Functional consequences of rapid evolution at *Drosophila* centromeres

Benjamin D. Ross

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2014

Reading Committee:

Harmit S. Malik
Celeste Berg
Willie Swanson

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Functional consequences of rapid evolution at *Drosophila* centromeres

Benjamin D. Ross

Chair of Supervisory Committee

Assistant Member Harmit S. Malik

Department of Molecular and Cellular Biology

Fred Hutchinson Cancer Research Center

Howard Hughes Medical Institute

Centromeres are loci on chromosomes that are essential for faithful inheritance of genomic information at every cell division in eukaryotic organisms. Centromeric chromatin is the foundation for one of the largest macromolecular structures in the cell – the kinetochore, which facilitates tension between the chromosome and the spindle pole during the act of chromosome segregation. Despite this vital function, centromeric DNA and genes that encode essential centromeric proteins evolve rapidly in plants and animals. However, the selective forces driving the rapid evolution of centromeres and the functional consequences of such rapid change are not well understood. I aimed to gain insight into the functional consequences of centromere evolution, using *Drosophila* as a model system. I discovered that rapid evolution has resulted in species-specific function of the centromeric histone variant CENP-A/Cid. Furthermore, I found rapid evolution had not only affected the primary sequence of kinetochore proteins in *Drosophila,* but also the composition of the kinetochore itself, since the young gene *Umbrea* gained essential centromere function after duplication. Finally, I found that divergence at *Drosophila* centromeres may have broad consequences for species, since some genes involved in speciation encode rapidly evolving centromeric proteins

Table of contents

List of Figures:

Acknowledgements:
My wife Sarah Stewart

For exceptional intellectual support and inspiration: Josh Bayes, Nitin Phadnis, Mia Levine, Maulik Patel, and other members of the Malik Lab, past and present.

My collaborators: Leah Rosin and Barbara Mellone at the University of Connecticut,
Axel Imhof and Andreas Thomae at the Ludwig Maximillians University Munich, Lara Morrison University of Washington.

List of Abbreviations

CENP: Centromere protein. Commonly used to describe proteins that localize to centromeres by immunofluorescence imaging.

LHR: Lethal Hybrid Rescue. Protein encoded by the *D. melanogaster* gene *Lhr. Lhr* mutants restore viability to male hybrids from crosses between *D. melanogaster* and *D. simulans.*

HMR: Hybrid Male Rescue. Protein encoded by the *D. melanogaster* gene *Hmr. Hmr* mutants restore viability to male hybrids from crosses between *D. melanogaster* and *D. simulans.*

PAML: Phylogenetic Analysis by Maximum Likelihood. Software tool used to compare rates of sequence change of a gene of interest across a phylogeny to different evolutionary models in order to determine if that gene evolves under purifying, neutral, or positive selection.

CID: Centromere Identifier. The name of the *Drosophila* homolog of the gene encoding the centromeric histone H3 variant.

HP1: Heterochromatin Protein 1. A family of conserved non-histone chromosomal proteins that play crucial roles in the maintenance and establishment of heterochromatin, but also influence transcription of euchromatic genes, among other roles.

Section 1: Functional consequences of rapid evolution at *Drosophila* centromeres

The centromere paradox

In the 1800's Gregor Mendel proposed a fundamental concept now taught to all biology students. Mendel believed that two alleles will randomly segregate from each other during reproduction and will be equally represented in the next generation. It is now known that the basic biology underlying Mendel's law of inheritance is the accurate and faithful transmission of genetic material during cell division in meiosis[1]. DNA packaged into chromosomes must be equally segregated between dividing cells to avoid aneuploidy, which can lead to birth defects[2] and cancer in animals[3]. This vital process, also crucial during mitosis, is governed by a locus on each chromosome called the centromere. Centromeres serve as the chromosomal attachment point for spindle fibers during cell division through the recruitment of a macromolecular protein complex called the kinetochore to centromeric DNA. When visualized by electron microscopy, the kinetochore appears to possess a tri-laminar architecture, with three distinct layers – the inner kinetochore proximal to the DNA, the middle kinetochore, and the fibrous corona or outer kinetochore, which directly interacts with spindle microtubules. Many literature reviews cover the basic biology and molecular composition of the kinetochore (see Cheeseman and Desai for a particularly fine review[4]). The crucial function of the centromere and kinetochorehas been conserved throughout hundreds of millions of years of evolution across eukaryotes[5], from fungi to plants and animals.

Rapid evolution of centromeric DNA across taxa

Despite such a high degree of functional conservation across eukaryotes, centromeric DNA varies widely among species [5]. This point can be illustrated by a comparison of just a few of the myriad examples of centromeres found in the literature[6]. The first centromeres characterized at the sequence level were the 'point' centromeres of *Saccharomyces*

cerevisiae[7]. One hundred and twenty-five base pairs (bp) of centromeric DNA was found to be necessary and sufficient to recruit and assemble centromeric chromatic, and the protein components of the budding yeast kinetochore complex. However, budding yeast is an exception, not the rule, even among other fungi. Centromeres in most multicellular organisms are more complex, composed of large AT-rich repetitive sequences. These repetitive sequences, also termed 'satellite' DNA due to their migration properties through cesium chloride gradients, were identified through early cloning and sequencing studies[8-10]. However, the repetitive nature of centromeric DNA presents a challenging problem, even for the most modern sequencing technology and assembly techniques. 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. Primate centromeres are composed of megabases of an AT-rich DNA sequence known as alpha-satellite. Alpha-satellite is a 171-bp 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 [11]. 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[12] 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 [13]. The higher-order array structure appears to be evolutionarily young, found 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 alpha-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 [14]. 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[15]. 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. Pericentric alpha-satellite monomers may therefore represent older centromeric satellites that were replaced by newly arisen variants in the middle of the centromere. In such a 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 pericentromeric region of the human X-chromosome, where the oldest alpha-satellite domains are the furthest from the current centromere [13]. 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 [16]. Surprisingly, centromeric satellite sequences are more divergent between species than are pericentric satellites [13]. 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 appears to be primarily composed of repetitive pentameric sequences interspersed with transposable elements, with eighty-five per cent of the centromeric sequence found to be AATAT and AAGAG satellites with very low sequence variation [9]. While these data come from analysis of a minichromosome, the sequence

composition of centromeric satellites seems to be invariant within species. However, the size of satellite arrays can vary dramatically within members of the same *Drosophila* species [17]**.** Furthermore, centromeric satellites can 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 [18, 19]. Furthermore, some satellite sequences 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 [20]. These observations between closely-related species have been born out by comparison across much broader evolutionary distance (for example, across the *Drosophila* phylogeny), as well as between strains of the same species[21]. The advent of dense genomic sampling of individuals from a single population of *D. melanogaster* opens the door to further test the hypothesis of dynamic intra-species evolution of satellite DNA[22].

These observations of centromeric DNA evolution are perhaps not so surprisingly on their own. This is because centromeres in all species (besides *S. cerevisiae*) are specified through epigenetic mechanisms – in other words, the proteins that bind at the centromere define the centromere, rather than the underlying repetitive sequences[23]. Since specific sequences do not seem to be required for centromere establishment, DNA could drift freely under relaxed selective pressure, subject to non-adaptive evolutionary forces[24]. In tacit support of this idea, some researchers have proposed that repetitive pericentromeric sequences exist as “buffers” for centromeres[25]. Centromeres that “drift” in location would be highly deleterious if moved over an active gene in euchromatin. However, pericentromeric repetitive DNA packaged into heterochromatin is typically devoid of active genes in primates (although this is not true in some plants or in *Drosophila*[26]), and therefore a “drifting” centromere would theoretically have no functional consequence on its chromatin environment.

Given the essential and conserved role of the centromere, and the epigenetic nature of centromere establishment and propagation, it is thus startling to realize that the amino acid sequences of many essential centromere and heterochromatin proteins evolve quite rapidly, even between closely related species of plants, primates, and flies[6, 27]. These proteins are typically encoded by single copy genes that are essential for chromosome segregation in every cell division in mitosis and meiosis. The puzzling disparity between functional conservation and divergence is referred to as the centromere paradox.

Rapid evolution of genes encoding centromeric proteins

Centromeres in most eukaryotes are epigenetically specified by the replacement of one of the four canonical histone proteins, histone H3, with a centromere-specific variant, CenH3/CENP-A, in centromeric nucleosomes. The discovery that the centromeric histone variant CenH3/CENP-A was rapidly evolving was startling. Most genes essential for important cellular processes are highly conserved, yet CENP-A was found to be very rapidly evolving between closely related species of *Drosophila[28]*. This result prompted investigation into the evolutionary trajectories of CENP-A in other lineages, as well as analysis of the few other bona fide centromeric proteins that have been identified and characterized. CENP-A rapid evolution was observed repeatedly in other lineages[29, 30], and appears to be a nearly universal trait in most organisms (*Tetrahymena* are a notable exception[31]). Furthermore, the important centromeric protein CENP-C also evolvea rapidly, in multiple lineages[30, 32]. CENP-A and CENP-A are both members of the inner kinetochore that forms the DNA-proximal surface to which other parts of the kinetochore assemble. The macromolecular complex of the kinetochore has traditionally been delineated by its appearance under electron microscopy, which revealed the presence of a trilaminar structure[4]. The inner kinetochore forms the base layer, making contact with the DNA and centromeric chromatin, while the outer kinetochore makes contact with the spindle

microtubules. The middle or central kinetochore is less distinct and occupies the region between the inner and outer layers. The fibrous corona extends outward from the outer kinetochore and is thought to contain regulatory and checkpoint proteins.

To characterize the extent to which rapid evolution has shaped the proteins at each architectural level of the kinetochore, Kevin Roach, a former graduate student in the Malik lab, analyzed the rate of evolution for a set of inner, middle, and outer kinetochore proteins encoded by primate genes using maximum likelihood methods implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) software package[33, 34]. He found kinetochore proteins that were rapidly evolving in each of the three laminar layers of the kinetochore. However, there was a statistically significant enrichment of rapid evolution at the inner kinetochore (Kevin Roach thesis, Janet Young and Harmit Malik personal communication), compared to the outer kinetochore and fibrous corona, although the feature of rapid evolution was not confined to the inner kinetochore. This asymmetrical distribution of evolutionary rates across kinetochore proteins provides strong evidence that the forces driving rapid evolution act at the interface with the centromeric DNA.

To gain further insight into the patterns of rapid evolution at the centromere, I curated a list of human genes encoding components of the inner (centromeric DNA proximal proteins, including the Constitutive Centromere Associated Network, or CCAN), the middle kinetochore, and the outer kinetochore (including components of the spindle checkpoint signaling apparatus), from recent reviews and primary literature[4]. In addition, this list included recently identified novel proteins that localize to the kinetochore, but have yet to be characterized with a defined function. These novel proteins were isolated through mass spectrometry analysis of mitotic chromosomes in chicken cells, and include a variety of protein domains including some with potential chromatin or DNA-associated functions[35].

Janet Young, a current staff scientist in the Malik Lab, was able to recapitulate some of Kevin's results, with my larger and more powerful dataset of genes from a greater number of primate species. Of note, inner kinetochore proteins known to bind to centromeric DNA (for example, CENP-T[36], or CENP-N[37]) or be involved in the CENP-A deposition or specification pathway were found to be particularly rapidly evolving. These results suggest that functional insight could be derived from evolutionary signatures for uncharacterized proteins that cytologically localize to centromeres. Many of the nearly one hundred kinetochore proteins identified in vertebrates lack detailed molecular characterization. Evolutionary analysis may be one way by which proteins of the kinetochore can be classified. To gain further detail into how kinetochore components co-evolve with one another, we are collaborating with the lab of Nathan Clark at the University of Pittsburgh to further characterize the architecture of evolution at the primate using evolutionary rate covariation, or ERC [38, 39]. This method exploits the observation that, over a phylogeny, the rate of evolution of a given protein will not be constant, but instead will vary to a certain degree. While most proteins will evolve independently with respect to each other, covariance of one or more proteins along the same branches of an evolutionary tree can indicate functional relatedness, including function in the same genetic pathway, function in the same molecular complex, or even direct physical interaction[38-40]. With this collaborative approach, we hoped to assemble a comprehensive view of the architecture of rapid evolution and coevolution at kinetochores.

Genetic conflict explains patterns of rapid evolution

Why would key constituents of the vital process of chromosome segregation evolve rapidly? Neutral models of drift are not sufficient to account for the evolutionary patterns observed for centromeric DNA elements and genes encoding centromeric proteins, since most mutations in essential centromeric components seem likely to be immediately deleterious and therefore

selected against[5]. What, then, can explain the centromere paradox? One selective pressure that can drive rapid evolution is genetic conflict, often found between two genetic elements with opposing interests. One common example of genetic conflict occurs between pathogens and their hosts, and often results in rapid evolution in factors that interact to promote the best interest of one party or the other. For instance, host anti-viral restriction factors typically bind to pathogen proteins to effect restriction. These restriction factors often evolve quite rapidly as they "chase" rapidly mutating viruses, or they evolve due to selective pressure to avoid blockade by a virus-encoded antagonist[41]. Another flavor of genetic conflict can come between elements that reside in the same genome (as opposed to elements encoded by opposing genomes), for example, between transposons and host machinery dedicated to restrict their transposition. Again, these restriction factors often evolve quite rapidly under positive selection, perhaps due to similar selective pressures as those faced by anti-viral factors (Richard McLaughlin and Antoine Molaro, Malik Lab, personal communication).

In contrast to these examples, the faithful segregation of chromosomes seems to offer no opportunity for exploitation by selfish elements. Indeed, in mitosis, chromosomes are equally distributed between daughter cells. Elaborate cellular checkpoints act to halt mitosis upon the sensing of tension imbalances that could lead to unequal segregation. In contrast, while male meiosis results in equal partitioning of chromosomes into sperm, selfish elements can act to alter Mendelian segregation in order to increase their own transmission to the next generation. These selfish elements can 'poison' either gametic development or embryonic viability, ensuring their own evolutionary success at the expense of other chromosomes. This phenomenon has been coined "meiotic drive"[42, 43]. Such post-meiotic dysfunction is seen in the Segregation Distorter system of *Drosophila*, the t-haplotype of mice, and the spore-killers of fungi [1,2]. Perhaps the best known system is the post-meiotic Segregation Distortion meiotic drive system in Drosophila, in which spermatids bearing "sensitive" chromosomes are killed while their

“insensitive” brethren survive[44], resulting in dramatic imbalances in the propagation of these chromosomes in the population.

A second violation of Mendelian inheritance occurs when selfish elements subvert the process of chromosome segregation. Mendelian inheritance results when both homologous chromosomes are represented in the progeny at approximately the same frequency. However, if a selfish element is able to skew the process of chromosome selection in its own favor, this results in biased inheritance known as meiotic drive [43]. The female version of meiosis is inherently asymmetric, revealing a unique opportunity for exploitation by selfish elements. Unlike in males, the products of the female meiotic divisions are not all passed on to the next generation. Indeed, only one of four products is left in the egg – the others are discarded into structures called polar bodies. In theory, each meiotic product would have an equal probability of being passed on to the next generation in the egg, as Mendelian segregation predicts. However, several lines of evidence suggest that this is in fact not the case – some chromosomes can behave selfishly to increase their probability of transmission to the egg, at the expense of their homologs[45-48]. While female meiotic drive as an evolutionary force was first proposed over 60 years ago[43], the molecular mechanisms and evolutionary consequences of drive remain mysterious. Insight into the molecular basis of female meiotic drive has been gained from detailed study of the few systems known. Cheating in female meiosis can occur in either of two meiotic divisions. One of the best-studied examples of selfish elements that subvert female meiosis emerged from the study of knob elements in maize chromosomes. Pioneering work from Marcus Rhoades revealed that in appropriate genetic backgrounds, knob elements can recruit microtubules [49]and direct their orientation in meiosis II, thereby increasing their chances of inclusion in the oocyte nucleus [50].

Centromere-mediated female meiotic drive

Cheating behavior can also occur during the first meiotic division of female meiosis, in which the position of a chromosome on the meiotic spindle determines its fate. Meiotic chromosomes on the cortical side of the spindle are fated to end up in the polar bodies, while interior chromosomes will be incorporated into the egg. Therefore, only interior chromosomes can be considered to be evolutionarily successful. When meiotic inheritance is Mendelian, the frequency at which a chromosome finds itself on the internal side of the meiosis I spindle is expected to be random. Indeed, this is often the case. In contrast, non-random positioning of a chromosome on the meiotic spindle would result in meiotic drive. Such bias is thought to explain the higher transmission of B chromosomes in grasshoppers and centromeric DNA expansions in monkeyflower[46, 47]. It is these cases of selfish behavior by centromeres during female meiosis that interested me the most as I began my thesis work.

Recent work on Robertsonian fusion chromosomes has provided startling insight into the cell biological mechanisms and selective forces underlying centromere evolution, revolutionizing our understanding of female meiotic drive. Indeed, these studies demand deeper description, since they are among the first to provide experimental support for much of the theory and predictions attempting to explain centromere evolution.

Robertsonian fusion chromosomes provide deeper insight into female meiotic drive

Telocentric chromosomes (with the centromere found at the chromosome end) can fuse to become metacentric (with the centromere in the middle). These fusion chromosomes (called Robertsonian fusions, or Rb) are found in many species and can exhibit biased transmission (meiotic drive) during female but not male meiosis[46, 51]. Curiously, the fused chromosome appears to be favorably transmitted in species like chickens and humans, yet appears to be disfavored in mice [52]. Furthermore, karyotypic surveys of mammalian species have found

non-random distributions of metacentric and telocentric chromosomes, suggesting that either one or the other are favored in separate species, and this preference has likely switched multiple times during mammalian evolution [52]. This implies that a 'mixed' karyotype consisting of both metacentric and telocentric chromosomes must be disfavored. From these studies, a case has emerged to support a role for alterations in centromere structure dictating success during female meiosis. Yet the cell biological evidence for this idea has been sparse, in part because laboratory models for studying chromosome segregation possess "homogenous" centromeres instead of mixed karyotypes. In addition, there has been no clear theoretical model for why metacentrics are favored in some lineages while telocentrics are favored in others.

Recent work leveraged a genetic and cell biological model system for Robertsonian chromosomes in mice to make key insights into how Rb transmission violates Mendelian inheritance in female meiosis[53, 54]. In mice whose telocentric chromosomes were transmitted at a higher frequency than Rb fusions, internal positioning on the meiotic spindle for the Rb or the telocentric chromosomes was found to be highly non-random, strongly correlating with inclusion in the egg. Antibody staining of meiotic chromosomes revealed that meiotic success was perfectly correlated with an increased abundance of both inner and outer kinetochore proteins. These analyses revealed that increased kinetochore protein levels at centromeres were directly associated with meiotic drive. Furthermore, in a different genetic background in which Rb fusion chromosomes were favored in female meiosis over telocentric, the opposite result held – Rb centromeres stained more intensely for both inner and outer kinetochore proteins. In both instances, differences in kinetochore protein recruitment were an intrinsic property of the centromeres themselves rather than reflecting different expression levels of kinetochore proteins. These findings suggest a model in which the evolutionary success of Rb chromosomes depends on its relative centromere strength compared to other chromosomes. If the genetic background contains largely 'weak' centromeres, then an Rb

chromosome is likely to be more successful. In contrast, the evolutionary success of an Rb chromosome would be diminished if it initially arose in a genetic background of 'strong' centromeres. Indeed, Rb fusions isolated from different populations of mice appear to be highly variable in their relative centromere strength[53, 54]. One plausible explanation for this diversity is that it is genetic variation — each Rb fusion possesses differential retention of the centromeric DNA from the original telocentric chromosome[55], thus influencing centromeric strength. Centromeres that can drive in female meiosis can therefore recruit more inner and outer kinetochore proteins than their "loser" homologs.

Why don't driving centromeres fix in populations? The frequency of Robertsonian fusions in different populations of wild mice in Europe can vary widely, with some populations found to exhibit a dramatic decrease in chromosome number due to fusions, while others lack fusions altogether[56]. In other geographical areas, Rb heterozygotes only appear in hybrid zones where individuals between two separate populations interbreed [56, 57]. If meiotic drive of Robertsonian appears so rampant, why have fusions not fixed across populations? One possibility is that heterozygosity for driving Robertsonians incurs fitness costs. Indeed, this is observed for male mice heterozygous for some Robertsonians [58], which have reduced fertility. Further support comes from monkeyflowers [47], where male plants bearing driving chromosomes exhibit decreased pollen viability, and from human data which indicate that male carriers of Robertsonian fusions are infertile [46, 51]. Why might this be so? It has been proposed that male meiosis is much less tolerant of errors than female meiosis[59]. Therefore, female meiotic drive could elicit costs in males, resulting in strong selection for suppressors or modifiers. Divergence in such modifiers between different populations could result in reproductive isolation. In this model, meiotic drive and its associated fitness costs are suppressed within populations but revealed in heterozygous hybrids[60, 61].

The centromere drive hypothesis

One explanation that unifies all of these disparate observations regarding centromere evolution is now broadly known as the centromere drive hypothesis[60, 61] (**Figure 1.1)**, which explains centromere evolution in the framework of genetic conflict. In this model, recurring and lineage-specific genetic conflict between selfish DNA elements and the kinetochore proteins that bind them drives the pervasive evolution of both. Centromeric DNA expansions that can recruit more centromere proteins can result in a stronger centromere able to drive against its homolog in meiosis I. These DNA expansions will be strongly selected and increase in frequency in populations, regardless of any associated fitness costs[62]. However, meiotic drive that reduces organismal fitness (for example, reducing male fertility as in the example of Robertsonian fusions[46], and in monkeyflowers[47]) will result in strong selection for alleles of genes that can act as suppressors of drive. The most likely candidates for female meiotic drive suppressors are centromeric and kinetochore proteins that physically bind to DNA, or heterochromatin proteins that normally flank the centromere region. Therefore, under the centromere drive hypothesis, centromeric proteins face two opposing evolutionary forces. On one hand, centromeric proteins face strong evolutionary constraint to facilitate error-free chromosome segregation during the countless rounds of cell divisions in organismal development. On the other hand, centromeric proteins are uniquely positioned to act as suppressors of centromere drive, and therefore evolve rapidly. While one strategy may be to evolve unique and discrete machinery for separate process of cell division (male or female meiosis, and mitosis) as seen for some components in male *Drosophila* (which lack synapsis and recombination[63]) and in *C. elegans* (which possesses as yet poorly understood differences in male meiotic kinetochore structure[64]), centromere protein evolution appears to be nearly universal in eukaryotes. Investigation into the selective pressure(s) and mechanism driving such unexpected evolution is certain to yield fundamental insights into basic biology.

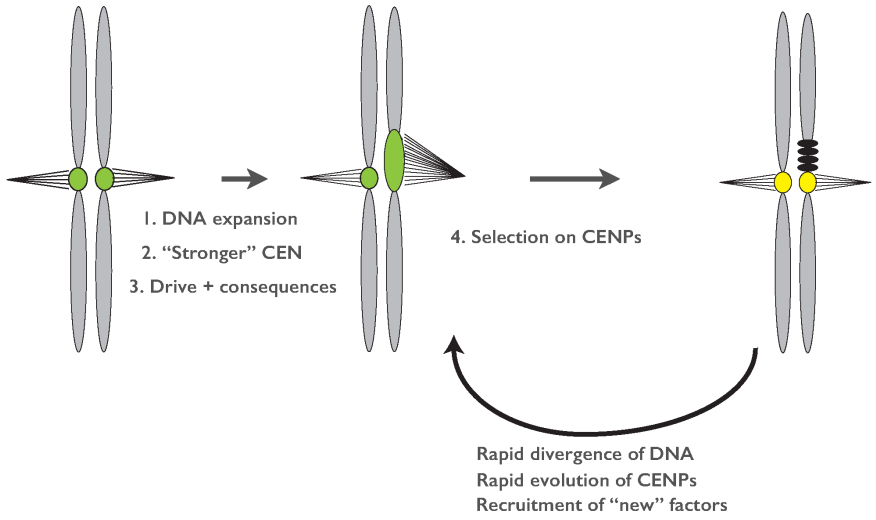


FIGURE 1.1 – The centromere drive hypothesis provides a stepwise model that explains the paradoxical observations of rapid evolution of centromeric DNA and centromeric proteins. This model posits that (1) centromeric DNA undergoes an expansion on one of the two homologous chromosomes (sister chromatids are not shown) during replication prior to the female meiotic divisions. (2) This expansion could recruit more centromeric proteins like CENP-A, more kinetochore proteins and greater binding by spindle microtubules, resulting in a "stronger" centromere (similar to that seen in Robertson fusion chromosome) that is able to drive to skew Mendelian ratios and increase in frequency in populations (3). This may incur fitness costs that lead to strong pressure to select variant alleles of centromere proteins that can suppress meiotic drive (4). Suppression could come in various forms, including alterations of DNA binding specificity that avoid the driving DNA expansion. Over evolutionary time, recurrent episodes of centromere drive and drive suppression would result in rapid evolution of centromeric DNA and centromeric proteins, and turnover in centromeric proteins.

In my dissertation work in the lab of Dr. Harmit Malik, I sought to gain insight into the functional consequences of centromere evolution through 3 separate lines of investigation addressing fundamental unanswered questions in the field. Taken together, my thesis work has provided a synergistic view of centromere divergence. **In Section 2**, I investigated how rapid evolution has altered the species-specific function of one essential member of the centromere, CENP-A/Cid in *Drosophila*. In **Section 3**, I dissected the molecular and evolutionary steps that a young duplicate gene called *Umbrea* underwent during a transition from functional redundancy to being required at the centromere for accurate chromosome segregation[65]. And finally**,** in **Section 4,** I investigated how functional divergence in the centromeric protein Lhr may have contributed to post-zygotic reproductive isolation during speciation between two species of *Drosophila.*

Section 2: Rapid divergence in function in the centromeric histone Cid in *Drosophila*

Many biologists consider sequence conservation of a gene to be synonymous with functional importance. Indeed, at some level, this rationale is at the heart of the use of model organisms to gain insight into human biology and physiology. For some genes, lack of sequence conservation can mean that there is a lack of selective pressure to maintain gene function. This is perhaps most easily observed after gene duplication, when one copy experiences pseudogenization (loss of function due to accumulation of mutations that are not selected against) due to functional redundancy. However, in a small class of genes, selective pressure can favor variants that alter the amino acid sequence, resulting in a lack of conservation with functional implications. During my thesis work in the Malik lab, I became very interested in centromere factors that exhibit such signatures of rapid evolution.

The centromeric histone variant specifies centromeres in many eukaryotes

Perhaps the most well-known centromeric protein is the histone variant CENP-A, also known as CenH3, or Cid (Centromere Identifier) in *Drosophila*. This molecule is the most upstream factor in the centromere specification pathway, through replacement of canonical histone H3 in centromeric nucleosomes[66-68]. Like other histones, CENP-A possesses an N-terminal tail domain and a core histone fold domain that mediates interactions with other core histone molecules in the nucleosome[66]. CENP-A function at the molecular level is still the object of heavy active research across many labs and multiple model systems. For example, while CENP-A clearly replaces H3 in centromeric nucleosomes, the precise composition of the other histone proteins, how DNA is wrapped around the centromeric nucleosome, and the mechanism for centromeric histone deposition are all areas of active and somewhat controversial research[23, 69, 70]. What is known is that CENP-A-containing nucleosomes provide one vital piece of the foundation for the establishment of the microtubule-binding kinetochore structure.

This essential activity occurs through protein interactions with other components of the inner centromere like CENP-C, which binds to a conserved hydrophobic motif at the extreme C-terminus of the histone fold domain[71], and is subsequently required for the function of other inner and middle kinetochore proteins.

CENP-A is both necessary and sufficient for centromere function

Multiple lines of evidence support the idea that centromeres are epigenetically defined by the presence of CENP-A[72-74], and not by intrinsic DNA-sequence specificity by CENP-A or other components of centromeric chromatin. For example, CENP-A (Cid, for Centromere Identifier in *Drosophila)* artificially tethered to euchromatic site on a chromosome arm was sufficient to recruit other kinetochore components that were propagated over subsequent cell cycles, and could direct accurate segregation during mitosis following laser-severing of the chromosome distal to the endogenous centromere[72]. Natural neocentromeres provide another example of the importance of CENP-A/Cid in the centromere specification pathway. These are sites (typically found on the chromosome arms) that were not centromeric in previous cell cycles and do not share the sequence characteristics of centromeric repetitive DNA. However, incorporation of CENP-A/Cid at these sites is sufficient to drive epigenetic establishment of centromere function through recruitment of the kinetochore[75-78]. Furthermore, overexpression of Cid in *Drosophila* cells and *in vivo* is sufficient to drive incorporation into euchromatic chromosome arms. This gain of function causes the formation of ectopic centromeres that recruit kinetochore proteins and results in catastrophic mitoses with multiple rounds of chromosome breakage[76, 79].

CENP-A is also necessary for centromere function. Ablation of CENP-A through genic mutations or RNAi causes lethality in *Drosophila* following depletion of maternally-deposited protein and mRNA during development[73, 80]. As expected given the role of CENP-A in

centromere and kinetochore function, lethality appears to be caused by the failure of cell proliferation due to chromosome missegregation during cell division. These phenotypes are similarly observed in yeast, mice, and in human cells, where removal of CENP-A function causes chromosome segregation defects, cell death, and lethality[74, 81-83].

Rapid evolution of CENP-A

In spite of this crucial role, CENP-A evolves rapidly across taxa and even between closely related species, and this rapid evolution is an example of the centromere paradox[28-30, 84]. For example, *Cid* in *Drosophila* has been shown to evolve rapidly between closely related species of the *melanogaster* species subgroup[28], by McDonald-Kreitman analysis, which detects an enrichment of non-synonymous amino acid changes between species, based on a background level of polymorphisms between species. More broadly, the N-terminal tail domain of *Cid* (and *CENP-A* genes in eukaryotes in general) evolves rapidly at both the sequence level and in terms of overall length[5]. In fact, the N-terminal tail is unalignable across even modest stretches of evolutionary time due to length differences. Sequence changes in the N-terminal tail may involve DNA-binding motifs, although this hypothesis lacks much empirical support[84]. *CENP-A* also evolves rapidly in primates and plants (as well as many other taxa) with some of the same characteristics of rapid evolution observed in *Drosophila[30, 32].*

Functional consequences of CENP-A rapid evolution

There are two countervailing hypotheses to explain the rapid evolution of *CENP-A*. One hypothesis posits that centromeric satellites evolve rapidly due to non-adaptive processes such as evolutionary drift, and *CENP-A* coevolves to preserve interactions with centromeric DNA over evolutionary time. While this hypothesis should be rigorously considered, it does not provide an explanation for why CENP-A would be under selective pressure to maintain binding to DNA sequences, when sequence itself does not define centromere localization. The opposing

viewpoint is that variant alleles of *CENP-A* that can suppress the deleterious consequences of centromere drive by centromeric satellites are favored during evolution, resulting in an increased rate of non-synonymous mutation relative to the background rate of mutation. Yet this hypothesis has little or no empirical support since functional analysis of the role of rapid evolution in *CENP-A* has been limited. Early cytological experiments found that, while *Cid* evolved rapidly between two closely related species *Drosophila melanogaster* and *D. simulans*, such rapid evolution did not impede the ability of a $Cid^{simulans}$-GFP fusion protein to localize to centromeres when overexpressed in *D. melanogaster* cultured cells[85]. However, a domain-swap of the rapidly-evolving Loop1 region of *Cid* between *D. melanogaster* and the more distantly-related species *D. bipectinata* found that the function of the Loop1 domain had diverged enough that centromere localization was compromised. This result implied that rapid evolution had altered the function of *Cid* between these species, at least at this gross level. Genetic experiments to assess the functional conservation of *CENP-A* between species have used exceeding broad evolutionary divergence, such that the effects of adaptive evolution may be masked. For example, Wieland and colleagues found that *CENP-A* from *S. cerevisiae* could functionally complement the cellular lethality phenotype of *CENP-A* mutations in human cells[86], concluding that functional and structural features of CENP-A are strictly conserved over evolution. In contrast to this study, researchers in plants have found that evolutionary sequence divergence has indeed altered function of *CENP-A*. In cultured cells of the tobacco plant, orthologous CENP-A GFP fusion proteins from other plant species were found to be unable to localize to endogenous centromeres[87], indicating that centromeric localization by CENP-A has species-specific attributes. These conclusions were greatly strengthened by genetic experiments in *Arabidopsis thaliana*, where *CENP-A* null mutants were rescued by the expression of either native CENP-A protein or CENP-A from other plant species, including *Brassica rapa*, spanning ~20 million years of divergence. These experiments found that evolutionary divergence resulted in comprised genetic rescue[88]. Interestingly, *B. rapa* CENP-

A could localize to *A. thaliana* centromeres, yet could not rescue *A. thaliana CENP-A* null mutants, suggesting that evolutionary divergence had resulted in species-specific functionality. In addition, a chimeric CENP-A fusion protein containing sequence from maize CENP-A revealed profound defects in the meiotic divisions, also in *A. thaliana* [89]. Finally, experiments in which CenH3 swaps were performed between different species of fungi revealed that the ability to complement *CenH3* mutations was correlated with phylogenetic distance[90]. Indeed, rapid evolution across fungi (albeit across greater than 300 million years of divergence between Archaeascomycetes, Hemiascomycetes, and Euascomycetes) seems to mirror that of rapid evolution seen in lineages like Drosophila – changes are enriched in the Loop1 region and the N-terminal tail[90].

From these studies, it seems that rapid divergence of CENP-A may impart species-specific function for the process of chromosome segregation, yet this conclusion appears to depend on the model system in which the analysis is performed. However, a strict hypothesis-driven approach to understanding the consequences of rapid evolution on CENP-A function has been lacking. I proposed that, in *Drosophila*, centromere-mediated female meiotic drive could have deleterious species-specific phenotypic consequences that impose a strong selective pressure for suppression[46, 47, 60], including meiotic defects affecting fertility and male sterility, and mitotic errors.

Evolutionary complementation of a homozygous-lethal mutation in *Cid* by orthologs

I aimed to test the effect of rapid evolution on Cid function by performing species-swaps into *D. melanogaster* (**Figure 2.1)**. In *Drosophila*, species-swaps have been used in other systems to understand the consequences of evolution in vital molecular pathways [91], revealing insight into the selection forces driving species-specific function. Evolutionary complementation relies on the ability to knock down the expression or function of the gene of interest in the model

system of choice, and attempt to functionally complement that gene by co-expressing transgenes derived from the orthologous gene from other species. In *D. melanogaster*, missense point mutations that disrupt *Cid* coding sequence are homozygous lethal during embryogenesis, and are also lethal in *trans* with each other[73, 80]. However, these mutations were generated through mutagenesis screening, and therefore contain other "background" mutations in their genome. I found that these mutations failed to complement a P-element insertion into the *Cid* gene (Cid^{GD4436}), suggesting that this insertion also ablates *Cid* function, yet is likely to have a "cleaner" genetic background, although ectopic P-element activity may also generate unwanted off-target mutations. This system seemed to be ideal for cross-species complementation experiments (**Figure 2.1, 2.2)**.

To generate transgenes for complementation, I recoded the *Cid* genes from *D. melanogaster*, *D. simulans,* and *D. yakuba* by changing synonymous codon positions. This recoding accomplished two things: First, it allowed me to change the codon bias of the *Cid* orthologs to appropriately favored codons in *D. melanogaster.* This was important because codon bias evolves rapidly across species of *Drosophila* [92, 93], and could result in differential expression or lack of expression of my transgenes in *D. melanogaster.* Second, it rendered these genes recalcitrant to RNAi, which requires 19bp or more homology for silencing. This was essentially a back-up strategy, to use if rescue of genic mutations in *Cid* was unsuccessful. In addition to recoding extant genes from closely related species, I used parsimony to reconstruct the most likely ancestral sequence of *Cid* from the common ancestor of *D. melanogaster* and *D. simulans*, and the common ancestor of *D. melanogaster* and *D. yakuba.* I then cloned these recoded genes into pattB vectors with a common promoter and 3' UTR derived from the *D. melanogaster Cid* gene. This strategy would endow each transgene with identical regulatory control. Since each transgene shared the same codon profile optimized for *D. melanogaster* expression, any differences in the ability of each transgene to rescue Cid^{GD4436} could be

attributed to differences in amino acid sequence (**Figure 2.1)**.

Generation of transgenic animals was performed by injection of pattB vectors into a strain of *D. melanogaster* expressing phiC31 recombinase in the female germline by The Best Gene (Chino Hills, CA). This strain (Bloomington Drosophila Stock Center #8622) possessed an attB recombination site referred to as a "landing pad" on the 3rd chromosome. Upon receipt of transgenic lines, I proceeded to generate "tester" rescue lines, by crossing the 3rd chromosome bearing my *Cid* rescue transgenes into the *Cid*^{GD4436} line **(Figure 2.1)**. To assay for the ability of each *Cid* rescue transgene to complement *Cid*^{GD4436}, I crossed balanced *Cid*^{GD4436} heterozygotes together and counted the number of progeny bearing or lacking balancer chromosomes **(Figure 2.1)**. A comparison of non-balancer flies to expected Mendelian ratios was used to determine degree of rescue.

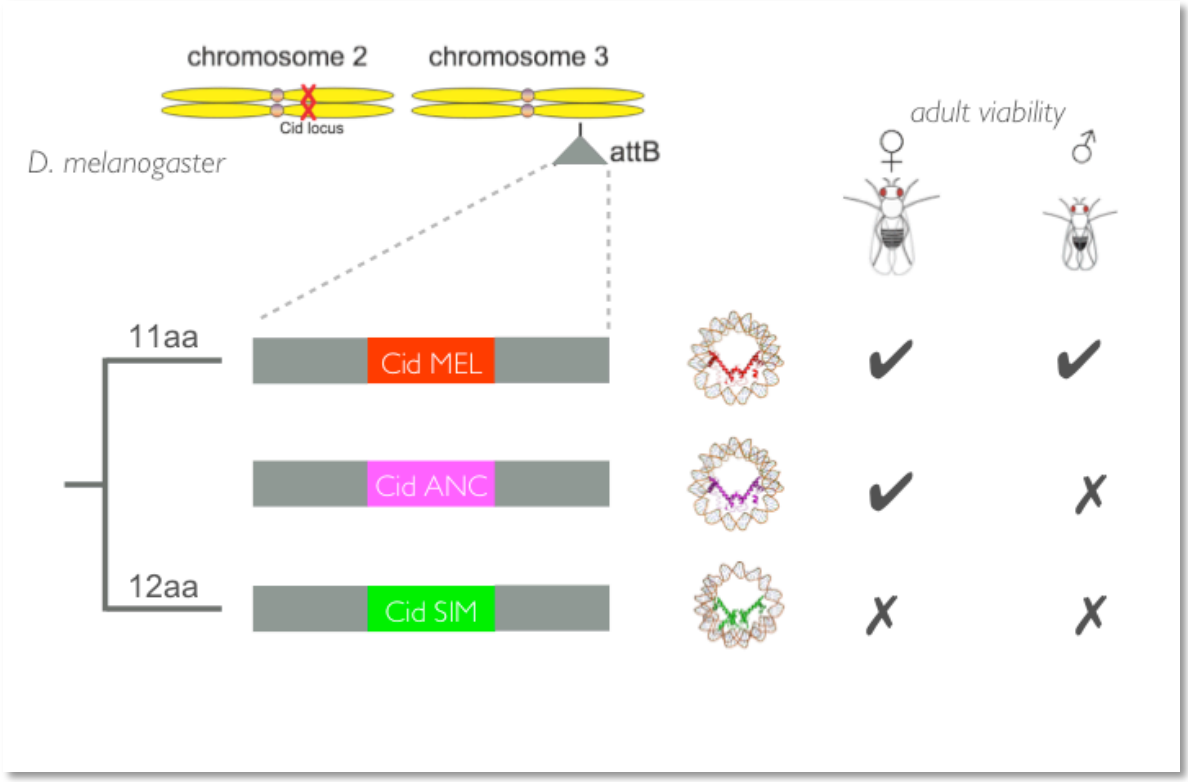


Figure 2.1 – Summary of transgenic rescue scheme for Cid mutations, and results from *Cid*^{melanogaster}, *Cid*^{simulans}, and *Cid*^{MSA} rescue of *Cid*^{GD4436}. Transgenic *Cid* rescue alleles (common 5' and 3' regulatory sequences in gray) were integrated on the 3rd chromosome and crossed to balanced *Cid*^{GD4436} mutants. Amino acid differences from ancestor shown along branches leading to *Cid*^{melanogaster} and *Cid*^{simulans}. Adult viability in *Cid*^{GD4436} homozygous adults was calculated by expected Mendelian frequency of balanced verus non-balanced flies. While *Cid*^{melanogaster} rescued full viability of both sexes, *Cid*^{simulans} largely failed to rescue, and *Cid*^{MSA} specifically failed to rescue males. See Figure 2.2 for rescue data.

Species-specific mitotic function of Cid

Cid^{melanogaster} appeared to rescue *Cid*^{GD4436} to full viability, based on expected Mendelian ratios from segregating balancer chromosomes with dominant visible phenotypes (curly wings) **(Figure 2.2)**. Furthermore, the sex-ratio of Cid^{melanogaster}-rescued flies appeared to be exactly 50:50 male-female, as expected. Startlingly, Cid^{simulans}-rescued flies appeared greatly underrepresented compared to expected Mendelian ratios, suggesting that Cid^{simulans} was not sufficient to rescue homozygous *Cid*^{GD4436} and that the function of *Cid* had been altered between *D. melanogaster* and *D. simulans* in just 2.5 million years of evolution **(Figure 2.2)**.

Indeed, Cid^{simulans}-rescued flies were nearly completely homozygous lethal, with both males and females equally affected **(Figure 2.2)**. This result was surprising, since Cid^{simulans} localizes appropriately to the centromeres of *D. melanogaster* cultured cells[85]. However, some male and female flies escaped lethality, surviving to adulthood. I found that female "escaper" flies exhibited fertility defects, while male flies did not (although the crude assay I did in hindsight was unlikely to reveal differences – it would have been much more sensitive to utilize the protocols of Hiraizumi or Wu, who used sperm exhaustion or sperm displacement experiments to obtain sensitive quantitative data for male fertility[94, 95]). For example, Cid^{simulans} homozygous females crossed to wild-type (Canton S) males produced fewer progeny than wild-type (Canton S) females or heterozygous females crossed to wild-type (Canton S) males. I predicted that these fertility defects could be the result of sex chromosome specific non-disjunction, which could lead to a skewed sex-ratio in the progeny of these animals – however, the progeny sex-ratio of rescued Cid^{simulans} homozygous females was 50:50.

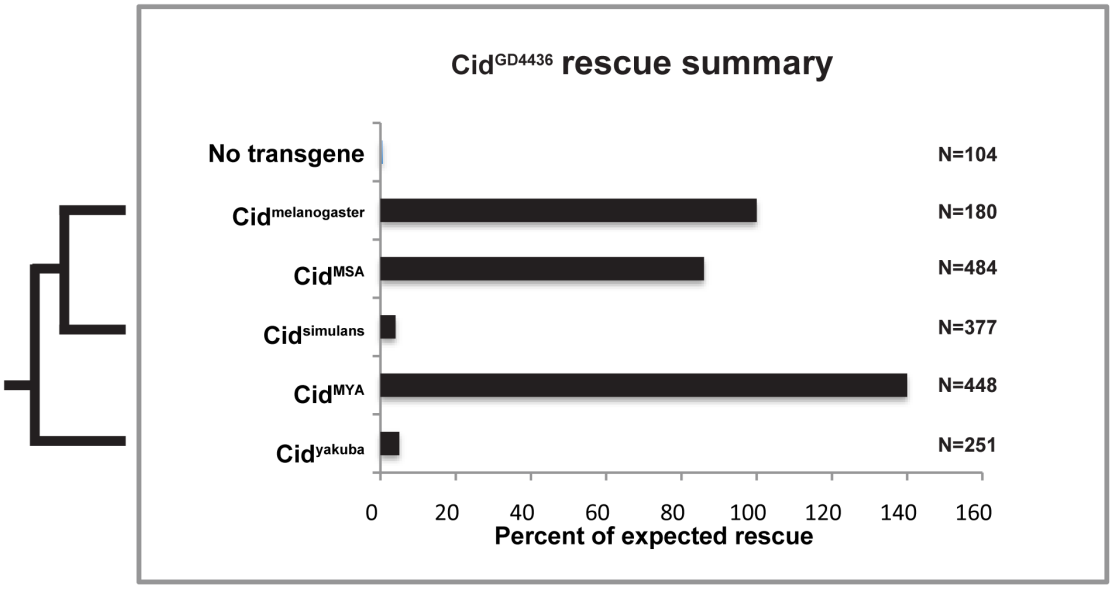


Figure 2.2 – Rescue of *Cid*^{GD4436} reveals dramatic differences in ability to complement *Cid* orthologs or ancestral alleles. Percent rescue was calculated by quantification of a co-segregating balancer chromosome based on expected Mendelian ratios. Failure to complement *Cid*^{GD4436} appears to be an independently derived trait in both the *D. simulans* and *D. yakuba Cid* orthologs. MSA = Melanogaster-Simulans-Ancestor. MYA = Melanogaster-Yakuba-Ancestor.

Cid^{simulans} cytology of lethal embryos

I found that by mating Cid^{simulans} males and females, I could generate a stable stock. This stock produced less than wild-type numbers of progeny each generation (roughly 50% less progeny, data not shown), suggesting that genetic rescue was stochastic. To understand the basis for Cid^{simulans} lethality, I stained embryos laid by homozygous Cid^{simulans} mothers crossed to homozygous Cid^{simulans} fathers, or control embryos laid by w1118 mothers crossed to homozygous Cid^{simulans} fathers. By staining with DAPI, I could visualize nuclei during the rapid mitotic divsions of the early embryo. While control cells exhibited uniformly spaced nuclei across the cortex of the syncytial embryo, Cid^{simulans} embryos displayed a propensity for disrupted nuclei. I found gaps in the cortical distribution of nuclei in Cid^{simulans} embryos **(Figure 2.3),** with an abundance of DNA in the interior syncytium of the embryo. This phenomenon is referred to as "nuclear fallout", and is indicative of an active (yet poorly understood) checkpoint mechanism active in the early embryo, whereby nuclei undergoing delayed or defective mitosis are displaced or discarded into the interior. I looked for embryos at the one-cell stage to understand if the defects observed in Cid^{simulans} embryos were manifest at the first mitotic division. Anti-phospho histone H3 antibodies mark mitotic chromosomes in metaphase or anaphase. While control embryos consistently exhibited proper metaphase alignment of chromosomes at the first mitotic division, Cid^{simulans} embryos often displayed lagging chromatin at anaphase that retained anti-phospho H3 staining **(Figure 2.4).** Later embryos exhibited the same phenomenon, yet with a greater abundance of chromosome fragments in and around the normal metaphase or anaphase configurations. These data indicate that Cid^{simulans} complementation causes mitotic defects manifest at the first mitotic division that might be the basis for lethality. Specific anti-Cid antibodies that recognize Cid^{simulans} as well as Cid^{melanogaster} will be very useful for dissecting the mechanistic basis for cortical fallout and lagging chromosomes during early mitotic divisions (**Figure 2.5)**. For example, Cid^{simulans} may not recognize a specific chromosome of *D. melanogaster.* Indeed, deletion or birth of large arrays of

satellite DNA has occurred between the *D. melanogaster* and *D. simulans* lineages[18-20, 96, 97], implying that the rescue phenotype that I observe may be due to species-specific features of centromeres that diverged following speciation. In this context, if $Cid^{simulans}$ lacks the ability to bind to centromeric DNA that it had not co-evolved with during its private evolutionary history, it would not be able to form a centromeric domain on one chromosome, and this may result in failure to properly segregate chromosomes during early mitosis.

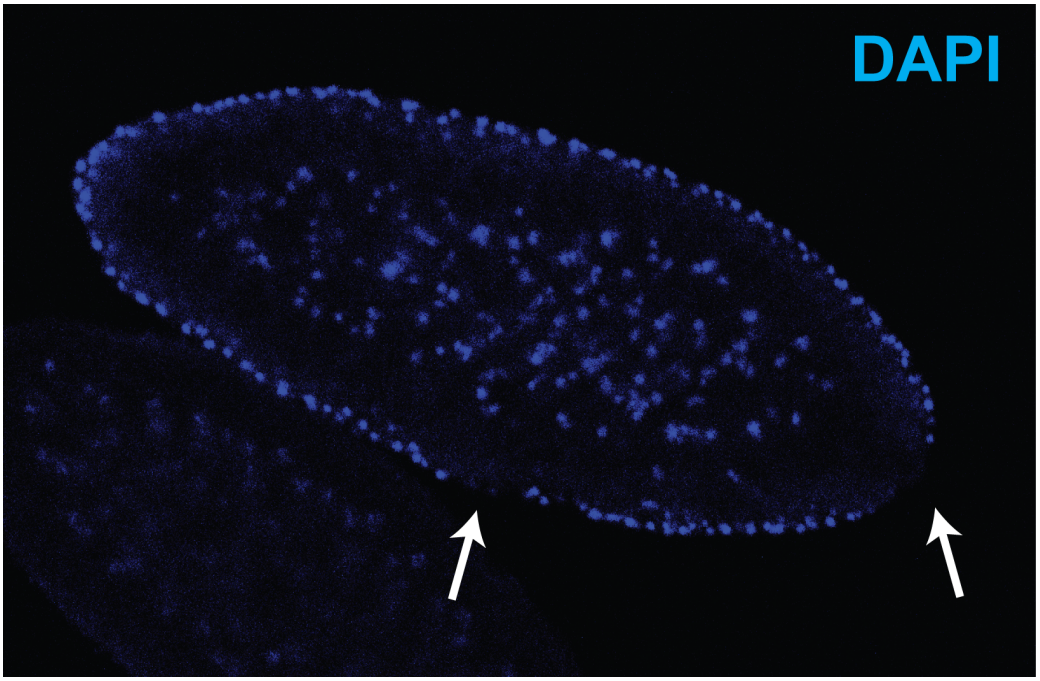


Figure 2.3 – Cid^{simulans} –rescued embryo exhibiting characteristic cortical "fallout" indicative of mitotic errors during the rapid syncytial divisions of *Drosophila* embryogenesis. White arrows indicate interruptions in the normal even patterning of nuclei as stained in blue by the DNA dye DAPI. Control embryos revealed no such cortical defects

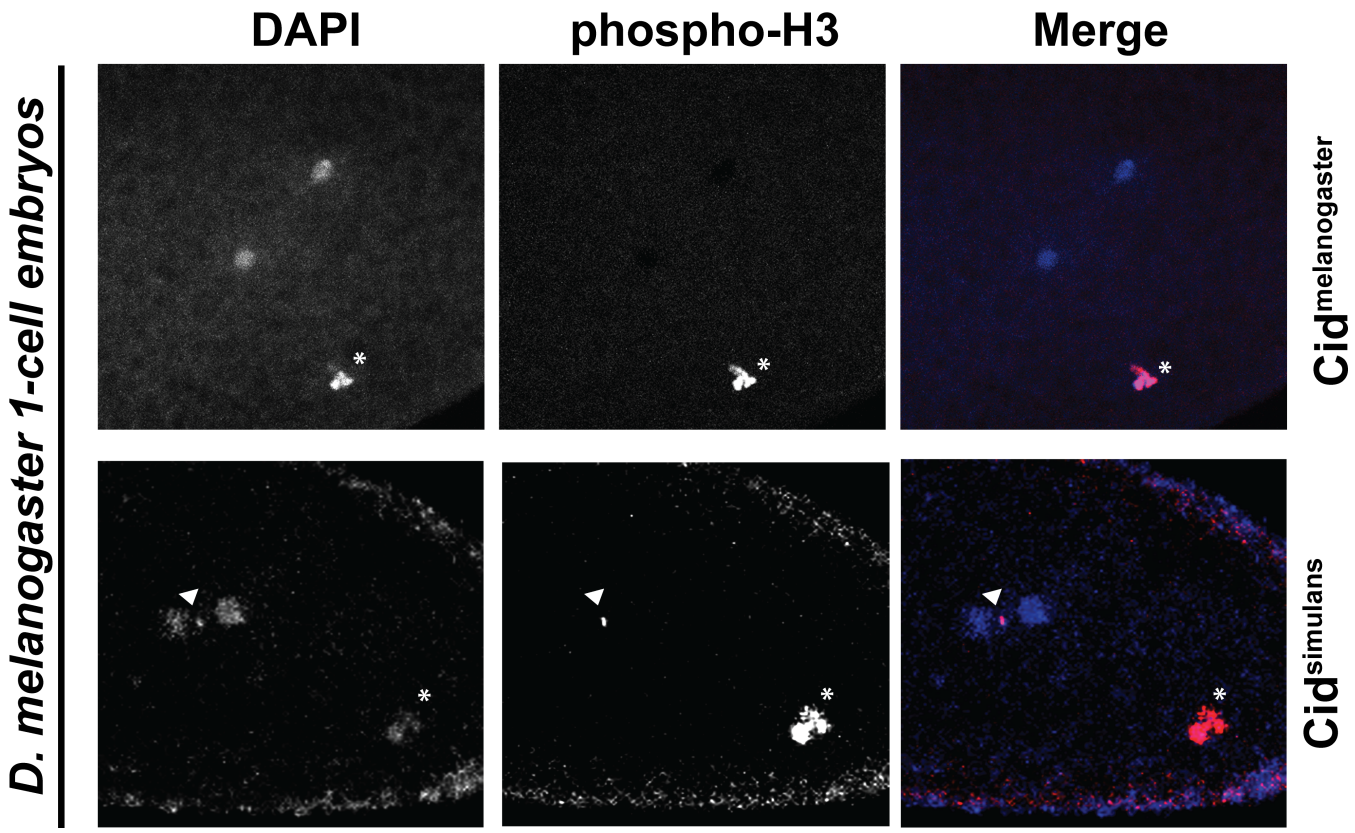


Figure 2.4 – Comparison between Cid^{melanogaster} (top panels) and Cid^{simulans} (bottom panels) –rescued embryos at the first mitotic division of embryogenesis. Embryo posterior is oriented to the left, anterior to the right. DAPI (gray in single channel image, blue in Merge) reveals the dividing chromosomes at anaphase of the first mitosis. Anti-phospho histone H3 staining reveals the mitotic chromosomes, which at this timepoint include all chromosomes of the dividing nucleus. The remnants of the meiotic products are visible in the bottom right, marked with an asterisk – these are the polar bodies, and stain positive for phospho-H3. Of note, a lagging chromosome (marked with a white arrowhead) appears between the anaphase dividing chromosomes in the Cid^{simulans} rescued embryo – no such lagging chromosome appears in the Cid^{melanogaster}-rescued embryos.

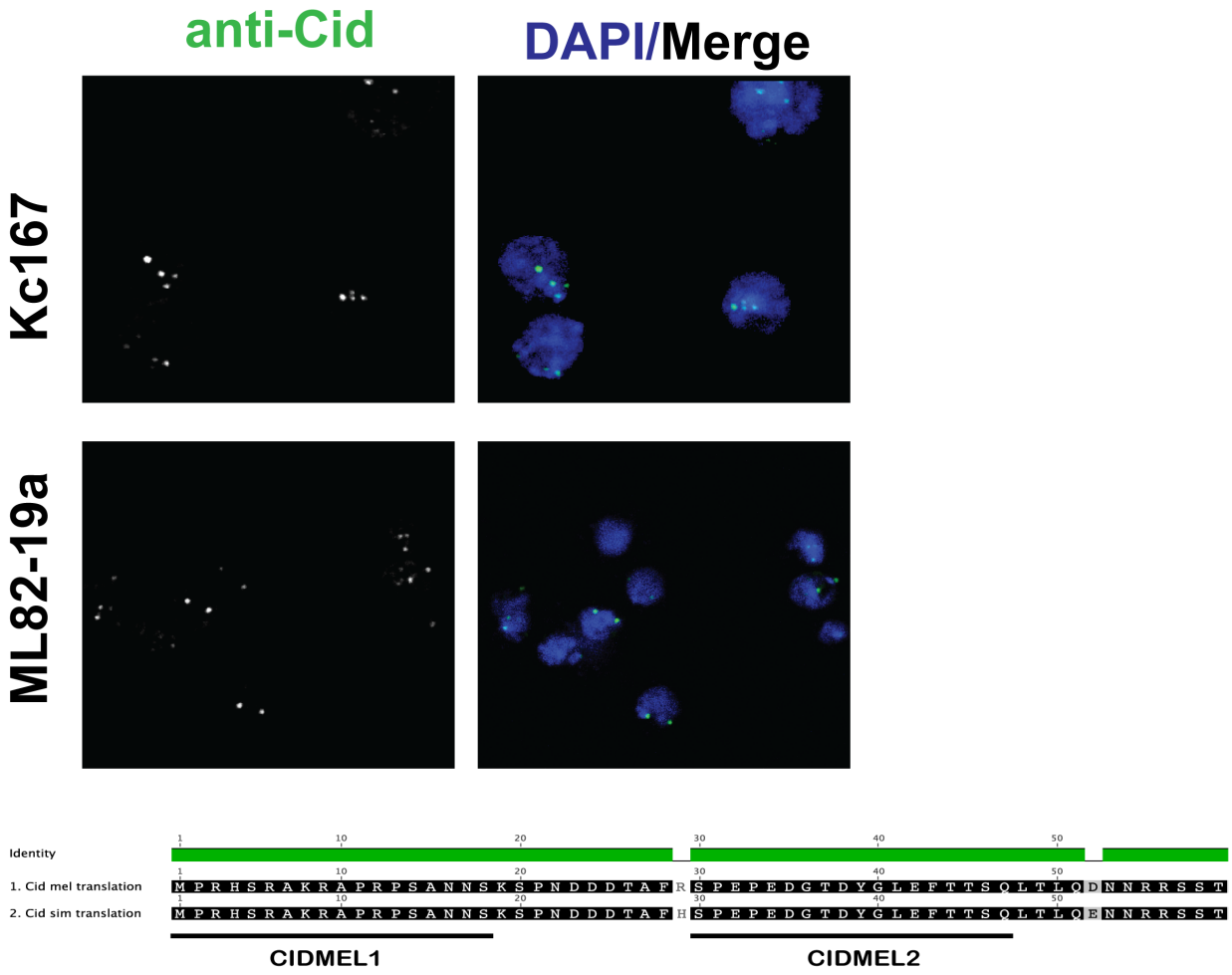


Figure 2.5 – Anti-Cid antibody staining (gray in left panel, green in Merge) in *D. melanogaster* (Kc167 cells) and *D. simulans* (ML82-19a cells) shows specificity for centromeres. Nuclei are indicated by DAPI staining (blue in Merge). As expected, anti-Cid recognizes puncta embedded within regions of intense DAPI staining, indicative of pericentric heterochromatin. Antibody used was CIDMEL1, raised against the peptide underlined in the alignment shown between $Cid^{melanogaster}$ and $Cid^{simulans}$. Both CIDMEL1 and CIDMEL2 show identical staining patterns and specificity and both antibodies should recognize Cid^{MSA} as well.

Male specific lethality of Cid-MSA rescue

Surprisingly, while *Cid*^{melanogaster} exhibited full rescue of *Cid*^{GD4436} and *Cid*^{simulans} failed to rescue, the ancestral allele *Cid*^{MSA} (encoding a protein that differed by only 11 amino acids from Cid^{melanogaster}) rescued females to full viability but failed to rescue males (**Figure 2.2** and **Figure 2.6)**. This result is remarkable because it suggests that the most recent evolutionary changes in the *Cid* gene, along the branch leading to extant *D. melanogaster*, were in response to male-specific selective pressure. This observation is precisely in line with predictions from the centromere drive hypothesis[5, 60] and with experiments from Robertsonian fusions[48, 52, 53] and monkeyflowers[47]. However, it is important to note that this *Cid*^{MSA} male-specific phenotype was due to mitotic defects during development and not meiotic problems. This result does not rule out additional meiotic problems; the few rescued males that eclosed however exhibited fertility on par with *Cid*^{melanogaster} rescued animals, although the same caveats apply to these results as for *Cid*^{simulans} rescue (assay lacked strong quantitation, see previous description). There are few developmental differences between males and females that could account for this sex-specific phenotype. One of these differences is the activity of the dosage compensation complex (DCC), specifically in males. This complex, which specifically binds to the X chromosome in males, acts to upregulate gene transcription of X-linked genes in order to equilibrate dosage relative to autosomal genes. Of particular interest, the genes that encode vital members of the DCC exhibit rapid evolution under positive selection between *D. melanogaster* and *D. simulans*[98, 99]. Another sex-specific feature is the Y-chromosome which, in *Drosophila,* is required for male fertility but not for male viability. Like the DCC, Y chromosomes in *Drosophila* evolve extremely rapidly even between strains of the same species, with manifold phenotypic consequences including global modulation of chromatin state[100-103]. I performed a simple experiment to simultaneously test the involvement of each of these process. Stocks of *D. melanogaster* exist in which two X chromosomes have been fused together by X-ray mutagenesis (BDSC stock #35). Females bearing this attached X-X

fusion chromosome (called C(1)RM, reverse metacentric) also carry a Y-chromosome (the Y-chromosome is required for male fertility but not for viability of males[1, 104], and its presence is tolerated by females, with some consequences for global gene expression[105]). This is because, upon mating to a wild-type male, offspring would either be tripo-X females (one paternal X, and the maternal attached X-X) which is lethal, X-X Y females, Y-Y lethal animals, or XY males with a maternally-derived Y chromosome. X-X Y females are perfectly fertile and viable. To test if the male-specific phenotype of Cid^{MSA} rescue was due to the presence of the Y chromosome or to the DCC, I crossed Cid^{MSA} alleles into the Cid^{GD4436} stock, in the presence of C(1)RM. Heterozygous *C(1)RM;* Cid^{GD4436}*/CyO;* Cid^{MSA} females were crossed to Cid^{GD4436}*/CyO;* Cid^{MSA} males. If an interaction between DCC and Cid^{MSA} was the basis for the male-specific phenotype, I predicted that males would still exhibit reduced viability. In contrast, if the male-specific phenotype was due to an interaction between Cid^{MSA} and the Y chromosome, I expected to see male viability and failure to rescue females. To control for genetic interactions between *C(1)RM* and Cid^{GD4436}, I analyzed $Cid^{melanogaster}$ rescue of Cid^{GD4436}in a *C(1)RM* background. To my surprise, these experiments gave clear results. While *C(1)RM* $Cid^{melanogaster}$ flies were viabile, Cid^{MSA} rescued XY male flies were also viable, while X-X Y females from the same cross exhibited decreased viability. These results provide strong evidence against a DCC-based mechanism for Cid^{MSA} male-specific lethality, and implicate the Y-chromosome in a synthetic lethal interaction with Cid^{MSA}. These results are exciting because the Y chromosome is largely composed of heterochromatic satellite DNA sequences, and is known to evolve extremely rapidly[100-103]. An important future experiment will be to generate Cid^{MSA} attached X-X females lacking the Y chromosome, as well as Cid^{MSA} XO males lacking the Y chromosome.

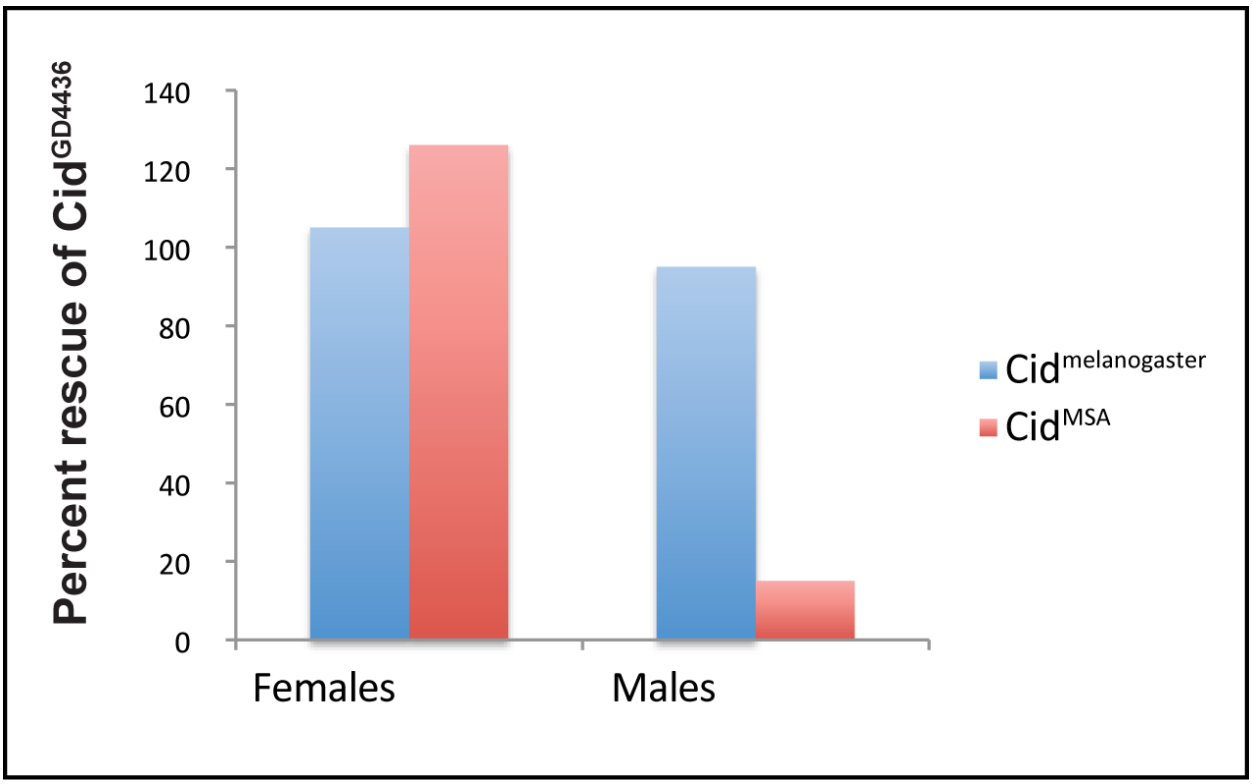


Figure 2.6 – Cid^{MSA} causes a male-specific failure to rescue *Cid^{GD4436}*. While female Cid^{MSA}-rescued flies are fully viable, Cid^{MSA} males are not rescued. Percent rescue was calculated by quantification of a co-segregating balancer chromosome based on expected Mendelian ratios. N = 484 progeny scored.

Genetic and epigenetic basis for Cid^{MSA} male lethality

The bulk of Cid needed for the rapid mitotic divisions that characterize early embryogenesis in *Drosophila* is deposited into the developing oocyte. However, recent work has found that paternal imprinting of Cid onto paternally deposited chromosomes is also required for proper mitoses in the early embryo[106]. In particular, paternal loading of Cid onto the Y chromosome appears to result in "overloading" of Cid on the Y centromere relative to the autosomes or the X chromosome. These results raise the possibility that epigenetic mechanisms contribute to the *Cid^{MSA}* male-specific rescue phenotype that I observed. To determine the contribution of genetic or epigenetic mechanisms to *Cid^{MSA}* male lethality, I leveraged the fact that *Cid^{MSA}* rescued females are viable. Therefore, by crossing rescued females to males of differing genotypes, I could create genetically identical rescued offspring from crosses of genetically different parents. I found that, indeed, male-specific lethality associated with rescue of *Cid^{GD4436}* by *Cid^{MSA}* was dependent on epigenetic inheritance of *Cid^{MSA}* from the parents **(Figure 2.7)**. For example, male-specific lethality occurred in all crosses in which Cid^{MSA} was present in the parental male, demonstrating that paternal-Cid^{MSA} was necessary for the male-specific effect. However, parental male Cid^{MSA} was not sufficient, since a cross in which a parental male possessing only Cid^{MSA} was crossed to a Cid^{melanogaster} female produced no lethality of any kind. One way to explain these results is to invoke a mechanism of lethality in which a Y chromosome is paternally loaded with Cid^{MSA}. This paternal epigenetic mark must thereafter be propagated through the rapid embryonic divisions for male lethality to manifest – perhaps due to an accumulation of mitotic errors (**Figure 2.8)**. Propagation would thus require maternal loading of Cid^{MSA} for full manifestation of the phenotype. One way by which this type of model could work is if *Cid* evolved to avoid binding to specific Y chromosomal satellite DNA elements (sequences that perhaps had some influence on meiotic drive, for example). In a setting in which Cid^{MSA} was present in a modern-day *D. melanogaster* genome, Cid^{MSA} could incorporate into a non-centromeric region, effectively setting up a dicentric scenario. As dicentric Y-chromosomes

went through mitosis in the rapidly dividing embryo, they could experience fragmentation and cells would eventually arrest leading to lethality. Females lacking the Y chromosome would theoretically be fine in this model, and indeed, that is consistent with the data. Cytological evaluation of Cid^{MSA} male or female embryos would provide evidence to support or reject this model. Alternatively, Cid^{MSA} may not be able to recognize *D. melanogaster* Y chromosomal sequences, and lethality may be due to mitotic dysfunction and to failure to segregate the Y during mitosis. Anti-Cid antibody staining of Y-bearing embryos should yield insight into which of these models is correct (**Figure 2.5).**

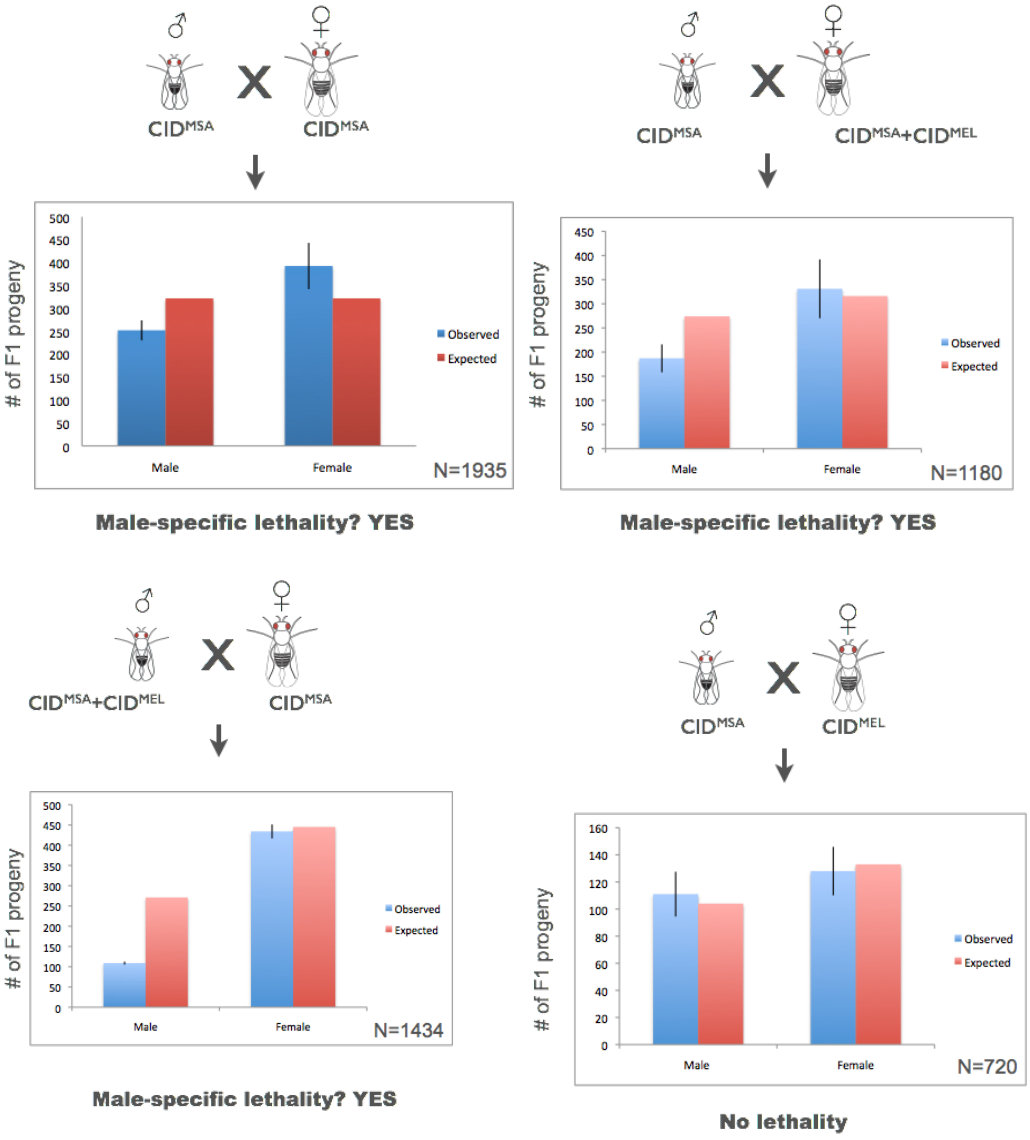


Figure 2.7 - Epigenetic effects on *Cid*^{MSA} male-specific lethality. Plots show quantification of male and female rescue compared to expected numbers (based on Mendelian ratios and number of balancer progeny of the cross). Parental genotypes are as follows: Cid^{MEL}= , Cid^{MSA}= , Cid$^{MSA+MEL}$= . Crosses were perfomed in triplicate, error bars reflect standard deviation. N= number of progeny counted.

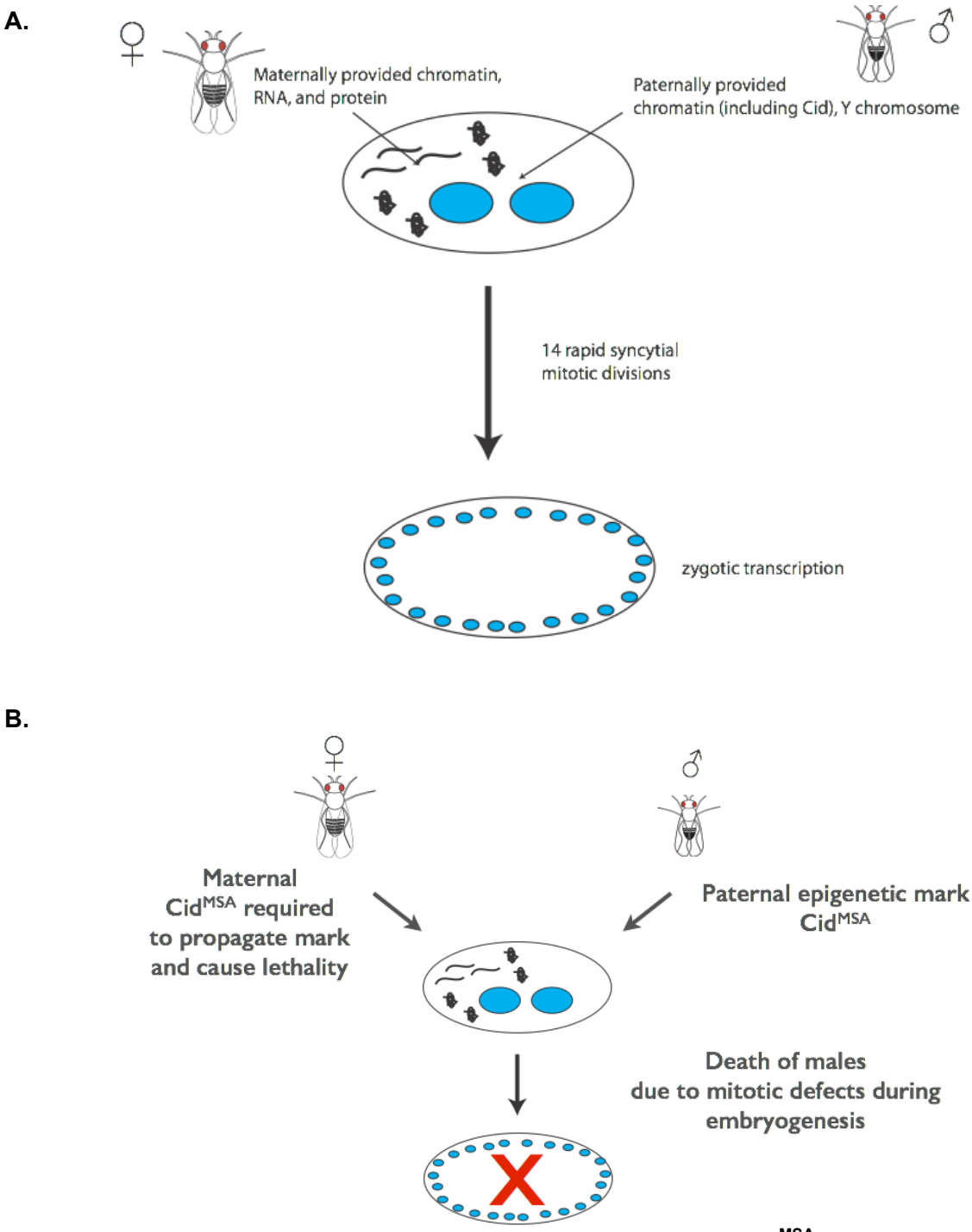


Figure 2.8 – Model for genetic and epigenetic contribution of Cid^{MSA} and the Y chromosome to male-specific lethality. A. Normal embryogenesis, with essential contributions in the form of RNA, protein, and chromatin coming from both parents. B. Cid^{MSA}-induced lethality requires genetic and epigenetic contributions from both parents.

Technical problems with replication of *Cid* rescue crosses

My experiments with *Cid* were put on hold while I completed my first paper (describing the birth of essential centromere function in the young gene *Umbrea*, see Section 2 for details). Upon revisiting the initial complementation analysis, I found that the male-specific phenotype of Cid^{MSA} had lessened, as had the lethality of $Cid^{simulans}$ rescue. This appeared to be due to contamination of the rescue stocks by non-mutant non-balancer second chromosomes, since I could isolate such chromosomes upon repeated backcrossing of my rescue lines to an external stock. Dishearteningly, I was unable to recreate the rescue lines from the original transgenic stocks by re-crossing each to the original P-element mutant line Cid^{GD4436}. Furthermore, I reinjected the same transgenes into different attB landing pad lines (BDSC #34760), re-crossed these new lines to Cid^{GD4436}, and again failed to observe the original genetic complementation that I had observed. This did not appear to be due to the acquisition of new mutations by the Cid^{GD4436} stock, since a previously published line expressing GFP-Cid under a native promoter[107] was able to rescue Cid^{GD4436} and trans-heterozygotes with *Cid* point mutations. Therefore, failure to rescue seemed to be a specific property of the transgenes themselves. It is important to note that the original transgenic stocks from which I generated my rescue lines had been maintained for over two years. Therefore, if slight overexpression of Cid was deleterious (as it is known to be in *Drosophila*[76], inactivating mutations could have been selected for in the stock vials. I tested if my transgenes produced mRNA, by performing RT-PCR with primers specific to the recoded sequence. These experiments revealed that *Cid* transgenes were expressed. However, because the transgenes produced proteins that were untagged, I could not assay for protein expression using antibodies and western blotting. Therefore, it remains possible that despite mRNA expression, my transgenes failed to rescue because protein was not being made. Further, I only assayed mRNA expression of whole flies. Since Cid is required in all proliferating tissues throughout development, it is possible that miss-expression or loss of expression in a particular cell type was responsible for the failure to rescue. However, the

newer transgenic lines should not have had time to accumulate inactivating mutations, yet behaved similarly to the old lines.

Replacement of *Cid* with orthologs by CRISPR-mediated homology-directed repair

To circumvent problems associated with replication of my original Cid^{GD4436} rescue experiments, I turned to the recently popularized genome-editing technology CRISPR-Cas9[108-110]. Instead of using transgenes to rescue null mutations in the endogenous *Cid* gene, homology-directed repair using dsDNA templates and Cas9 endonuclease activity would allow me to replace *Cid,* in its native locus, with tagged recoded *Cid* orthologs or ancestral genes. This approach addresses multiple separate issues with my previous genetic complementation strategy. First, by replacing only the coding region of *Cid* in its native context, all endogenous regulatory sequences are left intact. Therefore, the expression pattern of transgenes would be exactly the same as the endogenous allele. Tagging the transgenes so that the encoded protein is fused to FLAG would allow me to assay for protein expression and compare relative expression of the transgene to the native allele, in a heterozygote. An additional attribute of this approach is that at no point during the process does the *Cid* gene increase in dosage relative to wild-type. Finally, the genetic crossing scheme required to generate the "tester" line is dramatically shortened. These experiments are currently ongoing.

Experimental evolution in population cages

I am greatly interested in understanding the adaptive basis for *Cid* evolution, as it has played out in the *D. melanogaster* species group. My data on Cid^{MSA} complementation suggest that selective pressure acted on *Cid* for binding to (or avoid binding) Y chromosome specific DNA elements. With a gene-centric view, the adaptive steps that *Cid* experienced were each (or some) of the 11 amino acids differences leading to the modern-day $Cid^{melanogaster}$. What were the corresponding changes in the centromeric DNA that might have driven these changes?

One way to gain insight into this process would be to take populations of *Cid*^{MSA} flies that exhibit male-specific lethality and passage them in hopes of experimentally evolving suppression of the phenotype. Selection should depend on the population size, but in theory, suppression of the phenotype could occur rapidly. Whole-genome sequencing of Cid^{MSA} flies compared to evolved populations could yield important clues into the back-and-forth evolutionary dynamics between Cid and centromeric DNA. If male-specific lethality occurs due to improper binding of Cid^{MSA} to the Y chromosome, suppression may occur by selection on deletions of Y chromosomal sequences, which could be identified by the decreased abundance of certain repetitive sequences upon whole-genome sequencing[111]. These experiments await the generation of CRISPR gene-swap flies, and the repetition of my original observations.

Y-chromosome mapping experiment using Y deletion lines.

A complementary method for the identification of sequences of repetitive DNA on the Y-chromosome that interact synthetically with Cid^{MSA} would be to cross in Y-chromosomes bearing deletions of specific regions[112]. This approach, coupled with cytology using FISH probes, could allow for targeted identification of candidate regions that might be the basis for Cid^{MSA} male specific lethality. In this approach, I predict that most Y chromosome deletions will not have any effect on Cid^{MSA} male lethality. However, if specific regions of the Y chromosome are causal for the synthetic lethality with Cid^{MSA} (for example, a large stretch of a particular heterochromatic repetitive DNA sequence), I predict that deletion of these regions improves male viability. If specific regions are thus identified, FISH using known Y chromosomal repetitive DNA sequences[96] can be performed to attempt to identify specific sequences of interest. Further validation could involve ChIP-Seq with anti-Cid antibodies from flies bearing Y-chromosomes or cell lines with Y chromosomes (for instance, CME W1 cl.8+ DGRC #151) to identify sequences that are physically associated with Cid (see next section), or whole genome

sequencing on flies bearing the Y deletions of interest to assess whether the abundance of particular repetitive sequences is altered by the deletion compared to a wild type control[111].

ChIP-seq to identify differences in centromeric DNA binding between Cid orthologs

While in *Saccharomyces cerevisae* CenH3 has DNA-binding specificity due to localization to the genetically defined point centromere, in no other system is the DNA binding profile of CENP-A characterized. Indeed, CENP-A is known to localize to subsets of repetitive satellite DNA arrays that are larger than the known centromeric domain. Furthermore, overexpression of heterochromatin binding proteins causes encroachment into the centromeric domain[113]. The opposite is also true, suggesting that the boundary between heterochromatin and centromeric compartments is somewhat fluid. I propose to perform CHIP-seq experiments with anti-Cid antibodies to define the centromeric sequences in *D. melanogaster* and *D. simulans* cells (either from whole flies or from cultured cells from each species). DNA sequences identified in these experiments would be mapped back to the *D. melanogaster* or *D. simulans* genome assemblies, or assigned to the unassembled portions of these genomes, which largely consist of heterochromatic and centromeric sequences[111]. Upon establishment of the within-species binding profile for Cid, the next experiment would be to express orthologs in an inter-species context to look for species-specific difference in specificity. If differences are observed, the molecular basis of these differences can be mapped by expressing $Cid^{melanogaster}$-$Cid^{simulans}$ chimeras, or ancestral Cid^{MSA} and doing pulldowns, Furthermore, the validity and chromosomal location of the identified sequences can be confirmed by performing fluorescence *in situ* hybridization with the sequences of interest in the cells of interest – for example, in *D. melanogaster* cl.8+ cells. This approach would yield a comprehensive and unprecedented view into the evolution of centromeric DNA and the effects of rapid evolution of Cid on its association with centromeric DNA sequences.

Section 3: Gain of essential function in young genes as a consequence of rapid evolution at centromeres

Gene duplication and genetic innovation

Gene duplication has been widely thought to be one of the primary mechanisms by which phenotypic diversity is generated during the course of evolution[114, 115]. Indeed, examples abound in the literature, exemplifying the attractiveness of duplication and divergence as a way to understand differences between species. Yet major unanswered questions remain. For instance, relaxed selection is expected to follow the birth of a new duplicate gene, since immediately following duplication the daughter gene is redundant with its parent gene. Despite the expectation that relaxed selection should result in pseudogenization (acquisition of debilitating mutations and loss of function), extant genomes (including those of *Drosophila*, mice, and humans) are littered with expressed, putatively functional duplicate genes.

Further complicating the picture is a long-standing paradigm that genic essentiality parallels conservation[116]. From this viewpoint, genes that are broadly conserved between species and across taxa are highly likely to be required for some aspect of fitness that is measurable in the laboratory. It follows, then, that young genes born by gene duplication are less likely to be essential. This perspective is highly intuitive. For example, if a species diverged prior to the birth by duplication of a gene, the duplicate seems unlikely to be essential, simply because species exist that lack it.

Young genes can rapidly become essential in *Drosophila*

However, this viewpoint has recently been shown to be incorrect, at least in the model organism *Drosophila melanogaster*. Indeed, a high proportion of young genes are essential for development in *Drosophila melanogaster* [117], directly challenging the dogma about the

relationship between essentiality and conservation. By GO analysis, these genes appear to function in a diverse set of cellular pathways, including metabolism, defense against pathogens, chromatin, and germline-specific functions. In addition, RNAi-mediated knockdown of each gene revealed that lethality was induced at various points in development, including larval and pupal stages. For those genes with P-element mutations, the RNAi results were recapitulated, suggesting that lethality was not an off-target effect. These results suggest that no single common feature bestows essentiality to a new gene born by duplication.

Essentiality in a young gene could be explained by invoking two differing mechanisms long thought to act on genes following duplication. One explanation for how young genes become essential is that essential, ancestral functions could be partitioned between (old) parental and (young) daughter genes [118, 119] provides one explanation – this is known as subfunctionalization. One way that this subfunctionalization could occur is upon the acquisition of complementary inactivating mutations in one of two domains in each of the parent and daughter genes, potentially increasing molecular complexity [120]. Another example of subfunctionalization is through the complementary loss of tissue-specific enhancers – if the parent gene is expressed ubiquitously in the germline and soma prior to duplication, following subfunctionalization the parent and daughter gene could exhibit restriction of expression in either the germline or soma.

Essential neofunctionalization

More difficult to understand is the situation by which new genes become essential via emergence of novel function (neofunctionalization)[121]. Broad interest exists in the general scientific community (as well as the general public) in understanding how gene duplication can result in new gene function, as exemplified by the recent implication of human-specific duplicate genes (genes born subsequent to the split between humans and chimpanzees) in the evolution

of human-specific traits such as higher intelligence [122]. Yet neofunctionalization is poorly understood, despite some recent reports of neofunctionalization over "deep" evolutionary time[123]. New functions could result from partial duplication of ancestral genes (thereby liberating one functional domain from another), by gene fusion between two domains that previously had been unlinked, or by rapid amino acid changes that impart differences in function, perhaps through changing protein-protein interaction [124-127]. Further complicating the situation is the fact that no one process can entirely explain the pattern of duplication and divergence for all duplication events in a given genome. For example, a recent report suggests that subfunctionalization preceeds neofunctionalization in yeast [128], which experienced a historical whole-genome duplication event. In contrast, in *Drosophila*, neofunctionalization may account for the greatest proportion of preserved duplicate genes, as assessed by novel RNA expression patterns between duplicates[129], a phenomenon which has been previously observed on the single-gene level[130]. Past studies of gene duplication and divergence have been limited in the depth of their insight, often because of inherent limitations in making general statements based on genome-wide datasets without the ability to dissect function on the single-gene level. Therefore, the contribution of selective processes to the acquisition of essential function is unknown, as are the underlying molecular changes.

The Heterochromatin Protein 1 family as a model for genetic innovation

To gain insight into the process of neofunctionalization and gain of essential function, I turned to the Heterochromatin Protein 1 (HP1) gene family in *Drosophila.* HP1 genes are found in all eukaryotic genomes (although copy number and functionality appear to vary widely [131]), and these genes encode chromosomal proteins that are defined by the presence of two characteristic domains: an N-terminal "chromodomain" (CD) and a C-terminal "chromoshadow" domain (CSD), separated by a hinge region of variable length[131]. Other non-HP1 proteins possess homologous chromodomains, including the Polycomb group of transcriptional

repressors. CDs are often considered to be “readers” of the so-called “histone code” since they typically bind post-translational modifications on the N-terminal tail of histone proteins. However, recent work suggests that they not only bind histones but also function in dimerization[132] and can also regulate histone-tail binding by adjacent HP1 molecules[133]. In contrast to the CD, only HP1 family members possess the C-terminal CSD, which is structurally related to the CD with distinct characteristics[134]. While CDs appear to be specialized for binding specifically to histone tail modifications, CSDs appear to be much more promiscuous, mediating a breadth of protein-protein interactions including homo- and hetero-dimerization[134, 135]. In *Drosophila,* the HP1 gene family has experienced remarkable diversification, with members functioning in cellular processes as diverse as heterochromatin formation, euchromatic gene transcription, and genome defense against transposable elements[131, 136, 137]. The three “founding” member of the HP1 family are HP1A, HP1B, and HP1C, each of which have been well characterized and found to have diverse chromosomal localization patterns and functional roles, including binding to different partner proteins[131, 138-140]. Besides characterized functional diversification, HP1 genes have recurrently duplicated over the *Drosophila* phylogeny, with multiple full-length duplicates derived from each of the “canonical” HP1 members and retained in diverse lineages. Surprisingly, “partial” HP1 duplicate genes are also found throughout *Drosophila,* some of which have been retained for millions of years following duplication, and may be important for fitness, as signatures of purifying or positive selections are observed[136]. Furthermore, both CD-only and CSD-only duplicate genes are preserved in *Drosophila* genomes, suggesting that each domain possesses intrinsic functionality without the presence of the other domain.

Since HP1 genes encode chromosomal proteins, many of which function in regulating pericentric heterochromatin, former staff scientist Danielle Vermaak and former technician Mary Alice Hiatt undertook a cell biological approach to screen uncharacterized HP1 family members

for interesting cytological localization patterns. They uncovered many interesting and previously undescribed phenotypes. However, most interesting to me was their report of a partial HP1 possessing only a CSD, derived from the well-characterized *HP1B* gene. While HP1B localizes to heterochromatin and euchromatin, Mary Alice and Danielle found that this new gene (which they called *Umbrea*, derived from the Latin word for shadow), localized specifically to centromeres (**Figure 3.1).** Addtionally, *Umbrea* had been reported to be essential for fly development[131, 141-143]. To me, this gene offered a perfect test case to gain insight into mechanisms of neofunctionalization and gain of essential function, while shedding further light on rapid evolution at *Drosophila* centromeres.

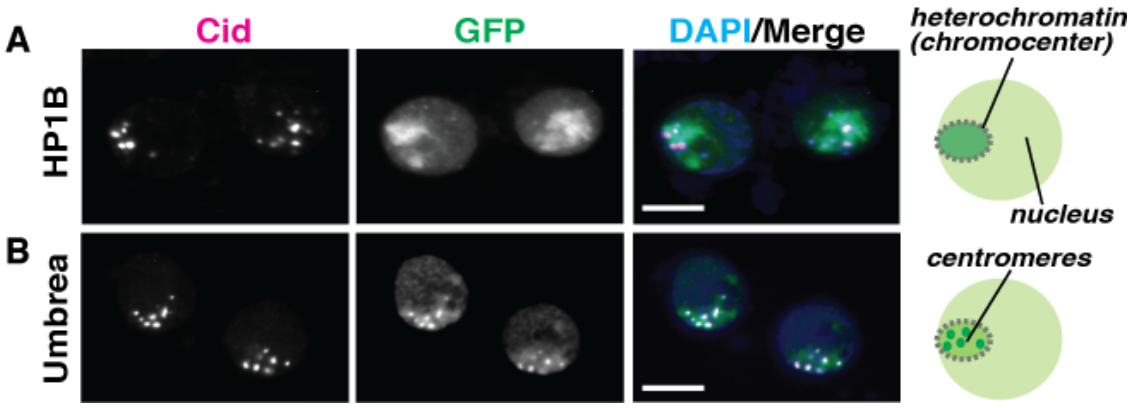


Figure 3.1 - Change in nuclear localization pattern between HP1B and Umbrea. GFP (green in merge) fused to either HP1B or Umbrea from *D. melanogaster* was expressed in *D. melanogaster* Kc cells by heat-shock induction. The sub-nuclear localization pattern of these fusion proteins was compared using costaining with anti-Cid antibody (pink in merge) to mark centromeres. DAPI staining of the nucleus shown in blue.

Umbrea gained localization to centromeres after duplication from *HP1B*

HP1B has been poorly defined at the functional level. Overexpression experiments indicated that it binds to an overlapping set of heterochromatic targets with HP1A, while knockdown experiments revealed decreased expression of heterochromatic genes[140]. These observations suggest that, like HP1A, HP1B plays a structural role in the pericentric heterochromatin. However, further observations indicate a broad role in euchromatic gene expression as well. Additionally complicating the issue, knockdown or mutant phenotypes for HP1B show that it is not essential for fertility or viability in laboratory conditions[140](personal communication, Nicole Riddle, Karpen Lab, http://www.drosophila-conf.org/2012/abstracts/text/f70493.htm).

I began by comparing the localization of HP1B and Umbrea in *D. melanogaster* Kc cells. GFP-tagged HP1B proteins from both *D. melanogaster* and *D. ananassae,* whose divergence predates birth of *Umbrea* [136], localized to pericentric heterochromatin and euchromatin (**Figure 3.1)**. In contrast, Umbrea-GFP localized to interphase centromeres, but not telomeres [141]. Thus, since birth, Umbrea acquired a new centromeric localization that might underlie its essential function.

To provide further evidence that Umbrea localized to centromeres *in vivo*, I raised polyclonal peptide antibodies in rabbits (Covance), targeting two predicted hydrophilic regions of Umbrea that corresponded to the N- and C-terminal tail domains. These antibodies proved to be specific for Umbrea by western blotting [65], using overexpressed tagged Umbrea in *D. melanogaster* Kc cultured cells **(Figure 3.2)**. Furthermore, use of anti-Umbrea antibodies in *D. melanogaster* adult (testes) and larval tissues (imaginal discs) revealed highly specific localization to centromeres, by costaining with anti-Cid [144] or anti-CENP-C[107] antibodies **(Figure 3.2)**. Specificity was further demonstrated by immunofluorescence imaging in Kc cells, which do not

natively express Umbrea. Transfection of constructs to overexpress tagged Umbrea into these Kc cells, revealed antibody staining specifically in transfected cells with no staining in untransfected cells.

The finding that Umbrea altered its localization pattern subsequent to its duplication from HP1B may not be in and of itself surprising. Indeed, in yeast, approximately a third of all duplicate genes retained following the yeast whole genome duplication event have evolved new subcellular localizations (and presumably new functions)[145]. However, the yeast whole-genome duplication event that spawned the duplicate genes analyzed in this study happened approximately 100 million years ago, a much greater period of divergence than that of *Umbrea*. The rapid alteration in sub-nuclear localization from heterochromatin to centromeres by Umbrea prompted me to further investigate the evolutionary and molecular basis for neofunctionalization.

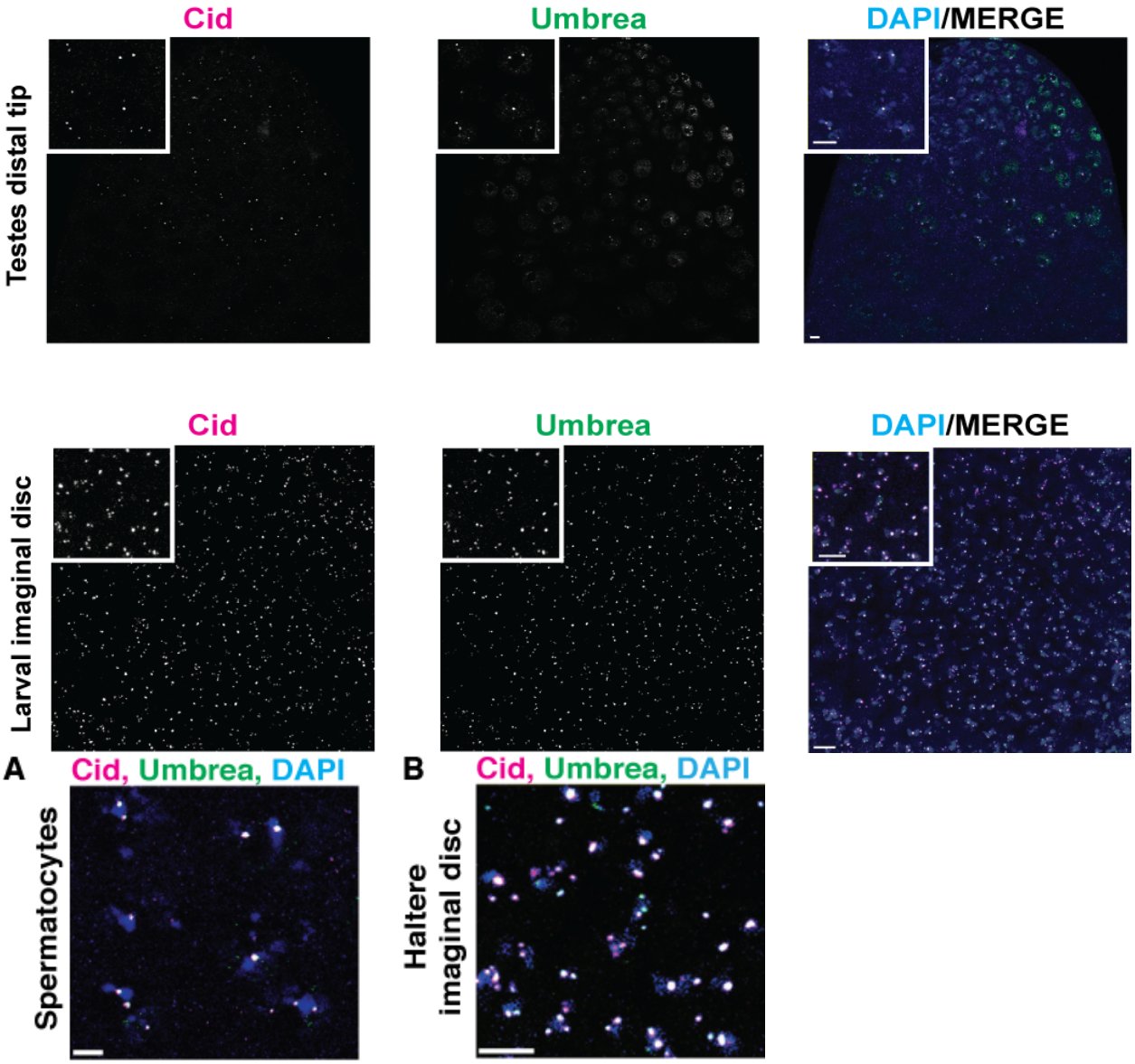


Figure 3.2 – Polyclonal antibodies specific for *D. melanogaster* Umbrea[65] confirm it's centromeric localization *in vivo.* Anti-Umbrea (gray, green in merge) colocalizes with anti-Cid antibody staining (gray, pink in merge) in *D. melanogaster* testes (top panels), or larval imaginal discs (middle panels). DAPI staining of nuclei in blue. Higher magnification images of tissues costained with anti-Umbrea and anti-Cid antibodies reveal the extent of colocalization in spermatocytes (A) or imaginal disc cells (B). Scale bar, 5 microns.

Umbrea* is an essential gene in *Drosophila melanogaster

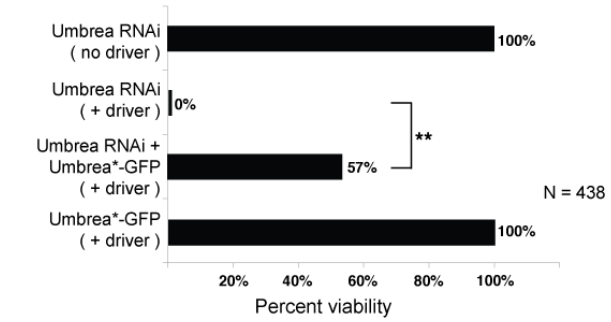
In contrast, previous reports concluded that *Umbrea* is essential in *D. melanogaster* [117, 141]. We confirmed and extended these results, by showing that *Umbrea* knockdown by RNAi resulted in 100% late larval-pupal lethality that could be rescued by an *Umbrea-GFP* fusion **(Figure 3.3A)**, in which all synonymous codon sites were recoded to be intransigent to RNAi. Genetic knockout experiments using P-element excision mutants **(Figure 3.3B)** further confirmed that *Umbrea* is essential in *D. melanogaster*. Previous studies had reported lethality of P-element disruption of *Umbrea* [117, 142] using the $HP6^{BG01429}$ allele. However, I found that this mutation failed to complement mutations in the *Dumpy* gene, when in *trans* **(Figure 3.3C)**. Furthermore, $HP6^{BG01429}$ homozygous mutants died as embryos, whereas *Umbrea* RNAi resulted in late larval-pupal lethality. These results did not rule out the essentiality of *Umbrea,* but necessitated the generation of additional, specific mutants. To this point, the P-element insertion in $HP6^{BG01429}$ was mobilized, resulting in two null mutants with homozygous late larval-pupal lethality (the same phenotype was observed when mutations were in *trans* over a deficiency for the locus. In summary, these data show that *Umbrea* is an essential gene in *D. melanogaster.*

Evolution, presence absence, inferring the history

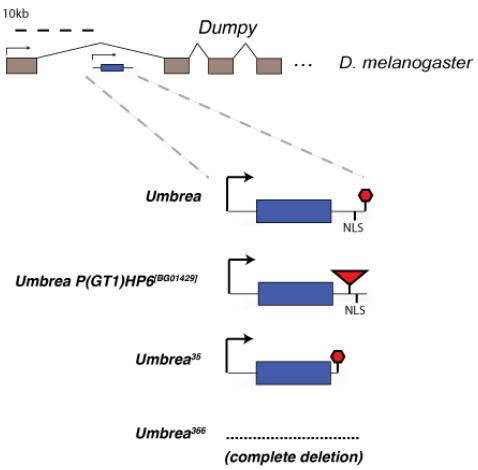
To precisely date the birth and subsequent changes in *Umbrea*, I expanded upon previous work by Danielle Vermaak and Mary Alice Hiatt. I used primers targeting highly conserved sequence in the *Dumpy* intron to amplify and sequence the syntenic *Umbrea* locus from 32 *Drosophila* species (**Figure 3.4)**. I confirmed FlyBase predictions and published data that the *HP1B* gene was preserved throughout the *Drosophila* genus [136]. In contrast, I found the *Umbrea* gene in only 20 of 32 species including the *melanogaster* subgroup (**Figure 3.5)**. Superimposing these data over a well-supported *Drosophila* phylogeny[146] allowed me to date the monophyletic origin of *Umbrea* to approximately 12-15 million years ago, following the split between *D.*

ananassae and the common ancestor of the *melanogaster*, *suzukii*, and *takahashii* subgroups. Four *Drosophila* lineages retain full-length *Umbrea* genes, including *D. gunungcola, D. lucipennis, D. elegans,* and *D. ficusphila*, two of which preserve an intact chromodomain (CD) [147] and ancestral residues essential for binding histone H3 tri-methyl lysine 9 (H3K9me) [148]. However, most extant *Umbreas* have lost their CDs, and retain only the chromoshadow domain (CSD), which mediates protein-protein interactions [149], including all species in the *melanogaster* and *takahashii* subgroups. Surprisingly, *Umbrea* has been lost to pseudogenization at least three independent times - in *D. eugracilis, D. fuyamai,* and in the common ancestor of the *suzukii* clade. Importantly, these pseudogenization events represent loss of *Umbrea* from lineages both with and without a CD. This observation implies that loss of the CD alone was not sufficient for the gain of centromere localization and essential function.

A



B



C

Complementation cross			F1 progeny Control	Trans-het.	N =	Complements?
P{GT1}HP6[BG01429] / Bal.	X	*Umbrea35 / Bal.*	83	89	172	Yes
P{GT1}HP6[BG01429] / Bal.	X	*Umbrea366 / Bal.*	80	94	174	Yes
P{GT1}HP6[BG01429] / Bal.	X	*dumpyivl / Bal.*	226	0	226	No
P{GT1}HP6[BG01429] / Bal.	X	*dumpyolvr / Bal.*	61	0	61	No
Umbrea35 / Bal.	X	*dumpyivl / Bal.*	167	157	324	Yes
Umbrea366 / Bal.	X	*dumpyivl / Bal.*	105	131	236	Yes
Umbrea35 / Bal.	X	*dumpyolvr / Bal.*	290	327	617	Yes
Umbrea366 / Bal.	X	*dumpyolvr / Bal.*	94	82	176	Yes
Umbrea35 / Bal.	X	*Umbrea366 / Bal.*	234	2	236	No

Figure 3.3 – *Umbrea* is an essential gene in *D. melanogaster.* I confirmed previous findings that *Umbrea* was required for development. Constitutively expressed RNAi resulted in 100% lethality at the late larval-pupal stage (A). This lethality could be specifically rescued by expression of an *Umbrea* transgene recoded to be resistant to RNAi knockdown. Further confirmation of *Umbrea* essentiality is shown by *Umbrea* mutant alleles derived from a P-element insertion into the C-terminus of the gene (B). This P-element disrupts *Dumpy* function as well as *Umbrea* function, as shown by transheterozygous complementation experiments. However, *Umbrea* excision alleles complement *Dumpy* mutants, suggesting that the homozygous lethality associated with these mutants is specific to *Umbrea.*

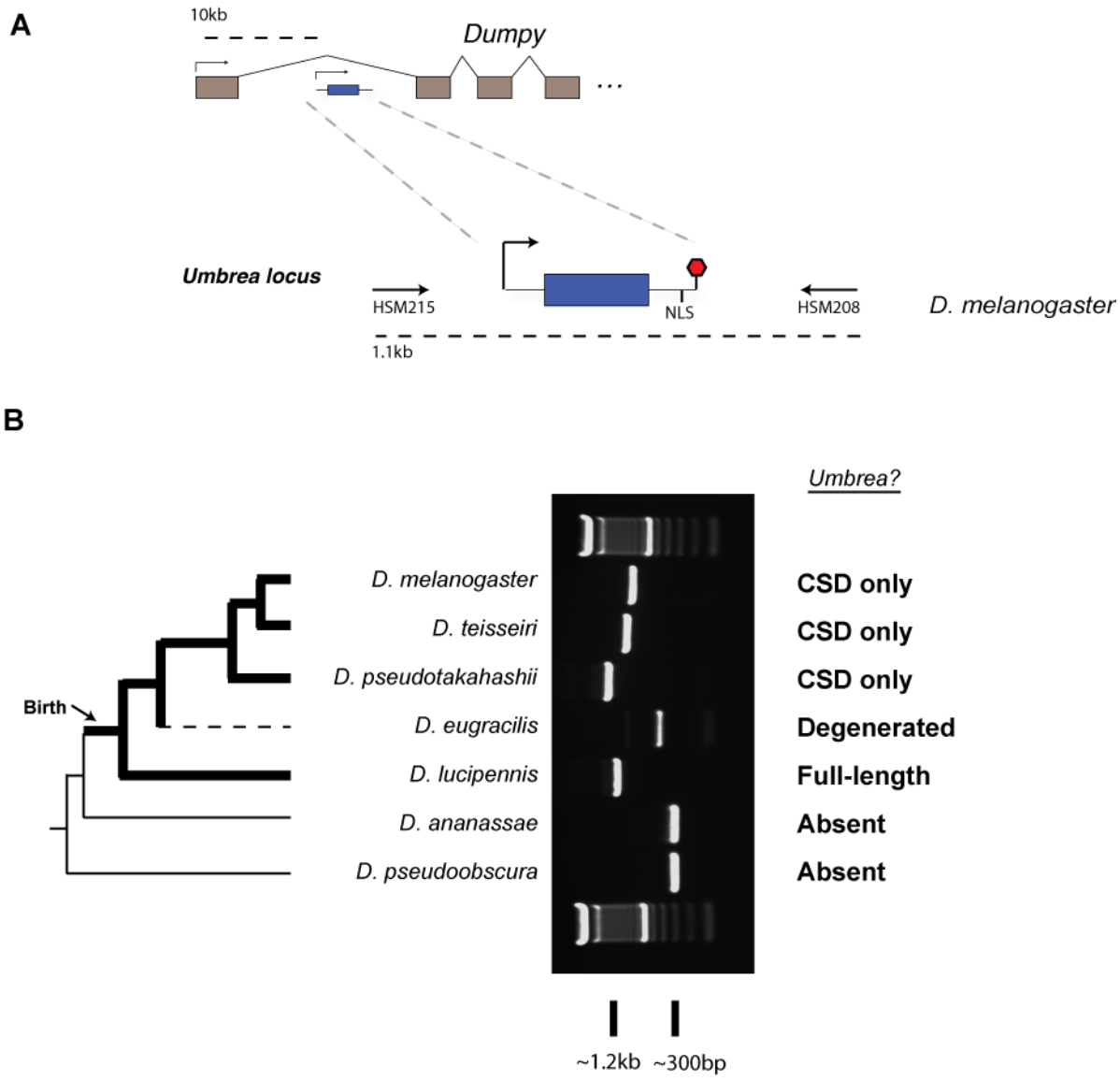


Figure 3.4 – *Umbrea* resides in an intron of the large *Dumpy* gene. Primers were designed to conserved sequence within the intron, spanning the *Umbrea* locus (A). These primers were then used to amplify the syntenic region from many species of *Drosophila*. The resulting bands exhibited size differences consistent with the presence or absence of *Umbrea*. These bands were sequenced using Sanger sequencing to confirm presence or absence and, if present, to identify *Umbrea* coding sequence (B).

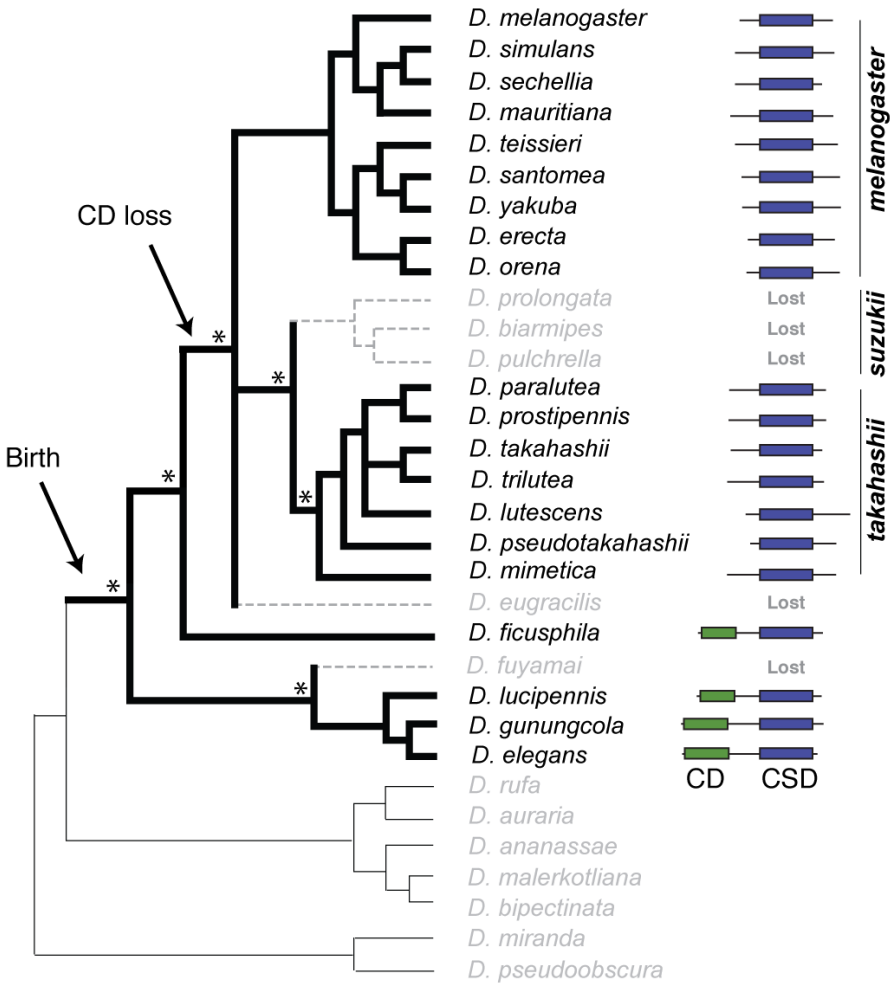


Figure 3.5 – Phylogeny depicting the presence or absence, and gene structure of *Umbrea.* Evolutionary relationships between species on the tree were based on analysis from a multilocus phylogeny[146]. Based on this, I mapped the birth of *Umbrea* to a single branch (indicated by arrow), and the loss of the ancestral chromodomain to another internal branch (arrow). Asterisks indicate phylogenetically important branchpoints with high confidence[146]. Based on this tree, *Umbrea* was lost at least three independent times.

Cell culture knockdown

Based on these results, I hypothesized that Umbrea functioned within cells to promote proper chromosome segregation during mitotic divisions in development. In order to test this hypothesis, I collaborated with Leah Rosin in Dr. Barbara Mellone's lab at the University of Connecticut. Leah incubated *D. melanogaster* S2 cells with dsRNA homologous to *Umbrea* coding sequence, or with control scrambled dsRNA. She compared fixed and stained cells from each class (control versus knockdown) and found that *Umbrea* depletion resulted in increased mitotic errors in knockdown cells compared to control cells ($p<.05$). To further define the extent of mitotic errors upon Umbrea knockdown, Leah took time-lapse movies of both *Umbrea* knockdown and control cells stably expressing histone H2B fused to GFP and tubulin fused to mCherry. This live imaging allowed her to determine the timing of errors in chromosome segregation during mitosis, as well as observe chromosome dynamics throughout the cell cycle. Examples of mitotic errors resulting from *Umbrea* depletion included delayed metaphase chromosome alignment, anaphase onset prior to congression, lagging anaphase chromosomes, and multipolar configurations. The identity (X, 2, 3, or 4, identified by physical morphology in cytological spreads) of the lagging chromosomes did not appear to be consistent across experiments or from cell to cell, suggesting that the defect is general and not specific to any one chromosome (for example, the X chromosome). These data suggest that Umbrea generally promotes proper chromosome segregation, in *D. melanogaster* cultured cells. It is important to note that *Umbrea* is expressed at extremely low levels in *D. melanogaster* S2 cells (FlyBase.org, Danielle Vermaak personal communication), yet knockdown reveals a strong phenotype. Other tissues *in vivo* exhibit far higher levels of *Umbrea* expression and may be more or less sensitive to depletion, perhaps contributing to the difference in penetrance in lethality during development compared to the *in vitro* knockdown in cell culture. An additional factor that may contribute to variability in phenotype is that RNAi phenotypes can be dependent on protein perdurance, which may vary between cell types.

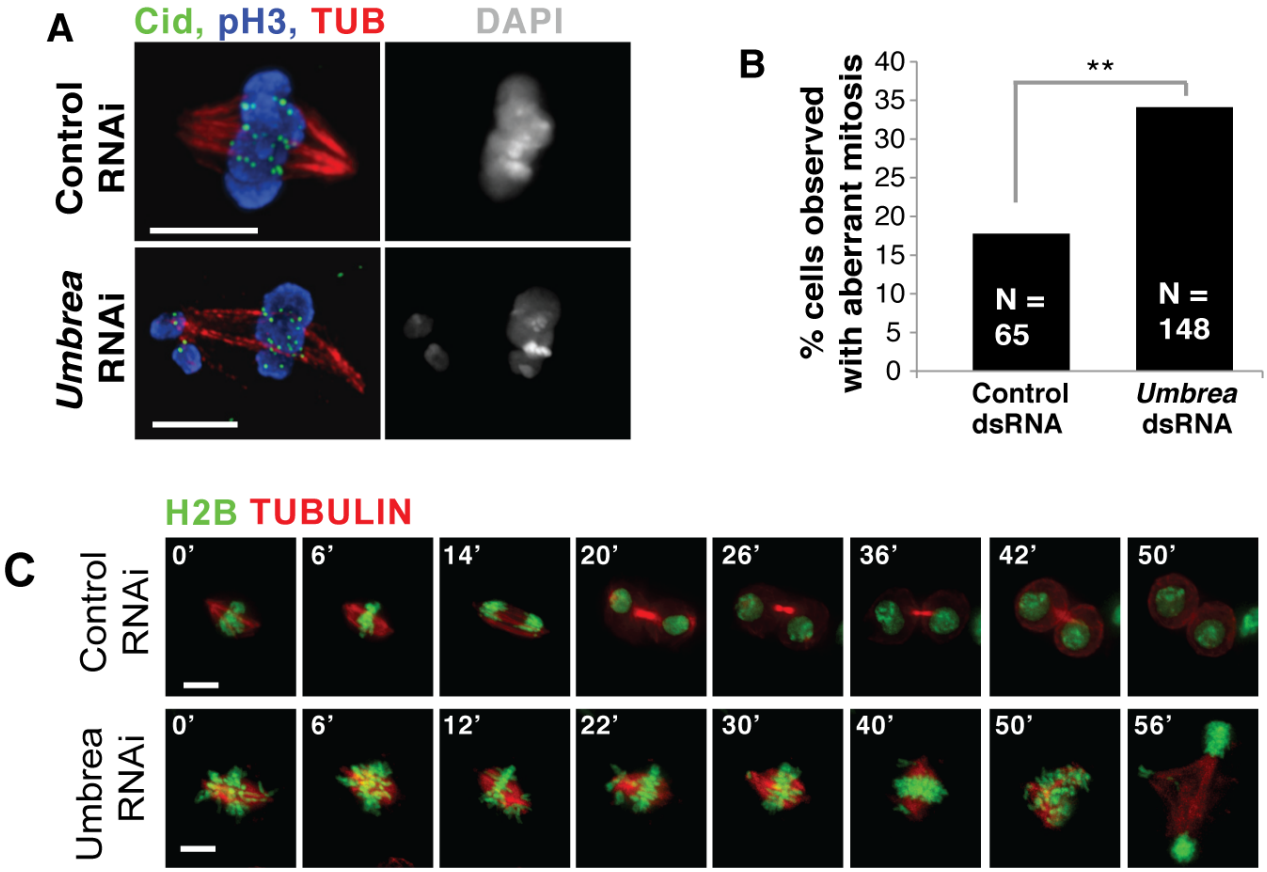


Figure 3.6 – Umbrea is required for error-free mitosis in *D. melanogaster* cultured cells. Cells were cultured in the presence of dsRNA homolgous to GFP control or to *Umbrea*, fixed, and stained for Cid (centromeres), pH3 (mitotic chromosomes), and tubulin to visualize spindle attachments (A). The frequency of mitotic errors (for instance, lagging chromosomes off the metaphase plate in (A)) was quantified (B), and revealed a statistically significant increase in observed errors in Umbrea depleted cells. Time-lapse imaging of control or Umbrea knockdown cells stably expressing GFP-histone H2B and RFP-Tubulin revealed further dynamic errors upon Umbrea depletion, including defects in spindle morphology, delayed anaphase, and lagging chromosomes.

Based on Leah's results, I predicted that Umbrea could function in one of several ways: upstream of Cid, downstream of Cid, or in a parallel pathway. Since *Umbrea* knockdown did not impair localization of the centromeric histone Cid (this was a qualitative, not quantitative observation), I infer that Umbrea functions downstream of Cid, or in an independent pathway prior to entry into mitosis. This inference is in line with my observation that Umbrea localizes to interphase centromeres but not mitotic centromeres. In contrast, Cid localizes to chromosomes during anaphase[107]. What the function of Umbrea actually is, will require further biochemical characterization of Umbrea interaction partners and genetic and cell biological characterization of the Umbrea depletion phenotype during development and in cultured cells.

The role of the chromodomain in the gain of centromere localization by Umbrea

I wished to understand how *Umbrea* gained centromere localization and essential function following duplication from *HP1B.* My phylogenetic studies of *Umbrea* indicated that a major event occurring after duplication was loss of the ancestral chromodomain (CD). HP1 chromodomains possess aromatic cages that bind with high affinity to the methylated tails of histone H3 molecules[150], thereby providing a "reader" for the epigenetic mark of heterochromatin. Furthermore, HP1 CDs play crucial roles in promoting heterochromatin spreading by facilitating CD-CD interactions between neighboring molecules on different nucleosomes[132], and even regulating the spreading by directly inhibiting the methyl binding ability of other HP1 molecules[133, 150].

To test the hypothesis that loss of the HP1B CD was the evolutionary event that allowed Umbrea to gain centromere localization. I first asked whether CD loss affects HP1B function. An HP1B-GFP fusion lacking the CD expressed in *D. melanogaster* cultured cells lost

heterochromatin localization, consistent with the requirement of HP1-CD for H3K9me binding [132, 151]. **(Figure 3.7,** compare with **Figure 3.1-3.2)** To investigate the role of CD loss in Umbrea neofunctionalization, I reconstructed an ancestral full-length state by fusing the HP1B CD and hinge to Umbrea-GFP, and expressed this fusion protein in *D. melanogaster* cultured cells. Addition of these HP1B domains to Umbrea reverted its localization from centromeres to heterochromatin, in a manner similar to localization seen for full-length HP1B **(Figure 3.7)**. These data suggest that loss of the ancestral CD was necessary for Umbrea to gain new function at centromere. Based on these observations, I propose that CD loss during *Umbrea* evolution was necessary but not sufficient for the gain of centromere localization and therefore essential function. In support of this idea, both full-length (in *D. fuyamai*) and CSD-only (in *D. eugracilis* and the *suzukii* clade) *Umbrea* genes have been lost, suggesting that these genes were never essential and likely pseudogenized following duplication due to genetic redundancy with *HP1B.* Our findings support a model of neofunctionalization following intermediate loss-of-function [152]. Increased dosage of HP1 proteins is known to be deleterious – overexpression of HP1A can lead to defects in chromosome segregation due to encroachment of heterochromatin into the centromere, as well as alterations in gene expression. Furthermore, overexpression of proteins (like full-length HP1s) that have the capacity to oligomerize [132] can lead to disorders of protein aggregation like Huntington's Disease in humans. Therefore, simple increased dosage of HP1B due to gene duplication may have incurred strong fitness costs, leading to rapid alterations including pseudogenization. In some lineages, pseudogenization proceeded to completion and loss of function for the entire protein (e.g *D. fuyamai*). In others, including perhaps the common ancestor of the *melanogaster* and *takahashii* subgroups, mutations only accumulated in the CD, resulting in loss (whether gradual or abrupt) of the chromodomain[65]. Interestingly, full-length *Umbrea* genes (from *D. gunungcola, D. ficusphila,* and *D. lucipennis*) encode proteins that either partially or fully retain residues in the CD known to be required for binding the H3K9me3 mark, suggesting that *Umbrea* in these lineages has

maintained part or all of the ancestral function of *HP1B.* In these lineages, preservation of *Umbrea* may have occurred through subfunctionalization of tissue expression or expression dosage.

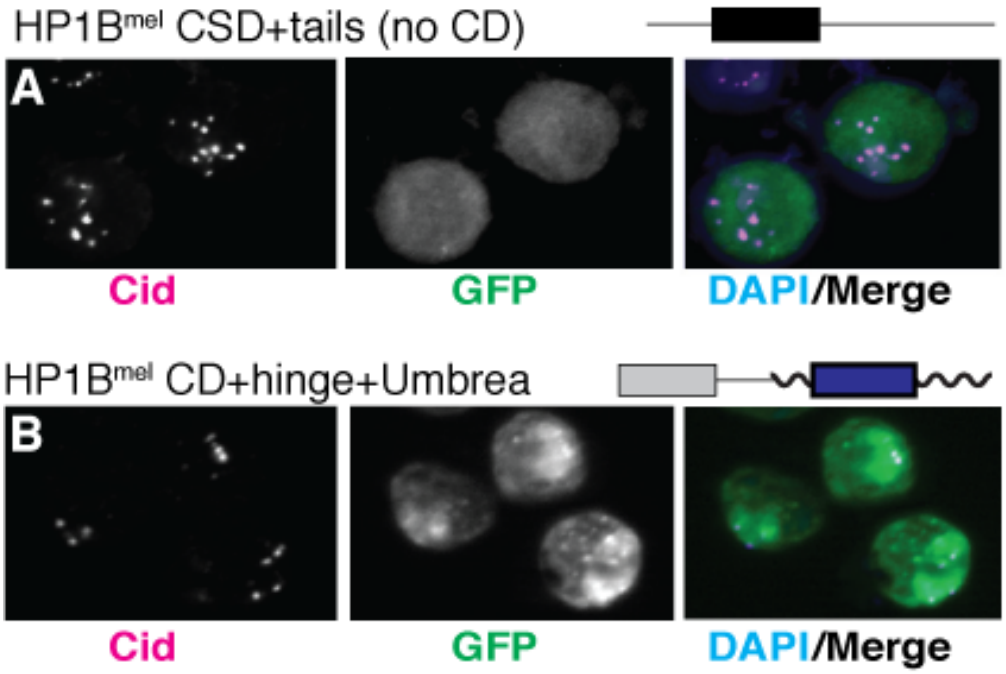


Figure 3.7 – Loss of the ancestral chromodomain was likely necessary but not sufficient for the gain of centromere localization by Umbrea. A GFP-HP1B fusion protein lacking the CD did not localize to centromeres (A). The addition of the HP1B CD to Umbrea^{melanogaster} resulted in delocalization of this fusion protein from centromeres to pericentric heterochromatin (B).

Role of the Umbrea CSD in gain of centromere localization

Since loss of the HP1B CD was necessary but not sufficient for the gain of centromere localization by Umbrea, I next investigated the consequences of evolutionary changes in the Umbrea-CSD. CSDs are only found in HP1-family proteins and mediate interactions with other HP1s [134] or proteins possessing degenerate PxVxL motifs [135, 153]. Interaction with PxVxL-containing peptides occur strictly in the context of CSD dimers, through interactions between a non-polar pit on the dimer interface and the hydrophobic peptide sequence. Mutations in amino acids that mediate PxVxL recognition alter specificity for different substrates in mouse cultured cells[135]. Therefore, I predicted that evolutionary change in the CSD leading to gain of centromeric localization could proceed either through mutations affecting dimerization or the CSD fold itself, or via mutation of the PxVxL recognition surface.

To understand what elements in the CSD of Umbrea had changed subsequent to its duplication from HP1B, I aligned amino acid sequences of diverse HP1B and Umbrea orthologs **(Figure 3.8)**. This analysis revealed that residues defining the CSD structural fold have been largely conserved between all Umbrea orthologs and HP1B, as well as those residues known to mediate dimerization between adjacent CSDs[134]. This result suggests that Umbrea possesses a bona fide CSD and that alteration of localization was not due to gross changes in protein structure.

In contrast, I found that three of the nine residues that mediate specificity for PxVxL-recognition[135] have changed along the branch leading to the *melanogaster* species subgroup (**Figure 3.8)**. These residues seemed to be ideal candidates for mediating the evolutionary gain of centromeric localization through new binding interactions. To test this hypothesis, I needed to find a minimal system with which to test necessity and sufficiency. I found that expression of a *D. melanogaster* GFP-Umbrea-CSD fusion protein (lacking the tail domains) could localize to

centromeres. However, this property was not shared among HP1B- or even other Umbrea-CSDs. For instance, neither 'parental' GFP-HP1B^{melanogaster}-CSD nor GFP-Umbrea^{pseudotakahashii}-CSD could localize to centromeric regions in *D. melanogaster* cells, although GFP-Umbrea^{pseudotakahashii}-CSD appeared to localize to some non-centromeric puncta in the nucleus. These data suggest that a discrete transition for centromere localization occurred in the Umbrea-CSD after divergence of the *melanogaster* and *takahashii* species subgroups, when changes in the PxVxL-specifying residues occurred. However, this was a correlation. To test causality of these residues for centromere localization, I generated a GFP fusion protein in which each of the three PxVxL interacting residues (C15, I57, F59) were reverted to the ancestral state **(Figure 3.8)**. This fusion protein, upon expression in *D. melanogaster* cells, delocalized the Umbrea^{melanogaster}-CSD from centromeres, but did not appear to alter the overall stability of the protein, since diffuse GFP signal was present throughout the nucleus. Moreover, replacement of the same residues in Umbrea^{pseudotakahashii}-CSD (which had not localized to centromeres when intact) to corresponding residues in Umbrea^{melanogaster} resulted in a gain of centromere localization. Therefore, I conclude that three residues in the PxVxL-recognition motif of Umbrea were necessary and sufficient for centromere localization in *D. melanogaster* cells. These data imply that the acquisition of these changes in Umbrea-CSD resulted in the gain of centromeric localization by Umbrea-CSD, which occurred in the common ancestor of the *melanogaster* species subgroup 5-7 million years ago. Consistent with this hypothesis, I found that GFP-Umbrea^{teisseiri} localizes to centromeres in *D. teissieri* cells, supporting the idea that centromere localization was a trait acquired in the common ancestor of *D. melanogaster* and *D. teisseiri* at the root of the *melanogaster* subgroup lineage. I suggest that centromeric localization may have also coincided with gain of essentiality, since *Umbrea* was lost three times prior to, but not after, CSD modification.

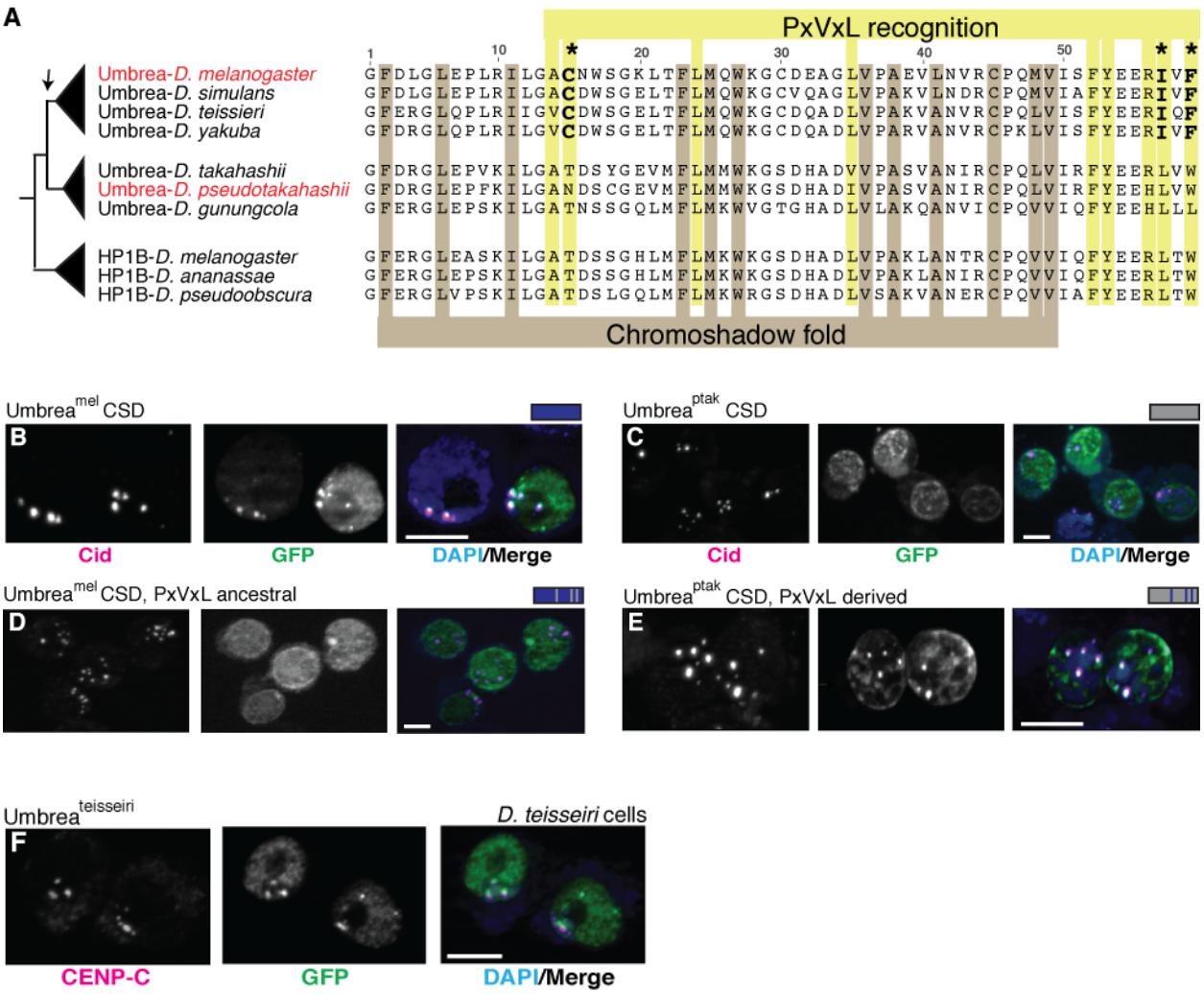


Figure 3.8 – An alignment of chromoshadow domains from HP1B and Umbrea proteins reveals features of conservation and divergence (A). The amino acid positions that form the chromoshadow fold itself are highly conserved between HP1B and Umbrea, suggesting that alteration to the protein fold itself was not responsible for the difference in nuclear localization. In contrast, three of the nine residues that govern protein-protein interaction via PxVxL motif recognition have diverged in the common ancestor of the *melanogaster* species subgroup (arrow). While the CSD of Umbrea^{melanogaster} tagged with GFP localizes to centromeres in *D. melanogaster* cultured cells (B), the CSD of Umbrea^{pseudotakahashii} does not (C). Mutation of the three divergent PxVxL recognition residues with the ancestral sequence from Umbrea^{pseudotakahashii} resulted in delocalization of the Umbrea^{melanogaster} CSD (D). Conversely, mutation to derived residues of PxVxL recognition caused Umbrea^{pseudotakahashii} to gain centromere localization (E), indicating

that mutation of PxVxL was both necessary and sufficient for the gain of centromere localization by Umbrea. These changes likely occurred in the common ancestor of *D. melanogaster* and *D. teisseiri*, since GFP-Umbrea^{teisseiri} localizes to centromeres in *D. teisseiri* cells (F).

Alteration of protein-protein interactions by Umbrea

To test the prediction that mutation of PxVxL recognition resulted in the ability of the Umbrea CSD centromere localization by the rewiring of protein interactions, I collaborated with Andreas Thomae in Axel Imhof's lab at the University of Munich. Andreas expressed FLAG-tagged Umbrea proteins in S2 cells, and performed proteomic analyses by mass spectrometry to identify proteins that co-immunoprecipitate with Umbrea **(Figure 3.9)**. We hypothesized that physical interaction partners of Umbrea would contain heterochromatin proteins or centromeric proteins, and that these proteins would contain degenerate PxVxL motifs. In confirmation of these predictions, Andreas found many chromatin factors that co-immunoprecipitated with Umbrea, including the heterochromatin proteins HP4/Hip, Su(var)2-HP2 and Su(var)3-9. These four proteins are classical suppressors of position effect varigation, implying that they are structural proteins required for the establishment, maintenance, and/or spread of heterochromatin. Indeed, Su(var)3-9 is the methyltransferase that propagates histone H3 lysine 9 methylation that recruits HP1A and is a hallmark of heterochromatin.

Another co-immunoprecipitating factor was the centromeric protein Cenp-C. In *Drosophila*, CENP-C is a constitutive component of the inner kinetochore and is interdependent with Cid for localization and function[154]. In mammalian cells, CENP-C binds directly to the centromeric nucleosome, through interactions with the tail of CENP-A and the histone H2A/H2B dimer, as well as making contact with centromeric DNA[71].

I found no overlap with protein partners of HP1B, which include the euchromatic proteins HP1C, Woc and Row [138], as well as several heterochromatin-associated proteins. These results suggest an extensive rewiring of the protein interaction network of Umbrea compared to HP1B and confirm our cell biological inferences.

Protein-protein interactions are thought to evolve as much as three times more slowly than the general rate of amino acid mutation[155], thus it is particularly strikingly how different the interactome profiles of HP1B and Umbrea are. It is perhaps notable that mutation of PxVxL recognition would potentially result in loss of some ancestral interactions, while simultaneously gaining a new set of protein partners. The strength of these new interactions could be subsequently modified through mutations of ancillary residues in the CSD[135]. While PxVxL modification clearly played a role in the gain of centromere localization by Umbrea, it remains likely that other differences in the chromoshadow domain or tails contributed to altered protein-protein interactions. Indeed, single mutations of the CSD appear to confer differential binding specificity for the *Drosophila* HP1A protein[156].

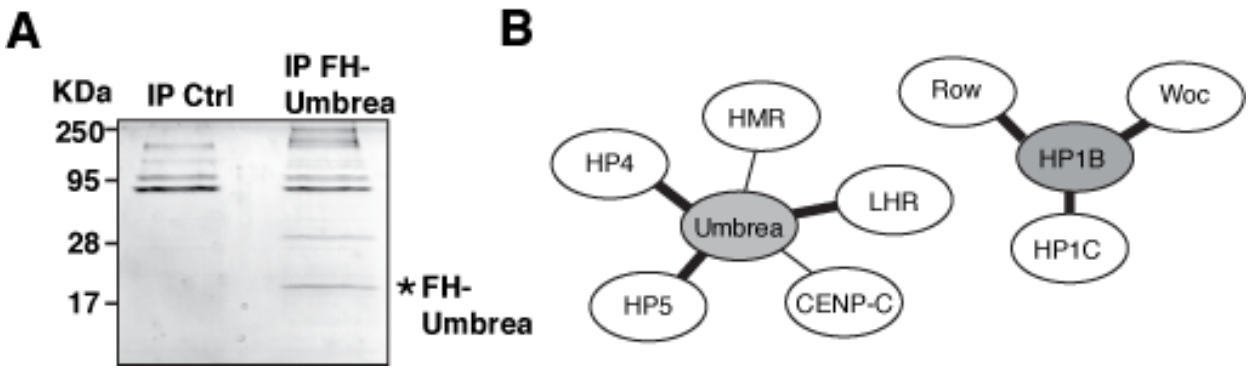


Figure 3.9 – Purification of protein complexes associated with tagged Umbrea proteins from *D. melanogaster* S2 cultured cells. Flag-HA tagged Umbrea was immunoprecipitated from cultured cells along with associated proteins. These proteins were subjected to trypsin digestion and mass spectrometry for identification. Protein interactors of Umbrea were enriched in heterochromatin associated proteins, centromeric proteins, and proteins that function to promote proper chromosome segregation in mitosis (B). The protein interaction profile of Umbrea revealed no overlap with that of HP1B[138], indicating that protein-protein interactions had been radically rewired during the course of Umbrea evolution.

Evolution of *Umbrea*

Neofunctionalization is associated with an increased rate of non-synonymous mutation[115, 129]. Indeed, the probability of neofunctionalization is increased with population size, since selection for new function could be more efficient[114]. In this sense, it should be expected that neofunctionalization could be more common in *Drosophila* than in mammals, which have smaller effective population sizes. Furthermore, many centromeric proteins in plants and animals evolve rapidly under positive selection. To understand the selective forces influencing the evolution of *Umbrea,* I tested for rapid evolution under positive selection using maximum likelihood methods implemented in the PAML software package[33, 34]. Furthermore, I tested for recurrent signatures of rapid evolution in the *HP1B* gene, over the same phylogenetic tree. I found that *HP1B* and *Umbrea* differed in what type of selective forces governed their evolution **(Figure 3.10)**. While *HP1B* did not show statistically significant evidence for positive selection, *Umbrea* did. Positive selection of *Umbrea* was both episodic and recurrent, with branches of accelerated evolution sprinked across the tree **(Figure 3.10).** These accelerated branches occurred both prior to and after the putative gain of centromere localization in the common ancestor of the *melanogaster* species subgroup. This result differs from previous examples of neofunctionalization, in which rapid evolution occurred at a single branch, followed by purifying selection[130, 157]. Furthermore, codons in both the short tail domains flanking the CSD and in the CSD itself evolved under statistically significant positive selection across the phylogeny. These features are consistent with a role for *Umbrea* in genetic conflict, since recurrent and lineage-specific positive selection is a hallmark of genes involved in pathogen defense[41] or centromere-drive [5, 14, 28, 41, 158]

Despite significant signals of positive selection via PAML, I found that *Umbrea* does not show statistically significant evidence for positive selection by the McDonald-Kreitman test, which draws power from polymorphism segregating within species to make inferences about whether

differences between species exceed expectations. *Umbrea* has an extremely high number of nonsynonymous fixed differences between *D. melanogaster* and *D. simulans*, but very few polymorphisms within species (Rf=30, Rp=8, Sf=9, Sp=5). This low occurance of polymorphism could potentially indicate that a recent selective sweep has taken place, although I did not pursue this possibility further.

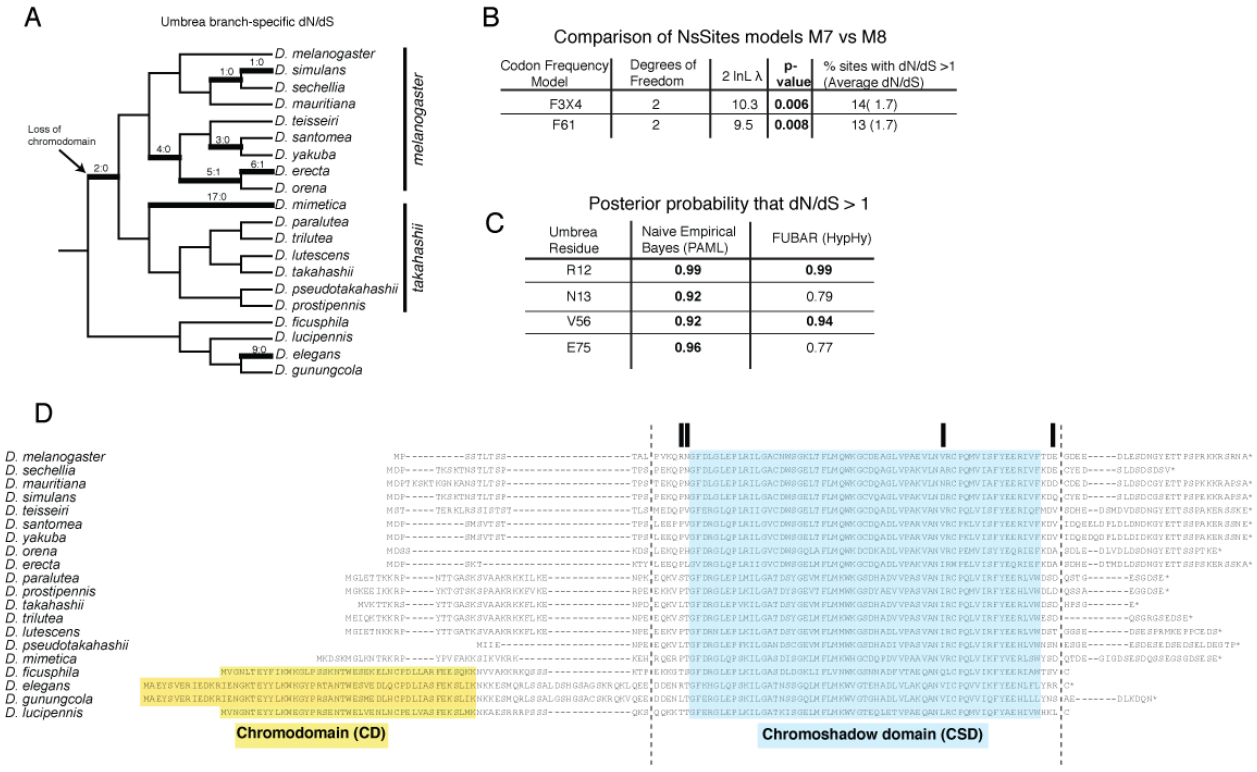


Figure 3.10 – Summary of evolutionary analysis of *Umbrea* using coding sequence alignments from the species indicated. Since *Umbrea* genes in this analysis included those both with and without chromodomains, only the chromoshadow domain and short regions of the tails were analyzed (alignment shown in (D)). Branch-specific dN/dS ratios (A) reveal that *Umbrea* has undergone episodic periods of accelerated amino acid evolution, including in lineages with and without chromodomains, and in lineages before and after the proposed gain of centromeric localization and essential function. On the whole-gene level, *Umbrea* shows a statistically significant signature of positive selection by comparing models M7 versus M8 in PAML (B). Sites identified using site-specific models identify three residues in the Umbrea tail domains and one in the chromoshadow domain as evolving under positive selection (C).

Species-specificity in centromere localization by Umbrea

My evolutionary analyses indicated that the most recent innovations in *Umbrea* occurred in the short tail sequences that flank the CSD **(Figure 3.10)**. The N-terminal tail domain was derived from a degenerated portion of the ancestral hinge region. The HP1 hinge domain in other systems is important in HP1 function through binding nucleic acids in a CD- and CSD-independent manner[159]. In contrast, the C-terminal tail of HP1 proteins is referred to as the C-terminal extension, or CTE[156], and is particularly variable in length. CTEs as short as only five amino acids can be functional. In HP1A in *Drosophila,* in vitro assays show that charged residues in the CTE play an important role in assisting in partner protein specificity and strength of binding [156]. Finally, the Umbrea CTE appears to contain a nuclear localization sequence (NLS, typically a strength of positively charged residues such as lysine or arginine), although the presence and position of the Umbrea NLS appears to be highly variable across orthologs. For instance, in the *takahashii* subgroup, the Umbrea N-terminal tail contains the NLS, whereas it is found in the CTE in the *melanogaster* subgroup. Of note, in the *melanogaster* subgroup, Umbrea from *D. sechellia* and *D. orena* appear to lack an NLS altogether. Nuclear and centromeric localization for these particular orthologs has not been tested. In summary, the Umbrea tail domains are evolutionarily dynamic and likely affect function.

I therefore tested how these changes contributed to *Umbrea* neofunctionalization, by expressing fusion proteins in *D. melanogaster* cultured cells and assessing centromeric localization. While HP1B^{mel}-CSD alone fused to GFP showed no discrete localization, addition of Umbrea^{mel}-tails was sufficient to confer centromere localization **(Figure 3.11)**. These data indicate that Umbrea may target centromeres using both the CSD, and the tail domains, separately. While the CSD is likely to target to centromeres via protein-protein interactions, Umbrea-tails may target centromeric nucleic acids, by analogy to the hinge region of mammalian HP1alpha [159], which binds to DNA. Since centromeric DNA sequence can diverge rapidly[19], I hypothesized that

evolution of the Umbrea-tails provides species-specific localization. I therefore tested whether Umbrea orthologs could localize to centromeres in *D. melanogaster* cells, by cloning *Umbrea* genes from closely related species and generating GFP-fusion proteins. Upon expression of these orthologous fusion proteins in *D. melanogaster* cultured cells, I found that evolutionary distance of the orthologous gene from *D. melanogaster* was correlated with its ability to localize to centromeres. For example, GFP-Umbrea^{simulans} localized well to *D. melanogaster* centromeres (**Figure 3.11).** However, GFP-Umbrea^{teisseiri} and GFP-Umbrea^{yakuba} did not, localizing instead to distinct foci adjacent to centromeres in the nucleus. These foci did not colocalize with telomeric markers, suggesting that they may be pericentric or embedded within euchromatin. This mislocalization is due to species-specificity, since Umbrea^{teisseiri} appropriately localizes to centromeres when expressed in its native context in *D. teissieri* cells(**Figure 3.11).** While positive selection of *Umbrea* preceded its centromeric localization, the mislocalization of Umbrea orthologs from the *melanogaster* species subgroup in *D. melanogaster* cells suggests that continued positive selection in the *melanogaster* species subgroup resulted in species-specific centromere targeting, reminiscent of CenH3/Cid in *Drosophila* [85].

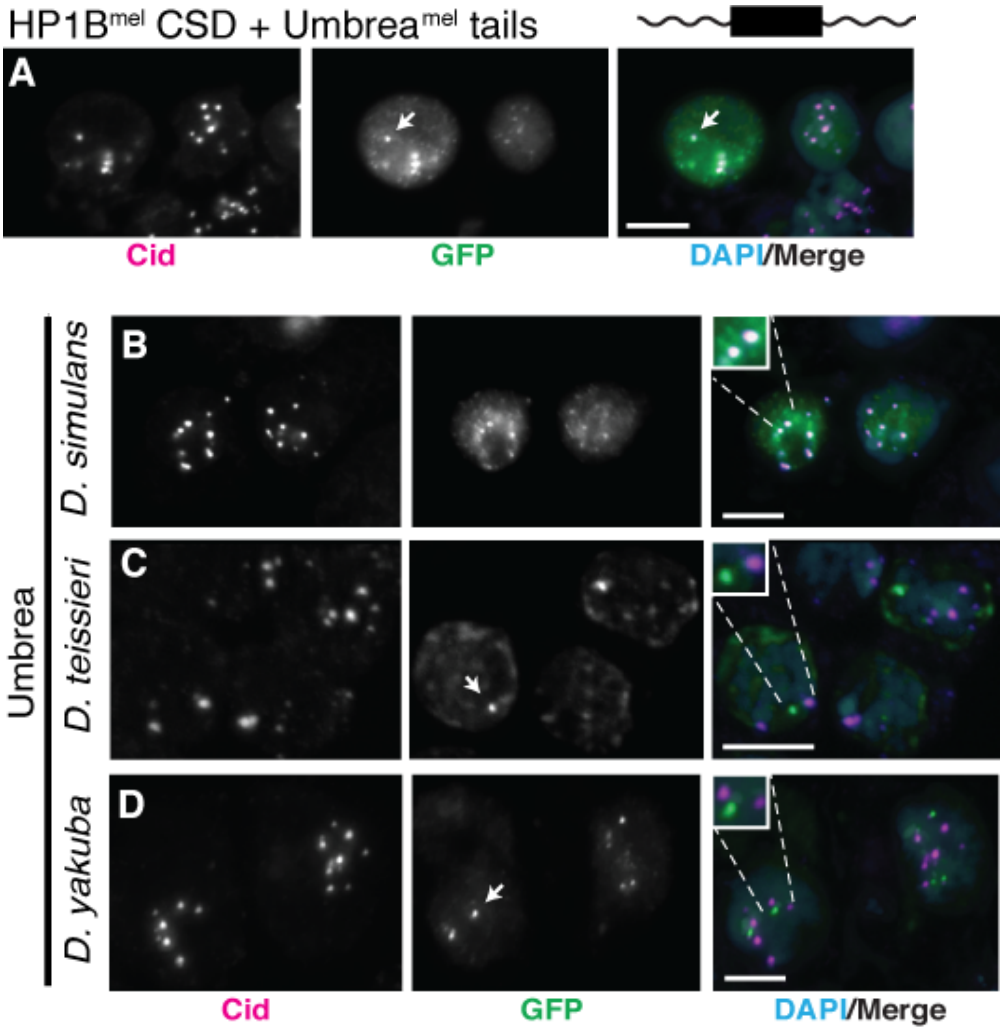


Figure 3.11 – Centromere targeting by Umbrea tail domains may confer species-specificity. Umbrea tails fused to GFP-HP1B-CD were sufficient to confer centromere localization (HP1B tails do not have this capability) (A), as visualized by antibody colocalization with Cid (pink), and DAPI (blue) in *D. melanogaster* cells. Rapid evolution of Umbrea (concentrated in the tail domains), resulted in species-specific targeting to *D. melanogaster* centromeres that was dependent on evolutionary divergence. Umbrea^{simulans} (B) colocalized with *D. melanogaster* centromeres, while Umbrea^{teisseiri} and Umbrea^{yakuba} did not, localizing instead to non-centromeric nuclear foci.

The CSD is epistatic to the tail domains in localizing Umbrea to the centromere

The Umbrea tail domains are sufficient to localize the HP1B CSD to centromeres. Since the tails evolve rapidly, I hypothesized that they were the primary determinants for mislocalization of Umbrea orthologs expressed in *D. melanogaster* cells (**Figure 3.11).** To test this idea, I created chimeric GFP fusion proteins in which the tail domains of Umbrea^{teisseiri} or Umbrea^{yakuba} were stitched to the Umbrea^{melanogaster} CSD, and expressed them in *D. melanogaster* cultured cells (data not shown). Surprisingly, all of these fusion proteins appeared to localize equally well to centromeres, suggesting that mislocalization of GFP-Umbrea orthologs is not solely dependent on the tail domains. Only one of the four residues evolving under recurrent positive selection lies in the CSD (instead of the tail domains), yet it is possible that this position confers species-specificity in centromere localization. An alternate explanation could be due to the fact that the residues evolving under positive selection in Umbrea were identified using species alignments that were broader than just the *melanogaster* species subgroup. This could implicate other positions even in the CSD as being required for species-specificity, and does not rule out the role of adaptive evolution for such positions, although alternate methods are needed for their identification.

What promotes lineage-specific gain of essential function?

A key question that is raised by the observation of the rapid gain of essential function in young duplicate genes in *Drosophila* is WHY these genes are essential[117]. *Drosophila* species that diverged prior to the birth of these genes do not require them for viability, nor do species that diverged after birth but prior to the gain of essential function (see the *takahashii* subgroup in the example of *Umbrea*). Some of these genes may have become essential due to non-adaptive processes. For example, one of the subunits of the hexameric ring complex of the V-ATPase proton pump was born by gene duplication from another subunit[120]. Loss of function of complementary domains in the parent and daughter gene rendered each essential for the

function of the complex as a whole. In this example, the mutations that rendered the daughter gene essential were not adaptive - in fact they were deleterious. Therefore, increased complexity arose simply due to the act of gene duplication and the acquisition of high probability inactivating mutations in one domain. A similar process could explain some of the young duplicate genes that are now essential for viability in *D. melanogaster.* However, this does not provide a satisfying explanation for the gain of essential function by *Umbrea.* Furthermore, many of these young *D. melanogaster* genes seem unlikely to have gained function by non-adaptive processes. Some appear to be related to pathogen defense. Others evolve under very high non-synonymous mutation rates indicative of positive selection. In the example of the V-ATPase duplication[120], the parent gene was essential. However, in *D. melanogaster*, young essential genes can arise from either essential or non-essential (as in the case of *Umbrea*) parental genes. Therefore, the gain of essential function appears to be independent of the function of the parent gene. I propose that another mechanism that would promote the lineage-specific acquisition of essential function in young genes is genetic conflict. Genetic conflicts are by their very nature lineage-specific. For example, species face different spectra of pathogens depending on their environmental niche, putting differential selective pressure on the immunity factors encoded by each species[41]. Since each separate genetic conflict involves two parties (pathogen and host, in this example), each conflict may resolve in an independent fashion. TRIM5alpha is a potent antiviral restriction factor in mammals and shows highly lineage-specific evolutionary patterns, including rampant gene duplication in rodents[160], and recurrent positive selection in specific domains in primates[161]. For centromere-mediated female meiotic drive, again, genetic conflict is likely to play out in a highly lineage-specific fashion. The molecular events leading to the expansion of centromeric satellite DNA arrays are by their very nature lineage-specific. The fusion of telocentric chromosomes to form Robertsonian chromosomes is highly idiosyncratic. Yet, after these rare events have occurred, the organism faces a new challenge and new selective pressure. Young duplicate genes are uniquely poised in these

scenarios, perhaps particularly so for intrinsic cellular processes where single-copy parental genes may face high evolutionary constraint (for example, genes encoding centromeric proteins). In the case of *Umbrea*, selection as a modifier of centromere drive could have inherently led to the acquisition of further protein-protein interactions and essentiality. This prediction, that genetic conflict is an intrinsic driver of the rapid gain of essential function in some genes, will require more experimental support on the single gene level for validation.

Umbrea provides an excellent test case for the hypothesis that genetic conflict drives the gain of new essential function. Indeed, implication of Umbrea function at the centromere (rapidly evolving in its own right) is significant. Further dissection of Umbrea function (through identification of the molecular basis for chromosome missegregation upon knockdown, for example) could shed light on what aspect of centromere biology is vulnerable to dynamic evolutionary processes.

The biological function of Umbrea

What could the biological function of Umbrea be? Removal of Umbrea function during development results in lethality at the larval-pupal transition. This phenotype is one characteristic of genes controlling vital cell cycle functions in Drosophila [162], in part because many genes controlling cell-cycle functions are heavily maternally deposited into the egg in preparation for the rapid embryonic divisions. Therefore, homozygous lethal mutations in cell cycle genes would only manifest their phenotype after the maternal contribution of wild-type protein was depleted which is thought to occur at the larval-pupal transition.

Further insight can perhaps be gained by an examination of the protein-protein interactions partners of Umbrea. Indeed, since Umbrea likely gained essential function through a gain in protein-protein interactions, examination of the function of Umbrea interactors is likely to be

informative. Many of the interaction partners predicted from genome-wide yeast two-hybrid screens, or by immunoprecipation and mass spectroscopy contain canonical or degenerate PxVxL CSD-binding motifs that could perhaps mediate interactions with Umbrea.

One of the most tantalizing interaction partners of Umbrea is the inner kinetochore protein CENP-C. This association could imply that Umbrea functions in a direct role in specifying or maintaining the structure of the centromere. However, this seems unlikely to be true. CENP-C is present at the centromere throughout the cell cycle, binds directly to CENP-A (Cid), and is required for maintaining stable CENP-A association at centromeres [71, 107]. In contrast, Umbrea-depletion does not appear to cause any reduction or change in localization of Cid during interphase or on mitotic chromosomes (personal communication, Leah Rosin, University of Connecticut). This lack of effect on Cid localization could suggest that Umbrea functions downstream of Cid to promote proper kinetochore function. Alternatively, Umbrea could act independently of kinetochore function at centromeres, perhaps by preventing the encroachment of pericentric heterochromatin into centromeres [113]. These separate possibilities could be distinguished by staining WT and Umbrea RNAi knockdown cells with antibodies that recognize various inner and outer kinetochore factors that are known to be dependent upon Cid incorporation (for example, the major microtubule binding protein NDC80[163]). A lack of differences in localization of outer kinetochore proteins like NDC80 upon ablation of Umbrea would provide some evidence to support Umbrea function in a parallel pathway (like centromere clustering, see below).

Function of Umbrea as a boundary element between heterochromatin and centromeres

In this vein, Umbrea may act as a type of boundary element factor, interacting with both heterochromatin proteins and centromere proteins to ensure the stability of each domain. For example, Umbrea could “block” interactions between heterochromatin proteins (HP1A, for

example) by acting as a dominant negative chromoshadow domain. For instance, Umbrea interacts with Su(var)3-9, the H3K9 methyltransferase that also interacts with HP1A[164]. Inhibition of this interaction by Umbrea would thereby prevent deleterious spreading of heterochromatin into centromeric domains [113]. Of further interest is the fact that CSDs can facilitate heterodimeric interactions as well as homodimeric interactions [134, 165], at least in *in vitro* experiments. This observation raises the possibility that Umbrea may heterodimerize with other HP1s, although it is important to note that no HP1s immunoprecipitated with Umbrea in our mass spectrometry experiments in cultured cells. This does not exclude the possibility that interactions with other HP1s via CSD heterodimerization could exist *in vivo*, and indeed, HP1A appears to form a heterotetramer with two of Umbrea's strongest interaction partners, Lhr and Hmr[166], although these data also derive from cultured cells. Alternatively, through interactions with centromeric proteins like CENP-C, Umbrea could dictate the extent of the centromeric domain by "recruiting" heterochromatin proteins. This function could be directly relevant to centromere drive suppression, by effectively blocking centromere protein recruitment onto centromeric satellite DNA expansions.

Recent proteomic characterization of proteins that interact with HP1A in *D. melanogaster* cells identified a number of factors that overlap with the protein partners of Umbrea[167]. Indeed, of 35 proteins identified in this study, 12 overlapped with the list of Umbrea interactors. Two of these factors are particularly interesting with respect to the idea that Umbrea could be involved in establishment of a boundary between heterochromatic and centromeric chromatin domains. The first protein, dADD1, possesses an ADD domain related to the ATP-dependent chromatin remodeling protein ATRX, and physically binds to the *Drosophila* homolog of ATRX, XNP. Umbrea shares this interaction with XNP. Interestingly, XNP primarily associates with a large block of satellite DNA on the X chromosome [168]. Furthermore, dADD1 is a bona fide suppressor of position effect varigation, indicating a functional role in the maintenance of

heterochromatin. The second Umbrea-interactor identified in the HP1 proteomic study was CG3680 (called HIPP1 for HP1 and insulator partner protein1). HIPP1 was the second most abundant protein associated with Umbrea in cultured cells, after only Umbrea self-association. HIPP1 localizes to pericentric heterochromatin and possesses a crotonase-fold which has been implicated in the acetylation of chromatin. Interestingly, HIPP1 interacts with all three members of the famous *gypsy* chromatin insulator comples – SU(HW), MOD(MDG4), and CP190. Umbrea shares the interaction with SU(HW). Since insulator elements have long been known to protect euchromatic regions from the spread of heterochromatin [169] and heterochromatin encroachment into centromeric regions is known to be deleterious [113], it seems plausible that Umbrea functions at heterochromatin-centromeres boundary through interactions with some or all of its identified interaction partners.

Interaction between Umbrea and nucleoplasmin

This idea of a boundary element could be further supported by the interaction between Umbrea and the protein nucleoplasmin or NLP. NLP localizes to centromeres and interacts with the chromatin insulator protein CTCF to function in promoting a poorly-understood phenomenon called centromere clustering. Centromeres cluster in peri-nucleolar bodies during interphase in many cell types in many organisms[170]. The consequences of disrupting centromere clustering by depletion of NLP or CTCF and its adapator molecule Modulo, appear to be pleiotropic, but could be the outcome of disruption of pericentric heterochromatin. For example, NLP knockdown results in overexpression of pericentric and heterochromatin transposable elements, as well as chromosome segregation errors during mitosis. These features of NLP biology resemble features of Umbrea biology – interphase localization to centromeres, protein-protein interaction, and knockdown phenotype. Further experimentation will be required to interrogate the interaction between Umbrea and NLP to see if they function in a common pathway.

Interaction between Umbrea and Lhr/Hmr

Besides CENP-C, the other identified interaction partners of Umbrea that have a characterized localization or function at the centromere are the proteins Lethal Hybrid Rescue (Lhr) and Hybrid Male Rescue (Hmr). These factors play an important role in enforcing species boundaries between *D. melanogaster* and *D. simulans*, since divergence in expression and primary sequence appear to be important in post-zygotic reproductive isolation in the form of hybrid male lethality[166, 171]. I pursued the basis for this functional divergence as part of my thesis, which is addressed in the following section. Of note, both Lhr and Hmr interact with NLP, suggesting that Umbrea and NLP/Lhr/Hmr function together in centromere clustering.

Interaction between Umbrea and Nipped-B

Another potential functional role for Umbrea could come from its interaction with Nipped B, a SCC2 homolog which acts to load cohesin molecules onto chromosomes to promote cohesion between sister chromatids prior to anaphase in mitotic cells[172]. Raquel Oliviera, an investigator at the Gulbenkian Institute in Portugal, is currently looking into a role for Umbrea in the Nipped B cohesion pathway. I sent her flies expressing GFP-Umbrea under the control of a UAS promoter. Interestingly, by examining neuroblast cells expressing GFP-Umbrea, Raquel found centromere localization of GFP-Umbrea at metaphase of mitosis. This result is at odds with all localization studies I had previously performed (using overexpressed fusion proteins, or anti-Umbrea antibodies), and suggests that the function of Umbrea could be cell-type specific. A further line of evidence supporting a functional role for an Umbrea-Nipped B interaction is that Nipped-B mutants die at the pupal-larval transition – the same stage of arrect caused by depletion of Umbrea. Furthermore, the human homolog Nipped-B-like interacts with the HP1 protein CBX5 via a canonical PxVxL motif[173], PVVVL. The corresponding positions (identified using an amino acid alignment with the human homolog) in *D. melanogaster* Nipped-B are

divergent: VRVCI, which could imply divergence in interaction between HP1A (homologous to CBX5) and Umbrea in Nipped B binding.

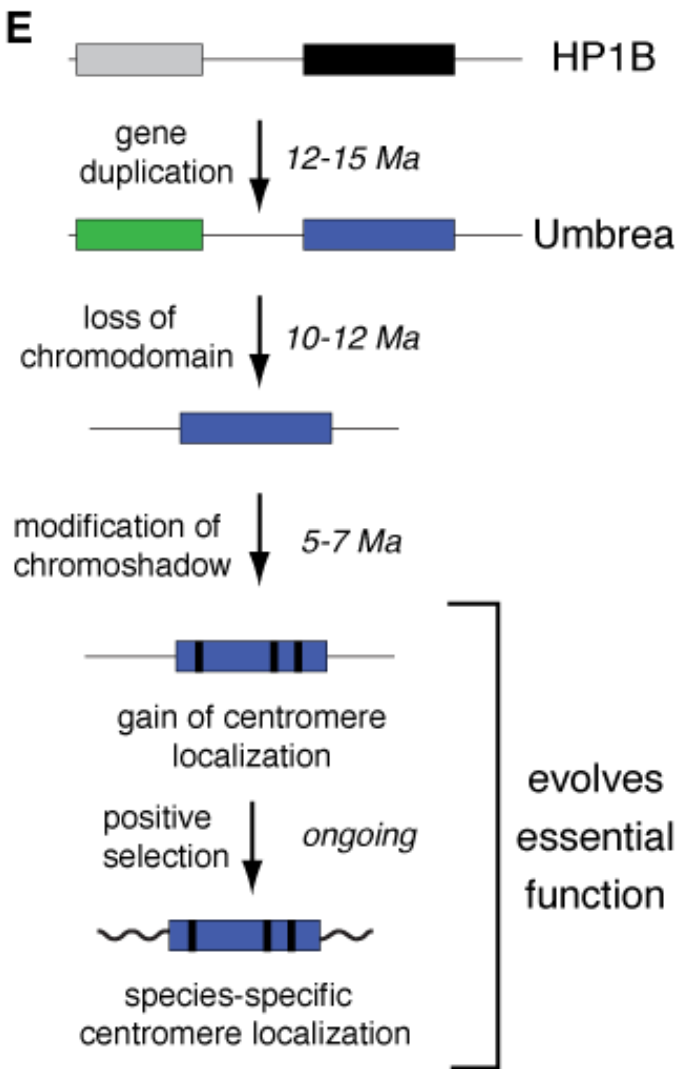


Figure 3.12– Model for the gain of essential centromere function by *Umbrea* following duplication from *HP1B.* Key events experimentally determined to be likely important for the stepwise gain of centromere function included the loss of the ancestral chromodomain 10-12 million years ago. Subsequent modification of the remaining chromoshadow domain along the branch leading to the *melanogaster* subgroup at 5-7 million years ago likely altered protein-protein interactions and led to centromere localization. Ongoing and recurrent positive selection in the Umbrea tail domains has led to species-specific localization within closely related species of the *melanogaster* subgroup.

Cross-species gene-editing of *Umbrea* using CRISPR-mediated homology directed repair

My studies have shown that Umbrea localizes to centromeres, and is essential for development in *D. melanogaster.* Furthermore, I have shown that rapid evolution (in particular in the tail domains) seems to confer species specificity to Umbrea localization, although there may be epistasis with the CSD. Left unanswered, however, is the question of what the consequences on development are for such rapid evolution. To test the functional consequences of rapid evolution in *Umbrea*, I propose to harness the power of homology-directed repair gene-swap technology via the CRISPR-Cas9 system. In essence, this approach allows for endogenous gene replacement, leaving intact the regulatory sequences required for native expression (see Cid project in previous section). This strategy is particularly important for *Umbrea*, because I had previously been unable to identify the noncoding sequences necessary for Umbrea expression (likely within other introns of the *Dumpy* gene) in prior transgene-rescue experiments. A particularly good test case would seem to be Umbrea^{simulans}, which appears to localize well to centromeres when expressed in *D. melanogaster* cells, but has many nonsynonymous mutations compared to Umbrea^{melanogaster}. I predict that replacing Umbrea^{melanogaster} with Umbrea^{simulans} will not affect development. Umbrea is predominantly expressed in the testes, and it is in the testes that the consequences of centromere drive are likely to be most severe. Therefore, I predict that *D. melanogaster* flies exclusively expressing Umbrea^{simulans} will be homozygous viable but male-sterile. An alternate outcome could be that, despite its localization to *D. melanogaster* centromeres, Umbrea^{simulans} does not possess cross-species functionality and flies die. This result could indicate that residues that are different between *D. melanogaster* and *D. simulans* other than those that localize Umbrea to centromeres are required for full function. This approach could yield insight into the selective forces driving *Umbrea* evolution. In addition, since testes-specific expression is likely to be the ancestral expression pattern of *Umbrea* following gene duplication (young genes in *D. melananogaster* and *D. pseudoobscura* are enriched for male- and testes-biased expression

patterns[129]), this approach could yield insight into the ancestral function of Umbrea, prior to its ubiquitous expression and essentiality.

Dissecting the gain of ubiquitous expression by *Umbrea*

Further insight into the origins of the gain of essential function by *Umbrea* could come from expression analyses. *Umbrea*^{melanogaster} is expressed ubiquitously at low levels, but is relatively highly expressed in the male germline. Since young genes often exhibit male-biased expression [129], it is highly likely that *Umbrea's* ancestral expression pattern was male-germline restricted despite its origination from the ubiquitously-expressed *HP1B*, and that ubiquitous expression of *Umbrea* evolved later. An alternate possibility is that *Umbrea* utilized enhancers regulating *Dumpy* expression immediately following duplication. *Dumpy* is expressed in a subset of tissue types, including the testes (at low levels), but is not ubiquitously expressed (Gelbart and Emmert 2013, FlyBase). It is tempting to speculate therefore that ubiquitous expression evolved simultaneous or subsequent to the gain of centromere localization mapped to the common ancestor of the *melanogaster* subgroup. To test this hypothesis, I propose performing RT-PCR analyses in key species of the *HP1B* and *Umbrea* phylogeny (see **Figure 3.5)**. While alterations in gene expression through the gain or loss of enhancer elements is widely documented and believed to be a major contributor to morphological and phenotypic evolution[146], few studies have examined alterations in gene expression following the evolutionary trajectory of a young duplicate gene. One way by which this can occur is through subfunctionalization of expression, in which the daughter gene retains part of the ancestral expression pattern (both *Umbrea* and *HP1B* lack introns, leaving open the possibility of non-reverse transcription dependent duplication mechanism). While this may hold for Umbrea, it seems equally plausible that a transitory male-restricted expression pattern intervened. Regardless, the question of *Umbrea* expression is an important one, since proteins that dimerize (like CSDs) or form oligomeric complexes are exquisitely sensitive to changes in

gene dosage or expression[174]. The proposed research into the origins of *Umbrea* ubiquitous expression could yield important insight into how regulatory evolution correlates with subcellular localization, and the gain of essential function.

Umbrea and the HP1 "revolving-door"

Umbrea is but one of a host of CSD-only HP1 duplicate genes in the genus *Drosophila*[136]. One curious observation made by Levine *et al.* is that testis-specific CSD-only genes appear in a "revolving door" evolutionary pattern of gene gain and loss in which, in any given *Drosophila* species, the total number of HP1 family members is constant but the genes themselves experience birth and death across the phylogeny.

One promising candidate to test the prediction that Umbrea-like factors have evolved convergently as part of the "revolving door" is the *D. pseudoobscura* gene *Skim. Skim* is a CSD-only gene that is present only in *D. pseudoobscura* and its closely related sister species and, similar to *Umbrea,* was born by duplication from *HP1B*. Unlike many of the other CSD-only genes in *Drosophila*, *Skim* shares characteristics of the derived PxVxL recognition motif with *Umbrea.* This is notable because I found that single amino-acid mutations from the ancestral position to the derived state were sufficient to drive centromere localization of Umbrea. Therefore, *Skim* represents a strong candidate for convergent evolution of centromere localization following gene duplication from *HP1B*. Lisa Kursel, a current graduate student in the Malik lab, tested the cytological localization of Skim by tagging it with GFP and expressing it in *D. pseudoobscura* cultured cells. She found that overexpressed GFP-Skim localized to subdomains of pericentric heterochromatin, by colocalization with anti-H3K9me3 (Lisa Kursel, personal communication). However, Skim did not appear to localize to discrete foci reminiscent of centromeres. Furthermore, when expressed in *D. melanogaster* cultured cells, GFP-Skim also localized to pericentric heterochromatin. This localization pattern, and the potential for the

convergent evolution of HP1B-derived CSD-only centromere localization, demand further investigation.

Identifying other young essential chromatin factors

Of the list of 195 young genes in *D. melanogaster*[117] tested for essentiality by RNAi knockdown during development, only 2 appear to be strong candidates to encode proteins that might bind DNA or function at chromosomes **(CG17802** and **CG3347).** Indeed, only one other besides *Umbrea* seems to be a candidate for similar centromeric neofunctionalization. *CG17802* is a young duplicate gene, born after the divergence of the *D. ananassae* ancestor from the *melanogaster* and *suzukii* subgroups. It possesses several C2H2 zinc fingers, and is in a cluster of other putatively related C2H2 genes, including its parent gene. Expression analysis shows that *CG17802* is expressed in a pattern tantalizingly similar to *Umbrea* – low but ubiquitous expression in most developing tissues, but expressed at high levels in the germline (Flybase, ModENCODE data). Unlike Umbrea, *CG17802* is expressed in the female germline at higher levels than the male germline. Furthermore, it is rapidly evolving in the *D. simulans* lineage, according to McDonald-Kreitman analysis[175], indicative of a potential role in genetic conflict. In support of this notion, and of particular interest, genome-wide yeast two-hybrid protein-protein interaction screens suggest that CG17802 interacts physically with several proteins that function in chromosome segregation in meiosis and mitosis [176]. One of these proteins is Shugoshin, a rapidly evolving protein that is essential during mitosis because it ensures that centromere cohesion is maintained until proper kinetochore-microtubule interactions are formed. Shugoshin is also critical for meiotic fidelity, by preventing sister chromatid centromeres from dissociating during the first meiotic division[177]. Two CG17802-interacting proteins are also C2H2 zinc finger proteins – the rapidly evolving protein Weckle, and the essential protein L(2)S5379.

I proposed to initially investigate CG17802 by confirming its essentiality for *D. melanogaster* development. Michelle Hays, a current graduate student in the Malik lab, performed an RNAi knockdown experiment during her rotation in the winter of 2013-14. She found that ubiquitously expressing dsRNA homologous to the coding sequence of *CG17802* throughout development resulted in a dramatic male-biased lethality. Indeed, compared to control flies, CG17802-knockdown males exhibited 100% lethality, while knockdown females survived. Interestingly, these knockdown females appeared to be sterile. I confirmed these results by repeating the same genetic cross. I found that 16% of expected CG17802 RNAi adults emerged from these crosses, and that the sex-ratio of RNAi adults was skewed towards more females – whereas 24% of expected RNAi females emerged from crosses, only 1% of expected males did (N=507 total F1 progeny scored). Furthermore, these females had compromised fertility (approximately 50% fewer progeny), compared to control females crossed to the same w^{1118} control males.

These results suggest that *CG17802* has evolved at least two distinct essential functions since birth, although the results do not distinguish between evolutionary models of duplicate gene preservation. One function may be male-specific. The interaction with Shugoshin implies that CG17802 functions during mitosis – therefore, the male-specific function could be dependent upon the Y chromosome. Alternatively, CG17802 could interact with the dosage compensation complex, which binds to the male X chromosome in order to establish transcriptional balance with the autosomes. However, CG17802 is clearly also important in females, so it may play a role in a common feature of development shared by both sexes. Furthermore, CG17802 may play a role in fertility, although this hypothesis requires further investigation. RNAi females exhibited compromised fertility, however, if CG17802 is required for development (for instance, in cell proliferation), the reduced fertility of these animals may be due to gross defects in tissue development. Alternatively, CG17802 could function specifically during female meiosis, as does

its interaction partner Shugoshin. The biological function, nuclear localization, and evolution of essentiality of CG17802 appear to be fruitful areas for future research.

Umbrea Materials and Methods (abbreviated from Ross et al, *Science* 2013)

Umbrea presence/absence

Genomic DNA was prepared from 32 species of Drosophila (Fig. 2A), using previously described methods [28]. Primers HSM208: CAACCAGTTCGCATGAAAATGCATAATCAATC and HSM215: CCCATTGAGCGTATCATTGGTGCATCTTCATCT were designed to sequences flanking *Umbrea* in the first intron of *Dumpy.* These sequences are conserved from *D. melanogaster* to *D. pseudoobscura* and therefore unambiguously amplified the *Umbrea* syntenic locus in all species tested. The presence or absence of *Umbrea* was identified by size differences on an agarose gel, and confirmed by sequencing. Umbrea sequences have been deposited at Genbank and at FlyBase.

Evolutionary analysis

Umbrea and HP1B protein sequences from different species were aligned using ClustalX [178]and manually adjusted for gaps. Nucleotide alignments that honored codon positions were made using Pal2Nal [179]. Phylogenies were created using DNA sequences and compared to published phylogenies [146]. Maximum-likelihood analysis of *Umbrea* and *HP1B* gene sequences was performed with codeml of the PAML software package [34], and web implementation of the HyPhy analysis package (www.datamonkey.com) [180]. dN/dS ratios between lineages were calculated in PAML using a free ratio model, allowing dN/dS variation along the branches of the phylogeny. To detect amino acids under selection in Umbrea and HP1B, we fit the multiple-alignment of amino acid sequences to either the F3x4 or the F61 codon frequency models. Likelihood ratio tests were performed by comparing *NS* sites models M1 (neutral) and M2 (selection), M7 (neutral, beta-distribution of omega <1) and M8 (selection, beta distribution, omega >1 allowed), or M8a (selection, beta distribution, omega >1 allowed) and M8. PAML analysis identified sets of amino acids in the *Umbrea* phylogeny with high posterior probabilities for positive selection by using Bayesian methods. We also performed

analyses of *Umbrea* genes from only the melanogaster species subgroup of species, or all Umbrea genes except those from the melanogaster species group to date the signature of positive selection.

_Kc_ cell cytology

All constructs used for cytology in *Kc* cells were generated by amplifying genomic DNA (*Umbrea* lacks introns) and directionally cloning into pENTR-D-TOPO (Invitrogen). Upon sequence verification, these entry clones were recombined via LR clonase (Invitrogen) reactions into Gateway Destination vector pHGW 1073 (http://www.ciwemb.edu/labs/murphy/Gateway%20vectors.html). *Kc* cells were seeded onto coverslips and transfected with FuGene (Roche) overnight with 2ug of plasmid DNA. Expression of fusion proteins was transiently induced by heat-shock at 37^{o}C for 45 minutes and cells were allowed to recover for 2 hours at room temperature. Nuclei were prepared by incubating cells in 0.5% sodium citrate and spinning away cytoplasm using a Shandon Cytospin3 at 1900 rpm. Nuclei were fixed for 15 minutes in 1X phosphate buffered saline plus 0.3% Tween-20 (PBST) plus 4% paraformaldehyde. Following fixation, nuclei were washed with PBST and then blocked for 45 minutes at room temperature with either PBG (PBST plus 0.2% cold water fish gelatin (Sigma) and 0.5% BSA (Sigma) or PBST plus 5% BSA. Primary antibodies diluted in PBG (see below) were incubated with nuclei for 1 hour at room temperature or overnight at 4^{o}C. Following primary antibody staining, nuclei were washed in block and incubated with Alexa Fluor secondary antibodies diluted in block.

Antibodies used

Anti-Cid (rabbit polyclonal, Abcam) 1:500; anti-Cid (rabbit polyclonal, Henikoff et al 2000); anti-Cid (chicken polyclonal, gift from Karpen lab), 1:1000; anti-CENP-C (rabbit polyclonal, gift from

Lehner lab), 1:5000; anti-HOAP: (rabbit polyclonal, gift from Theurkauf lab), 1:1000; anti-HipHop: (rabbit polyclonal, gift from Rong lab), 1:500

Umbrea transgenesis for RNAi rescue

Umbrea genes were synthesized by GenScript that were re-encoded such that all synonymous sites were changed, leaving the amino acid sequence intact. Transgenes were cloned into pENTR D TOPO, then recombined using Gateway (Invitrogen) technology into expresion vector pUASp-GFP, with miniwhite rescue. Transgenesis by standard embryo injection was performed by The Best Gene Inc. (Chino Hills, CA, US).

Drosophila maintenance and strains

All *Drosophila* strains were maintained on standard molasses-cornmeal medium in uncrowded conditions. The following strains were obtained from the Bloomington Drosophila Stock Center for use in complementation analysis of *Dumpy* and *Umbrea* : dp^{lv1} b^{1}/SM5 (Stock number: 278), dp^{olvR}/SM5 (Stock number: 280), w^{1118}; net^{1} $P[w^{+mGT=GT1}][HP6^{BG0142}]$ $dp^{BG01429}$/In(2LR)Gla, wg^{Gla-1} Bc^{1} (Stock number: 12747)

The following Vienna Drosophila RNAi Center (VDRC) line was used in *in vivo* RNAi knockdown experiments: w^{1118}; P[GD4434]v13074/CyO (Stock number: v13074)

The following GAL4 driver lines were obtained from the Bloomington Drosophila Stock Center: y^{1} w*; P[Act5C-GAL4-w]E1/CyO (Stock number: 25374)

RNAi knockdown of Umbrea in S2 cells

dsRNA was prepared using the MegaSCRIPT® T7 kit (Applied Biosystems) following the manufacturer's instructions. Templates were generated by PCR from genomic DNA (for

Umbrea) and from the pCopia-LAP-CID plasmid (for the control dsRNA), using the following primers:

Umbrea forward, TAATACGACTCACTATAGGGCGCCCAGCTCCACTTTGAC,

Umbrea reverse, TAATACGACTCACTATAGGGCGCATTTCGTGATCGTTTCTT, scrambled forward, TAATACGACTCACTATAGGGCAAGAGCTTGGCGGCGAAT, scrambled reverse, TAATACGACTCACTATAGGGCCGCGGGTTCCTTCCGGTA.

The dsRNA was transfected into the cells using DOTAP liposomal transfection reagent (Roche). 10^6 logarithmically growing S2 cells were plated in 1 ml of serum medium in a 6-well plate.10µg dsRNA were transfected with DOTAP (Roche), following the manufacturer's instructions. After 24hr, the DOTAP containing medium was replaced with new serum containing medium, and cells were incubated for 4 additional days. Samples were taken on day 5 and were subjected to indirect IF and time-lapse analysis.

Immunofluorescence (IF) on fixed S2 cells

Cells were prepared for immunofluorescence as previously published[65, 181]. Briefly, cells were settled onto slides and fixed with 3.7% paraformaldehyde in PBS plus 0.1% Triton X-100. Blocking was performed using 5% milk, and antibodies were incubated in block overnight at 4°C. The primary antibodies used were 1:1000 chicken anti-CID [73], mouse anti-α-tubulin (Sigma-Aldrich), and rabbit anti-phospho-H3S10 (PH3; Millipore). All of the secondary antibodies (Cy5 anti-chicken, 546 anti-mouse and 488 anti-rabbit, Invitrogen) were diluted 1:500 in 5% milk in PBST and incubated for 45 min at room temperature. After three 5-min washes in PBST, cells were mounted in SlowFade® Gold anti-fade reagent (Invitrogen) containing DAPI. All images were taken on a Personal DeltaVision (DV) microscope (Applied Precision, LLC) and deconvolved using softWoRx (with iteration set to conservative, 5 cycles; Applied Precision, LLC). Images were taken as z stacks of 0.4- or 0.5-µm increments using 60× and 100× oil-

immersion objectives. Mitotic cells were scored as defective or normal based on chromosome and spindle morphology compared with control cells, and the p-value was calculated using the Fisher's Exact Test.

Time-lapse analysis of mitosis

Time-lapse videos were performed using a Personal DV microscope using a 60× oil-immersion objective. S2 cells expressing mCherry-tubulin and H2B-GFP (gift of G. Goshima) were mounted using the hanging drop method [76] 5 days after RNAi treatment. Images of cells in prophase/prometaphase were taken every 2 to 4 min until cytokinesis or for up to 90 min in cases where cells were experiencing delays in mitosis. 14 control and 22 Umbrea RNAi videos of randomly selected cells in prophase were made. Videos were deconvolved and quick-projected, and the time in minutes in the still images reflects the actual elapsed time during image acquisition.

Quantitative Real-time PCR

For reverse transcription, total RNA was extracted from the S2 cells 5 days after RNAi using the RNeasy kit (QIAGEN). Genomic DNA was removed from the sample with RQI RNase Free DNase (Promega). 2 μg RNA was used as the template for cDNA synthesis, which was performed using the SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen). The cDNA was then used as a template for quantitative real-time PCR, which was run using a BioRad iCycler iQ™. Briefly, the program was run in reaction volumes of 15 μl containing the following reagents: 7.5 μl EXPRESS SYBR® GREEN**ER**™ SuperMix with Premixed ROX (Invitogen), 6.0 μl molecular grade water, 0.5 μl cDNA template, and 1 μl forward and reverse primer mix. The following primers were used for qPCR:

Umbrea forward: CTTCCGTTTGGTTTTAGATATCGT

Umbrea reverse: AAACACTTGACAAAACGTGACAAT

Actin5C forward: GATCTGTATGCCAACACCGT

Actin5C reverse: ATGGCCGAATTCTCAGTGGA

Each cDNA/primer set combination was run in triplicate. Cycle threshold (CT) values for Actin5C and Umbrea in the scrambled RNAi sample and Umbrea RNAi sample were used to calculate the $\Delta\Delta$CT values for Umbrea expression.

Purification and Mass Spectrometry

Protein purification and mass spectrometry were performed as published[65, 135]. In brief, chromatin was extracted from large cultures of S2 cells expressing Flag-tagged proteins. Immunoprecipitation was performed using anti-FLAG M2 Sepharose beads (Roche). Protein complexes were separated on a 15% PAA gel. The entire lane was then cut out and subjected to trypsin digestion before analysis on a Ultimate 3000 HPLC system (LC Packings Dionex).

Proteins were identified and quantified using the Andromeda algorithm of the MaxQuant 1.3.3.X software package[180]. Identified proteins were considered as interaction partners if their MaxQuant intensities displayed a greater than 8-fold enrichment compared to control anti-FLAG purifications from L2-4 nuclear extracts not expressing any FLAG-tagged protein. Protein levels were estimated in the Umbrea pulldown using the intensity based absolute quantification values (iBAQ), which corresponds to the sum of peak intensities of all peptides matching to a specific protein divided by the number of theoretically observable peptides.

Section 4: Speciation as a consequence of rapid evolution at centromeres

Rationale

Darwin provided an explanation for evolution by natural selection that modern experimentation has proven to be correct in a plethora of systems, from the molecular level to the organismal level. Darwin was puzzled however, by another phenomenon – that of speciation, which he called the "mystery of mysteries". Speciation is the splitting of one species into two distinct species, and is thought to occur through the establishment of reproductive isolating barriers. These barriers can come in two flavors – mechanisms that prevent species from mating (the evolution of divergent mating plumage, for example, and mechanisms that act after mating resulting in reduced viability or fertility in the hybrid offspring (hybrid infertility or inviability) [182]. Decades of research on speciation have uncovered several principles that seem to hold true across taxa [182]. For example, the heterogametic sex (males in XY species like humans, or females in ZW species like birds) is usually most strongly affected by hybrid phenotypes. Additionally, the X chromosome appears to evolve loci that are involved in hybrid phenotypes at a much fast rate than the rate of incompatibility evolution on the autosomes [182]. These hybrid phenotypes inevitably involve the evolution of reduced fitness. Yet, as Darwin hypothesized, natural selection should act to remove mutations that result in reduced fitness. How, then, do barriers to reproduction evolve between species? One model that explains this phenomenon is the Dobzhansky-Muller model [182, 183], which offers the prediction that incompabilities arise as a by-product of divergence between two populations, whether that divergence is driven by neutral evolution or selection. All that is required is that two (or more) interacting loci diverge differentially in each population **(Figure 4.1)**. The key is that each species adapts to the changes wrought in each locus separately. Yet, in a hybrid, the two divergent loci are brought together for the first time. Natural selection has not yet had a chance to act on this new interaction and negative epistasis can result, leading to stereotypical hybrid dysfunction. It is

not hard to imagine the DM model applying to divergence between centromere drive suppressors and driving centromeric satellite elements, between populations. In this scenario, two populations with a common ancestor would experience independent bouts of centromere drive and selection for suppression **(Figure 4.1)**. In a hybrid, derived from a cross between the two populations, mismatches between co-evolved centromeric satellite DNAs and centromeric proteins such as CENP-A could lead to hybrid dysfunction in the form of mitotic or meiotic defects in chromosome segregation (for example). Yet, the link between centromeric divergence and speciation has remained obscure.

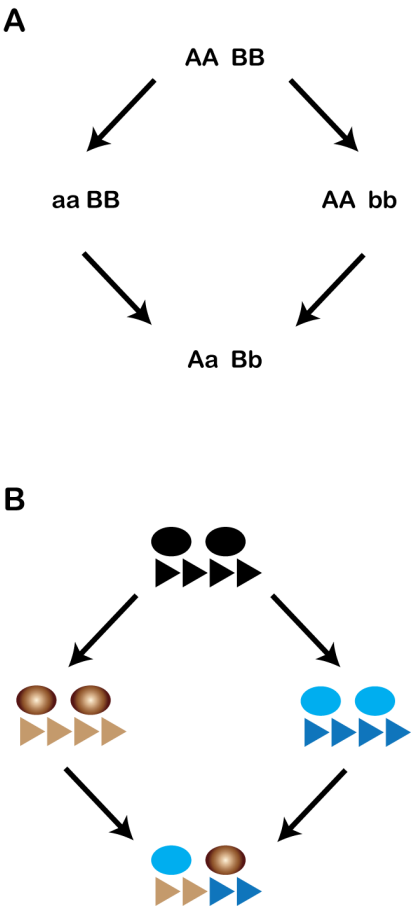


Figure 4.1 – The Dobzhansky-Muller model for hybrid incompatibilities[182]. (A) Consider two loci in an ancestral population, *A* and *B*. Upon separation of individual populations, new alleles, a and *b*, arise in each population. Natural selection has therefore tested the functionality of *a* with *B*, and *b* with *A*, respectively. However, *a* and *b* have not been tested for compatibility in the same genome. In a hybrid, incompatibility between *a* and *b* could lead to hybrid dysfunction. (B) Divergence between centromeric satellite DNA elements (repeating triangles) and centromeric proteins (ovals) such as CENP-A is conceptually similar. Upon splitting of an ancestral population into two, lineage-specific divergence of centromeres and drive suppressors could result in incompatibility upon formation of hybrids.

Speciation research in *Drosophila*

Crosses between closely-related sibling species of *Drosophila melanogaster* have provided a model for much of what is known of the genetic and molecular basis of speciation, based on early genetical work initiated by pioneering researchers in the field (including Alfred Sturtevant)[184] and more recent work on the molecular cloning of speciation genes [20, 97, 171, 185-187], and the characterization of the mechanism of hybrid dysfunction[47, 96, 181, 188]. Early in my thesis work in the Malik lab, I became interested in applying principles of genetic conflict and rapid evolution to gain insight into the selective forces and molecular mechanisms that underlie the speciation process. I chose to study hybrids resulting from successful matings between *D. melanogaster* females and *D. simulans* males are strictly female [184]. The reciprocal cross is somewhat more variable in the penetrance of the hybrid phenotype, but in many crosses, only male hybrids survive, while females die as embryos. It is clear the genetic basis of the two reciprocal crosses is different, as mutations that rescue lethality in one direction of the cross fail to rescue in the other [20, 97, 185-187, 189].

Much progress has been made on understanding the genes and mechanisms involved in speciation from the *D. melanogaster* female X *D. simulans* male cross. Hybrid males from this cross die at late larval stages or as early pupae[184]. Early work from Sturtevant, and later work from Muller and Pontecorvo, found that three components, encoded by the *D. melanogaster* X chromosome, the *D. simulans* second chromosome, and the *D. simulans* third chromosome, were all required for full hybrid male inviability[190, 191]. Early cytological work to understand the molecular basis of male-specific hybrid inviability concluded that larval arrest was the result of mitotic defects[192]. Hybrid male larval exhibited tissues of reduced size, suggesting that cell proliferation was hindered. This was supported by the induction of male clones in hybrid females through X-nondisjunction – these clones were also found to be of reduced size. However, later work on the basis of this classic speciation system suggested that

mechanisms other than the control of the mitotic divisions might be at play[193, 194]. This hypothesis was based in large part on the only biological differences between males and females in *Drosophila* (see section 1 for another treatment of this subject) – the Y chromosome, which is dispensible for viability but required for male fertility, and dosage compensation, which in *Drosophila* equalizes the expression of the X chromosome between the sexes based on the X-autosome ratio. Intriguingly, the complex responsible for up-regulating gene expression on the male-X (the dosage compensation complex or DCC) is rapidly evolving between *D. melanogaster* and *D. simulans*[98, 99], as are the DCC binding sites on the X chromosome[195].

To gain insight into the genetic basis of *D. melanogaster-D. simulans* hybrid male inviability, decades of research had focused on the isolation of mutants that specifically rescued hybrids to viability. Mutations were identified through the isolation of "rescue" strains of either *D. melanogaster* or *D. simulans* from the wild. This approach saturated quickly, with multiple mutations mapping to the X chromosome of *D. melanogaster* and the 2nd chromosome of *D. simulans*[187, 189]. Although effective, this strategy may have failed to identify the missing 3rd component (on the *D. simulans)* third chromosome because rescue mutations may be lethal in a pure species context. What has been learned from the identification of the two rescue genes? The rescue mutants in *D. simulans* mapped to a gene called Lethal Hybrid Rescue or Lhr, while the *D. melanogaster* mutants mapped to a gene called Hybrid Male Rescue or Hmr. Early work on Lhr suggested that it was a heterochromatin-localizing protein[143, 171, 196] that was not essential for viability or fertility, since null mutants could be isolated from wild populations[189].

Initial theories of hybrid dysfunction proposed that hybrid incompatibility genes were species-specific transcription factors that caused gene expression dysregulation in F1 hybrids[197]. This model may still hold for some systems. However, the transcription factor model was shown

to not be true in at least one case, in *D. melanogaster- D. simulans* hybridizations, since microarray data revealed no extensive gene misregulation in lethal male F1 hybrids compared to F1 hybrid males rescued by mutations in *Lhr* or *Hmr* [198]. Excitingly, both Lhr and Hmr are rapidly evolving in the *melanogaster* species group[171, 175], although there are some reports that rapid evolution does not affect function, at least as measured by conservation of protein-protein interactions[199]. I first looked for evidence to support these observations, by performing my own evolutionary analyses of *Lhr* and *Hmr.*

Evolution of *Hmr*

I looked for evidence of recurrent positive selection in *Hmr* in sequence data from 7 species of *Drosophila*, spanning approximately 12-15 million years of evolution (*D. melanogaster, D. simulans, D. sechellia, D. mauritiana, D. yakuba, D. erecta,* and *D. ananassae*), using maximum likelihood methods implemented in the PAML software package[33, 34]. This analysis utilized 1194 alignable codons (out of 1414 total in the $Hmr^{melanogaster}$ gene). Using site-based methods, I found significant support for positive selection ($p < .001$). However, no single codon was identified as statistically significant. By free-ratio analysis, several branches exhibited elevated dN/dS, indicative of positive selection.

In parallel with the PAML analysis, I analyzed the same alignment with the online Datamonkey software package (HyPhy). These approaches revealed evidence of recurrent positive selection on a subset of codons in *Hmr*. For example, FEL analysis revealed 20 codons that met a significance threshold of .9. FUBAR analysis identifies 2 codons that met the significance threshold, 625 (posterior probability = .96), and 1052 (posterior probability = .92). Other analyses using Datamonkey were not significant, including PARRIS, which is typically considered to be conservative. Codons identified as evolving under recurrent positive selection

were distributed throughout the length of the gene, and did not seem to be enriched in any one of the MADF DNA- or chromatin-binding domains[200].

Evolution of *Lhr*

To confirm previous reports of positive selection on the *Lhr* gene in *Drosophila,* I performed McDonald-Kreitman analysis using sequences from at least 10 isogenic African strains each of *D. melanogaster* and *D. simulans* deposited in GenBank by the Barbash lab. McDonald-Kreitman analysis is a powerful method used to test the hypothesis of neutral evolution in a gene of interest[201]. If a gene evolves under neutrality, the ratio of substitutions that result in the alteration of amino acid sequence (replacement changes) to synonymous mutations between species should be equal to the ratio of replacement to synonymous polymorphisms within species. Using the DNASP software package, I found a strong signature of positive selection in *Lhr* (Sf:Rf versus Sp:Rp, 20:50 versus 16:12 (Sf = synonymous differences fixed between species, Rf = replacement differences fixed, Sp = Synonymous differences polymorphic within species, Rp = replacement polymorphisms)) **(Figure 4.2)**.

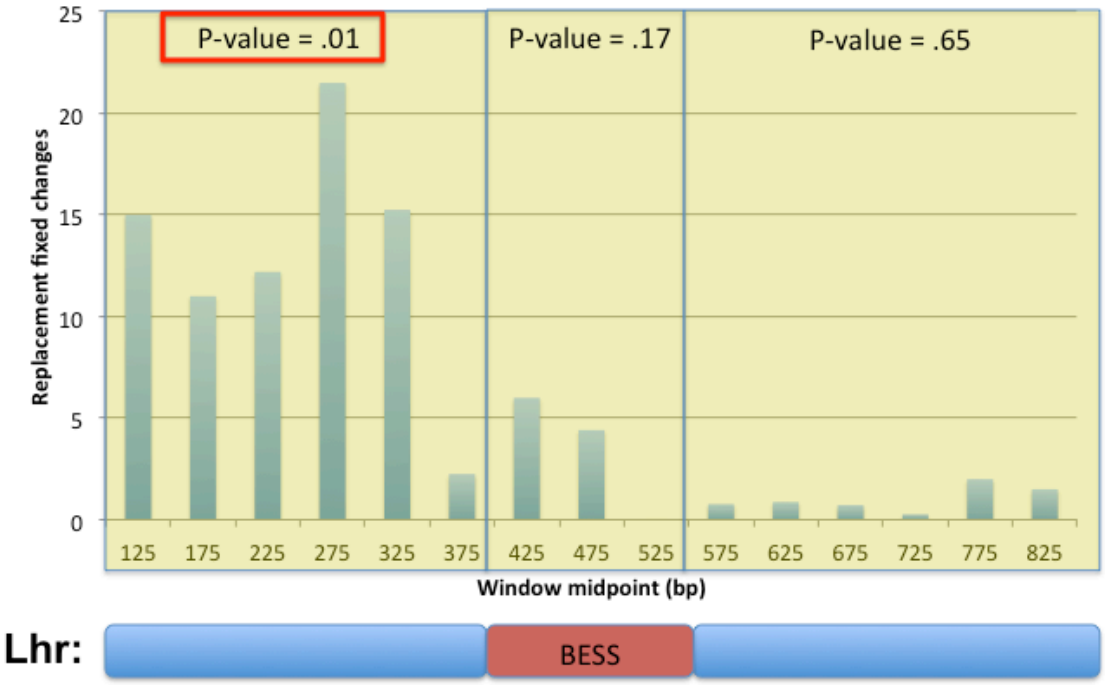


Figure 4.2 – Summary of fixed replacement differences between *Lhr*^{melanogaster} and *Lhr*^{simulans} mapped over domain windows. P-values refer to McDonald-Kreitman analysis, showing that the N-terminal leucine zipper domain of *Lhr* is rapidly evolving, while the BESS domain and C-terminal region domain are slowly evolving.

To look for positive selection across a broader phylogeny, I ran PAML and Datamonkey on 295 alignable codons of *Lhr* sequence from 320 total codons between 8 species (*D. melanogaster, D. simulans, D. sechellia, D. mauritiana, D. yakuba, D. erecta, D. biarmipes*, and *D. eugracilis*). By Datamonkey, most tests are negative, including PARRIS and REL. In PAML, a comparison of the data to models M7 and M8 reveals that positive selection does not best explain the data (p>.3). These analyses suggest that *Lhr* is not rapidly evolving under positive selection as a general rule across *Drosophila* species. In contrast, my McDonald-Kreitman analysis suggests that *Lhr* positive selection is restricted to divergence between the *D. melanogaster* and *D. simulans* lineages.

Divergence in Lhr causes species-specific localization to heterochromatin

Considering that Lhr had been shown to localize to heterochromatin in cultured cells and *in vivo[143, 171, 199, 202{Filion, 2010 #287]},* such rapid coding sequence evolution is suggestive of genetic conflict (as discussed previously) between Lhr and heterochromatic satellite DNA sequence. Thus, divergence at or near centromeres perhaps due to recurrent drive and suppression might promote speciation.

At the time, my hypothesis was that *D. melanogaster- D. simulans* hybrids die because of dominant mislocalization of the $Lhr^{simulans}$/$Hmr^{melanogaster}$ "poison" in the nucleus of cells. To test this hypothesis, I began by tagging $Lhr^{melanogaster}$ and $Lhr^{simulans}$ and expressed these fusion proteins in *D. melanogaster* male and female cells.

Cell biology of Lhr

I hypothesized that $Lhr^{simulans}$ conferred lethality upon hybrids through a dominant mislocalization in the nucleus of male cells. To test this hypothesis, I used heat-shock induced overexpression, and found that an $Lhr^{melanogaster}$ GFP-fusion protein localized specifically to heterochromatin in female *D. melanogaster* Kc cells but diffusely in "male" S2 cells (S2 cells lack the Y chromosome but, due to aneuploidy, have male-like X to autosome ratio and express the dosage compensation complex) **(Figure 4.3)**. This male-specific localization pattern differed from $Lhr^{simulans}$, which localized to heterochromatin in both Kc and S2 cells. These results confirmed previous findings from Adam Waite, a prior Malik-lab rotation student, who was the first to look at Lhr ortholog localization in *D. melanogaster* cultured cells. It is important to note that (similar to the observation that $Lhr^{simulans}$ exhibits a genetic dominant phenotype in hybrids) these *D. melanogaster* cells express native $Lhr^{melanogaster}$ (Flybase, ModENCODE RNA expression data). Therefore, the dominant ability of $Lhr^{simulans}$ to localize to heterochromatin in the presence of $Lhr^{melanogaster}$ mirrors its genetic phenotype of dominant male sterility.

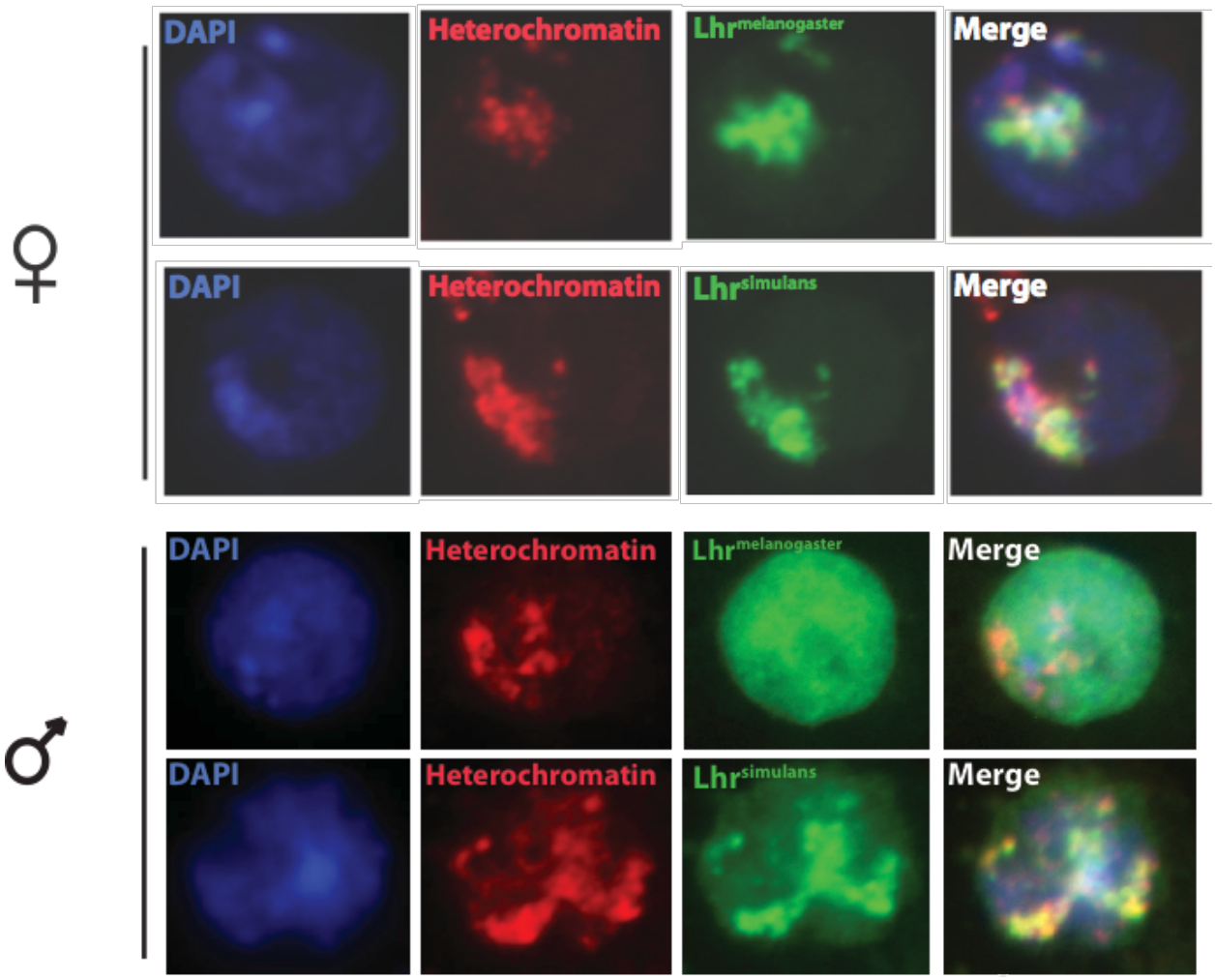


Figure 4.3 – Cell biology of Lhr divergence reveals dominant mis-localization of GFP-Lhr^{simulans} in *D. melanogaster* male cells. Heat-shock induction of GFP-Lhr expression in *D. melanogaster* female Kc167 cells and male S2 cells. Cells are stained with DAPI to reveal DNA and anti-H3K9me3 to indicate heterochromatin. While Lhr^{melanogaster} localizes to heterochromatin in female cells but not male cells, Lhr^{simulans} localizes to heterochromatin in both. These data imply that Lhr^{simulans} misregulation in male cells contributes to hybrid male inviability.

How did this localization pattern evolve? One way by which the cellular localization patterns of proteins change is through the gain or loss of protein-protein interactions. Indeed, this is what I hypothesize to have occurred in the evolution of Umbrea[65]. However, *Lhr* evolution in the *melanogaster* subgroup has not affected its ability to form protein-protein interactions with at least 2 of its interaction partners – HP1 and Umbrea[199]. The Lhr BESS domain may be important for protein-protein interactions – indeed, the BESS domain and C-terminus of Lhr contain several degenerate PxVxL HP1-interacting motifs required for interaction with HP1A and Umbrea[199]. Therefore, the BESS domain and C-terminus were unlikely to be involved in the difference in localization.

Instead, I hypothesized that the rapidly evolving region of Lhr was required for the sex- and species-specific localization pattern in cultured cells. To test this, I generated Lhr GFP-fusion chimeras in which the N-terminal leucine zipper region was derived from either $Lhr^{melanogaster}$ or $Lhr^{simulans}$, and the BESS domain and C-terminus was derived from the reciprocal species, and expressed these fusion proteins in either male or female *D. melanogaster* cells (data not shown). These fusions revealed that the N-terminal rapidly evolving region of Lhr was necessary and sufficient for sex-specific localization to heterochromatin in male cells. The N-terminus of Lhr contains a leucine zipper domain that is characteristic of DNA-binding transcription factors[203]. This homology implies that Lhr has DNA binding properties, and that sex-specific DNA binding underlies the localization patterns that I observed.

Sex-specific localization of Lhr is an ancestral trait

I next asked if the property of $Lhr^{simulans}$ to localize to sex-specific heterochromatin in *D. melanogaster* cells was a derived or ancestral characterized. To test this, I cloned *Lhr* from *D. sechellia,* a sister species to *D. simulans*, and *D. yakuba*, a basally-branching member of the *melanogaster* subgroup, and expressed them in *D. melanogaster* cells as GFP-fusion proteins.

I found that Lhr^{sechellia} shared the property of heterochromatin localization in male *D. melanogaster* cells with Lhr^{simulans}, while Lhr^{yakuba} appeared diffuse (data summarized in **Figure 4.4)**. Both orthologous fusion proteins localized to heterochromatin in female *D. melanogaster* cells. These data indicate that the property of Lhr^{simulans} to mis-localize to heterochromatin in *D. melanogaster* male cells evolved as a derived gain-of-function characteristic, in the common ancestor of *D. simulans* and *D. sechellia.* While it is also possible that two independent loss of function events occured in the *D. yakuba* and *D. melanogaster* lineages, it is more parsimonious that a single gain-of-function event occurred along the *D. simulans* lineage.

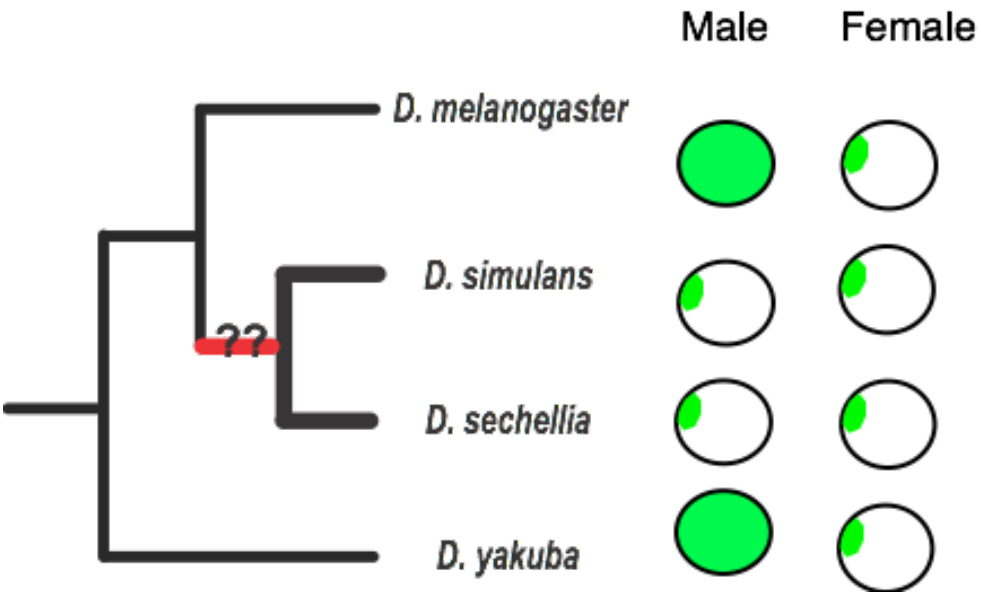


Figure 4.4 – Summary of GFP-Lhr ortholog localization in *D. melanogaster* cells. All Lhr orthologs localize to heterochromatin in female Kc167 cells. However, only Lhr^{simulans} and Lhr^{sechellia} localize to heterochromatin in S2 cells, while Lhr^{melanogaster} and Lhr^{yakuba} localize diffusely. This suggests that the ancestral state of localization of Lhr in male cells (but not female cells) was diffuse.

Recent investigation of the cellular basis for hybrid dysfunction

Recent work has likely uncovered the basis for Lhr-Hmr mediated male hybrid lethality in crosses between *D. melanogaster* females and *D. simulans* males. Thomae *et al.* used cytological, proteomic, and biochemical approaches to show for the first time that Lhr and Hmr are centromeric proteins that (like Umbrea) interact with both heterochromatin proteins and centromeric proteins like CENP-C[166]. Furthermore, they proposed a model for hybrid inviability that invokes dosage imbalances between stoichiometric components of the Lhr-Hmr complex, perhaps due to increased expression of *D. melanogaster Hmr*. However, by itself, increased expression of Hmr does not fully explain the sex-specificity of the hybrid phenotype – an additional mechanism must be invoked. Thomae *et al.* suggest that the dosage-compensation complex may play an additional role, although no genetic evidence supports this hypothesis [193]. The Y-chromosome could play a role, although again there is a lack of evidence to support a role for the Y chromosome in hybrid male inviability[204]. Lhr also regulates the transcription of species-specific satellite DNA elements and transposable elements[202]. At first glance, these results might indicate that the rapid evolution of *Lhr* was driven by a genetic conflict not with centromeres but instead with selfish repetitive elements like transposons. However, de-repression of transposable elements may be the result of altered chromatin state, and might not indicate a specific role in silencing transposable elements. For instance, knockdown of *HP1* also de-represses transposable elements, yet is incontrovertibly involved in general chromatin structure[205]. Therefore, Lhr may function in maintenance or establishment of chromatin at or near centromeric heterochromatin.

Future Directions

My work on Lhr has shown that rapid evolution of the N-terminal leucine zipper region is likely to be the basis for differential sex-specific localization to heterochromatin. It is clear that this result needs to be reexamined, now that Lhr (and Hmr) are known to be centromeric[166], perhaps by using more precise titration of inducible fusion proteins, or by using antibodies specific to each ortholog. I predict that sex-specific differences in Lhr localization to centromeres would be observed between orthologs. For instance, I predict that $Lhr^{simulans}$ localizes to centromeres in both male cells and female cells, while $Lhr^{melanogaster}$ only localizes to centromeres in female cells. If these predictions hold, fine-scale mapping of the localization determinant (for example, to the N-terminal domain), should yield insight into the biological function of Lhr. I predict that Lhr functions in a similar pathway as Umbrea, given the link to common binding partners and the demonstrated physical interaction between Umbrea and Lhr (see previous Section for further details).

A link between such biochemical and cytological results and the real phenotype of interest (hybrid lethality) awaits genetical tests. For instance, expression of $Lhr^{simulans}$ in *D. melanogaster* flies should induce male lethality, but only in the presence of the *D. simulans* third chromosome factor, which has not been cloned[191].

Finally, it is intriguing to speculate on the selective forces that might drive differential localization of a centromeric factor like Lhr. As discussed in previous Sections for Cid and Umbrea, a role in promoting Y-chromosome-specific chromosome segregation could underlie differences in localization or binding between the sexes. In partial support of this idea, Lhr has been found to regulate the expression of many small RNAs, including repetitive sequences found in the pericentric heterochromatin[202]. Of particular interest are some sequences found on the Y chromosome. However, it is clear that Lhr mutant females also show differences in the

abundance of certain classes of small RNAs in the ovary. Indeed, Lhr mutant females show clear fertility defects, consistent with a failure to properly regulate the expression of repetitive elements like transposons. If there is a male-specific role for Lhr (or Hmr) in a pure-species context, it remains to be discovered

Broad thoughts on the link between speciation and genetic conflict

It is intriguing to note that Lhr/Hmr and Umbrea are interaction partners[65, 166]. I have already devoted some attention to the idea that these proteins could function at centromere-heterochromatin boundary factors (see Umbrea section). Given the centromere drive hypothesis, it is particularly tempting to note the link between genetic conflict and speciation. Two of the other cloned speciation genes implicated in hybrid male dysfunction in *Drosophila* evolve rapidly. Divergence of the gene *Overdrive* causes meiotic drive and hybrid male sterility between subspecies of *D. pseudoobscura*[188], by unknown mechanisms. The function and cytological localization of Overdrive are not known, but might not be unexpected should further investigation reveal that Overdrive localize to heterochromatin. The other famous speciation gene in *Drosophila* is *Odysseus*[181], which causes hybrid male sterility between sister species in the *melanogaster* subgroup. Odysseus is known to localize to heterochromatin, with differential binding to the Y chromosome likely to be causal for the sterility phenotype[181]. Finally, a pericentromeric satellite repeat underlies hybrid female lethality, in the other direction of the *D. melanogaster-D. simulans* cross[20, 96]. These features may define a general rule for speciation, at least in terms of post-zygotic hybrid dysfunction.

The mechanism of action of many of the meiotic drive systems observed in nature, or in the lab, is not known. However, meiotic drive is intimately linked to speciation[188, 206-208]. It is tempting to speculate that centromere drive and post-meiotic gamete killing drive are related. Indeed, one way to restore Mendelian balance following an episode of centromere-mediated

female meiotic drive would be to evolve a gamete killing mechanism similar to that of the Responder drive system in *Drosophila*[42]. There are a number of fascinating links between these processes. For instance, sperm bearing Responder-sensitive chromosomes are killed following meiosis (killing appears unrelated to the meiotic divisions), while insensitive chromosomes are preserved. Sensitivity in this system is determined by the abundance of a pericentromeric satellite DNA array called the Responder array. Extrapolating further from this analogy, the expansion of the Responder array could have triggered centromere drive during female meiosis leading to an over-representation of this chromosome in populations. Indeed, cryptic meiotic drive systems appear to be rampant in populations and are only unleashed upon outcrossing[206]. The functional and evolutionary relationships between male and female meiotic drive systems await deeper molecular and genetic understanding of each, in a tractable model system.

Lhr project Materials and Methods

Evolutionary analysis

Sequence analysis of *Lhr* and *Hmr* was performed as in previous section (see Umbrea methods), with one exception. McDonald-Kreitman analysis was perfomed using the software package DNASP[209], on Clustal-generated sequence alignments.

Cytology

See Part 2 (Umbrea section) Methods for details about cloning, transfection of *D. melanogaster* cultured cells, and immunofluorescence cytology.

Grand conclusions

My thesis research has been devoted to a deeper understanding of rapid evolution at centromeres. I have found that rapid evolution dramatically altered the functionality of the centromeric histone Cid in *Drosophila.* While *Cid* homologs are found in many organisms, I found that rapid evolution at centromeres could induce the incorporation of the products of young genes into the chromosome segregation machinery. These results indicate that centromeres functionally diverge between species, in lineage specific ways. I found that this centromere divergence might have consequences for speciation, through the dissection of sex-specific localization of a heterochromatin protein (now known to be centromeric). While my results may be confined to *Drosophila,* the general results may be broadly applicable. For example, centromere proteins evolve across the eukaryotic tree of life, and gene duplication as an avenue to genetic innovation in the face of selective pressure is a general mechanism thought to drive novelty in all organisms. Many of the centromeric proteins identified in vertebrates have no homolog in *Drosophila*, or in other taxa. Rapid sequence divergence may have made these genes unidentifiable. It could be, however, that gene turnover dynamics have resulted in dramatically different composition of centromere proteins. Centromeres represent one of the most puzzling paradoxes in biology. Investigation of the dynamic aspects of centromeres will lead to insight that is missed by a myopic focus on the highly conserved.

Curriculum vitae

Benjamin D. Ross

EDUCATION

PhD in Molecular and Cellular Biology Sept. 2014
University of Washington – Seattle, WA
BA in Biochemistry 2005
Lewis & Clark College – Portland, OR

RESEARCH EXPERIENCE

Ph.D. Candidate 2008 – present
University of Washington – Seattle, WA
Dept. of Molecular and Cellular Biology, Dept. of Basic Sciences, and Fred Hutchinson Cancer Research Center
Advisor: Dr. Harmit Malik

- Functional consequences of rapid evolution of centromeres in *Drosophila*

Post-baccalaureate Intramural Research Fellowship 2005 – 2008
National Institutes of Health – Bethesda, MD
Advisor: Dr. Andy Golden

- Chromatin dynamics during oogenesis in *C. elegans*

AWARDS/HONORS

Society for Molecular Biology and Evolution Travel Award 2014
American Society for Cell Biology Travel Award 2013
Society for Molecular Biology and Evolution Travel Award 2013
Society for Molecular Biology and Evolution Travel Award 2012
Keystone Travel Award (Evolution and Development Meeting) 2011
National Science Foundation Graduate Research Fellowship 2010
NIH Genome Training Grant (UW) 2009
NIH Post-baccalaureate Intramural Research Training Award 2005

PUBLICATIONS

Janet A. Young, **Benjamin D. Ross**, and Harmit S. Malik. The architecture of rapid evolution across the primate kinetochore. *(In prep.)*

Benjamin D. Ross and Harmit S. Malik. Species-specific function of the adaptively evolving centromeric histone in *Drosophila*. (*In prep.)*
Functional characterization of the genetic consequences of rapid evolution of the centromeric histone CenH3/Cid, between closely-related species of Drosophila. Ancestrally reconstructed Cid or ortholog transgenes used to complement loss-of-function mutations in D. melanogaster, resulting in sex-specific lethality and mitotic catastrophe.

Benjamin D. Ross and Harmit S. Malik. (2014). Genetic Conflicts: Stronger Centromeres Win Tug-of-War in Female Meiosis**.** *Current Biology.*
Comment on Chmatal et al. (2014), Current Biology.

Benjamin D. Ross, Leah Rosin, Andreas W. Thomae, Mary Alice Hiatt, Danielle Vermaak, Aida Flor A. de la Cruz, Axel Imhof, Barbara G. Mellone, Harmit S. Malik (2013) Stepwise Evolution of Essential Centromere Function in a Drosophila Neogene. *Science* 7:340, 1211-1214 (Faculty of 1000 Recommended article, highlighted by ScienceDaily and Nature Reviews Genetics).
Identification of the molecular basis of neofunctionalization in a newly essential gene in Drosophila. The gene Umbrea lost an ancestrally important domain and rewired its protein-interaction network through amino acid alteration of a binding interface to transition to centromeric localization and function in mitosis. Rapid evolution resulting in species-specific localization suggests that genetic conflict may promote the gain of essential function.

Roach, K. C., **Ross, B. D**. and Malik, H. S**.** (2012) Rapid evolution of centromeres and centromeric/ kinetochore proteins. in *"Evolution in the Fast Lane: Rapidly Evolving Genes and Genetic Systems"* (eds: Rama Singh, Jianping Xu & Rob Kulathinal), pp. 83-93, Oxford University Press.

Roach, K. C., **Ross, B. D**. and Malik, H. S. (2011) Adaptive Evolution of Centromeric Proteins. (Review) In: *Encyclopedia of Life Sciences* [doi: 10.1002/9780470015902.a0022868], John Wiley & Sons Publishing Co.

Kathryn K. Stein, Jessica E. Nesmith, **Benjamin D. Ross**, and Andy Golden. (2010) Functional Redundancy of Paralogs of an Anaphase Promoting Complex/Cyclosome Subunit in Caenorhabditis elegans Meiosis. Genetics 2010 186:1285-1293.
Genetic mapping and characterization of two redundant paralogs of a key meiotic subunit of the APC/C, in C. elegans.

PRESENTATIONS

Society for Molecular Biology and Evolution, San Juan, PR 2014
Functional consequences of centromere evolution *(Invited talk)*

American Society for Cell Biology, New Orleans, LA 2013
Evolution of essential centromere function *(Invited talk)*

Society for Molecular Biology and Evolution, Chicago, IL 2013
Evolution of essential centromere function *(Invited talk)*

HHMI conference: Evolution and Development, Bethesda, MD 2012
Evolutionary cell biology reveals the emergence of essential function in a *Drosophila* neogene *(poster)*

Society for Molecular Biology and Evolution, Dublin, Ireland 2012
Uncovering the molecular basis of the centromere paradox *(Invited talk)*

Annual *Drosophila* Research Conference, Chicago, IL 2012
Emergence of mitotic function in the *Drosophila* neogene *Umbrea (poster)*

International *Drosophila* Heterochromatin Meeting, Gubbio, Italy 2011
Centromeric neofunctionalization of the heterochromatin protein Umbrea *(Invited talk)*

Keystone Conference: Evolution and Development, Lake Tahoe, CA 2011
Emergence of essential centromeric function in a *Drosophila* neogene *(Invited talk)*

International *C. elegans* Meeting 2007

C. elegans VRK-1 is required for meiotic progression and chromatin organization through MAP kinase signaling *(poster)*

International *C. elegans* Meeting 2006
RNAi knockdown of *C. elegans* VRK-1 results in meiotic defects and penetrant lethality *(poster)*

TEACHING

Lakeside High School, Seattle WA:
Invited Lecture May 2013

University of Washington:
Biology 302 (Cell and Molecular Biology) Fall 2010
Biology 355 (Cell and Molecular Biology lab) Winter 2011

MENTORING

Lara Morrison (undergraduate researcher) 2011 – 2013
Functional consequences of centromere protein evolution in Drosophila
Lisa Kursel (MCB rotation student) 2013
Functional characterization of a new HP1 gene in Drosophila pseudoobscura
Meara Davies (MCB rotation student) 2012
Analysis of uncharacterized MADF/BESS domain proteins in Drosophila
Claire Gonzalez (MCB rotation student) 2011
Analysis of uncharacterized MADF/BESS domain proteins in Drosophila
Jen Cech (MCB rotation student) 2011
Analysis of uncharacterized MADF/BESS domain proteins in Drosophila

COMMITTEES

UW Dept. of Molecular and Cellular Biology Steering Committee 2012 – 2013
(FHCRC student representative)

FHCRC Weintraub Graduate Award selection committee 2012 – 2013

PEER REVIEW

Reviewer for papers submitted to:

Chromosoma

Developmental Cell

eLIFE

PROFESSIONAL MEMBERSHIPS

American Society for Cell Biology

Genetics Society of America

International Society for Molecular Biology and Evolution

THESIS COMMITTEE

Dr. Harmit Malik (FHCRC, Basic Sciences)

Dr. Robert Eisenman (FHCRC, Basic Sciences)

Dr. Celeste Berg (UW, Genome Sciences)

Dr. Richard Gardner (UW, Pharmacology)

Dr. Willie Swanson (UW, Genome Sciences)

REFERENCES CITED

1. Morgan, T.H. (1910). Sex Limited Inheritance in Drosophila. Science *32*, 120-122.
2. Compton, D.A. (2011). Mechanisms of aneuploidy. Current opinion in cell biology *23*, 109-113.
3. Gordon, D.J., Resio, B., and Pellman, D. (2012). Causes and consequences of aneuploidy in cancer. Nature reviews. Genetics *13*, 189-203.
4. Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. Nature reviews. Molecular cell biology *9*, 33-46.
5. Malik, H.S., and Henikoff, S. (2009). Major evolutionary transitions in centromere complexity. Cell *138*, 1067-1082.
6. Roach, K.C., Ross, B.D., and Malik, H.S. (2012). Rapid evolution of centromeres and centromeric/kinetochore proteins, (Oxford University Press).
7. Bloom, K.S., and Carbon, J. (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell *29*, 305-317.
8. Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional Drosophila centromere. Genome research *13*, 182-194.
9. Sun, X., Wahlstrom, J., and Karpen, G. (1997). Molecular structure of a functional Drosophila centromere. Cell *91*, 1007-1019.
10. Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. American journal of human genetics *37*, 524-532.
11. Haaf, T., Warburton, P.E., and Willard, H.F. (1992). Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell *70*, 681-696.
12. Cellamare, A., Catacchio, C.R., Alkan, C., Giannuzzi, G., Antonacci, F., Cardone, M.F., Della Valle, G., Malig, M., Rocchi, M., Eichler, E.E., et al. (2009). New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. Molecular biology and evolution *26*, 1889-1900.

13. Rudd, M.K., Wray, G.A., and Willard, H.F. (2006). The evolutionary dynamics of alpha-satellite. Genome research *16*, 88-96.
14. Malik, H.S., and Henikoff, S. (2002). Conflict begets complexity: the evolution of centromeres. Current opinion in genetics & development *12*, 711-718.
15. Choo, K.H. (2001). Domain organization at the centromere and neocentromere. Developmental cell *1*, 165-177.
16. Schueler, M.G., and Sullivan, B.A. (2006). Structural and functional dynamics of human centromeric chromatin. Annual review of genomics and human genetics *7*, 301-313.
17. Bachmann, L., and 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.
18. Lohe, A.R., and Brutlag, D.L. (1986). Multiplicity of satellite DNA sequences in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America *83*, 696-700.
19. Lohe, A.R., and Brutlag, D.L. (1987). Identical satellite DNA sequences in sibling species of Drosophila. Journal of molecular biology *194*, 161-170.
20. Sawamura, K., Fujita, A., Yokoyama, R., Taira, T., Inoue, Y.H., Park, H.S., and Yamamoto, M.T. (1995). Molecular and genetic dissection of a reproductive isolation gene, zygotic hybrid rescue, of Drosophila melanogaster. Idengaku zasshi *70*, 223-232.
21. Bosco, G., Campbell, P., Leiva-Neto, J.T., and Markow, T.A. (2007). Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics *177*, 1277-1290.
22. Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The Drosophila melanogaster Genetic Reference Panel. Nature *482*, 173-178.
23. Black, B.E., and Cleveland, D.W. (2011). Epigenetic centromere propagation and the nature of CENP-a nucleosomes. Cell *144*, 471-479.
24. Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. Nature *299*, 111-117.

25. Burrack, L.S., and Berman, J. (2012). Flexibility of centromere and kinetochore structures. Trends in genetics : TIG *28*, 204-212.
26. Yasuhara, J.C., and Wakimoto, B.T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. Trends in genetics : TIG *22*, 330-338.
27. Roach, K.C., Ross, B.D., and Malik, H.S. (2011). Adaptive evolution of centromeric proteins. In Encyclopedia of Life Science. (John Wiley & Sons Publishing Co.).
28. Malik, H.S., and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics *157*, 1293-1298.
29. Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. The Plant cell *14*, 1053-1066.
30. Schueler, M.G., Swanson, W., Thomas, P.J., and Green, E.D. (2010). Adaptive evolution of foundation kinetochore proteins in primates. Molecular biology and evolution *27*, 1585-1597.
31. Elde, N.C., Roach, K.C., Yao, M.C., and Malik, H.S. (2011). Absence of positive selection on centromeric histones in Tetrahymena suggests unsuppressed centromere: drive in lineages lacking male meiosis. J Mol Evol *72*, 510-520.
32. Talbert, P.B., Bryson, T.D., and Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. Journal of biology *3*, 18.
33. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences : CABIOS *13*, 555-556.
34. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution *24*, 1586-1591.
35. Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves Fde, L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., et al. (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. Cell *142*, 810-821.
36. Nishino, T., Takeuchi, K., Gascoigne, K.E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I.M., and Fukagawa, T. (2012). CENP-T-W-S-X

forms a unique centromeric chromatin structure with a histone-like fold. Cell *148*, 487-501.

37. Carroll, C.W., Silva, M.C., Godek, K.M., Jansen, L.E., and Straight, A.F. (2009). Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nature cell biology *11*, 896-902.
38. Clark, N.L., Alani, E., and Aquadro, C.F. (2012). Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome research *22*, 714-720.
39. Clark, N.L., Alani, E., and Aquadro, C.F. (2013). Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. Genetics *193*, 529-538.
40. Clark, N.L., and Aquadro, C.F. (2010). A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Molecular biology and evolution *27*, 1152-1161.
41. Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annual review of genetics *46*, 677-700.
42. Lyttle, T.W. (1991). Segregation distorters. Annual review of genetics *25*, 511-557.
43. Sandler, L., and Novitski, E. (1957). Meiotic Drive as an Evolutionary Force. The American Naturalist *91*, 105-110.
44. Larracuente, A.M., and Presgraves, D.C. (2012). The selfish Segregation Distorter gene complex of Drosophila melanogaster. Genetics *192*, 33-53.
45. Burt, A., and Trivers, R. (2006). Genes in conflict : the biology of selfish genetic elements, (Cambridge, Mass.: Belknap Press of Harvard University Press).
46. 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.
47. Fishman, L., and Saunders, A. (2008). Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science *322*, 1559-1562.

48. Pardo-Manuel de Villena, F., and Sapienza, C. (2001). Nonrandom segregation during meiosis: the unfairness of females. Mamm Genome *12*, 331-339.
49. Yu, H.G., Hiatt, E.N., Chan, A., Sweeney, M., and Dawe, R.K. (1997). Neocentromere-mediated chromosome movement in maize. The Journal of cell biology *139*, 831-840.
50. Rhoades, M.M. (1942). Preferential Segregation in Maize. Genetics *27*, 395-407.
51. Daniel, A., Hook, E.B., and Wulf, G. (1989). Risks of unbalanced progeny at amniocentesis to carriers of chromosome rearrangements: data from United States and Canadian laboratories. Am J Med Genet *33*, 14-53.
52. Pardo-Manuel de Villena, F., and Sapienza, C. (2001). Female meiosis drives karyotypic evolution in mammals. Genetics *159*, 1179-1189.
53. Chmatal, L., Gabriel, S.I., Mitsainas, G.P., Martinez-Vargas, J., Ventura, J., Searle, J.B., Schultz, R.M., and Lampson, M.A. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. Curr Biol *24*, 2295-2300.
54. Ross, B.D., and Malik, H.S. (2014). Genetic Conflicts: Stronger Centromeres Win Tug-of-War in Female Meiosis. Curr Biol *24*, R966-968.
55. Nanda, I., Schneider-Rasp, S., Winking, H., and Schmid, M. (1995). Loss of telomeric sites in the chromosomes of Mus musculus domesticus (Rodentia: Muridae) during Robertsonian rearrangements. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *3*, 399-409.
56. Britton-Davidian, J., Catalan, J., da Graca Ramalhinho, M., Ganem, G., Auffray, J.C., Capela, R., Biscoito, M., Searle, J.B., and da Luz Mathias, M. (2000). Rapid chromosomal evolution in island mice. Nature *403*, 158.
57. Gunduz, I., Lopez-Fuster, M.J., Ventura, J., and Searle, J.B. (2001). Clinal analysis of a chromosomal hybrid zone in the house mouse. Genetical research *77*, 41-51.
58. Hauffe, H.C., and Searle, J.B. (1998). Chromosomal heterozygosity and fertility in house mice (Mus musculus domesticus) from Northern Italy. Genetics *150*, 1143-1154.

59. Hunt, P.A., and Hassold, T.J. (2002). Sex matters in meiosis. Science *296*, 2181-2183.
60. Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science *293*, 1098-1102.
61. Henikoff, S., and Malik, H.S. (2002). Centromeres: selfish drivers. Nature *417*, 227.
62. Crow, J.F. (1991). Why is Mendelian segregation so exact? BioEssays : news and reviews in molecular, cellular and developmental biology *13*, 305-312.
63. McKee, B.D., Yan, R., and Tsai, J.H. (2012). Meiosis in male Drosophila. Spermatogenesis *2*, 167-184.
64. Shakes, D.C., Wu, J.C., Sadler, P.L., Laprade, K., Moore, L.L., Noritake, A., and Chu, D.S. (2009). Spermatogenesis-specific features of the meiotic program in Caenorhabditis elegans. PLoS genetics *5*, e1000611.
65. Ross, B.D., Rosin, L., Thomae, A.W., Hiatt, M.A., Vermaak, D., de la Cruz, A.F., Imhof, A., Mellone, B.G., and Malik, H.S. (2013). Stepwise evolution of essential centromere function in a Drosophila neogene. Science *340*, 1211-1214.
66. Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.Y., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232-235.
67. Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. The Journal of cell biology *127*, 581-592.
68. Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proceedings of the National Academy of Sciences of the United States of America *97*, 7266-7271.
69. Henikoff, S., Ramachandran, S., Krassovsky, K., Bryson, T.D., Codomo, C.A., Brogaard, K., Widom, J., Wang, J.P., and Henikoff, J.G. (2014). The budding yeast Centromere DNA Element II wraps a stable Cse4 hemisome in either orientation in vivo. eLife *3*, e01861.

70. Quenet, D., and Dalal, Y. (2014). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. eLife *3*, e03254.
71. Kato, H., Jiang, J., Zhou, B.R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T.S., Straight, A.F., and Bai, Y. (2013). A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. Science *340*, 1110-1113.
72. Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 is sufficient for centromere formation. Science *334*, 686-690.
73. Blower, M.D., and Karpen, G.H. (2001). The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nature cell biology *3*, 730-739.
74. Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E., et al. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. Nature cell biology *15*, 1056-1066.
75. Scott, K.C., and Sullivan, B.A. (2014). Neocentromeres: a place for everything and everything in its place. Trends in genetics : TIG *30*, 66-74.
76. Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Developmental cell *10*, 303-315.
77. Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., Saccani, S., and Heun, P. (2011). Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nature cell biology *13*, 799-808.
78. Ketel, C., Wang, H.S., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009). Neocentromeres form efficiently at multiple possible loci in Candida albicans. PLoS genetics *5*, e1000400.
79. Schittenhelm, R.B., Althoff, F., Heidmann, S., and Lehner, C.F. (2010). Detrimental incorporation of excess Cenp-A/Cid and Cenp-C into Drosophila centromeres is prevented by limiting amounts of the bridging factor Cal1. Journal of cell science *123*, 3768-3779.

80. Blower, M.D., Daigle, T., Kaufman, T., and Karpen, G.H. (2006). Drosophila CENP-A mutations cause a BubR1-dependent early mitotic delay without normal localization of kinetochore components. PLoS genetics *2*, e110.
81. Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Molecular cell *25*, 309-322.
82. Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proceedings of the National Academy of Sciences of the United States of America *97*, 1148-1153.
83. Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes & development *9*, 573-586.
84. Malik, H.S., Vermaak, D., and Henikoff, S. (2002). Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proceedings of the National Academy of Sciences of the United States of America *99*, 1449-1454.
85. Vermaak, D., Hayden, H.S., and Henikoff, S. (2002). Centromere targeting element within the histone fold domain of Cid. Molecular and cellular biology *22*, 7553-7561.
86. Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S., and Hemmerich, P. (2004). Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. Molecular and cellular biology *24*, 6620-6630.
87. Nagaki, K., Terada, K., Wakimoto, M., Kashihara, K., and Murata, M. (2010). Centromere targeting of alien CENH3s in Arabidopsis and tobacco cells. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *18*, 203-211.
88. Ravi, M., Kwong, P.N., Menorca, R.M., Valencia, J.T., Ramahi, J.S., Stewart, J.L., Tran, R.K., Sundaresan, V., Comai, L., and Chan, S.W. (2010). The rapidly

evolving centromere-specific histone has stringent functional requirements in Arabidopsis thaliana. Genetics *186*, 461-471.

89. Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M., and Chan, S.W. (2011). Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana. PLoS genetics *7*, e1002121.
90. Baker, R.E., and Rogers, K. (2006). Phylogenetic analysis of fungal centromere H3 proteins. Genetics *174*, 1481-1492.
91. Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS biology *10*, e1001428.
92. Powell, J.R., and Moriyama, E.N. (1997). Evolution of codon usage bias in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *94*, 7784-7790.
93. Vicario, S., Moriyama, E.N., and Powell, J.R. (2007). Codon usage in twelve species of Drosophila. BMC Evol Biol *7*, 226.
94. Hiraizumi, Y. (1990). Negative segregation distortion in the SD system of Drosophila melanogaster: a challenge to the concept of differential sensitivity of Rsp alleles. Genetics *125*, 515-525.
95. Wu, C.I. (1983). Virility Deficiency and the Sex-Ratio Trait in DROSOPHILA PSEUDOOBSCURA. I. Sperm Displacement and Sexual Selection. Genetics *105*, 651-662.
96. Ferree, P.M., and Barbash, D.A. (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS biology *7*, e1000234.
97. Sawamura, K., and Yamamoto, M.T. (1993). Cytogenetical localization of Zygotic hybrid rescue (Zhr), a Drosophila melanogaster gene that rescues interspecific hybrids from embryonic lethality. Molecular & general genetics : MGG *239*, 441-449.
98. Levine, M.T., Holloway, A.K., Arshad, U., and Begun, D.J. (2007). Pervasive and largely lineage-specific adaptive protein evolution in the dosage compensation complex of Drosophila melanogaster. Genetics *177*, 1959-1962.

99. Rodriguez, M.A., Vermaak, D., Bayes, J.J., and Malik, H.S. (2007). Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *104*, 15412-15417.
100. Lemos, B., Araripe, L.O., and Hartl, D.L. (2008). Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science *319*, 91-93.
101. Lemos, B., Branco, A.T., and Hartl, D.L. (2010). Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proceedings of the National Academy of Sciences of the United States of America *107*, 15826-15831.
102. Sackton, T.B., Montenegro, H., Hartl, D.L., and Lemos, B. (2011). Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *108*, 17046-17051.
103. Zhou, J., Sackton, T.B., Martinsen, L., Lemos, B., Eickbush, T.H., and Hartl, D.L. (2012). Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *109*, 9941-9946.
104. Carvalho, A.B., Lazzaro, B.P., and Clark, A.G. (2000). Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. Proceedings of the National Academy of Sciences of the United States of America *97*, 13239-13244.
105. Lemos, B., Branco, A.T., Jiang, P.P., Hartl, D.L., and Meiklejohn, C.D. (2014). Genome-wide gene expression effects of sex chromosome imprinting in Drosophila. G3 *4*, 1-10.
106. Raychaudhuri, N., Dubruille, R., Orsi, G.A., Bagheri, H.C., Loppin, B., and Lehner, C.F. (2012). Transgenerational propagation and quantitative maintenance of paternal centromeres depends on Cid/Cenp-A presence in Drosophila sperm. PLoS biology *10*, e1001434.

107. Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol *17*, 237-243.
108. Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029-1035.
109. Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics *196*, 961-971.
110. Gratz, S.J., Wildonger, J., Harrison, M.M., and O'Connor-Giles, K.M. (2013). CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. Fly *7*, 249-255.
111. Krassovsky, K., and Henikoff, S. (2014). Distinct chromatin features characterize different classes of repeat sequences in Drosophila melanogaster. BMC genomics *15*, 105.
112. Kennison, J.A. (1981). The Genetic and Cytological Organization of the Y Chromosome of DROSOPHILA MELANOGASTER. Genetics *98*, 529-548.
113. Melcher, M., Schmid, M., Aagaard, L., Selenko, P., Laible, G., and Jenuwein, T. (2000). Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. Molecular and cellular biology *20*, 3728-3741.
114. Lynch, M., and Katju, V. (2004). The altered evolutionary trajectories of gene duplicates. Trends in genetics : TIG *20*, 544-549.
115. Ohno, S. (1970). Evolution by gene duplication, (Berlin, New York,: Springer-Verlag).
116. Miklos, G.L., and Rubin, G.M. (1996). The role of the genome project in determining gene function: insights from model organisms. Cell *86*, 521-529.
117. Chen, S., Zhang, Y.E., and Long, M. (2010). New genes in Drosophila quickly become essential. Science *330*, 1682-1685.

118. Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. J Mol Evol *49*, 169-181.
119. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531-1545.
120. Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W. (2012). Evolution of increased complexity in a molecular machine. Nature *481*, 360-364.
121. Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. Nature reviews. Genetics *11*, 97-108.
122. Dennis, M.Y., Nuttle, X., Sudmant, P.H., Antonacci, F., Graves, T.A., Nefedov, M., Rosenfeld, J.A., Sajjadian, S., Malig, M., Kotkiewicz, H., et al. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell *149*, 912-922.
123. Weng, J.K., Li, Y., Mo, H., and Chapple, C. (2012). Assembly of an evolutionarily new pathway for alpha-pyrone biosynthesis in Arabidopsis. Science *337*, 960-964.
124. Zhang, J., Dean, A.M., Brunet, F., and Long, M. (2004). Evolving protein functional diversity in new genes of Drosophila. Proceedings of the National Academy of Sciences of the United States of America *101*, 16246-16250.
125. Zhang, J., Zhang, Y.P., and Rosenberg, H.F. (2002). Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet *30*, 411-415.
126. Deng, C., Cheng, C.H., Ye, H., He, X., and Chen, L. (2010). Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proceedings of the National Academy of Sciences of the United States of America *107*, 21593-21598.
127. Katju, V. (2012). In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. International journal of evolutionary biology *2012*, 341932.

128. He, X., and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics *169*, 1157-1164.
129. Assis, R., and Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *110*, 17409-17414.
130. Ding, Y., Zhao, L., Yang, S., Jiang, Y., Chen, Y., Zhao, R., Zhang, Y., Zhang, G., Dong, Y., Yu, H., et al. (2010). A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. PLoS genetics *6*, e1001255.
131. Vermaak, D., and Malik, H.S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. Annual review of genetics *43*, 467-492.
132. Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Molecular cell *41*, 67-81.
133. Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., et al. (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. Nature *496*, 377-381.
134. Cowieson, N.P., Partridge, J.F., Allshire, R.C., and McLaughlin, P.J. (2000). Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Curr Biol *10*, 517-525.
135. Thiru, A., Nietlispach, D., Mott, H.R., Okuwaki, M., Lyon, D., Nielsen, P.R., Hirshberg, M., Verreault, A., Murzina, N.V., and Laue, E.D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. The EMBO journal *23*, 489-499.
136. Levine, M.T., McCoy, C., Vermaak, D., Lee, Y.C., Hiatt, M.A., Matsen, F.A., and Malik, H.S. (2012). Phylogenomic Analysis Reveals Dynamic Evolutionary History of the Drosophila Heterochromatin Protein 1 (HP1) Gene Family. PLoS genetics *8*, e1002729.

137. Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.
138. Abel, J., Eskeland, R., Raffa, G.D., Kremmer, E., and Imhof, A. (2009). Drosophila HP1c is regulated by an auto-regulatory feedback loop through its binding partner Woc. PloS one *4*, e5089.
139. Smothers, J.F., and Henikoff, S. (2001). The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. Molecular and cellular biology *21*, 2555-2569.
140. Zhang, D., Wang, D., and Sun, F. (2011). Drosophila melanogaster heterochromatin protein HP1b plays important roles in transcriptional activation and development. Chromosoma *120*, 97-108.
141. Joppich, C., Scholz, S., Korge, G., and Schwendemann, A. (2009). Umbrea, a chromo shadow domain protein in Drosophila melanogaster heterochromatin, interacts with Hip, HP1 and HOAP. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *17*, 19-36.
142. Mukai, M., Hayashi, Y., Kitadate, Y., Shigenobu, S., Arita, K., and Kobayashi, S. (2007). MAMO, a maternal BTB/POZ-Zn-finger protein enriched in germline progenitors is required for the production of functional eggs in Drosophila. Mech Dev *124*, 570-583.
143. Greil, F., de Wit, E., Bussemaker, H.J., and van Steensel, B. (2007). HP1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. The EMBO journal *26*, 741-751.
144. Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved organization of centromeric chromatin in flies and humans. Developmental cell *2*, 319-330.
145. Marques, A.C., Vinckenbosch, N., Brawand, D., and Kaessmann, H. (2008). Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol *9*, R54.

146. Prud'homme, B., Gompel, N., Rokas, A., Kassner, V.A., Williams, T.M., Yeh, S.D., True, J.R., and Carroll, S.B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature *440*, 1050-1053.
147. Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.
148. Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.
149. Murzina, N., Verreault, A., Laue, E., and Stillman, B. (1999). Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Molecular cell *4*, 529-540.
150. Canzio, D., Larson, A., and Narlikar, G.J. (2014). Mechanisms of functional promiscuity by HP1 proteins. Trends in cell biology *24*, 377-386.
151. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.
152. Ohta, T. (1988). Evolution by gene duplication and compensatory advantageous mutations. Genetics *120*, 841-847.
153. Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol *10*, 27-30.
154. Orr, B., and Sunkel, C.E. (2011). Drosophila CENP-C is essential for centromere identity. Chromosoma *120*, 83-96.
155. Qian, W., He, X., Chan, E., Xu, H., and Zhang, J. (2011). Measuring the evolutionary rate of protein-protein interaction. Proceedings of the National Academy of Sciences of the United States of America *108*, 8725-8730.
156. Mendez, D.L., Kim, D., Chruszcz, M., Stephens, G.E., Minor, W., Khorasanizadeh, S., and Elgin, S.C. (2011). The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. Chembiochem : a European journal of chemical biology *12*, 1084-1096.

157. Kohl, K.P., Jones, C.D., and Sekelsky, J. (2012). Evolution of an MCM complex in flies that promotes meiotic crossovers by blocking BLM helicase. Science *338*, 1363-1365.
158. Malik, H.S., and Bayes, J.J. (2006). Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. Biochemical Society transactions *34*, 569-573.
159. Meehan, R.R., Kao, C.F., and Pennings, S. (2003). HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. The EMBO journal *22*, 3164-3174.
160. Tareen, S.U., Sawyer, S.L., Malik, H.S., and Emerman, M. (2009). An expanded clade of rodent Trim5 genes. Virology *385*, 473-483.
161. Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proceedings of the National Academy of Sciences of the United States of America *102*, 2832-2837.
162. Gatti, M., and Baker, B.S. (1989). Genes controlling essential cell-cycle functions in Drosophila melanogaster. Genes & development *3*, 438-453.
163. Przewloka, M.R., Zhang, W., Costa, P., Archambault, V., D'Avino, P.P., Lilley, K.S., Laue, E.D., McAinsh, A.D., and Glover, D.M. (2007). Molecular analysis of core kinetochore composition and assembly in Drosophila melanogaster. PloS one *2*, e478.
164. Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. The EMBO journal *21*, 1121-1131.
165. Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., and Worman, H.J. (1997). Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. The Journal of biological chemistry *272*, 14983-14989.

166. Thomae, A.W., Schade, G.O., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A pair of centromeric proteins mediates reproductive isolation in Drosophila species. Developmental cell *27*, 412-424.
167. Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P.V., and Kuroda, M.I. (2014). Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. Genes & development *28*, 1445-1460.
168. Schneiderman, J.I., Sakai, A., Goldstein, S., and Ahmad, K. (2009). The XNP remodeler targets dynamic chromatin in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *106*, 14472-14477.
169. Giles, K.E., Gowher, H., Ghirlando, R., Jin, C., and Felsenfeld, G. (2010). Chromatin boundaries, insulators, and long-range interactions in the nucleus. Cold Spring Harbor symposia on quantitative biology *75*, 79-85.
170. Padeken, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.J., Kremmer, E., and Heun, P. (2013). The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. Molecular cell *50*, 236-249.
171. Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D.A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science *314*, 1292-1295.
172. Rollins, R.A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Molecular and cellular biology *24*, 3100-3111.
173. Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochemical and biophysical research communications *331*, 929-937.
174. Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proceedings of the National Academy of Sciences of the United States of America *109*, 14746-14753.

175. Begun, D.J., Holloway, A.K., Stevens, K., Hillier, L.W., Poh, Y.P., Hahn, M.W., Nista, P.M., Jones, C.D., Kern, A.D., Dewey, C.N., et al. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS biology *5*, e310.
176. Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.
177. Watanabe, Y. (2005). Shugoshin: guardian spirit at the centromere. Current opinion in cell biology *17*, 590-595.
178. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.
179. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research *34*, W609-612.
180. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.
181. Bayes, J.J., and Malik, H.S. (2009). Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science *326*, 1538-1541.
182. Coyne, J.A.a.O., H.A (2004). Speciation, (Sunderland, MA: Sinauer Associates).
183. Orr, H.A. (1996). Dobzhansky, Bateson, and the genetics of speciation. Genetics *144*, 1331-1335.
184. Sturtevant, A.H. (1920). Genetic Studies on DROSOPHILA SIMULANS. I. Introduction. Hybrids with DROSOPHILA MELANOGASTER. Genetics *5*, 488-500.
185. Sawamura, K., Taira, T., and Watanabe, T.K. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. I. The maternal hybrid rescue (mhr) gene of Drosophila simulans. Genetics *133*, 299-305.
186. Sawamura, K., Watanabe, T.K., and Yamamoto, M.T. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. Genetica *88*, 175-185.

187. Barbash, D.A., Roote, J., and Ashburner, M. (2000). The Drosophila melanogaster hybrid male rescue gene causes inviability in male and female species hybrids. Genetics *154*, 1747-1771.
188. Phadnis, N., and Orr, H.A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science *323*, 376-379.
189. Watanabe, T.K. (1979). A gene that rescues the lethal hybrids between *D. melanogaster* and *D. simulans*. Japanese Journal of Genetics *54*, 325-331.
190. Muller, H.J., and Pontecorvo, G. (1940). Recombinants between Drosophila Species the F1 Hybrids of which are Sterile
. Nature *10 August*, 199-200.
191. G, P. (1943). Viability interactions between chromosomes of Drosophila melanogaster and Drosophila simulans. Journal of Genetics *45*, 51-66.
192. Orr, H.A., Madden, L.D., Coyne, J.A., Goodwin, R., and Hawley, R.S. (1997). The developmental genetics of hybrid inviability: a mitotic defect in Drosophila hybrids. Genetics *145*, 1031-1040.
193. Barbash, D.A. (2010). Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. Genetics *184*, 313-316.
194. Pal Bhadra, M., Bhadra, U., and Birchler, J.A. (2006). Misregulation of sex-lethal and disruption of male-specific lethal complex localization in Drosophila species hybrids. Genetics *174*, 1151-1159.
195. Bachtrog, D. (2008). Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in Drosophila. Genetics *180*, 1123-1129.
196. Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.
197. Nei, M., and Zhang, J. (1998). Molecular origin of species. Science *282*, 1428-1429.
198. Barbash, D.A., and Lorigan, J.G. (2007). Lethality in Drosophila melanogaster/Drosophila simulans species hybrids is not associated with

substantial transcriptional misregulation. Journal of experimental zoology. Part B, Molecular and developmental evolution *308*, 74-84.

199. Brideau, N.J., and Barbash, D.A. (2011). Functional conservation of the Drosophila hybrid incompatibility gene Lhr. BMC Evol Biol *11*, 57.
200. Barbash, D.A., Awadalla, P., and Tarone, A.M. (2004). Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. PLoS biology *2*, e142.
201. McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature *351*, 652-654.
202. Satyaki, P.R., Cuykendall, T.N., Wei, K.H., Brideau, N.J., Kwak, H., Aruna, S., Ferree, P.M., Ji, S., and Barbash, D.A. (2014). The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS genetics *10*, e1004240.
203. Hirst, J.D., Vieth, M., Skolnick, J., and Brooks, C.L., 3rd (1996). Predicting leucine zipper structures from sequence. Protein engineering *9*, 657-662.
204. Yamamoto, M.T. (1992). Inviability of hybrids between D. melanogaster and D. simulans results from the absence of simulans X not the presence of simulans Y chromosome. Genetica *87*, 151-158.
205. Khurana, J.S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. The Journal of cell biology *191*, 905-913.
206. McDermott, S.R., and Noor, M.A. (2010). The role of meiotic drive in hybrid male sterility. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *365*, 1265-1272.
207. Tao, Y., Hartl, D.L., and Laurie, C.C. (2001). Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *98*, 13183-13188.
208. Zanders, S.E., Eickbush, M.T., Yu, J.S., Kang, J.W., Fowler, K.R., Smith, G.R., and Malik, H.S. (2014). Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. eLife *3*, e02630.

209. Rozas, J., and Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics *15*, 174-175.