3, CENP-B, and CENP-C from Primate RNA

CenH3 and CENP-C were amplified and sequenced from total RNA with Superscipt III One-step RT-PCR kit (Invitrogen) by using the PCR and sequencing primers shown in the Primer Table (Appendix I). RT-PCR products were sequenced directly, except in a few cases where they were first cloned into the TOPO TA cloning vector (Invitrogen), followed by sequencing of as least three independent clones. Sequence reads were assembled into full length transcripts. The CenH3, CENP-B, and CENP-C cDNA sequences for human, *M. mulatta* (Rhesus macaque), *P. abelli* (Orangutan), *C. jacchus* (Marmoset) were obtained from the UCSC genome browser. The CenH3 cDNA sequences for *Papio anubis* (Baboon), *Colobus guereza* (Colobus Monkey), *Aotus nancymaae* (Ma's Night Monkey), *Callicebus moloch* (Dusky Titi), *Lemur catta* (Ring Tailed Lemur) were obtained from Willie Swanson[5]. Phylogenies were constructed using neighbor joining (clustalX) and Bayesian methods (Mr Bayes). Both trees agreed with the accepted primate phylogeny[136].

Sequence Analysis

For proteins with a dN/dS value greater than one in two independent comparison a maximum likelihood analysis was performed with codeml in the paml 4 software package. To detect selection multiple alignments were fitted to either the F3×4 or 1/61 models of codon frequencies. Likelihood ratio tests of the data were performed by using different sets of site-specific (NS sites) models as follows: M1 (two-state, neutral, dN/dS > 1 not allowed) to M2 (three-state, selection, dN/dS > 1 allowed); and M7 (beta distribution, dN/dS > 1 not allowed) to M8 (beta distribution but dN/dS > 1 allowed). Limiting the analysis to five primates resulted in low power to detect positive selection that was sensitive to codon substitution model. To increase power to detect selection, for complexes that had one member reach statistical significance in the 5 primate analysis, the codeml simulations were rerun using an alignment of all primate sequence available. For those proteins which had significant statistical support for positive selection I list codons with a posterior probably of selection greater than 95%.

III. The Scm3/HJURP chaperone constrains the evolution and functional divergence of centromeric histones

Centromeres in virtually all eukaryotes are characterized by the presence of specialized chromatin, comprised of both canonical and centromeric nucleosomes [33,137]. The latter have several features that distinguish them from canonical nucleosomes including the replacement of canonical histone H3 with the centromeric H3 (CenH3) variant [30]. CenH3 not only defines centromeric chromatin but also serves as a scaffold to recruit other kinetochore proteins and is essential for chromosome segregation [33]. Since CenH3 targeting to ectopic chromosomal locations is sufficient to recruit additional kinetochore proteins and propagate functional centromere identity through cell division [138], CenH3 has been proposed to be responsible for maintaining centromere identity.

Canonical H3 proteins are highly conserved over eukaryotic evolution. In contrast, CenH3s evolve rapidly in both their histone fold domain and N-terminal tails, [5,30,37,39,41,84] so much so that their N-terminal tails cannot even be considered homologous across different taxa. In both yeast and human CenH3s, the Loop1 and alpha-helix2 regions of the histone fold domain have been shown to comprise a unique CENP-A Targeting Domain (CATD), which is sufficient to confer centromeric targeting even to a canonical histone H3 fold domain [1].

More recently, this domain has also been shown to mediate important protein-protein interactions of CenH3 with other proteins, including the Scm3/HJURP chaperone, the CENP-N protein and the ubiquitin ligases that ensure exclusively centromeric localization of CenH3s [1,52,54,106,139-141].

Previously hypotheses proposed that the rapid evolution of CenH3s is driven by the selective pressure to suppress 'cheating' centromeres that exploit asymmetric female meiosis for their own propagation, even at a cost to host fitness [7,11]. This evolutionary pressure should be common to all taxa that harbor asymmetric female meiosis (in which only one of four meiotic products is retained) but not those that only carry out symmetric meiosis. Consistent with this hypothesis, evolutionary studies have revealed strong evidence of adaptive evolution of CenH3 genes in plants, insects, and mammals, which undergo asymmetric meiosis, but not fungi, which only carry out symmetric meiosis [5,37,39,84,142].

Despite this common pressure to suppress 'cheating' during asymmetric meiosis, the functional consequences and patterns of CenH3 evolution vary dramatically. In plants and insects, both the N-terminal tail and the Loop1 region of the CATD are hotspots of rapid evolution [37,41,84]. Replacement of just the rapidly evolving Loop1 portion of the CATD in *D. melanogaster* CenH3 with the orthologous domain from *D. bipectinata* abrogated centromeric localization [38]. Similarly, CenH3 alleles from related plant species are not sufficient to confer full CenH3 function in *Arabidopsis* species [41]. In contrast to this lack of cross-

complementation over even short evolutionary distances in plants and flies, *Saccharomyces cerevisiae* CenH3 is able to complement the near-complete depletion of human CenH3 [143]. These observations are difficult to reconcile, particularly since primate CenH3s do evolve rapidly whereas fungal CenH3s do not [2,5,142]. I hypothesize that protein-protein interactions, which are shared between primate and fungal CenH3s but not with CenH3s from other taxa, may help explain this apparent paradox.

In order to explain the observed patterns of cross-complementation, I began by extending the previous analysis of primate CenH3 evolution [5]. In plants and insects, both the N-terminal tail and the Loop1 region of the CATD are hotspots of rapid evolution [39,41,84]. Replacement of just the rapidly evolving Loop1 portion of the CATD in *D. melanogaster* CenH3 with the orthologous domain from *D. bipectinata* abrogated centromeric localization [38]. Similarly, CenH3's HFD domain, containing the rapidly evolving Loop1, is sufficient to confer centromere localization in *Arabidopsis* species [41]. I sequenced additional primate *CENP-A* genes, expanding the alignment of the *CENP-A* gene to 19 primate species and increasing the power to detect positive selection. I then performed maximum-likelihood analyses to detect positive selection. Consistent with the previous study [5], I found strong evidence of positive selection on the whole protein ($p<0.001$) and in particular the N-terminal tail ($p<0.001$) (Figure 3.1A). Moreover, I could show that there is unequivocal

evidence of positive selection acting on the histone-fold domain ($p<0.01$). Surprisingly, however, I found no evidence of positive selection acting on the CATD alone ($p\sim0.5$) (Figure 3.1A). In particular, the Loop1 and alpha-helix2 have only three nonsynonymous changes over the ~60 million year evolution of primates. This is in stark contrast to the 30 synonymous changes in the CATD and the numerous insertions/deletions and non-synonymous changes in the corresponding N-terminal regions (Figure 3.1B). Since the CATD is sufficient to confer centromeric localization onto even canonical H3 [1], this suggests that all primate CENP-A's may retain the ability to localize to each other's centromeres. These findings in primate CENP-A are in stark contrast to previous studies in both *Drosophila* and *Arabidopsis* species, where the CenH3s had a hotspot of positive selection in their Loop1 region [39,84], and CenH3s were unable to cross-localize [38,41].

I considered the possibility that the absence of positive selection in the CATD of primate CenH3s may reflect an evolutionary constraint that is absent in both plant and *Drosophila* CenH3s. Centromeric histone chaperones present one such candidate. In *Saccharomyces cerevisiae,* CenH3 requires a specialized chaperone, Scm3*,* to localize to centromeres [52,53]. Structural studies have found that Scm3 binds the CATD of CenH3; mutations in the CATD result in loss of Scm3 binding and defects in centromeric localization [51,54]. Recent studies have discovered that primates encode an ortholog of Scm3 called HJURP [50]. Like Scm3*,* HJURP directly binds the CenH3 CATD and is required for centromeric localization of CenH3[40,51,55,144]. Further, retargeting HJURP is

sufficient to recruit CenH3 to non-centromeric DNA, leading to ectopic kinetochore formation [56]. Thus, in both yeast and human cells, the CenH3-chaperone interaction is critical for chromosome segregation, hence this interaction surface (the CATD in CenH3) would be expected to be constrained in its evolution, which is what I have found in primate genomes. In contrast to fungi and other eutherian mammals, *Drosophila* and *Arabidopsis* genomes reveal no orthologs of Scm3/HJURP [50], thus there is no *a priori* expectation of the same degree of selective constraint acting on CATD from either *Drosophila* or *Arabidopsis* CenH3s. I therefore hypothesized that it is the presence or absence of Scm3/HJURP orthologs that results in differences in both the evolutionary pressures and functional divergence of CenH3.

To directly test this hypothesis, I investigated the functional properties of CenH3s from a wide evolutionary range of eukaryotic genomes that possessed or lacked a Scm3/HJURP chaperone. *SCM3* orthologs are present widely (but not universally) in both fungi and animals, but not in any other lineages. I therefore propose that the Scm3/HJURP chaperone arose in the common ancestor of fungi and animals (*Opisthokonta*) [145,146] likely after the split from *Amoebozoa* (e.g. *Dictyostelium discoideum* lacks an apparent ortholog) (Figure 3.2B). I selected two CenH3s from taxa preceding the birth of Scm3/HJURP- the plant *Arabidopsis thaliana* [37], and the ciliated protozoan *Tetrahymena thermophila* [128]. I also selected CenH3s from two divergent fungi, an ascomycete *S. cerevisiae* and a basidiomycete *Cryptococcus neoformans* (I did not record any loss events of *SCM3* from fungal genomes studied). Scm3 is also

present in a basally branching relative of metazoans, the choanoflagellate *Monosiga brevicollis* but has been apparently lost in several animal lineages, represented by the nematode *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster.* Among vertebrates, *HJURP* appears to be ubiquitously present in eutherian mammals (represented by *H. sapiens*), and is also found in birds and amphibians (represented by *Xenopus tropicalis*). However, all sequenced fish genomes (represented by zebrafish *Danio rerio*) appear to lack an *HJURP* ortholog.

A previous study had shown that substituting the CATD of *S. cerevisiae* CenH3 (CSE4), into a H3 histone fold domain (HFD) together with the CSE4 N-terminal tail and four amino acid changes in the C-terminus of the HFD could complement a *cse4*Δ mutation[1] (schematized in Figure 3.2A). Using the same approach and CATD boundary definitions, I decided to test CATD domains from diverse CenH3s for their ability to genetically perform the same function in budding yeast (Figure 3.2A). I created diploid yeast strains heterozygous for the wildtype *CSE4* gene and a modified chimeric *cse4/H3* gene in the allelic locus. Following meiosis, the two genes are expected to segregate in a 2:2 Mendelian fashion into tetrads. If the chimeric gene is able to complement *cse4*Δ*,* I expect to see four viable products of meiosis. If however, the chimeric gene is unable to complement, I expect to only see two viable spores per tetrad. When I tested the *S. cerevisiae* CATD in this assay, as expected, I observed *cse4*Δ complementation (Figure 3.2B). This complementation was also observed in the case of CenH3 CATDs from *C. neoformans, M. brevicollis, H. sapiens* and *X.*

laevis (I were able to obtain *M. brevicollis* CenH3 sequence from its genome sequence [147]). However, I observed no complementation when CenH3 CATDs from *A. thaliana, T. thermophila, D. melanogaster, C. elegans and D. rerio* were tested. Overlaying the presence or absence of the Scm3/HJURP chaperone onto these results revealed a perfect congruence between chaperone presence and ability to complement *cse4*Δ (Figure 3.2). Thus, chaperone presence is a better predictor of CATD complementation than evolutionary relatedness of the CenH3s.

To functionally analyze these evolutionarily diverse CATD domains in another system where the Scm3/HJURP chaperone had been retained, I turned to human tissue culture cells. In contrast to the genetically-defined 'point' centromeres of *S. cerevisiae*, human centromeres are large and epigenetically defined, consisting of repeated arrays of alpha-satellites. A previous analysis had shown that a human Histone H3 with a (human CenH3) CENP-A CATD domain (hereafter referred to as $H3^{CATD}$) accurately localizes to human centromeres [48]. I used a similar construct to analyze the centromere-localizing properties of CATD domains from a panel of CenH3s from genomes with or without their chaperones. Chimeric $H3^{CATD}$ were tagged with an N-terminal 3XFlag tag and transiently transfected into HEK293T cell lines (Figure 3.3). I found that all $H3^{CATD}$ chimeras from species with *SCM3* orthologs (*M. brevicollis, S. cerevisiae, X. laevis, H. sapiens*) were able to localize to centromeres (marked by an anti-CENP-A antibody that does not recognize the 3XFlag tagged proteins). In contrast, $H3^{CATD}$ chimeras lacking Scm3/HJURP chaperones either had a diffuse localization pattern (*A. thaliana, D. melanogaster*) or localized adjacent but not coincident to centromeres (*D. rerio* – Figure 3.4). The *D. rerio* CATD localization is especially reminiscent of results from a previous study of heterologous CenH3s in *D. melanogaster* cells, where the heterologs localized to heterochromatin but not to centromeres [148]. Thus, consistent with the *cse4*Δ

complementation analysis (Figure 3.2), successful localization to human centromeric regions (Figure 3.4) is completely predicted by the presence or absence of a cognate centromeric Scm3/HJURP chaperone rather than by evolutionary relatedness.

The Loss of HJURP Releases the Fish CATD to Evolve Rapidly

By virtue of its preservation in the *X. laevis* and mammalian genomes, I could infer that the common ancestor of the fish, amphibian and mammalian lineages must have possessed the HJURP chaperone. Closer examination of multiple completely sequenced fish genomes (Figure 3.5) did not reveal an *HJURP* ortholog. Thus, it appears that the common ancestor of fishes lost the HJURP chaperone. This presented an opportunity to examine the evolutionary constraints acting on a CenH3 lineage that 'recently' lost its cognate HJURP chaperone. Using publicly available genome and transcriptome sequences, I investigated whether, after losing *HJURP*, CenH3 CATD had started to rapidly evolve in fishes (Figure 3.6). I found significant evidence of positive selection in fish CenH3s when the whole protein was analyzed ($p<0.05$) just like in primates (Figure 3.1) and eutherian mammals in general[128]. However, in contrast to mammals, I found significant evidence of positive selection acting on the CATD of fish CenH3s ($p<0.01$), in particular residues in the Loop1 of the CATD. Thus, it appears that, released from the constraint of maintaining chaperone binding, the Loop1 region of fishes has evolved similarly to *Drosophila* and *Arabidopsis* CenH3s [39,84]. Based on this pattern, I predict that fish CenH3s will show the same lack of cross-localization seen in Drosophila and Arabidopsis lineages. I

also did not find strong evidence of positive selection acting on fish CenH3s in their N-terminal tail. However, this is likely attributable to the lack of power to detect positive selection due to the large number of indels (Figure 3.5). Together, my analyses reveal that the presence of the chaperone is predictive not only of CenH3 cross-complementation abilities but also of patterns of evolutionary constraint acting on their CATD domains.

My findings that fish CenH3 Loop1 regions evolve rapidly just like in Drosophila and plant lineages, still do not reveal why a CenH3 could not simultaneously have both positively selected Loop1 as well as a Scm3/HJURP interacting alpha-helix2. Structural and evolutionary studies have proposed that alpha-helix2 is the critical region for both HJURP and Scm3 interaction whereas the Loop1 region was not required for centromere localization or chaperone binding [54,140,149,150]. Yet primate CenH3 Loop1 domains show a surprising lack of diversity, consistent with selective constraint, in primate Loop1.

This suggested that primate Loop1 evolution is still constrained by HJURP binding, even if it does not involve a direct contact. To test this possibility, I utilized functional differences between the zebrafish and *Xenopus* CATD to investigate the role of Loop1 and the alpha-helix2 separately; while *X. laevis* CATD is capable of providing *cse4*Δ complementation, *D. rerio* CATD is not (Figure 3.2B). I created chimeric CATD domains of CenH3s from *D. rerio* and *X. laevis.* However, neither of the two chimeras (*D. rerio* Loop1 *X. laevis* alpha-

helix2 or *X. laevis* Loop1 *D. rerio* alpha-helix2) could rescue a *cse4*Δ mutation (Figure 3.2). The inability of a 'functional' *X. laevis* alpha-helix2 to complement *cse4Δ* when paired with *D. rerio* Loop1, indicates that even if Loop1 does not provide direct contacts for Scm3/HJURP binding, it nonetheless is constrained by the presence of the chaperone. Loss of *Scm3* permits the acquisition of Loop1 mutations that are now no longer compatible for binding to the chaperone, perhaps due to an epistatic effect that compromises alpha-helix2 binding to Scm3.

Despite conserved and essential function, centromeres and centromere binding proteins evolve rapidly across many animal and plant taxa. This recurrent rapid evolution is a hallmark of ongoing genetic conflict. I previously hypothesized that asymmetric 'female' meiosis provides a unique landscape for an intragenomic conflict [11,20]. Unlike in male meiosis, where all four meiotic products become gametes, only one of the four products of female meiosis is retained in the oocyte, while the other three products are discarded as evolutionary dead ends. Therefore, female meiosis provides opportunity for loci on homologous chromosomes to compete for evolutionary success. As the genetic loci responsible for ensuring chromosome segregation, centromeres are ideally positioned to cheat during meiosis and selfishly increase their own transmission into the oocycte. Cheating by centromeres can be deleterious to the rest of the genome and select for new alleles of CenH3 that suppress this centromere drive [7,11,20]. Repeated cycles of centromere-drive and its suppression can explain the rapid evolution of both centromeric DNA and CenH3.

In this context, it is important to emphasize that there is significant evidence for positive selection acting on primate CenH3 N-terminal tails and histone fold domains ([5], above). Therefore, I conclude that simply maintaining HJURP interaction and centromeric localization during mitosis may not be sufficient for the full repertoire of centromeric functions. For instance, having a correctly adapted CenH3 may be especially crucial during meiosis, which may have divergent requirements than mitosis [151]; notably, meiotic function has not been assessed in previous or my heterologous 'swap' experiments.

Despite its explanatory power, the centromere-drive model still left unanswered the question of why CenH3 evolution resulted in rapid functional divergence in *Drosophila* and *Arabidopsis* but not in primates despite the same evolutionary pressures. My results suggest that an ancient but idiosyncratic co-evolutionary relationship between the CATD domain of CenH3 and the homologous CenH3-interacting domain of Scm3/HJURP provides an explanation for this diverse pattern. CATD domains of CenH3s from species with an Scm3/HJURP ortholog retain the ability to localize and functionally complement each other over long evolutionary distances, whereas those from species lacking Scm3/HJURP orthologs are unable to complement. The latter category includes taxa, such as plants, that diverged prior to the birth of Scm3/HJURP or those who subsequently lost Scm3/HJURP such as insects and fish. Thus, while genetic conflicts between centromeric DNA and CenH3s helps explain the rapid evolution of CenH3s, co-evolution with a Scm3/HJURP chaperone helps explain the observed pattern of complementation and CATD evolution.

The CATD domain of CenH3s has been previously defined both as the centromeric localization determinant as well as the Scm3/HJURP binding domain [1,40,54,140]. Further dissection and crystallographic studies have implicated the alpha-helix2 as being sufficient for Scm3/HJURP to discriminate between CenH3 and canonical H3 [150]. It appears that HJURP is tolerant of significant variation in Loop1. For instance, chimeric CATD domains with a shorter Loop1 from canonical H3 and CenH3 alpha-helix2 can still be bound by HJURP and localized to the centromere [150]. Furthermore, there is significant variation in Loop1 in many CenH3s that still retain the functional property of binding Scm3/HJURP. Yet, I find that Loop1 can negatively influence the CenH3-chaperone interaction. For instance, the *D. rerio* Loop1 paired with a 'functional' alpha-helix2 is still incapable of binding Scm3 to complement *cse4*Δ, suggesting that constraint to avoid such negative epistasis still shapes Loop1 evolution, even though Loop1 may not represent a direct contact with Scm3/HJURP.

Scm3/HJURP is not the only partner to potentially constrain the evolution of CenH3 CATD. I considered the alternate possibility whether any other interactors might explain the divergent patterns of CenH3 function and evolution. First, I considered another kinetochore protein essential for CenH3 localization, CENP-N, which also recognizes CATD [106]. Three observations suggest that CENP-N orthologs do not explain the evolutionary and functional divergence of CenH3s. First, CENP-N orthologs can be readily identified in mammals, birds, amphibians and fish [152], despite fish CenH3s not sharing the same evolutionary history or functional complementarity of other vertebrates. Second,

although I can trace CENP-N orthologs in fungi, the homologous domain shared by fungi and metazoan CENP-N proteins is not sufficient for CenH3 interaction [106,152] unlike the Scm3/HJURP homologous domain. Finally, I were unable to find choanoflagellate CENP-N orthologs, even though *M. brevicollis* CenH3 shares the functional and evolutionary constraint of fungal and mammal CenH3s. Thus, both because of the lack of a conserved CenH3 interacting domain as well as an incongruent phylogenetic distribution, I conclude that CENP-N is unlikely to explain the idiosyncratic pattern of CATD function and evolution.

S. cerevisiae CenH3 CATD is also recognized by Psh1, an E3 ubiquitin ligase required for removal of CenH3 from ectopic sites [139,141]. Proteasome dependent degradation of ectopically deposited CenH3 is thought to be generally important to ensure proper centromere localization. However, there are no obvious *Psh1* orthologs in metazoans. As yet, the only known metazoan E3-ubiquitin ligase that functions in a manner analogous to Psh1 is *D. melanogaster* Partner of Paired (Ppa) [153] even though it is not a direct ortholog of *Psh1*. Ppa also interacts with the CenH3 CATD and yet does not constrain the evolution and functional divergence of *Drosophila* CenH3s. This could reflect fundamental differences between Ppa and Psh1 binding to the CATD or reflect the fact that E3-ubiquitin ligases may not shape the disparate patterns of CenH3's functional divergence. This conclusion can be revisited as more E3 ubiquitin ligases with analogous functions are identified, particularly in mammalian genomes.

Canonical H3, CenH3 and other histone variants require chaperones to be deposited onto DNA [154]. How is CenH3 still deposited at centromeric DNA in taxa that have undergone independent losses of Scm3/HJURP? One possibility is that CenH3 may have evolved to be deposited through a completely novel pathway. However, another, more likely, possibility is that the evolution of a new chaperone may have facilitated the loss of the ancestral *SCM3/HJURP*. This possibility is supported by the function and evolution of the CAL1 protein in *D. melanogaster* [57,58,155,156]. CAL1 appears to serve an analogous role to Scm3/HJURP in terms of CenH3 deposition [57,58]. This dipteran specific chaperone possesses a domain similar to the Scm3 homology domain within CAL1's CenH3 interacting region, leading to the hypothesis that CAL1 is the product of convergent evolution [156]. Although it is still not known what domain of CenH3 interacts with CAL1, it is interesting to note that CAL1-binding has not constrained the CATD domain's evolution in the same way as Scm3/HJURP. Either the CAL1 interaction domain in *Drosophila* CenH3s does not involve the CATD, or the interaction interface of CAL1 is more limited or labile, as evidenced by *Drosophila* CenH3 CATD's rapidly evolution and functional divergence [38,84]. Since CAL1 likely convergently evolved a Scm3 homology domain, it may not present exactly the same structural and evolutionary constraints as the divergently inherited Scm3/HJURP orthologs from a common ancestor.

The extent of coevolution between the CATD and Scm3 over hundreds of millions of years of evolution is remarkable. This is especially striking when one considers that there is no recognizable homology outside the CATD-interacting

domains in *SCM3* from different taxa. Likewise, there is ample evidence of CenH3 rapid evolution throughout animal and plant lineages. However, the CATD interaction interface has been functionally and evolutionarily maintained throughout this diversification, except in those lineages that lost or never possessed the chaperone, obviating their CATDs of such constraint. In such an instance of a deep protein-protein interaction coevolution, it is nonetheless possible that these two protein domains could co-evolve independently down each lineage, such that they maintain binding to each other but only within their own lineage or species. If this were the case, I would not necessarily expect to My findings that highly divergent CATD domains still maintain functional interactions with both yeast and human chaperones suggest that the interaction interface between CenH3 and *Scm3*/HJURP has remained functionally unchanged despite hundreds of millions of years of evolution [157]. I infer that any mutations resulting in even subtle reduction or loss of interaction affinity were presumably so deleterious that they were not tolerated in the entire co-evolutionary history of these two protein domains. This is the likely reason why functional complementation has been successful in lineages as divergent as yeast and human [143]. Thus, despite the backdrop of genetic conflict and rapid evolution, the CATD-Scm3 interaction also presents a powerful paradigm of coevolution of an ancient protein-protein interaction.

Materials and Methods

Phylogenetic and Evolutionary Analysis of CenH3

Primate CenH3 sequence for *H. sapiens, P. troglodytes, G. gorilla, P. pygmaeus, N. leucogenys, M. mulatta, P. anubis, C. aethiops, M. talapoin, T. francoisi, C. guereza, A. geoffroyi, A. nancymaae, C. jacchus, S. boliviensis, M. murinus,* and *O. garnetti* were obtained from either my sequencing, genome sequencing projects or a previous study [5]. I additionally obtained CenH3 sequence from *M. talapoin, T. francoisi, C. pygmaea, L. lagotricha* via RT- PCR using RNA isolated from individual cell lines (obtained from Coriell repositories). RT-PCR was carried out using the Invitrogen Superscript III One-Step RT-PCR with Platinum Taq kit (Invitrogen) and Primers KR83 and KR87 (primer sequences are in Appendix Table 1). Products were gel purified and directly sequenced using ABI BigDye 3.0. Sequences obtained by this study have been deposited into Genbank (accession numbers are forthcoming).

CenH3 sequences were aligned using ClustalX with manual refinement of gap positions [158]. A phylogenetic tree was created, and bootstrapped using the Neighbor-joining method [159]. To identify rapid evolution, nested PAML models M7 and M8 were compared. Model M7 allows codons to have dN/ dS values according to a beta distribution (two parameters). I tested for significance of diversifying selection by comparing twice the difference in log likelihoods between the M7 and M8 Nsites models with two degrees of freedom [103].

To detect remote homology, I also carried out protein database searches using PSI-BLAST and HHPred [160,161].

Plasmid Construction

The *cse4-hph* plasmids were constructed by amplifying 500 bp downstream of *CSE4* with primers KR327 and KR328 with engineered *Sac*I and *Spe*I sites and cloned into the same sites of the pAG32 vector [162]. *H3*^{CATD} fragments were PCR amplified from pSB1538 in two reactions: first using KR250 and a species specific primer designed to modify the Loop1 of the CATD, the second using KR251 and a species specific primer designed to modify alpha-helix2 (all primers are reported in the Appendix). These two fragments were ligated using overlap PCR with primers KR250 and KR251. An additional 200 bp downstream of *CSE*4 was amplified using primers KR327 and KR329 then ligated to the *H3*^{CATD} fragment with overlap PCR using primers KR250 and KR329 engineered with *Hin*dIII and *Bam*HI sites and cloned into the same sites of the *cse4-hph* plasmid.

To create the constructs used for cytology, *HISTH3* was amplified using KR285 and KR286. To modify the CATD, HISTH3 was amplified in two fragments: first using KR285 and a species specific primer designed to modify Loop1 of the CATD, the second using KR286 and a species specific primer design to modify alpha-helix (Appendix Table 1). These two fragments were ligated using overlap PCR with primers KR285 and KR286 engineered with EcoRI and Knp1 and cloned into p3xFLAG-Myc-CMV™-24 (Sigma-Aldrich)

Tetrad Analysis

All yeast strains are detailed in Appendix Table 2. Strains SZY55 and SZY56 were mated and sporulated, selecting for NATMX4 and KANMX4 resistant haploids, KRY100 (MATa) and KRY101(MATα). Strain KRY102 was generated from a cross between KRY100 and KRY101. *H3*^{CATD} plasmids were digested with *Hin*dIII and *Spe*I and transformed into KRY102. After selection for HPHMX4 resistance, replacement of *CSE4* with *H3*^{CATD} was verified by PCR and sequencing. These strains were sporulated on plates for 36 hours, then dissected on YPD plates. The genotypes of the spore colonies were verified via replica-plating.

Cytology

Human 293T cells were cultured in Dulbecco's Modified Eagle Medium (Gibco) with 10% Fetal Bovine Serum, 50u/ml Penicillin, and 50µg/ml Streptomycin, at 37˚C and 5% CO_2. Cells were grown to confluence, trypsinized and counted. 2x10^5 cells in 2.5mL of DMEM complete media were added to each well of a 6 well plate containing a glass cover slip. 24 hours later 2.5ng of 3xFlag-*H3*^{CATD} plasmid was transfected using TransIT®-LT1 transfection reagent (Muris) according to manufacturer instructions. Cells were allowed 24 hours for expression before being fixed with 4% paraformaldehyde for 15 minutes and quenched with PBS plus 0.2% Tween (PSBT) at room temperature. 293T cells were blocked with PBG, 0.2% (w/v) cold water fish gelatin (Sigma G-7765). Centromeres were visualized with anti-CENP-A (1:1000 dilution) and *H3*^{CATD} was detected with anti-Flag (1:1000 dilution) primary antibodies. AlexaFluor 488 anti-

rabbit (Invitrogen, diluted 1:100) and anti-mouse AlexaFluor 568 (Invitrogen, diluted 1:1,000) were used for secondary detection. DNA was stained with DAPI and coverslips were mounted in Vectashield (Vector Labs, Burlingame, CA). Slides were observed under a Zeiss AxioPlan inverted scope, deconvolved, and processed with AxioVision software (Zeiss, Thornwood, NY).

IV. Rapid Evolution of HJURP, a Centromeric Histone Chaperone

HJURP was only recently characterized as the mammalian ortholog of CenH3 chaperone Smc3[50]. HJURP is not only a histone chaperone, essential for proper localization, but a molecular chaperone, binding the CenH3 histone fold domain and assisting in proper folding[150,163]. My work has demonstrated that the presence of an Scm3/HJURP chaperone constrains the evolution of CenH3, potentially limiting its ability to suppress centromere drive. HJURP recognizes specific residues in the CATD of CenH3, with the interaction between HJURP and CenH3 limiting the evolution of the CATD, reducing CenH3's ability to suppress centromere drive[51,55,144,149,150]. The presence of such a chaperone and constraint of the CATD could disadvantage the genome when confronted with driving centromere drive. However, HJURP could provide a new target for selection limit the deleterious effects of drive. A new suppressor of centromere drive will likely recognize centromeric DNA or chromatin, be a critical component of centromere function, and rapidly evolving.

The full localization determinants of HJURP are not yet known, as there is as yet no definitive pathway for *in vitro* centromere specification and kinetochore formation. Since DNA sequence itself was revealed to be insufficient to specify centromere function, it has been hypothesized that centromeres are epigenetically defined. There is overwhelming evidence that CenH3 are an essential and defining epigenetic mark of active centromeres but CenH3 is dependent on the functions of a number of other centromere binding proteins,

including a direct binding by HJURP. A fungal ortholog of HJURP, Scm3 has a nonspecific DNA binding domain that shows preference for AT-rich DNA and *de novo* kinetochore formation at sights of ectopic HJURP localization suggests that centromeric histone chaperones interact with centromeric DNA and act upstream of all other the kinetochore proteins, including CenH3[56,163]. However, HJURP requires the Mis18 complex of proteins to localize to endogenous centromeres, raising questions about how critical DNA recognition is for HJURP localization[56].

HJURP's essential role in centromere specification is supported, as mislocalizing HJURP to chromosome arms resulted in the relocalization of the kinetochore[56]. HJURP-LacI constructs were able to localize CenH3 to LacO arrays[56]. The ectopic localization of HJURP-LacI and CenH3 was sufficient to form *de novo* kinetochores at the LacO[56]. Critically, both the inner and outer kinetochore is recruited to these ectopic LacO sites. Overexpression of CenH3 leads to incorporation of CenH3 throughout the chromosome, likely through HJURP-independent mechanisms, however while overexpression leads to the relocalization of CENP-C, but these regions do not recruit additional CenH3 binding proteins or support kinetochore formation[164]. It is not known why recruitment of CenH3 to the LacO array by HJURP-LacI is successful at reconstituting centromere and kinetochore activity, while HJURP-independent ectopic kinetochore formation is unsuccessful. There may be some additional essential kinetochore requirement fulfilled by HJURP, perhaps through its molecular chaperone role. It may also be result of an increased density of CenH3

at the LacO arrays relative to untargeted deposition into general chromatin by overexpression. Finally, the kinetochore activity of the LacO array may be due to an as yet unrecognized HJURP role in the assembly of CenH3 nucleosomes. Regardless, HJURP's central role in centromere specification led me to investigate its evolutionary history and role as a suppressor of centromere drive.

HJURP has undergone repeated bouts of rapid evolution

While there is no evidence of HJURP directly recognizing or binding centromeric DNA, both the upstream Mis18 complex, and downstream CenH3 genes of kinetochore formation pathway are rapidly evolving. In addition, evolutionary studies of CenH3 were able to identify previously unknown centromere localization determinants[38,84]. I sought to characterize whether HJURP was likewise shaped by centromere drive and gain potential insights into the chaperone's centromere determinants. I amplified HJURP from a combination of cDNA and genomic DNA to obtain full-length sequence from 15 primates.

I used codeml to perform a maximum likelihood analysis on the Primate HJURP. The 16 primate sequences allowed me to determine that models allowing for positive selection were significantly more likely than neutral models ($p<0.0001$). Despite the strong support for positive selection, only five codons reached a posterior probability greater than 95% (Figure 4.1A). However, there were 13 codons with high posterior probability greater than 80% of selection and 61 amino had a posterior probably of selection greater than 50%. Very little is known about the function of HJURP domains outside the overlapping Scm3

homology domain and CenH3 binding domain, and selected sites are spread throughout the proteins[50,55]. The chaperone is functionally conserved; suggesting that selection to suppress cheating centromeres is driving HJURP's rapid evolution[51-53,144,165]. However, in contrast to CenH3, there is no clustering of selection in the primary amino acid sequence, lending little insight into the localization determinants of HJURP. The lack of any predicted tertiary structure makes it impossible to rule out a clustering of selected sites in the folded structure. Further research into the structure of HJURP may help illuminate clustering of selected sites and potential localization determinants.

The SCM3 homology domain is the only recognized region of homology linking orthologous chaperones[50]. Unsurprisingly, the SCM3 homology domain also contains the CenH3 binding motifs[55,149]. The general homology of the SCM3 domain and the functional conservation of the CATD binding partner, lead me to investigate whether this protein interaction domain is rapidly evolving. My analysis of the 51-codon region reveled a high level of conservation ($p \sim 1.0$ for positive selection, Figure 4.1B). Further, the critical CenH3 binding determinant, the TLTY Box, had no nonsynonymous changes but six synonymous changes, highlighting the close conservation and co-evolution of the partner domains of HJURP and CenH3.

The repetitive nature of centromeric DNA made analysis of centromeric sequences difficult but even with the limited data available major changes in primate centromere structure and function. All simian centromeric satellites

investigated thus far are based on 171-bp of alpha-satellites repeats. Centromeres of hominoids closely related to *H. sapiens* consist of multimeric arrays of alpha satellites[166-172]. These highly homogeneous arrays can consist of five to 22 diverged alpha-satellites, repeated from 2 Mb to 5 Mb in length [167,168,173]. Even among great apes, higher order arrays of alpha-satellites are not universal. *N. leucogenys* centromeres consist of tandem monomeric alpha-satellite repeats similar to Cercopithecidae primates' centromeres[9,174,175]. Platyrrhini primate *C. jacchus* has yet another repeat, dimeric arrays, of alpha satellites, showing that yet another centromere organization and an independent occurrence of higher order array evolution[9]. The single prosimian's *(Daubentonia madagascariensis*) centromeres analyzed consisted of novel DMA1 and DMA2 satellites, 146 base pairs and 268 base pairs respectively, with no homology to any other primates' centromeric satellite [176].

Repeated evolution of novel alpha-satellite sequences and arrays gave me the opportunity to investigate whether HJURP's rapid evolution is also episodic and repeated, consistent with driving novel centromeric satellites. Codeml is able to test each lineage of the primate HJURP phylogeny individually for rapid evolution. While this analysis is able to test each lineage separately, it is able to calculate only whole protein dN/dS values, diluting any signal of positive selection, but multiple lineages have dN/dS ratios greater than one. The rapid evolution appears to be ancient and repeated, with whole protein dN/dS values greater than 1 spread throughout the Catarrhini branches. (Figure 4.2). It

is particularly tantalizing to note that there were almost 16 nonsynonymous changes and no synonymous changes at the base of the Hominidae lineage, just as arrays of alpha-satellites are beginning emerge as the dominant centromeric repeat.

HJURP is a putative a new suppressor of centromere of centromere drive.

Centromere drive poses a serious challenge for the genome. The deleterious effects of cheating centromeres can significantly reduce male fertility, both while driving centromeres are heterozygous and after they have swept to fixation[21,24-26]. In addition, XY systems pose a particular challenge. The X chromosome might drive unchecked because the Y chromosome never has the opportunity for selection by being passed through asymmetric meiosis. Segregation distortion by a driving X chromosomes in an XY system would threaten a dangerous sex ratio imbalance. Previous studies have shown that CenH3's rapid evolution has lead to changes in satellite recognition and localization[37,38,41,177]. This rapid evolution and change in localization was focused on the CATD region, which is free to evolve only in the absence of a HJURP ortholog[2,37-39]. Selection would disfavor the creation of a chaperone interaction that reduced CenH3, and the genome's, ability to relieve the negative

effects of centromere drive unless there was a ready substitute.

HJURP's rapid evolution suggests that the chaperone is able to supplement, and perhaps supplant, CenH3 as the means to restrict centromere drive and specify centromeric location in primates. The chaperone restricts the evolution of CenH3's critical localization determination domain but provides another avenue involved in restricting centromere drive. In the organisms such as primates where the chaperone is present, the chaperone is rapidly evolving but organisms where there is not a known chaperone, the CenH3 is rapidly evolving. Because of the functional requirements of the protein interactions between chaperones and CenH3, I have never seen both the chaperone and CenH3 CATD rapidly evolving in the same lineages.

Materials and methods

Sequencing of HJURP from Primate RNA

HJURP was amplified and sequenced from total RNA with Superscipt III One-step RT-PCR kit (Invitrogen) by using the PCR and sequencing primers shown in Primer Table (Appendix I). RT-PCR products were sequenced directly, except in a few cases where they were first cloned into the TOPO TA cloning vector (Invitrogen), followed by sequencing of as least three independent clones. Sequence reads were assembled into full length transcripts. Additional cDNA sequences for *P. alebii* (Orangutan), *C. jacchus* (Marmoset) were obtained from the UCSC genome browser. HJURP coding sequences from *H. sapiens, M. mulatta* , and *C. jacchus* were used to query the nr, wgs, and hgts NCBI databases to *Otolemur garenttii* (Galago)*, Microcebus murinus* (Mouse Lemur), *Tarsius syrichta* (Tarsier) [134].

Selection Analysis

Translational alignments of primate sequences were created using Geneious with hand alignments of gaps[135]. Phylogenies, which conformed to the accepted primate phylogeny were constructed using neighbor joining (clustalX). A maximum likelihood analysis was performed with codeml in the PAML4 software package. To detect selection, multiple alignments were fitted to either the F3×4 or 1/61 models of codon frequencies. Likelihood ratio tests of the data were performed by using different sets of site-specific (NS sites) models as follows: M1 (two-state, neutral, dN/dS > 1 not allowed) to M2 (three-state, selection, dN/dS >

1 allowed); and M7 (beta distribution, dN/dS > 1 not allowed) to M8 (beta distribution but dN/dS > 1 allowed). Whole protein synonymous changes per synonymous site (dS)/replacement changes per replacement site (dN) ratios for the tree were calculated by a free-ratio model, which allows dN/dS to vary along different branches.

V. Conclusions

The rapidly increasing availability of genomic sequence data and the increase availability of sequencing technologies is enabling ever more robust use of evolutionary analysis to direct my functional characterization both genetic pathways and proteins. Evolutionary analysis has been used to direct the functional characterization of a variety of proteins. It has long been recognized that most proteins are highly conserved both functionally and at the sequence level. The landscape and strength of that conservation can point to critical amino acids, aiding in both functional characterization and prediction. There is another, rarer, class of proteins that remains functionally conserved but evolves rapidly at the sequences level. In these cases, evolutionary theory suggests that rapid evolution is driven by genetic conflict and analysis of the rapidly evolving regions has given us insight into nature of the genetic conflict. Until recently, the cost and difficulty of sequencing genes from many related species has confined these analyses to single genes or families of genes. I took advantage of the expanding amount of sequence data available in primates, in addition to extensive work to characterize the fundamental proteins of the kinetochore, to leap past the analysis of the small number of genes previously investigated and investigate the whole system.

The centromere drive conflict is restricted to the inner kinetochore

Investigation of the effects conflict between centromeres and centromere binding proteins thus far has been restricted to centromeric histones. Some of this has been due to the paucity of inner kinetochore proteins available in most species, in addition to the hypothesized central role of CenH3 in centromere specification and kinetochore function. In addition, many kinetochore proteins directly bind CenH3 and almost all are dependent on the centromeric histone for localization. This left open the question of what role, if any, the other proteins of the kinetochore might play in suppressing centromere drive. Selection was hypothesized to act on only DNA binding proteins to suppress the deleterious effects of driving centromeres but I have shown that positive selection acts widely on the inner kinetochore. Proteins with known DNA binding activity, such as CENP-T or KNL2, are rapidly evolving, as well as proteins from the CENP-O complex, without any hypothesized DNA binding activity. The surprising breadth of selection did not extend to the outer kinetochore or fibrous corona. Unlike the inner kinetochore, where many proteins have not been discovered outside of mammals, the outer kinetochore and fibrous corona is well conserved across taxa. Kinetochore proteins outside of the inner kinetochore display the same conservation of sequence in primates. I can now say that while the inner kinetochore evolution is driven by positive selection to suppress centromere drive, the outer kinetochore and fibrous corona have been preserved by their essential role in chromosome segregation.

The notable exception to the conservation of the outer kinetochore is CASC5, which links the kinetochore and microtubules. The CENP-O complex of the inner kinetochore also is rapidly evolving and regulates microtubule kinetochore interactions. Neither CASC5 nor CENP-O have any DNA binding activity but are among the most rapidly evolving kinetochore proteins, suggesting that there maybe an additional pathway for suppression of centromere drive. Meiotic positioning is believed to be critical for success in asymmetric meiosis but no one has proposed a mechanism by which regulation of microtubule binding or dynamics could suppress centromere drive. This data suggests that the kinetochore has an entirely novel pathway to suppress driving centromeres. Further testing of this hypothesis will be difficult as changes in microtubule dynamics are more challenging to quantify than changes in localization, and maybe present only in meiosis.

Resolving conflicting observations about CenH3 evolution

CenH3's function and evolution had been extensively studied, but conflicting evidence had begun to arise. CenH3 is rapidly evolving in every species with asymmetric meiosis studied thus far, however, complementarity between species varied dramatically. Rapid of evolution of CenH3 in plants and diptera abrogated localization in related species, but despite the rapid evolution of both CenH3 and centromeres, CenH3 is able to localize throughout the primate lineage. Even a distantly related fungal CenH3 is able to localize primate centromeres, comprised of satellites divergent in sequence and organization. I reconciled these observations by analyzing the evolution of primate CenH3 in

light of recent evidence identifying key targeting elements and the phylogenic distribution of a specialized chaperone, HJURP. This work directly identifies and addresses the selective pressures acting on the histone fold domain of CenH3; selection to innovate to suppress centromere drive and purifying selection to retain chaperone binding and interaction. It provides evidence that fundamental kinetochore proteins can be born and lost, and the rest of the kinetochore can adapt to new roles in centromere recognition. The HJURP ortholog was lost in *Drosophila* with the birth of a new chaperone but CenH3 is no longer exclusively dependent on the new chaperone, rather CenH3, the chaperone, and CENP-C are now interdependent.

Future directions

My work provides two clear avenues for further research on the effects of centromere drive and the organization of the kinetochore. My findings on the co-evolution of CenH3 and HJURP point to the importance of HJURP in specifying centromere location and function, however, we know little about how HJURP recognizes centromeric DNA or chromatin. This is a clear departure from the previous hypothesis that CenH3 was the proximate determinant of centromere specification. Further research into how HJURP, and Mis18, on which HJURP is dependent, specify centromere function is necessary but the choice of model system is likely to be of critical importance. Several model systems lack HJURP orthologs and in those system CenH3 appears to take a more direct role in localization, which is just one indication that centromere specification may evolutionarily labile.

My survey of the evolutionary history of the kinetochore also starts to provide both a more detailed mechanism of centromere cheating during drive and a potential novel new path to suppress centromere drive. It has been hypothesized that larger or more 'active' centromeres recruited more microtubules to gain advantage position and transmission but there has been little direct evidence of this. I found that kinetochore proteins that provide attachment sites or regulate microtubule binding are among the most rapidly evolving proteins of the kinetochore suggesting that microtubule binding plays a key role in centromere's cheating. It also expands the previous hypothesis, only proteins that directly interact with centromeric or pericentromeric DNA are candidates to suppress centromere drive. The wide spread rapid evolution of microtubule regulators strongly suggests that they are suppressing centromere drive or involved in a unrecognized conflict. Perhaps the most unexpected outcome of this research is the recognition of the greatly expanded impact of the centromere drive confliction on the kinetochore and genome

VI. References

1. Black BE, Jansen LET, Maddox PS, Foltz DR, Desai AB, et al. (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell 25: 309-322.

2. Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3: 18.

3. Sekulic N, Bassett EA, Rogers DJ, Black BE (2010) The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. Nature 467: 347-351.

4. Malik HS, Henikoff S (2009) Major Evolutionary Transitions in Centromere Complexity. Cell 138: 1067-1082.

5. Schueler MG, Swanson W, Thomas PJ, Program NCS, Green ED (2010) Adaptive evolution of foundation kinetochore proteins in primates. Molecular Biology and Evolution 27: 1585-1597.

6. van Valen L (1973) A new evolutionary law. Evolutionary Theory 1: 1-30.

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

8. Haaf T, Warburton PE, Willard HF (1992) Integration of human α-satellite DNA into simian chromosomes: Centromere protein binding and disruption of normal chromosome segregation. Cell 70: 681-696.

9. Cellamare A, Catacchio CR, Alkan C, Giannuzzi G, Antonacci F, et al. (2009) New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. Molecular Biology and Evolution 26: 1889-1900.

10. Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. Genome Research 16: 88-96.

11. Malik HS, Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. Current Opinion in Genetics & Development 12: 711-718.

12. Schueler MG, Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin. Annual review of genomics and human genetics 7: 301-313.

13. Sun X, Wahlstrom J, Karpen G (1997) Molecular Structure of a Functional Drosophila Centromere. Cell 91: 1007-1019.

14. Bachmann L, Sperlich D (1993) Gradual evolution of a specific satellite DNA family in Drosophila ambigua, D. tristis, and D. obscura. Molecular Biology and Evolution 10: 647-659.

15. Lohe AR, Brutlag DL (1987) Identical satellite DNA sequences in sibling species of Drosophila. Journal of Molecular Biology 194: 161-170.

16. Sawamura K, Fujita A, Yokoyama R, Taira T, Inoue YH, et al. (1995) Molecular and genetic dissection of a reproductive isolation gene, zygotic

hybrid rescue, of Drosophila melanogaster. The Japanese Journal of Genetics 70: 223-232.

17. Lee H-R, Zhang W, Langdon T, Jin W, Yan H, et al. (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc Natl Acad Sci USA 102: 11793-11798.

18. Bensasson D, Zarowiecki M, Burt A, Koufopanou V (2008) Rapid Evolution of Yeast Centromeres in the Absence of Drive. Genetics 178: 2161-2167.

19. Dover G (2002) Molecular drive. Trends Genet 18: 587-589.

20. Henikoff S, Malik HS (2002) Centromeres: selfish drivers. Nature 417: 227.

21. Malik HS (2005) Mimulus finds centromeres in the driver's seat. Trends in Ecology & Evolution 20: 151-154.

22. Hewitt G (1973) Variable transmission rates of a B-chromosome in Myrmeleotettix maculatus (Thumb.) (Acrididae: Orthoptera). Chromosoma 40: 83-106.

23. Pardo-Manuel de Villena F, Sapienza C (2001) Transmission ratio distortion in offspring of heterozygous female carriers of Robertsonian translocations. Hum Genet 108: 31-36.

24. Fishman L, Saunders A (2008) Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science 322: 1559-1562.

25. Fishman L, Willis JH (2005) A novel meiotic drive locus almost completely distorts segregation in mimulus (monkeyflower) hybrids. Genetics 169: 347-353.

26. Daniel A (2002) Distortion of female meiotic segregation and reduced male fertility in human Robertsonian translocations: consistent with the centromere model of co-evolving centromere DNA/centromeric histone (CENP-A). Am J Med Genet 111: 450-452.

27. Cheeseman IM, Desai A (2008) Molecular architecture of the kinetochore-microtubule interface. Nature Reviews Molecular Cell Biology 9: 33-46.

28. Hori T, Amano M, Suzuki A, Backer CB, Welburn JP, et al. (2008) CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. Cell 135: 1039-1052.

29. Foltz DR, Jansen LE, Black BE, Bailey AO, Yates JR, 3rd, et al. (2006) The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol 8: 458-469.

30. Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nat Struct Biol 10: 882-891.

31. Earnshaw WC, Migeon BR (1985) Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. Chromosoma 92: 290-296.

32. Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proceedings of the National Academy of Sciences 88: 3734-3738.

33. Panchenko T, Black BE (2009) The Epigenetic Basis for Centromere Identity. Centromere 48: 1-32.

34. Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, et al. (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Current Biology 7: 901-904.

35. Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proceedings of the National Academy of Sciences of the United States of America 103: 3238-3243.

36. Malik HS, Vermaak D, Henikoff S (2002) Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proc Natl Acad Sci USA 99: 1449-1454.

37. Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. Plant Cell 14: 1053-1066.

38. Vermaak D, Hayden HS, Henikoff S (2002) Centromere targeting element within the histone fold domain of Cid. Molecular and Cellular Biology 22: 7553-7561.

39. Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol 21: 1712-1718.

40. Tachiwana H, Kagawa W, Shiga T, Osakabe A, Miya Y, et al. (2011) Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476: 232-235.

41. Ravi M, Kwong PN, Menorca RMG, Valencia JT, Ramahi JS, et al. (2010) The rapidly evolving centromere-specific histone has stringent functional requirements in Arabidopsis thaliana. Genetics 186: 461-471.

42. Conde e Silva N, Black BE, Sivolob A, Filipski J, Cleveland DW, et al. (2007) CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. J Mol Biol 370: 555-573.

43. Camahort R, Shivaraju M, Mattingly M, Li B, Nakanishi S, et al. (2009) Cse4 Is Part of an Octameric Nucleosome in Budding Yeast. Molecular Cell 35: 794-805.

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

45. Black BE, Cleveland DW (2011) Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. Cell 144: 471-479.

46. Henikoff S, Furuyama T (2010) Epigenetic Inheritance of Centromeres. Cold Spring Harbor Symposia on Quantitative Biology.

47. Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, et al. (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Developmental Cell 10: 303-315.

48. Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, et al. (2004) Structural determinants for generating centromeric chromatin. Nature 430: 578-582.

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

50. Sanchez-Pulido L, Pidoux AL, Ponting CP, Allshire RC (2009) Common ancestry of the CENP-A chaperones Scm3 and HJURP. Cell 137: 1173-1174.

51. Foltz DR, Jansen LE, Bailey AO, Yates JR, 3rd, Bassett EA, et al. (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137: 472-484.

52. Stoler S, Rogers K, Weitze S, Morey L, Fitzgerald-Hayes M, et al. (2007) Scm3, an essential Saccharomyces cerevisiae centromere protein

required for G2/M progression and Cse4 localization. Proceedings of the National Academy of Sciences 104: 10571-10576.

53. Camahort R, Li B, Florens L, Swanson SK, Washburn MP, et al. (2007) Scm3 Is Essential to Recruit the Histone H3 Variant Cse4 to Centromeres and to Maintain a Functional Kinetochore. Molecular Cell 26: 853-865.

54. Zhou Z, Feng H, Zhou B-R, Ghirlando R, Hu K, et al. (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. Nature 472: 234-237.

55. Shuaib M, Ouararhni K, Dimitrov S, Hamiche A (2010) HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proceedings of the National Academy of Sciences 107: 1349-1354.

56. Barnhart MC, Kuich PHJL, Stellfox ME, Ward JA, Bassett EA, et al. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. The Journal of Cell Biology 194: 229-243.

57. Mellone BG, Grive KJ, Shteyn V, Bowers SR, Oderberg I, et al. (2011) Assembly of Drosophila centromeric chromatin proteins during mitosis. PLoS Genet 7: e1002068.

58. Schittenhelm RB, Althoff F, Heidmann S, Lehner CF (2010) Detrimental incorporation of excess Cenp-A/Cid and Cenp-C into Drosophila

centromeres is prevented by limiting amounts of the bridging factor Cal1. J Cell Sci 123: 3768-3779.

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

60. Politi V, Perini G, Trazzi S, Pliss A, Raska I, et al. (2002) CENP-C binds the alpha-satellite DNA in vivo at specific centromere domains. Journal of Cell Science 115: 2317-2327.

61. Trazzi S, Perini G, Bernardoni R, Zoli M, Reese JC, et al. (2009) The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. PLoS ONE 4: e5832.

62. Du Y, Topp CN, Dawe RK (2010) DNA Binding of Centromere Protein C (CENPC) Is Stabilized by Single-Stranded RNA. PLoS Genet 6: e1000835.

63. Orr B, Sunkel CE (2011) Drosophila CENP-C is essential for centromere identity. Chromosoma 120: 83-96.

64. Casola C, Hucks D, Feschotte C (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Molecular Biology and Evolution 25: 29-41.

65. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in

alphoid DNA, a human centromeric satellite. The Journal of Cell Biology 109: 1963-1973.

66. Ohzeki J-I, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. The Journal of Cell Biology 159: 765-775.

67. Amano M, Suzuki A, Hori T, Backer C, Okawa K, et al. (2009) The CENP-S complex is essential for the stable assembly of outer kinetochore structure. The Journal of Cell Biology 186: 173-182.

68. Perpelescu M, Fukagawa T (2011) The ABCs of CENPs. Chromosoma: 1-22.

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

70. Fujita Y, Hayashi T, Kiyomitsu T, Toyoda Y, Kokubu A, et al. (2007) Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Developmental Cell 12: 17-30.

71. Petrovic A, Pasqualato S, Dube P, Krenn V, Santaguida S, et al. (2010) The MIS12 complex is a protein interaction hub for outer kinetochore assembly. The Journal of Cell Biology 190: 835-852.

72. Cheeseman IM, Chappie JS, Wilson-Kubalek EM, Desai A (2006) The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. Cell 127: 983-997.

73. Cheeseman IM, Hori T, Fukagawa T, Desai A (2008) KNL1 and the CENP-H/I/K complex coordinately direct kinetochore assembly in vertebrates. Molecular Biology of the Cell 19: 587-594.

74. Powers AF, Franck AD, Gestaut DR, Cooper J, Gracyzk B, et al. (2009) The Ndc80 Kinetochore Complex Forms Load-Bearing Attachments to Dynamic Microtubule Tips via Biased Diffusion. Cell 136: 865-875.

75. Desai A, Rybina S, M√ºller-Reichert T, Shevchenko A, Shevchenko A, et al. (2003) KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in C. elegans. Genes & Development 17: 2421-2435.

76. Williams B, Leung G, Maiato H, Wong A, Li Z, et al. (2007) Mitch - a rapidly evolving component of the Ndc80 kinetochore complex required for correct chromosome segregation in Drosophila. Journal of Cell Science 120: 3522-3533.

77. Lampert F, Westermann S (2011) A blueprint for kinetochores - new insights into the molecular mechanics of cell division. Nat Rev Mol Cell Biol 12: 407-412.

78. Miranda JJ, De Wulf P, Sorger PK, Harrison SC (2005) The yeast DASH complex forms closed rings on microtubules. Nat Struct Mol Biol 12: 138-143.

79. Cheeseman IM, Enquist-Newman M, M√ºller-Reichert T, Drubin DG, Barnes G (2001) Mitotic Spindle Integrity and Kinetochore Function Linked by the Duo1p/Dam1p Complex. The Journal of Cell Biology 152: 197-212.

80. Buttrick GJ, Millar JB (2011) Ringing the changes: emerging roles for DASH at the kinetochore-microtubule Interface. Chromosome Res 19: 393-407.

81. Sardar HS, Luczak VG, Lopez MM, Lister BC, Gilbert SP (2010) Mitotic kinesin CENP-E promotes microtubule plus-end elongation. Curr Biol 20: 1648-1653.

82. Bayes JJ, Malik HS (2008) The evolution of centromeric DNA sequences. Encyclopedia of Life Sciences.

83. Earnshaw WC, Rothfield N (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91: 313-321.

84. Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157: 1293-1298.

85. Amor DJ, Kalitsis P, Sumer H, Choo KHA (2004) Building the centromere: from foundation proteins to 3D organization. Trends in Cell Biology 14: 359-368.

86. Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC (1996) Surprising deficiency of CENP-B binding sites in African green monkey

alpha-satellite DNA: implications for CENP-B function at centromeres. Molecular and Cellular Biology 16: 5156-5168.

87. Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, et al. (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. The Journal of Cell Biology 141: 309-319.

88. Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, et al. (1998) The cenpB gene is not essential in mice. Chromosoma 107: 570-576.

89. Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL, Vogel KG, et al. (1998) Centromeric protein B null mice are viable with no apparent abnormalities. Developmental Biology 201: 135-143.

90. Tanaka K, Li Chang H, Kagami A, Watanabe Y (2009) CENP-C Functions as a Scaffold for Effectors with Essential Kinetochore Functions in Mitosis and Meiosis. Developmental Cell 17: 334-343.

91. Sugimoto K, Yata H, Muro Y, Himeno M (1994) Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. J Biochem 116: 877-881.

92. Yang CH, Tomkiel J, Saitoh H, Johnson DH, Earnshaw WC (1996) Identification of overlapping DNA-binding and centromere-targeting

domains in the human kinetochore protein CENP-C. Molecular and Cellular Biology 16: 3576-3586.

93. Sugimoto K, Kuriyama K, Shibata A, Himeno M (1997) Characterization of internal DNA-binding and C-terminal dimerization domains of human centromere/kinetochore autoantigen CENP-C in vitro: role of DNA-binding and self-associating activities in kinetochore organization. Chromosome Res 5: 132-141.

94. Trazzi S, Bernardoni R, Diolaiti D, Politi V, Earnshaw WC, et al. (2002) In vivo functional dissection of human inner kinetochore protein CENP-C. J Struct Biol 140: 39-48.

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

96. Milks K, Moree B, Straight A (2009) Dissection of CENP-C-directed Centromere and Kinetochore Assembly. Molecular Biology of the Cell.

97. Przewloka MR, Venkei Z, Bolanos-Garcia VM, Debski J, Dadlez M, et al. (2011) CENP-C Is a Structural Platform for Kinetochore Assembly. Curr Biol 21: 399-405.

98. Régnier V, Vagnarelli P, Fukagawa T, Zerjal T, Burns E, et al. (2005) CENP-A Is Required for Accurate Chromosome Segregation and Sustained Kinetochore Association of BubR1. Molecular and Cellular Biology 25: 3967-3981.

99. Okada M, Cheeseman IM, Hori T, Okawa K, Mcleod IX, et al. (2006) The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat Cell Biol 8: 446-457.

100. Izuta H, Ikeno M, Suzuki N, Tomonaga T, Nozaki N, et al. (2006) Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. Genes to Cells 11: 673-684.

101. Gascoigne KE, Takeuchi K, Suzuki A, Hori T, Fukagawa T, et al. (2011) Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. Cell 145: 410-422.

102. Comeron JM (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 15: 763-764.

103. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591.

104. Suzuki A, Hori T, Nishino T, Usukura J, Miyagi A, et al. (2011) Spindle microtubules generate tension-dependent changes in the distribution of inner kinetochore proteins. The Journal of Cell Biology 193: 125-140.

105. Nishino T, Takeuchi K, Gascoigne KE, Suzuki A, Hori T, et al. (2012) CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. Cell 148: 487-501.

106. Carroll CW, Silva MC, Godek KM, Jansen LE, Straight AF (2009) Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat Cell Biol 11: 896-902.

107. Carroll CW, Milks KJ, Straight AF (2010) Dual recognition of CENP-A nucleosomes is required for centromere assembly. The Journal of Cell Biology 189: 1143-1155.

108. McClelland SE, Borusu S, Amaro AC, Winter JR, Belwal M, et al. (2007) The CENP-A NAC/CAD kinetochore complex controls chromosome congression and spindle bipolarity. EMBO J 26: 5033-5047.

109. Fukagawa T, Mikami Y, Nishihashi A, Regnier V, Haraguchi T, et al. (2001) CENP-H, a constitutive centromere component, is required for centromere targeting of CENP-C in vertebrate cells. EMBO J 20: 4603-4617.

110. Alonso A, Fritz B, Hasson D, Abrusan G, Cheung F, et al. (2007) Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. Genome Biology 8: R148.

111. Liu S-T, Hittle JC, Jablonski SA, Campbell MS, Yoda K, et al. (2003) Human CENP-I specifies localization of CENP-F, MAD1 and MAD2 to kinetochores and is essential for mitosis. Nat Cell Biol 5: 341-345.

112. Mikami Y, Hori T, Kimura H, Fukagawa T (2005) The Functional Region of CENP-H Interacts with the Nuf2 Complex That Localizes to Centromere during Mitosis. Molecular and Cellular Biology 25: 1958-1970.

113. Nishihashi A, Haraguchi T, Hiraoka Y, Ikemura T, Regnier V, et al. (2002) CENP-I is essential for centromere function in vertebrate cells. Developmental Cell 2: 463-476.

114. Sugata N, Li S, Earnshaw WC, Yen TJ, Yoda K, et al. (2000) Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere--kinetochore complexes. Human Molecular Genetics 9: 2919-2926.

115. Park JE, Erikson RL, Lee KS (2011) Feed-forward mechanism of converting biochemical cooperativity to mitotic processes at the kinetochore plate. Proc Natl Acad Sci U S A 108: 8200-8205.

116. Kang YH, Park CH, Kim T-S, Soung N-K, Bang JK, et al. (2011) Mammalian Polo-like Kinase 1-dependent Regulation of the PBIP1-CENP-Q Complex at Kinetochores. Journal of Biological Chemistry 286: 19744-19757.

117. Hua S, Wang Z, Jiang K, Huang Y, Ward T, et al. (2011) CENP-U cooperates with Hec1 to orchestrate kinetochore-microtubule attachment. J Biol Chem 286: 1627-1638.

118. Saito A, Muro Y, Sugiura K, Ikeno M, Yoda K, et al. (2009) CENP-O, a protein localized at the centromere throughout the cell cycle, is a novel target antigen in systemic sclerosis. J Rheumatol 36: 781-786.

119. Toso A, Winter JR, Garrod AJ, Amaro AC, Meraldi P, et al. (2009) Kinetochore-generated pushing forces separate centrosomes during bipolar spindle assembly. J Cell Biol 184: 365-372.

120. Hori T, Okada M, Maenaka K, Fukagawa T (2008) CENP-O class proteins form a stable complex and are required for proper kinetochore function. Mol Biol Cell 19: 843-854.

121. Kim IS, Lee M, Park KC, Jeon Y, Park JH, et al. (2012) Roles of Mis18α in Epigenetic Regulation of Centromeric Chromatin and CENP-A Loading. Molecular Cell 46: 260-273.

122. Liu D, Vleugel M, Backer CB, Hori T, Fukagawa T, et al. (2010) Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. The Journal of Cell Biology 188: 809-820.

123. Welburn JPI, Vleugel M, Liu D, Yates Iii JR, Lampson MA, et al. (2010) Aurora B Phosphorylates Spatially Distinct Targets to Differentially Regulate the Kinetochore-Microtubule Interface. Molecular Cell 38: 383-392.

124. Shepperd LA, Meadows JC, Sochaj AM, Lancaster TC, Zou J, et al. (2012) Phosphodependent Recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 Kinase Maintains the Spindle Checkpoint. Curr Biol 22: 891-899.

125. Kiyomitsu T, Obuse C, Yanagida M (2007) Human Blinkin/AF15q14 Is Required for Chromosome Alignment and the Mitotic Checkpoint through Direct Interaction with Bub1 and BubR1. Developmental Cell 13: 663-676.

126. Bolanos-Garcia Victor M, Lischetti T, Matak-Vinković D, Cota E, Simpson Pete J, et al. (2011) Structure of a Blinkin-BUBR1 Complex Reveals an

Interaction Crucial for Kinetochore-Mitotic Checkpoint Regulation via an Unanticipated Binding Site. Structure 19: 1691-1700.

127. Ruchaud S, Carmena M, Earnshaw WC (2007) Chromosomal passengers: conducting cell division. Nat Rev Mol Cell Biol 8: 798-812.

128. Elde N, Roach K, Yao M-C, Malik H (2011) Absence of Positive Selection on Centromeric Histones in Tetrahymena Suggests Unsuppressed Centromere-Drive in Lineages Lacking Male Meiosis. Journal of Molecular Evolution: 1-11.

129. Maddox PS, Hyndman F, Monen J, Oegema K, Desai A (2007) Functional genomics identifies a Myb domain–containing protein family required for assembly of CENP-A chromatin. The Journal of Cell Biology 176: 757-763.

130. Moree B, Meyer CB, Fuller CJ, Straight AF (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. The Journal of Cell Biology 194: 855-871.

131. Dambacher S, Deng W, Hahn M, Sadic D, Frohlich J, et al. (2012) CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3: 101-110.

132. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, et al. (2011) The UCSC Genome Browser database: update 2011. Nucleic Acids Res 39: D876-882.

133. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996-1006.

134. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

135. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, et al. (2011) Geneious v5.3, Available from http://www.geneious.com.

136. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, et al. (2011) A Molecular Phylogeny of Living Primates. PLoS Genet 7: e1001342.

137. Blower MD, Sullivan BA, Karpen GH (2002) Conserved Organization of Centromeric Chromatin in Flies and Humans. Developmental Cell 2: 319-330.

138. Mendiburo MaJ, Padeken J, F√ºl√∂p S, Schepers A, Heun P (2011) Drosophila CENH3 Is Sufficient for Centromere Formation. Science 334: 686-690.

139. Hewawasam G, Shivaraju M, Mattingly M, Venkatesh S, Martin-Brown S, et al. (2010) Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. Mol Cell 40: 444-454.

140. Cho U-S, Harrison SC (2011) Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. Proceedings of the National Academy of Sciences 108: 9367-9371.

141. Ranjitkar P, Press MO, Yi X, Baker R, MacCoss MJ, et al. (2010) An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. Mol Cell 40: 455-464.

142. Baker RE, Rogers K (2006) Phylogenetic analysis of fungal centromere H3 proteins. Genetics 174: 1481-1492.

143. Wieland G, Orthaus S, Ohndorf S, Diekmann S, Hemmerich P (2004) Functional Complementation of Human Centromere Protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. Mol Cell Biol 24: 6620-6630.

144. Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, et al. (2009) HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137: 485-497.

145. Wainright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260: 340-342.

146. Steenkamp ET, Wright J, Baldauf SL (2006) The Protistan Origins of Animals and Fungi. Molecular Biology and Evolution 23: 93-106.

147. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, et al. (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451: 783-788.

148. Henikoff S, Ahmad K, Platero JS, van Steensel B (2000) Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci U S A 97: 716-721.

149. Hu H, Liu Y, Wang M, Fang J, Huang H, et al. (2011) Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. Genes & Development 25: 901-906.

150. Bassett EA, DeNizio J, Barnhart-Dailey MC, Panchenko T, Sekulic N, et al. (2012) HJURP Uses Distinct CENP-A Surfaces to Recognize and to Stabilize CENP-A/Histone H4 for Centromere Assembly. Developmental Cell 22: 749-762.

151. Clarke L, Baum MP (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol 10: 1863-1872.

152. Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biol 7: R23.

153. Moreno-Moreno O, Medina-Giro S, Torras-Llort M, Azorin F (2011) The F box protein partner of paired regulates stability of Drosophila centromeric histone H3, CenH3(CID). Curr Biol 21: 1488-1493.

154. Bronner C, Hamiche A (2012) Cancer cell death and selection: unexpected putative roles for pRb2/p130, BORIS and CTCF in endoplasmic stress response maintained by the T-antigen. Cell Cycle 11: 2052.

155. Erhardt S, Mellone BG, Betts CM, Zhang W, Karpen GH, et al. (2008) Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation. J Cell Biol 183: 805-818.

156. Phansalkar R, Lapierre P, Mellone BG (2012) Evolutionary insights into the role of the essential centromere protein CAL1 in Drosophila. Chromosome Res 20: 493-504.

157. Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proceedings of the Royal Society B: Biological Sciences 273: 1867-1872.

158. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin Fß, Higgins DG (1997) The CLUSTALX Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. Nucleic Acids Research 25: 4876-4882.

159. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

160. Altschul SF, Koonin EV (1998) Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. Trends Biochem Sci 23: 444-447.

161. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33: W244-W248.

162. Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast 15: 1541-1553.

163. Xiao H, Mizuguchi G, Wisniewski J, Huang Y, Wei D, et al. (2011) Nonhistone Scm3 binds to AT-rich DNA to organize atypical centromeric nucleosome of budding yeast. Mol Cell 43: 369-380.

164. Van Hooser AA, Ouspenski II, Gregson HC, Starr DA, Yen TJ, et al. (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. Journal of Cell Science 114: 3529-3542.

165. Bernad R, Sanchez P, Rivera T, Rodriguez-Corsino M, Boyarchuk E, et al. (2011) Xenopus HJURP and condensin II are required for CENP-A assembly. The Journal of Cell Biology 192: 569-582.

166. Mitchell AR, Gosden JR, Miller DA (1985) A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma 92: 369-377.

167. Nusbaum C, Mikkelsen TS, Zody MC, Asakawa S, Taudien S, et al. (2006) DNA sequence and analysis of human chromosome 8. Nature 439: 331-335.

168. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. (2005) The DNA sequence of the human X chromosome. Nature 434: 325-337.

169. Haaf T, Willard HF (1997) Chromosome-specific alpha-satellite DNA from the centromere of chimpanzee chromosome 4. Chromosoma 106: 226-232.

170. Miller DA, Sharma V, Mitchell AR (1988) A human-derived probe, p82H, hybridizes to the centromeres of gorilla, chimpanzee, and orangutan. Chromosoma 96: 270-274.

171. Warburton PE, Haaf T, Gosden J, Lawson D, Willard HF (1996) Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. Genomics 33: 220-228.

172. Haaf T, Willard HF (1998) Orangutan alpha-satellite monomers are closely related to the human consensus sequence. Mamm Genome 9: 440-447.

173. Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC (1997) Human centromeric DNAs. Human Genetics 100: 291-304.

174. Musich PR, Brown FL, Maio JJ (1980) Highly repetitive component alpha and related alphoid DNAs in man and monkeys. Chromosoma 80: 331-348.

175. Pike LM, Carlisle A, Newell C, Hong SB, Musich PR (1986) Sequence and evolution of rhesus monkey alphoid DNA. J Mol Evol 23: 127-137.

176. Lee H-R, Hayden KE, Willard HF (2011) Organization and Molecular Evolution of CENP-A–Associated Satellite DNA Families in a Basal Primate Genome. Genome Biology and Evolution 3: 1136-1149.

177. Nagaki K, Terada K, Wakimoto M, Kashihara K, Murata M (2010) Centromere targeting of alien CENH3s in Arabidopsis and tobacco cells. Chromosome Res 18: 203-211.

VI. Appendix

Table of Primers

Table of Primers 1 of 3

Primer	Sequence	Gene	Species
JAK25	CCACATCAAGCTGTTCTTCAGCTGG	CENP-C	Primates
JAK97	ATATTTAAGSTTGGTTKATCTTTCATC	CENP-C	Primates
KR150	GCGGCCGCATTATGGTTTACCCATACGACGTTCCAGACTACGCT ATGGCTGCGTCSGGTCTGGATCATC	CENP-C	Primates
KR170	CTCGAGCGGCCGCTTAACTATTGTGATGCATTGGTAATTC	CENP-C	Hominoids
KR171	CTCGAGCGGCCGCTTATAGTGTCATGACGYACTGTTAATTC	CENP-C	New World Monkeys
KR172	CTCGAGCGGCCGCTTAAMCTCCAGAACCTTCAGCATT	CENP-C	Hominoids
KR173	CTCGAGCGGCCGCTTAACTGTTGYGATGCATTGGTAATTC	CENP-C	Old World Monkeys
KR174	CTCGAGCGGCCGCTTAACTGTTGTGAGGCATTGGTAATTC	CENP-C	P. abelii
KR175	CTCGAGCGGCCGCTTAACYTCCAGAACCTTCAGCATT	CENP-C	Old/New World Monkeys
KR250	GGAAAGAAGCTTATGTCAAGTAAACAACAATG	CenH3	
KR251	CTAGATAAATTGACCTCTTAATC	CenH3	
KR252	GCTTCTGCTGCCTCTTGTAGGGCCAATAGGGCCTGGGCTTGCCA ATTGAAGTCCACACCACGAGTGAATTTAACACAGATTTCTCTGA CCAATCTTTGGA	CenH3	H. sapiens
KR253	TTGGCCCTACAAGAGGCAGCAGAAGCATTTCTAGTTCATCTCTT TGAGGACGCCTATCTCCTCACCTTACATGCAGGCCGTGTCACTA TCCAAAAGAAGG	CenH3	H. sapiens
KR254	CGCACGACTCCTGCATGGCCAATAGGGCGCCTTCGGTGACCCTT AGCGGCTCGTCGTCGCTGTACTTCACGATGATTTCTCTGACCAA TCTTTGGAAAGG	CenH3	D. melanogaster
KR255	TTGGCCATGCAGGAGTCGTGCGAGATGTACTTGACGCAGCGGCT CGCCGACTCCTACATGCTAACCAAGCATCGCAATCGTGTCACTA TCCAAAAGAAGG	CenH3	D. melanogaster
KR260	AGCTTCAGTAGTAGTTTGGATAGCCAAAAGAGCTCCAGGAGTTA TTCTAAGACCATCGTCGAATAAGCACATGATTTCTCTGACCAAT CTTTGGAAAGG	CenH3	D. discoideum
KR261	AAGAGAAGAAGAAGGCTAGTTGCTTCAGTGATTCGCAGACTTAG AAATTCTGTTGATTTTATGATTCCTCGTCTCCCTTTCCAAAGAT TGGTCAGAGAAA	CenH3	D. discoideum
KR262	CTGCCGCCTCTTGAAGAGCAACAAGAGCTTCAGCTGTCCAACGA TTGATTTGGGGAGGGGCCAACATATGGGTGATTTCTCTGACCAA TCTTTGGAAAGG	CenH3	A. thaliana
KR263	GTTGCTCTTCAAGAGGCGGCAGAAGATTACTTGGTTGGTTTGTT CTCAGATTCAATGCTCTGTGCTATCCATGCAAAGCGTGTCACTA TCCAAAAGAAGG	CenH3	A. thaliana
KR264	TTGCTTCCTGTAACGCCATAATCGCCATTGACTGCCAACGTAAA TCCTGATCTTTAGTTGTAAACTCGTCTGTGATTTCTCTGACCAA TCTTTGGAAAGG	CenH3	S. cerevisiae
KR267	GTTTTACTTCAAACAGTTGTAGAAGATTATATGGTTTCATTTTT TGAAGATGCGAATGCTTGTGCTCTCCATGCTAAACGTGTCACTA TCCAAAAGAAGG	CenH3	T. thermophila
KR268	GCCTCTGCAGCCTCCTGTAGAGCCATTAAAGCATATCCTTGCCA CATCATGTGCTCTCGGCTGAACATCTGACAGATTTCTCTGACCA ATCTTTGGAAAG	CenH3	D. rerio
KR269	GGCTCTACAGGAGGCTGCAGAGGCGTTCATGGTTCGTCTGTTTT CTGATGCCAACCTCTGTGCCATTCACGCCAAAAGCGTGTCACTA TCCAAAAGAAGG	CenH3	D. rerio
KR285	GAAAGGGAATTCTGCCCGAACCAAGCAGACTGCGCGCAAG	CenH3	
KR286	GAAAGGGGTACCTCAGGCCCGCTCCCCGCGGATA	CenH3	
KR327	CAGTTGGCTAGAAGATTAAGAGGTCAATTTATCTAGTATTTGTG CTTCAGGGCTG	CenH3	
KR328	GAAAGGACTAGTGGTGTTTCTGGAGCAGCGCGA	CenH3	

KR329	GGAAAGGTCGACCCTGCAGAAGTATCCCTTAATCAGTC	CenH3	

Table of Primers 2 of 3

Primer	Sequence	Gene	Species
KR337	TCGGCAGCCTCTTGCAGTGCCTCAATGGCCTTGGCTTGCCACCG GAGCGGGGCAGTGTGATAAAAGTCTTGGGCGATTTCTCTGACCA ATCTTTGGAAAG	CenH3	M. brevicollis
KR338	GGCACTGCAAGAGGCTGCCGAAGCCTACATTGTTCGCCTCTTGG AGGACGGCAATCTTTGCGCCATCCATGCCAAGCGTGTCACTATC CAAAAGAAGGAT	CenH3	M. brevicollis
KR339	TGCTTCTTGTAATGCCATGATCGCCGAACTTTGCCATCGATATT CGCCCACCTCATCAGAGCCTACGTTCATGGCGATTTCTCTGACC AATCTTTGGAAA	CenH3	C. neoformans
KR340	AAATGCTTCTGCTGCCTCTTGTAGGGCCAATAGGGCCTGGGCTT GCCAATTGAAGTCCACACCACGAGTGAATTTAACACAGATCTCG CGCATCAGCCGC	CenH3	H. sapiens
KR341	ATTGGCCCTACAAGAGGCAGCAGAAGCATTTCTAGTTCATCTCT TTGAGGACGCCTATCTCCTCACCTTACATGCAGGCCGGGTCACC ATCATGCCTAAG	CenH3	H. sapiens
KR342	GGCATTACAAGAAGCAGCAGAGGCTTTTCTTGTTCATCTCTTTG AAGATGCGAATTTATGTGCCATCCATGCAAAGCGTGTCACTATC CAAAAGAAGGAT	CenH3	C. neoformans
KR344	ATCTTCTGCCGCCTCTTGAAGAGCAACAAGAGCTTCAGCTGTCC AACGATTGATTTGGGGAGGGGCCAACATATGGGTGATCTCGCGC ATCAGCCGCTGG	CenH3	A. thaliana
KR347	CATGGCATTACAAGAAGCAGCAGAGGCTTTTCTTGTTCATCTCT TTGAAGATGCGAATTTATGTGCCATCCATGCAAAGCGGGTCACC ATCATGCCTAAG	CenH3	C. neoformans
KR348	TTCGCTTGCTTCCTGTAACGCCATAATCGCCATTGACTGCCAAC GTAAATCCTGATCTTTAGTTGTAAACTCGTCTGTGATCTCGCGC ATCAGCCGCTGG	CenH3	S. cerevisiae
KR349	TATGGCGTTACAGGAAGCAAGCGAAGCGTATCTGGTAGGATTAT TGGAACATACAAACCTCTTGGCGCTGCATGCAAAACGGGTCACC ATCATGCCTAAG	CenH3	S. cerevisiae
KR349	ATTGGCCATGCAGGAGTCGTGCGAGATGTACTTGACGCAGCGGC TCGCCGACTCCTACATGCTAACCAAGCATCGCAATCGGGTCACC ATCATGCCTAAG	CenH3	D. melanogaster
KR350	AGTACATCTCGCACGACTCCTGCATGGCCAATAGGGCGCCTTCG GTGACCCTTAGCGGCTCGTCGTCGCTGTACTTCACGATGATCTC GCGCATCAGCCG	CenH3	D. melanogaster
KR351	TGAGGCACTGCAAGAGGCTGCCGAAGCCTACATTGTTCGCCTCT TGGAGGACGGCAATCTTTGCGCCATCCATGCCAAGCGGGTCACC ATCATGCCTAAG	CenH3	M. brevicollis
KR354	GGCTTCGGCAGCCTCTTGCAGTGCCTCAATGGCCTTGGCTTGCC ACCGGAGCGGGGCAGTGTGATAAAAGTCTTGGGCGATCTCGCGC ATCAGCCGCTGG	CenH3	M. brevicollis
KR356	GAACGCCTCTGCAGCCTCCTGTAGAGCCATTAAAGCATATCCTT GCCACATCATGTGCTCTCGGCTGAACATCTGACAGATCTCGCGC ATCAGCCGCTGG	CenH3	D. rerio
KR360	CGCCGCTTCTTGAAGAGCACTGATGGCGTCAGAACGAATACGGC AGTCGGCGCCAAATGGAGTGGAAGTCTGCATGATTTCTCTGACC AATCTTTGGAAA	CenH3	C. elegans
KR361	GATGGCGTCAGAACGAATACGGCAGTCGGCGCCAAATGGAGTGG AAGTCTGCATGATTTCTCTGACCAATCTTTGGAAACGTGTCACA CTCATGACAACG	CenH3	C. elegans
KR390	ATGCTGGGTMCGCTGCGCGCCATG	HJURP	Hominoids
KR391	ATGGAGAGCGAGGACGTG	HJURP	C. aethiops
KR391	WYGGAGGRCGAGGGCATGG	HJURP	New World Monekys
KR392	CTTTAAGGGCAGAGAAGTCAATC	HJURP	A. geoffroyi
KR393	TCTCTTCTGAGTTGGGCTGTG	HJURP	Primates
KR394	CTCTTCCATCCTGTAAGAYGTG	HJURP	Primates

KR395	GATKCYGGGTCCGCTGCGC	HJURP	Old World Monkeys

Table of Primers 3 of 3

Primer	Sequence	Gene	Species
KR396	CTCTTCCATCCTGTAAGCCATG	HJURP	P. anubis
KR397	ATGGAGAGCGAGGACGTGG	HJURP	Old World Monkeys
KR398	CTTTAAGGGCAGAGAAGTCAACC	HJURP	Primates
KR399	TCTCTTCGGAGTTGGGCTGTG	HJURP	P. abelii
KR400	CTCTTCCGTCCCGTAAGATGTG	HJURP	S. boliviensis
KR55	CCTCCGCGGCGTGCCATGG	CENP-A	Primates
KR55	CCTCYGCGGCGTGYCATGG	CENP-A	Primates
KR56	CCTTAGGTAACYGGCCGCGG	CENP-A	Primates

Table of Strains

Table of Yeast Strains 1 of 2

STRAIN	Genotype
SZY55	SK1 strain; *MATα, ho::hisG, ura3, his3del::KANMX4, zip2del::NAT*
SZY56	SK1 strain; *MATa, ho::hisG, ura3Δ::NATMX4, his3Δ::KANMX4*
KRY100	SK1 strain; *MATa, ho::hisG, ura3Δ::NATMX4, his3Δ::KANMX4*
KRY101	*SK1 Strain; MATα, ho::hisG, ura3Δ::NATMX4, his3Δ::KANMX4*
KRY102	SK1 Strain; *MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4*
KRY110	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Hsapiens-CATD*
KRY112	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Hsapiens-CATD*
KRY115	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Scerevisiae-CATD*
KRY122	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Dmelanogaster-CATD*
KRY123	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Mbrevicollis-CATD*
KRY125	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Mbrevicollis-CATD*
KRY127	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Cneoformans-CATD*
KRY130	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Cneoformans-CATD*
KRY132	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Drerio-CATD*
KRY135	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Drerio-CATD*
KRY136	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Athaliana-CATD*

Table of Yeast Strains 2 of 2

STRAIN	Genotype
KRY142	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Ttermophila-CATD*
KRY145	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Xlaevis-CATD*
KRY146	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Xlaevis-CATD*
KRY148	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Celegans-CATD*
KRY150	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Celegans-CATD*
KRY153	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Xlaevis-loop1/Drerio-αhelix2*
KRY155	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Xlaevis-loop1/Drerio-αhelix2*
KRY157	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Drerio-loop1/Xlaevis-αhelix2*
KRY158	*SK1 Strain; MATa/MATα, ho::hisG/ho::hisG, ura3Δ::NATMX4/ura3Δ::NATMX4, his3Δ::KANMX4/his3Δ::KANMX4,CSE4/cse4Δ::cse4Drerio-loop1/Xlaevis-αhelix2*