GENETIC & FUNCTIONAL ANALYSIS OF HYBRID INCOMPATIBILITY GENES IN DROSOPHILA

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Tawny Nicole Cuykendall
August 2015

© 2015 Tawny Nicole Cuykendall

ALL RIGHTS RESERVED

GENETIC AND FUNCTIONAL ANALYSIS OF HYBRID INCOMPATIBILITY GENES IN DROSOPHILA

Tawny Nicole Cuykendall, Ph.D.

Cornell University 2015

Reproductive isolating mechanisms maintain species boundaries by preventing gene flow and act either before or after mating. Hybrid incompatibility (HI) is an example of a post-zygotic barrier and includes hybrid lethality and sterility. Mating of *D. melanogaster* females with *D. simulans* males, which diverged ~3 million years ago, produce viable but sterile daughters, and sons which die as 3rd instar larvae. Lethality is caused by an epistatic interaction in the hybrid background between *D. melanogaster Hybrid male rescue (Hmr)* and *D. simulans Lethal hybrid rescue (Lhr)*.

There is evidence that additional unknown factors contribute this lethal interaction in hybrids. In collaboration with members of the Barbash lab, I performed a screen using the Bloomington Deficiency Kit to identify putative HI factors. We found that there are no additional major-effect HI genes in the *D. melanogaster* autosomal genome, however, additional HI factors on the X chromosome are required for lethality.

HI genes are typically characeterized by rapid evolution driven by selection.

Likewise, *Hmr* exhibits extensive coding sequence divergence between *D. melanogaster* and *D. simulans*. In order to identify the forces driving selection, it is necessary to first describe the intraspecific function of *Hmr*. I used immunofluorescence and fluorescent in situ hybridization (FISH) to characterize the nuclear localization of the encoded protein and found that Hmr is heterochromatic and localizes to specific satellite sequences. Using functional genomics techniques, I showed that Hmr regulates the transcript abundance of transposable elements (TEs), which comprise a large portion of heterochromatin. In addition, genome-wide profiling of H3K9me3 enrichment and RNA Polymerase II occupancy in *Hmr* mutants showed that Hmr likely silences TEs post-transcriptionally.

The heterochromatic properties of Hmr, as well as its role in TE regulation, are largely conserved between the orthologs. However, the *D. melanogaster* and *D. simulans* genomes have diverged in TE content, density, and activity. TEs are highly dynamic and rapidly evolving, and it is possible that the sequence evolution of TEs and other rapidly evolving repetitive elements has contributed to the divergence of *Hmr.*

BIOGRAPHICAL SKETCH

Tawny Nicole Cuykendall was born in Tarzana, CA, but grew up in the wholesome mid-western state of Iowa. As a young girl, she dreamed of being a primatologist and following in the footsteps of her favorite scientist, Dr. Jane Goodall. Before she really understood what it meant, she was fascinated that human and chimpanzee DNA are 98% identical and would share this information with whoever would listen. By middle school, she was quite captivated by genetics and loved drawing punnett squares, inspecting her sisters' ear lobes, and asking her parents if they could curl their tongues. However, she never fancied herself a scientist. She entered the University Iowa in 2001 with an undeclared major, but after learning genetics from Professor Joseph Frankel, she decided on Biology with a focus in Genetics and Biotechnology. After graduation, she stayed at the UI and worked for four years in the lab of Dr. Douglas Houston where she studied early embryonic development in *Xenopus laevis.* She gained valuable research experience in the Houston lab and joined the graduate field of Genetics, Genomics, and Development at Cornell University in 2009 with the goal of studying evolutionary genetics in Drosophila. She chose to do her thesis research in Dr. Daniel Barbash's lab where she studied the genetics of reproductive isolation in Drosophila for the next five years.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Dr. Daniel Barbash and thank him for his dedication in helping me to become a better scientist, for his patience when I was stuck, for taking the time each week to meet with me, and for his helpful comments on this thesis and other manuscripts. I would like to thank my committee members Dr. Andy Clark and Dr. Kenneth Kemphues for being supportive, sharing their vast knowledge and insight, and for never missing even one of my student seminars.

I would like to thank all the members of the Barbash lab: Dr. Erin Kelleher for sitting with me before my A exam and calming me down, for her support through the years, and for helping me start on my path to learning bioinformatics, Dr. Shamoni Maheshwari, Dr. Heather Flores and Dr. Satyaki Rajavasireddy for teaching me the ropes of the lab, Shuqing Ji for her invaluable help in troubleshooting protocols, Daniel Zinshteyn, Michael McGurk, and Kevin Wei for spending hours with me during practice talks, for helping me learn bioinformatics, and for their boisterous and entertaining conversations.

I would like to thank Carol Bayles for her assistance in the Microscopy and Imaging Facility. I would like to thank Dr. Margarida Moreira who helped me analyze my first genome-wide data set and for teaching me how to use Linux. I would like to thank the Lis lab for reagents and for help in doing ChIP-seq. I would like to thank the Department chair Dr. Eric Alani for always being a friendly face and the rest of the department for their support.

Finally, I would like to thank my family: my mom for teaching me to believe in myself, for her endless support and for being a source of inspiration, my dad for his unyielding optimism and for teaching me the most important thing I needed was perseverance, my sisters Maren and Shannon for always having my back and somehow putting things into perspective, and my boyfriend Stavros Nikolaou for inspiring me daily, supporting me in all my endeavors, and making my final years at Cornell ones I will never forget.

TABLE OF CONTENTS

HYBRID INCOMPATIBILITY IN DROSOPHILA MELANOGASTER/DROSOPHILA SIMULANS INTERSPECIFIC HYBRIDS 1
1.1 Genetics of Hybrid Incompatibility 1
1.2 Intraspecific function of Hmr 9
1.3 The piRNA pathway 11
1.4 Hmr functional divergence 13

A SCREEN FOR F1 HYBRID MALE RESCUE REVEALS NO MAJOR-EFFECT HYBRID LETHALITY LOCI IN THE *D. MELANOGASTER* AUTOSOMAL GENOME . 16
2.1 Introduction 16
2.2 Materials and Methods 22
2.2.1 Fly Stocks and Crosses 22
2.2.2 Genome Coverage 22
2.2.3 RNAi Construct Design and Knockdown 23
2.2.4 Reverse transcription polymerase chain reaction (RT-PCR) 25
2.2.5 Molecular Breakpoint Determination 25
2.3 Results 27
2.3.1 D. melanogaster Lhr rescues F1 hybrid male lethality 27
2.3.2 Lack of evidence for additional major-effect HI loci in the D. melanogaster genome 27
2.3.3 Characterization of 75A-76D region 31
2.3.4 Characterization of the 65D-E region 31
2.3.5 Characterization of molecular breakpoints by NGS of rescuing 3L deficiencies 33
2.3.6 siRNA was ineffective in silencing in D. simulans msl-3 36
2.3.7 mel-Hmr does not cause lethality to X_{sim} hybrid males. 39
2.4 Discussion 41
2.4.1 Major-effect versus minor-effect HI genes 41
2.4.2 Challenges with using shmiRs to knockdown gene expression. 42
2.4.3 The potential role of X chromosome genes 43

THE *HMR* AND *LHR* HYBRID INCOMPATIBILITY GENES SUPPRESS A BROAD RANGE OF HETEROCHROMATIC REPEATS 46
3.1 Introduction 46
3.2 Materials and Methods 51
3.2.1 Construction of the Lhr^{KO} mutant 51
3.2.2 Hmr transgenes 52
3.2.3 Transgenic Fly Lines 56
3.2.4 Drosophila strains 57
3.2.5 Fertility assays 57
3.2.6 Hatch rate assays 58
3.2.7 Crosses for generating Hmr genotypes for RNA-Seq of ovarian mRNA 58
3.2.8 Crosses for generating pure-species and hybrid samples for RNA-Seq of larvae 59
3.2.9 Preparation of protein lysates for semi-quantitative Western blots. 60
3.2.10 Anti-Lhr antibodies and Western blots 60
3.2.11 Co-immunoprecipitation 61
3.2.12 RT-PCR and qRT-PCR assays 62
3.2.13 qPCR of HeT-A DNA copy number 62
3.2.14 RNA-Seq samples 63
3.2.15 RNA-Seq analysis 64

3.2.16 Small RNA sequencing and analysis *65*
3.2.17 Immuno-fluorescence and Immuno-FISH *67*
3.2.18 Yeast two-hybrid assays *67*
3.2.19 Data Access *68*
3.3 Results 69
3.3.1 Lhr and Hmr form a complex with HP1a *69*
3.3.2 Lhr is required for female fertility *75*
3.3.3 Lhr and Hmr are required to repress transposable elements *77*
3.3.4 Lhr and Hmr affect expression of heterochromatic genes *81*
3.3.5 Lhr and Hmr mutants have long telomeres *81*
3.3.6 Satellite DNA transcripts are upregulated in Lhr and Hmr mutants. *83*
3.3.7 siRNA and piRNA patterns are largely normal in Lhr^{KO} *88*
3.3.8 Comparing Lhr function in D. simulans and D. melanogaster *92*
3.3.9 Comparison of Hmr ortholog function *95*
3.3.10 Transposable elements are upregulated in hybrids *99*
3.4 Discussion 102
3.4.1 Lhr and Hmr interact with HP1a *102*
3.4.2 Rapidly evolving heterochromatin proteins and repetitive DNA variation *104*
3.4.3 Hmr and Lhr repress transposable elements *105*
3.4.4 Hmr and Lhr regulate telomeres *106*
3.4.5 Hmr and Lhr regulate species-specific satellite DNAs *107*
3.4.6 Is the adaptive evolution of Hmr and Lhr driven by diverging heterochromatic repeats? *108*
3.4.7 Repeat load, adaptation and hybrid incompatibilities *110*

ANALYSIS OF THE MECHANISM OF HMR TE REPRESSION 114

4.1 Introduction 114
4.2 Materials and Methods 120
4.2.1 Fly stocks *120*
4.2.2 ChIP *120*
4.2.3 IP western *121*
4.2.4 ChIP-qPCR *122*
4.2.5 ChIP-Seq *122*
4.2.6 Sequencing data preprocessing *123*
4.2.7 ChIP-seq data analysis *124*
4.2.8 RNA-seq data re-analysis *127*
4.3 Results 128
4.3.1 Pol II Rpb3 and H3K9me3 pull down efficiencies *128*
4.3.2 ChIP-seq data quality *131*
4.3.3 Hmr^{3} and Hmr^{+} backgrounds have diverged *134*
4.3.4 The major mode of Hmr TE regulation is not transcriptional *136*
4.3.5 Hmr does not modify chromatin structure at TEs *141*
4.3.6 Hmr regulates H3K9me3 deposition on the X chromosome *144*
4.4 Discussion 147
4.4.1 Analysis pipelines lead to distinct conclusions *147*
4.4.2 Telomeric TE copy number increases in Hmr mutants *149*
4.4.3 Mechanism of Hmr TE repression *150*
4.4.4 Does Hmr function within the piRNA pathway? *152*
4.4.5 Hmr functions at the telomere *154*
4.4.6 Does the effect on the X chromosome indirectly regulate TE sequences genome-wide? *156*

HMR CHROMATIN IMMUNOPRECIPITATION 157
5.1 Introduction 157
5.2 Methods 160
5.2.1 Fly stocks 160
5.2.2 qRT-PCR 160
5.2.3 ChIP 161
5.2.4 Western blot and IP western 162
5.2.5 ChIP-qPCR 163
5.3 Results and Discussion 164
5.3.1 Hmr-HA transgenes are expressed 164
5.3.2 Anti-Hmr antibodies 164
5.3.3. Hmr is prone to degradation and pull-down is weak 167
5.3.4 Absence of Hmr enrichment at TEs 175
5.3.5 Conclusions 175

CHAPTER 1

HYBRID INCOMPATIBILITY IN DROSOPHILA MELANOGASTER/DROSOPHILA SIMULANS INTERSPECIFIC HYBRIDS

1.1 Genetics of Hybrid Incompatibility

Several different isolating mechanisms contribute to speciation. These mechanisms are classified as either pre-mating or post-mating. The lethality and sterility of interspecific hybrid progeny is an example of the latter and is referred to as hybrid incompatibility (HI). Genes contributing to HI, known as HI genes, have been identified in species ranging from *Arabidopsis* to mouse. The Dobzhansky-Muller (D-M) model posits that HI is an indirect consequence of lineage-specific evolution and that the incompatibility arises due to negative epistatic interactions among alleles in the hybrid background (Dobzhansky 1937; Muller 1940). The simplest form of the D-M model invokes two loci. For example, the ancestral genotype *aabb* diverges between two independently evolving lineages, whereby each lineage fixes a new allele, giving rise to two derived genotypes, *AAbb* and *aaBB*. The incompatible interaction arises between the '*A'* and the '*B'* derived alleles in the hybrid.

HI genes are frequently characterized by rapid divergence driven by selection (Presgraves 2010; Maheshwari and Barbash 2011). A fundamental goal of speciation genetics is to determine what is driving natural selection at these loci. Ecological adaptation and internal genetic conflicts are two evolutionary mechanisms that have been proposed to cause divergence of HI genes (Maheshwari and Barbash 2011).

Strikingly, several recently characterized HI genes exhibit heterochromatin-associated functions. Because selfish DNA elements, such as transposons and satellite sequences, make up the bulk of heterochromatin, this suggests that incompatibility could be an indirect consequence of genetic conflicts with the host (Presgraves 2010; Maheshwari and Barbash 2011). For example, *Zhr* is a tandemly repeated 359-bp species-specific monomer located in pericentric heterochromatin of the X chromosome in *D. melanogaster* (Sawamura *et al.* 1993). Ferree and Barbash (2009) showed that lethality is caused by mis-segregation of the *Zhr* satellite sequence in hybrids (Ferree and Barbash 2009). Another example is the rapidly evolving *OdsH*, which encodes a homeodomain protein (Ting *et al.* 1998). OdsH protein has diverged in its heterochromatic localization properties between the hybridizing species and mislocalizes to the heterochromatic Y chromosome in hybrids (Bayes and Malik 2009).

Matings between *D. melanogaster* females and *D. simulans* males produce sterile hybrid females and invariantly lethal hybrid sons, which die as 3rd instar larvae. Brideau *et al.* (2006) showed that hybrid lethality in this cross is due to the epistatic interaction between *D. melanogaster Hmr* (X chromosome) and *D. simulans Lhr* (chromosome 2), which form a pair of D-M interacting genes (Brideau *et al.* 2006). Loss-of-function mutations in *D. melanogaster Hmr* or *D. simulans Lhr* rescue lethality (Watanabe 1979; Hutter and Ashburner 1987). Thus, the wild-type activity of both *Hmr* and *Lhr* causes lethality. Like other HI genes, both *Hmr* and *Lhr* have evolved rapidly under positive selection (Barbash *et al.* 2004; Brideau *et al.* 2006). Notably, both proteins localize to

heterochromatin, suggesting that divergence in both genes could be a result of conflict-driven evolution.

The hybrid lethal function of *Lhr* has functionally diverged between *D. melanogaster* and *D. simulans* (Brideau *et al.* 2006). However, recent tests have shown that this divergence is less clear than previously thought (Maheshwari and Barbash 2012). Transgenic lines of *D. melanogaster* carrying either *D. melanogaster* or *D. simulans Lhr* transgenes were generated using ϕC31 site-specific integration (Groth *et al.* 2004) and crossed to *D. simulans* males carrying the Lhr^1 hybrid rescue allele. The hybrid lethal activity of each transgene was assayed by testing for complementation of *D. simulans* Lhr^1 (Maheshwari and Barbash 2012). Surprisingly, despite their extensive sequence divergence, both transgenes demonstrated hybrid lethal activity by suppressing hybrid male rescue, indicating that hybrid lethality is a shared ancestral function of *Lhr* (Maheshwari and Barbash 2012). Correspondingly, there is no divergence in the heterochromatic localization of Lhr, nor does it mislocalize in hybrids (Maheshwari and Barbash 2012). Moreover, expression of *D. simulans Lhr* in the *D. melanogaster* pure species background is not lethal (Brideau *et al.* 2006), suggesting that HI is more complex than the simple two locus model.

There are two lines of evidence demonstrating functional divergence of *Hmr* with respect to its hybrid function. First, patroclinous exceptional hybrid males, resulting from production of nullo X oocytes in the *D. melanogaster* mother, are viable, suggesting either that *D. simulans Hmr* does not exhibit hybrid lethal activity, or that there are other

factors on the *D. melanogaster* X chromosome that enhance the lethal activity of *Hmr*. Second, P-element transgenes expressing *D. melanogaster Hmr*, but not *D. simulans Hmr*, suppress rescue by a loss of function *D. melanogaster Hmr* allele (Barbash *et al.* 2003, 2004). Thus, only *D. melanogaster Hmr* has hybrid lethal activity.

Hmr and *Lhr* are considered major-effect hybrid lethality genes because loss-of-function mutations in *D. melanogaster Hmr* or *D. simulans Lhr* fully suppress developmental lethality (Watanabe 1979; Hutter and Ashburner 1987). However, additional factors likely contribute to lethality in this system. GAL4/UAS driven expression of *D. simulans Lhr* in a *D. melanogaster* pure species background is insufficient to cause lethality (Brideau *et al.* 2006). Therefore, the interaction between *Hmr* and *Lhr* alone cannot account for F1 hybrid male lethality. Furthermore, experiments done nearly 70 years ago by Pontecorvo suggest that lethality involves interactions between loci on the *D. melanogaster* X *(Hmr)*, the *D. simulans* 2nd chromosome *(Lhr)*, and the *D. simulans* 3rd chromosome (Pontecorvo G 1943).

Several screens of the *D. melanogaster* genome have been done to search for additional HI genes in *D. melanogaster/D. simulans* hybrids (Coyne *et al.* 1998; Presgraves 2003; Matute *et al.* 2010). Coyne *et al*. (1998) crossed *D. melanogaster* stocks carrying deficiencies (deletions) to *D. simulans* males and assayed F1 hybrid female viability. This screen was designed to identify genes that cause lethality when hemizygous in a hybrid background and are therefore recessive. Approximately 50% of the *D. simulans* genome was screened; no regions causing unconditional lethality were

identified. Matute *et al.* (2010) repeated the screen and increased coverage by 30%, identifying 10 regions which decrease hybrid viability (Matute *et al.* 2010). Cattani and Presgraves (2012) re-examined the sibling species' X chromosome for incompatibilities and mapped a HI factor in the pericentric heterochromatin which they called *heterochromatin hybrid lethal* (*hhl*). They then generated *D. melanogaster* compound-X females carrying X chromosome duplications transposed on the Y (*Dp*(*1;Y*)) and crossed them to *D. mauritiana* males to screen for dominant lethals on the X (Cattani and Presgraves 2012). The only expected viable progeny from this cross are X^{mau}*/Dp(1:Y)*. Only one *Dp(1;Y)* was lethal in males and subsequent mapping refined the region to 9F12-10A6 (Cattani and Presgraves 2012). Matute and Gavin-Simith (2014) also attempted to map HI factors on the X chromosome using *Y*-linked duplications of X_{mel} in X_{sim} hybrids and identified two lethal duplications, one causing larval lethality spanning 9C-10B and a region spanning 4C-4D which caused pupal lethality, concluding that lethality is caused by a dominant-recessive interaction between the X chromosomes (Matute and Gavin-Smyth 2014). However, the region spanning 9C-10B had already been discovered more than a decade previously and shown to significantly reduce viability of F1 patroclinous exceptional sons (inheriting either X_{sim}, X_{sec}, or X_{mau}), (Barbash *et al*. 2000; Orr and Irving 2000), as well as X_{mel}/X_{mau} and X_{mel}/X_{sim} hybrid females (Barbash *et al.* 2000), presumably due partially to the presence of *Hmr.* This latter result indicates that lethality is caused by an interaction between dominant acting genes within the duplication and dominantly acting genes in *D. simulans* and *D. mauritiana* that are either autosomal or X-linked. (Cuykendall *et al.* 2014).

Presgraves (2003) performed a more sensitive screen to identify recessive-recessive interactions between the X chromosome and the autosomes by crossing *D. melanogaster* females with *D. simulans* males carrying the *Lhr* rescue mutation and assaying the viability of rescued F1 hybrid males (Presgraves 2003). Approximately 70% of the genome was screened, resulting in the identification of 40 regions (20 lethal, 20 semilethal), that when hemizygous in hybrids, cause lethality in rescued males, concluding that recessive-recessive HI is the most common type of interaction (Presgraves 2003). Among the genes mapped within these regions were two nucleoporins, *Nup96* (Presgraves 2003) and *Nup160* (Tang and Presgraves 2009). However, a major caveat of the aforementioned screens designed to identify recessively acting HI factors is that the lethal regions could potentially contain genes that are haploinsufficient or recessive lethals in hybrids (Barbash 2011). Furthermore, the genes identified are recessive and their effects are only visible in the F2 generation; therefore, they are not expected to affect F1 hybrid male viability.

However, both *Hmr* and *Lhr* are dominant and exert their effects in the F1 generation. Therefore, in collaboration with other members of the Barbash lab, I carried out a screen designed to identify putative HI genes in *D. melanogaster* with effects similar to *Hmr* and *Lhr*, that when removed, are dominant suppressors of hybrid lethality. We used the Bloomington Deficiency kit, which contains systematic chromosomal deletions, and screened 89% of the *D. melanogaster* autosomal genome. In order to increase sensitivity, we assessed viability throughout development, not just at the adult stage.

For example, pharates are fully formed adults that fail to eclose from the pupal case. Because hybrid males normally die as 3rd instar larvae, survival beyond the larval stage is considered rescue. *D. melanogaster* females carrying a deletion were crossed to *D. mauritiana* males. *Hmr* mutations rescue hybrid males with all three sibling species of *D. melanogaster*, however, *D. mauritiana* produces the highest level of rescue (Hutter and Ashburner 1987). *D. mauritiana* and *D. simulans* diverged relatively recently, within the past 250,000 years, (Kliman *et al.* 2000; McDermott and Kliman 2008) and therefore are expected to share HI genes, however, it is possible that lineage-specific HIs have evolved.

Maheshwari and Barbash (2012) showed that deletions in *D. melanogaster* spanning *Lhr* rescue hybrid males to the pharate stage (Maheshwari and Barbash 2012). This result supports the previously mentioned conclusion that the hybrid lethal function of *Lhr* is ancestral (Maheshwari and Barbash 2012). It is important to stress that rescue between *Lhr* orthologs is not equivalent: loss of the *D. simulans* ortholog rescues males to the adult stage, while loss of the *D. melanogaster* ortholog rescues to the pharate stage. Maheshwari and Barbash (2012) elegantly showed that this disparity in hybrid lethal activity is due to the asymmetrical expression of the *Lhr* orthologs in the hybrid background (Maheshwari and Barbash 2012).

Loss-of-function mutations in major-effect HI genes completely suppress developmental lethality. While *Lhr* is considered major-effect in *D. simulans*, it is considered as a minor-effect gene in *D. melanogaster.* Therefore, our deficiency screen has the potential to

detect major-effect HI genes in the sister species. We identified several regions that produce weak rescue (Cuykendall *et al.* 2014). The highest level of rescue was observed on chromosome 3L in a region spanning the gene *male-specific lethal-3* (*msl-3*), which is involved in dosage compensation. Like *Hmr* and *Lhr*, *msl-3* is rapidly evolving due to positive selection (Rodriguez *et al.* 2007). Interestingly, components of the dosage compensation complex (DCC) have been shown to mislocalize in hybrids (Pal Bhadra *et al.* 2006), however, a recent study showed Msl-2 is retained on the X in hybrids (Thomae *et al.* 2013). Furthermore, transcription from the X chromosome is not preferentially affected in hybrids suggesting that dosage compensation is not a direct cause of hybrid lethality (Wei *et al.* 2014). Attempts to specifically knock-down *D. simulans msl-3* in hybrids using RNAi were unsuccessful due in part to nonspecific effects (Cuykendall *et al.* 2014). Therefore, the effect of *msl-3* in hybrid lethality remains unknown. However, use of the clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system to assay mutations in *D. simulans* candidate HI, such as *msl-3*, should prove effective.

While several other deletions rescued to the pharate stage, the putative HI genes responsible could not be mapped, presumably because this phenotype is vulnerable to background effects. We concluded that there are no additional major-effect HI loci in the *D. melanogaster* autosome. However, crosses between *D. melanogaster* compound-X females homozygous for ϕ{*mel-Hmr-HA}* with *D. simulans* produced viable hybrid sons carrying the paternal X chromosome, indicating that additional factors on the *D. melanogaster* X are required for lethality (Cuykendall *et al.* 2014).

1.2 Intraspecific function of Hmr

According to the D-M model, the hybrid lethal function of *Hmr* is an indirect consequence of lineage-specific divergence. Therefore, in order to understand the evolutionary forces driving this divergence, characterization of the intraspecific molecular function(s) of *Hmr* is required.

Loss-of-function mutations in *D. melanogaster Hmr* reduce fertility and have a variable effect on viability (Aruna *et al.* 2009). Co-immunoprecipitation studies suggest that Lhr and Hmr interact in a complex with HP1a (Satyaki *et al.* 2014). The interaction between Hmr and Lhr is striking because it parallels the genetic interaction of *Hmr* and *Lhr* in hybrids. Furthermore, immunofluorescence using epitope-tagged *Hmr* and *Lhr* transgenes has shown that both proteins localize to pericentric heterochromatin in *D. melanogaster* (Maheshwari and Barbash 2012; Satyaki *et al.* 2014). However, Hmr and Lhr do not exhibit identical localization patterns (S. Prasad, unpublished data), indicating that while Hmr and Lhr participate in a shared function, they likely have additional unique roles.

Hmr encodes a 1,430 amino acid chromatin-binding protein containing four MADF domains (Barbash *et al.* 2003; Maheshwari *et al.* 2008). The MADF domain is homologous to the DNA-binding domain in the Adf1 transcription factor (England *et al.* 1992), suggesting a possible role in gene regulation for Hmr (Maheshwari *et al.* 2008). Fluorescent in situ hybridization (FISH) experiments on ovaries and embryos showed that Hmr associates with a subset of satellite sequences, including the pericentromeric

satellite dodeca (AATAACATAG) on 2L and 3L (Abad *et al.* 1992), GA-rich repeats, and the 2L3L satellite (Satyaki *et al.* 2014). In addition to satellite sequences, transposable elements (TE) comprise a large component of heterochromatin. Analysis of steady-state transcript levels in ovaries from *Hmr* mutants and Hmr^{+} identified 55 TE families that were upregulated at least 2 fold in *Hmr* mutants, indicating that Hmr functions as a repressor of TE transcript abundance (Satyaki *et al.* 2014). Both germline (e.g. *HeT-A, copia*) and somatic (*gypsy*) TEs were misregulated (Satyaki *et al.* 2014). Among the most highly derepressed TEs were the telomeric TEs, *HeT-A*, *TART*, and *TAHRE*. In addition, a small subset of satellite sequences also displayed increased transcription in the *Hmr* mutant (Satyaki *et al.* 2014).

The telomeric TEs *HeT-A, TAHRE,* and *TART* showed the highest derepression in both *Hmr* and *Lhr* mutants (Satyaki *et al.* 2014). *Drosophila* maintain telomere length by successive rounds of transposition of the telomeric retrotransposons at chromosome ends (Pardue and Debaryshe 2011). Therefore, precise regulation of the telomeric TEs is critical. The distal end of the telomere is a protein-rich structure called the cap. HP1a localizes to the cap and interacts with the HP1/ORC associated protein (HOAP) and the HP1-HOAP interacting protein (HipHop) to protect the chromosome ends and preserve genome stability (Andreyeva *et al.* 2005; Gao *et al.* 2010; Raffa *et al.* 2011). Hmr and Lhr interact in a complex with HP1a and also colocalize with HP1a at telomere ends (Satyaki *et al.* 2014). Furthermore, both *Hmr* and *Lhr* mutants possess extended telomeres compared to wild-type controls (Satyaki *et al.* 2014). These observations strongly implicate both Hmr and Lhr in telomere function.

1.3 The piRNA pathway

TE derepression is a hallmark of piRNA pathway mutants. Active TEs endanger the host genome on several levels. Transposition can cause deleterious mutations by inserting into functional DNA (Zachar and Bingham 1982). In addition, the double-strand breaks required for transposition can destabilize the genome (Orsi *et al.* 2010). The piRNA pathway protects eukaryotic genomes from unsolicited TE activity by invoking small-RNA mediated silencing. Small antisense RNAs (23-30 nt), known as piRNAs, originate from regions in the genome containing mostly degenerate TE sequences that represent a broad range of TE families (Brennecke *et al.* 2007). These regions (piRNA clusters) are transcribed and the resulting transcript is processed and then complexed with PIWI proteins, which then target TEs for post-transcriptional gene silencing.

The PIWI family proteins Piwi, Aubergine (Aub), and Argonuate3 (AGO3) are central players in piRNA-mediated silencing in the germline. Antisense piRNAs bound to Piwi-clade Argonaute proteins target homologous sequences for degradation. An additional role for Piwi in recruiting HP1a and establishing H3K9 methylation at TE sequences has also been described (Klenov *et al.* 2011; Wang and Elgin 2011; Sienski *et al.* 2012; Le Thomas *et al.* 2013; Rozhkov *et al.* 2013). However, Klenov *et al.* (2014) assayed H3K9me3 and HP1a enrichment at TEs in ovaries lacking nuclear Piwi, and found that only a subset of TEs showed differential enrichment (Klenov *et al.* 2014), suggesting that the heterochromatinization of the majority of TEs is independent of Piwi.

An essential component of the piRNA pathway is piRNA biogenesis. Primary piRNA transcripts are transcribed from piRNA clusters. Recently, a complex comprised of Rhino (Rhi), Cutoff (Cuff) and Deadlock (Del) was shown to be required for transcription from dual-strand clusters (Mohn *et al.* 2014). Dual-strand clusters exhibit transcription from both strands and are active primarily in the germline (Siomi *et al.* 2011). Primary piRNA transcripts are processed in the cytoplasm. Aub and AGO3 function in a piRNA amplification loop called the 'ping pong' cycle. TE degradation provides sense piRNAs, which are exploited by the 'ping pong' cycle and function to guide cleavage of piRNA cluster transcripts, thereby amplifying the signal. Loss of Aub or AGO3 causes a severe reduction in piRNA pools (Li *et al.* 2009; Malone *et al.* 2009).

A recent screen using RNAi was done to identify additional germline piRNA pathway components (Czech *et al.* 2013). Using RNAi, thousands of genes were screened and among them, 74 were identified as having a significant effect on TE transcript abundance (Czech *et al.* 2013). In the primary screen, only 4 TEs were assayed for derepression (*HeT-A, TAHRE, blood, burdock*). Interestingly, *Hmr* was not identified as having a strong effect on the transcription of these elements. However, transcriptome analysis of ovaries dissected from *Hmr* mutants showed massive derepression of both *HeT-A* and *TAHRE,* representing a 30 fold increase in transcript abundance (Satyaki *et al.* 2014). This discrepancy could be due to incomplete knockdown by Czech *et al.* 2013, difference in tissue assayed (ovaries vs. whole fly), or method used to measure TE transcript abundance (RNA-seq vs. RT-qPCR).

Similar to *Hmr* and *Lhr,* several piRNA pathway proteins exhibit signatures of adaptive evolution. It is unclear whether Hmr and Lhr function in concert with known piRNA pathway proteins or are part of a parallel pathway. *Aub* and *AGO3* mutants exhibit mislocalization of Vasa, which is required for secondary piRNA biogenesis, as well as piRNA pool depletion (Li *et al.* 2009; Malone *et al.* 2009). However, Vasa localization is not disrupted in *LhrKO* ovaries, nor is the piRNA pool aberrant (Satyaki *et al.* 2014), suggesting that Lhr, and by proxy, Hmr, are not involved in piRNA biogenesis. Thomae *et al.* (2013) used tandem-affinity purification in *D. melanogaster* Schneider cells (S2) to identify protein interactors with Hmr and Lhr. None of the identified factors are known piRNA-effector proteins, however, among the proteins shown to interact with Hmr were *Stonewall* (*Stwl*), *cenp-c*, and *mod(mdg4)* (Thomae *et al.* 2013). Notably, all three of these proteins displayed significant TE derepression in the germline screen by Czech *et al.* (2013).

1.4 Hmr functional divergence

Both *Hmr* and *Lhr* exhibit considerable divergence in coding sequence (CDS) congruent with positive selection (Barbash *et al.* 2004; Brideau *et al.* 2006). While the hybrid lethal function of *Hmr* shows clear divergence between *D. melanogaster* and *D. simulans* (Barbash *et al.* 2004), divergence of *Lhr* is less explicit (Maheshwari and Barbash 2012). However, according to the D-M model, hybrid lethality is an indirect consequence of lineage-specific evolution. This leads directly to the hypothesis that *Hmr* orthologs have diverged in their heterochromatic properties. Expression of a *D. simulans Hmr* transgene in *D. melanogaster* maintains heterochromatic localization (Satyaki *et al.*

2014), indicating that the *D. simulans* ortholog has not diverged to the extent that it does not recognize heterochromatic sequences in *D. melanogaster*.

Compatible with this result, *Hmr* orthologs have not diverged significantly with respect to TE regulatory function (Satyaki *et al.* 2014). Parallel *mel-Hmr-FLAG* and *sim-Hmr-FLAG* transgenes were crossed into an *Hmr* mutant background and ovarian TE transcript levels were compared. Only a small fraction of TE families are differentially regulated between the orthologs, displaying increased expression in Hmr^{-}; ϕ{*sim-Hmr-FLAG*}/+ relative to Hmr^{-}; ϕ{*mel-Hmr-FLAG*}/+. This indicates an inability of *D. simulans* Hmr to fully complement Hmr^{-}. Among the differentially expressed TEs are 3 putatively active elements (*Helena, Doc6*, and *BS*) which are candidates for co-evolution with *Hmr* (Satyaki *et al.* 2014).

The question remains as to what is driving the adaptive evolution of *Hmr* and *Lhr*. Because heterochromatic repeats are not subject to strong constraint, they are generally rapidly evolving, leading to the hypothesis that the evolution of repetitive sequences is driving divergence at *Hmr* and *Lhr*. However, the observation that only a few TE families are differentially regulated between *Hmr* orthologs does not support the hypothesis that *Hmr* is evolving in response to the collective evolution of repetitive sequences. It remains possible, though, that divergence is driven largely by a small number of sequences (e.g. *BS, Doc6*) (Satyaki *et al.* 2014). Alternatively, selection could be acting on an unknown function of *Hmr* (Satyaki *et al.* 2014).

The question also arises of whether the TE regulatory function of Hmr and Lhr is connected to hybrid lethality. Transcriptomes from hybrid male larvae carrying an *Hmr*$^{-}$ rescue allele and lethal *Hmr*$^{+}$ male larvae were sequenced (Satyaki *et al.* 2014). While the lethal *(Hmr*$^{+}$*)* hybrid males exhibit increased TE misexpression, the fitness effect of TE misexpression in hybrids is unknown (Castillo and Moyle 2012). However, there is little overlap between misregulated TEs in the rescued hybrid male and *Hmr* mutant ovaries (Satyaki *et al.* 2014). Furthermore, rescued hybrids have decreased TE expression compared to lethal hybrids, therefore, the presence of *Hmr* leads to increased expression (Satyaki *et al.* 2014). Consistent with the D-M model, the hybrid lethal activity of *Hmr* is likely an indirect consequence of lineage-specific evolution of *Hmr* likely driven by its interaction with rapidly evolving repetitive sequences.

This thesis details experiments done to 1) identify additional genes involved in HI, 2) describe the intraspecific molecular function of *Hmr*, and 3) investigate the functional divergence of the encoded protein between *D.melanogaster* and *D.simulans*. The second chapter is a paper published in *G3* (2014) 4: 2451–2460 detailing the results of the deficiency screen described above. The third chapter is a paper published in *PloS Genet* 10(2): e1004240, which describes both my work on *Hmr* as well as the work of my colleague, Satyaki PRV, on the intraspecific function of *Lhr*. The major finding of this paper is the shared function between *Hmr* and *Lhr* in silencing TEs. The fourth chapter is unpublished and aims to disentangle the molecular mechanism through which Hmr elicits TE regulation. The final chapter consists of ongoing work to identify the binding sites of the Hmr protein genome-wide.

CHAPTER 2

A SCREEN FOR F1 HYBRID MALE RESCUE REVEALS NO MAJOR-EFFECT HYBRID LETHALITY LOCI IN THE *D. MELANOGASTER* AUTOSOMAL GENOME[1]

2.1 Introduction

Speciation requires the evolution of reproductive isolating barriers that prevent the production of viable and fertile offspring between groups of individuals. Several isolating mechanisms maintain these barriers and are classified as either pre-mating or post-mating. Hybrid incompatibility (HI), the lethality and sterility of interspecific hybrid progeny, is an example of the latter. The Dobzhansky-Muller (D-M) model posits that HI is an indirect consequence of lineage-specific evolution and arises from negative epistatic interactions among alleles in the hybrid background. The simplest form of the D-M model invokes two loci. For example, the ancestral genotype $a_1a_1b_1b_1$, where a_1 and b_1 are the ancestral alleles of genes *a* and *b*, diverges between two independently evolving lineages, whereby each lineage fixes a new allele, giving rise to two derived genotypes, $a_2a_2b_1b_1$ and $a_1a_1b_2b_2$. The incompatible interaction arises between the 'a_2' and the 'b_2' derived alleles in the hybrid. A fundamental question arising from this model is whether HI is caused by a simple interaction between 2 loci, as illustrated by this general example of a D-M interaction, or rather by complex multilocus interactions.

[1] This is a manuscript published as Tawny N Cuykendall, P. Satyaki, Shuqing Ji, Derek M Clay, Nathaniel B Edelman, Alexandra Kimchy, Ling-Hei Li, Erin A Nuzzo, Neil Parekh, Suna Park, and Daniel A Barbash (2014) *G3.* 4: 2451–2460. Author contributions are as follows: T.N.C. P.S. & D.A.B. conceived and designed the experiments. T.N.C. P.S. S.J. D.M.C N.B.E. A.K. L.H.L E.A.N N.P S.P. performed the crosses. S.J. designed the RNAi constructs. T.N.C. analyzed data for tables and figures. T.N.C. & D.A.B. wrote the paper. Supplmental material available online: DOI: 10.1534/g3.114.014076.

Matings between *D. melanogaster* females and *D. simulans* males produce sterile hybrid females and invariantly lethal hybrid sons, which die as 3rd instar larvae (Barbash 2010b). Brideau et al. 2006 showed that hybrid lethality in this cross is due in part to the epistatic interaction between *D. melanogaster Hmr* and *D. simulans Lhr*, as posited by the D-M model (Brideau *et al.* 2006). *Hmr* and *Lhr* are characterized as major-effect HI genes because loss of function mutations in *D. melanogaster Hmr* or *D. simulans Lhr* suppress hybrid male lethality (Watanabe 1979; Hutter and Ashburner 1987). Both *Hmr* and *Lhr* are evolving rapidly due to positive selection in both the *D. melanogaster* and *D. simulans* lineages, suggesting functional divergence of the orthologs (Brideau *et al.* 2006; Maheshwari *et al.* 2008). Rescue by *Lhr* is asymmetric; only elimination of *D. simulans Lhr* rescues lethality to produce viable adult males, suggesting functional divergence of the *Lhr* coding sequence with respect to hybrid lethal activity (Brideau *et al.* 2006).

However, functional divergence of *Lhr* is more complex than originally proposed based on its asymmetry of rescue. Transgenic lines of *D. melanogaster* expressing either *D. melanogaster* or *D. simulans Lhr* transgenes were generated and the hybrid lethal activity of each ortholog was assayed by testing for complementation (i.e. suppression) of the *D. simulans Lhr*^{1} hybrid rescue mutation (Maheshwari and Barbash 2012). Despite their extensive sequence divergence, both transgenes suppressed rescue, indicating that hybrid lethal activity is an ancestral function of *Lhr* (Maheshwari and Barbash 2012). Further experiments showed that *D. melanogaster Lhr* is expressed at a lower level in hybrids compared to *D. simulans Lhr,* suggesting that *D. simulans Lhr*

may have higher hybrid lethal activity because it is expressed at a higher level in hybrids. Consistent with this interpretation, two *D. melanogaster Lhr*$^{-}$ deletions produced weak rescue to the pharate male stage (7-21% of total deficiency-carrying progeny) (Maheshwari and Barbash 2012).

Although *Hmr* and *Lhr* are major-effect hybrid lethality genes, additional factors likely contribute to lethality. Experiments performed by Muller and Pontecorvo nearly 70 years ago (1940, 1943) suggest that F1 hybrid male lethality involves interactions between loci on the *D. melanogaster* X (*Hmr*), the *D. simulans* 2nd chromosome (*Lhr*), and the *D. simulans* 3rd chromosome (Muller and Pontecorvo 1940; Pontecorvo G 1943). More recently, Brideau et al. (2006) found that expression of *D. simulans Lhr* in a *D. melanogaster* background is not lethal, demonstrating that the interaction between *D. melanogaster Hmr* and *D. simulans Lhr* is insufficient to cause lethality (Brideau *et al.* 2006). Taken together, these studies strongly suggest that additional factors contribute to lethality.

Several screens of the *D. melanogaster* genome have searched for additional HI genes in *D. melanogaster/D. simulans* hybrids (Coyne *et al.* 1998; Presgraves 2003; Matute *et al.* 2010). Coyne et al. (1998) crossed *D. melanogaster* stocks containing deficiencies (deletions) to *D. simulans* males and assayed F1 hybrid female viability. This screen was designed to identify genes that cause lethality when hemizygous in a hybrid background. These regions could potentially contain genes that are haploinsufficient or recessive lethal in hybrids. The screen covered just under 50% of the *D. simulans*

genome and did not find any regions that caused unconditional lethality. Matute et al (2010) repeated this screen with coverage increased to 79.4% of the genome, and identified 10 regions which cause lethality when hemizygous in hybrid females (Matute *et al.* 2010). Presgraves (2003) performed a more sensitive screen by crossing *D. melanogaster* females with *D. simulans* males carrying the *Lhr* hybrid rescue mutation and assaying the viability of rescued F1 hybrid males (Presgraves 2003). This screen, unlike that of Coyne et al. (1998), can identify recessive-recessive interactions between the X chromosome and the autosomes. He screened ~70% of the *D. simulans* genome and found 40 non-overlapping regions (20 lethal, 20 semi-lethal), that when hemizygous in hybrids, cause lethality in rescued males, concluding that recessive-recessive hybrid incompatibility is the most common type of interaction. The genes mapped within these regions include two nucleoporins, *Nup96* (Presgraves 2003) and *Nup160* (Tang and Presgraves 2009).

However, the genes identified in the above screens act recessively and are therefore not expected to affect F1 hybrid male viability. In contrast, both *Hmr* and *Lhr* are dominant and their presence causes lethality in the F1 generation. Here we use the Bloomington Deficiency Kit to systematically screen the vast majority of the *D. melanogaster* autosomal genome for genes with effects similar to *Hmr* and *Lhr* by crossing deficiency-carrying females to *D. mauritiana* males (Figure 2.1). *Hmr* mutations rescue hybrid males with all 3 of the sibling species of *D. melanogaster*, but rescue best with *D. mauritiana* compared to *D. simulans* and *D. sechellia* (Hutter and Ashburner 1987). Because *D. mauritiana* and *D. simulans* are very closely related, having

diverged within the past 250,000 years (Kliman *et al.* 2000; McDermott and Kliman 2008), we expect that most hybrid lethality genes are shared by the two species, but it is possible that lineage-specific HIs have evolved. For example, *Nup96*-dependent lethality is specific to *D. simulans* and *D. sechellia* (Barbash 2007). To further increase sensitivity, we screened for rescue to the pharate adult stage, since *D. melanogaster* deletions of *Lhr* rescue to this stage (Maheshwari and Barbash 2012). The difference in strength of rescue when deleting *D. melanogaster* versus *D. simulans Lhr* appears to be due to a higher expression level of *D. simulans Lhr* in hybrids (Maheshwari and Barbash 2012). Based on these previous findings with *Lhr*, we suggest that our screen also has the potential to uncover genes where the *D. mauritiana* (or *D. simulans*) allele contributes more to hybrid lethality than the *D. melanogaster* allele.

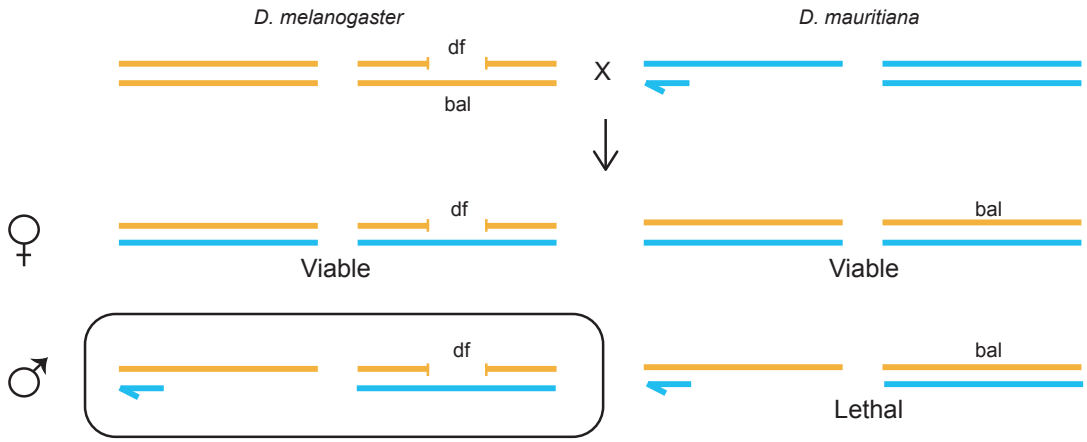


Figure 2.1. Screen design. *D. melanogaster* females from the Bloomington Deficiency Kit were crossed to *D. mauritiana* males. Female progeny inheriting either the deficiency chromosome (df) or the balancer chromosome (bal) are viable, though sterile. Males inheriting the balancer chromosome are invariably lethal. However, if the deficiency deletes a hybrid lethality gene we expect to observe rescue of this class of males (circled). We consider rescue to be survival to the pharate adult stage or beyond.

2.2 Materials and Methods

2.2.1 Fly Stocks and Crosses

D. melanogaster female flies from the Bloomington Deficiency Kit were crossed to at least two different lines of *D. mauritiana* (Figure 2.1). The *D. mauritiana* lines included $w^1 f^2$, W139, and two different isofemale lines (105 and 207). Crosses were set up at 18°C with ~20 females and ~25 males, flipped every 3-4 days for two weeks, and progeny scored until the last fly eclosed. Pharates were then dissected to determine sex and, where possible, genotype. Rescue was calculated by dividing the number of rescued male pharates by the number of females carrying the deficiency chromosome. A *D. melanogaster Lhr* deletion stock (*Df(2R)k08901*, which we will refer to as *Df(2R)Lhr*$^-$*)* was used as a positive control for pharate rescue (Maheshwari and Barbash 2012).

2.2.2 Genome Coverage

The proportion of the genome covered by the screen was calculated based on either the known or estimated molecular breakpoints of the deficiencies. Approximately 40% of the deficiencies are not mapped molecularly. For these, we estimated cytological breakpoints using GBrowse on FlyBase and *D. melanogaster Gene Models/Evidence* (*R5.48*) to convert to molecular estimates. When there was uncertainty in the cytological location of a breakpoint, we took the average of the extremes of the described range. We then determined regions of overlap among deficiencies and

counted each region (i.e. base pair) only once to arrive at the total number of base-pairs screened. This was done for each chromosome arm separately and then divided by the total number of base-pairs in the chromosome to determine the percentage of the chromosome arm covered.

2.2.3 RNAi Construct Design and Knockdown

We used the shmiR system of RNA interference with the Valium20 vector to attempt to knock down gene expression in hybrids (Haley *et al.* 2008; Ni *et al.* 2011). We initially designed 21 bp siRNAs, either with a single mismatch at position 2 (thought to ensure that only the antisense siRNA is loaded into the slicer complex), or without any mismatches as in (Ni *et al.* 2011) (Table 2.1). After 3 constructs failed to knockdown *D. simulans msl-3*, we made several design modifications. First, we increased the siRNA length to 22 bp because most shRNAs made from 3 different shRNA constructs expressed in S2 cells were 22 bp in length (Ni *et al.* 2011), and most microRNAs from the endogenous miR1 locus are also 22 bp (Ruby *et al.* 2007). Additionally, 22 bp siRNAs were found to have better silencing than shorter siRNAs (Wu *et al.* 2011). Second, mismatches between the guide and passenger strands were included at positions 2 and 11 (Haley *et al.* 2008) in order to mimic the endogenous structure of miR1. Note, however, that sequencing data suggests that mismatches are not necessary to achieve preferential accumulation of the guide versus passenger strands (Ni *et al.* 2011). A, C and G were mismatched with the same nucleotide; U was mismatched with C (Wu *et al.* 2011).

Table 2.1. ShmIR constructs and summary of results

Construct	Length	Mismatches	AS sequence	Knock-down assayed by RT-PCR
sim-Lhr-shRNA-577	21	Mismatch at bp 2	TAGATTCATTGCTAACACCAT	Yes, knockdown.
sim-msl-3-shRNA-579	21	Mismatch at bp 2	TATTGTGATAGAAGGTCTCGG	Yes, no knockdown.
sim-msl-3-shRNA-599	21	None	TAACATAGTTCTCCCTGTCGA	N/A; lethal to both sexes in *D. melanogaster*
sim-msl-3-shRNA-600	21	None	TAGTACCTTGACCATATTCCG	Yes - lethal to *D. melanogaster* males
sim-msl-3-shRNA-601	21	None	TAGCGCCGTCATCACTTGCAG	Yes, no knockdown.
sim-msl-3-shRNA-633	22	Mismatches at bps 2 and 11	TGAATGGGACCAAGTTAGTCAC	N/A; lethal to both sexes in *D. melanogaster*
sim-msl-3-shRNA-634	22	Mismatches at bps 2 and 11	TCTCCCGTGTGGAGTGGATCCA	Yes, no knockdown.
sim-msl-3-shRNA-635	22	Mismatches at bps 2 and 11	TCGCACATGGGCATCGACCGAT	N/A; semi-lethal to both sexes in *D. melanogaster*

ShmIR, short hairpin microRNA; AS, anti-sense; RT-PCR, reverse-transcription polymerase chain reaction; N/A, not applicable.

2.2.4 Reverse transcription polymerase chain reaction (RT-PCR)

cDNA synthesis was performed as in (Maheshwari and Barbash 2012). RT-PCR for *Lhr* used the primers GTAGCTTTCTCTTGGCGCTCTT and GTAAGTGAACTGAAGCTGCGTTGG, which span a fixed indel between *D. melanogaster* and *D. simulans Lhr*, amplifying products of 278 bp and 326 bp from *D. melanogaster* and *D. simulans*, respectively. RT-PCR for *D. melanogaster msl-3* (*mel-msl-3*) used the primers AGGAAAAACCCCCGTCCGGA and GCGTGCTGTTTGCCTAGTACCTT. RT-PCR for *sim-msl-3* used the primers AGGAGAAACCCCCGCCACCC and GCGTGCTGTTTGCCTAGTACCTT.

2.2.5 Molecular Breakpoint Determination

Df(3L)BSC27 (6867) was outcrossed to a sequenced wild-type DGRP strain (NC486). *Df(3L)BSC33* (6964) is balanced over *TM2*, and we did not outcross this stock to NC486 because *Ubx* was not sufficiently expressed to accurately genotype the progeny. DNA was isolated from approximately ten 3-day old females using the Qiagen DNeasy Blood and Tissue Kit according to the supplemental insect protocol. The Epigenomics Core Facility at Weill Cornell Medical Center constructed the libraries and performed paired-end sequencing.

Raw reads (50 bp) were aligned separately to the reference genome (Dmel r5) using the default settings of BWA 0.6.2. Three different methods were used to analyze the sequences: split-read mapping, paired-end mapping, and depth of coverage. The split-read mapping was performed by Pindel (Ye *et al.* 2009), which identifies paired reads in

which only one read maps to the reference genome. This mapped read serves as an anchor point for the detection of the other read. Depending upon the value of the expected sizes of CNVs, Pindel then splits the unmapped read into 2 or 3 fragments and maps them separately. Paired-end analysis was done using Hydra and Delly (Quinlan *et al.* 2010; Rausch *et al.* 2012), which identify read-pairs that do not map to the reference genome with the expected size and orientation and then clusters them based on which CNV they support. This approach successfully identified the deletion in *Df(3L)BSC27*. Lastly, depth of coverage was assessed using BEDTools to calculate the coverage per 50-base-pair interval and plotted in R. Breakpoints can be determined to within 500 base pairs with this method, which confirmed the predicted breakpoints in *Df(3L)BSC33*.

2.3 Results

2.3.1 *D. melanogaster Lhr* rescues F1 hybrid male lethality

We recently reported that *D. melanogaster Lhr* has weak hybrid lethal activity (Maheshwari and Barbash 2012). Two different Lhr^{-} deletions produced ~7%-21% rescue of males to the pharate adult stage in crosses to *D. mauritiana* (Maheshwari and Barbash 2012). We therefore used $Df(2R)Lhr^{-}$ as a positive control for hybrid male rescue in our screen and observed rescue with two different stocks of *D. mauritiana*, averaging 7%-8% at 18°C (Table 2.2).

However, as reported in (Maheshwari and Barbash 2012), we also found that other Lhr^{-} deletions (*Df(2R)BSC44* and *Df(2R)BSC161*), did not rescue males to the pharate stage. We conclude that our screen is vulnerable to false negatives due to background effects in the genome that affect our ability to detect weak rescue. This finding complicates mapping efforts because stocks that do not rescue are uninformative.

2.3.2 Lack of evidence for additional major-effect HI loci in the *D. melanogaster* genome

We screened 278 stocks from the Bloomington Deficiency Kit, covering approximately 89% of the autosomal genome (Supporting Information, Table S1). The average number of female progeny per cross was 238, with a slight bias for progeny carrying the deficiency chromosome (*t*-test, p-value<0.001). However, the numbers varied widely among crosses due to variable mating efficiency. 72.5% and 52.4% of crosses

produced a minimum of 50 and 100 deficiency-carrying females, respectively, summed over replicates. The percentage of each chromosome arm covered is shown in Table 2.2.

Table 2.2 Genome coverage of screen

Arm	All stocks screened	Sensitivity	
		≥50 Df-carrying females obtained	≥100 Df-carrying females obtained
2L	92.4%	52.8%	21.9%
2R	81.1%	56.3%	23.9%
3L	86.5%	83.8%	61.9%
3R	94.0%	91.2%	91.0%
Total of chromosome 2 and 3	**88.9%**	**72.5%**	**52.4%**

We did not find any regions in the *D. melanogaster* genome, which when removed, produce viable hybrid males (a single live male was found in one replicate with *Df(3L)XDI98*, but not in 3 other replicates). Therefore, within the regions of the genome that we screened, there are likely no additional major-effect *D. melanogaster* hybrid lethality genes. However, we found two adjacent deficiencies on 3L which give rescue comparable to positive controls with Lhr^{-} deficiencies, *Df(3L)BSC27* (65D4-65E6) and *Df(3L)BSC33* (65E10-65F6) (Table 2.3). We also found weak rescue with deficiencies spanning 75A6-76D5 (Table 2.3). Further characterization of these two regions is described below. Three additional regions, two on 3L and one on 3R, also produced pharate males, but were not pursued for further study due to low or variable rescue,

and/or the paucity of additional deficiency stocks (Table 2.3). For example, *Df(3L)emc-E12* produced 6.7% rescue with *D. mauritiana* W139 but no rescue with strain iso 105. Thus, we conclude that rescue to the pharate stage is a rare event and is vulnerable to background effects.

Table 2.3 Regions that rescue hybrid males to the pharate adult stage.

Region	Deficiency (% Rescue)	Molecular breakpoints[a]		Inferred molecular breakpoints[b]	
Lhr	Df(2R)BSC49 (21.4%)			12738807	13290649
	Df(2R)k08901 (7.21%/8.31%)	13309963	13340212		
	Df(2R)BSC44			13166788	13309036
	Df(2R)BSC161	13192288	13372333		
61A-62E5	Df(3L)emc-E12 (6.7%)			206780	885293
	Df(3L)R-G7 (3.0%)			1863545	2541764
	Df(3L)Ar14-8			641337	1615040
	Df(3L)BSC181	1688724	1841694		
65D-66C5	Df(3L)BSC27 (5.6%/10.8%)	6935985	7149104		
	Df(3L)RM5-2[c]			6999777	7879617
	Df(3L)BSC33 (12.8%)	7271620	7319021		
	Df(3L)GN24			3922651	5203390
	Df(3L)XDI98			3967594	4134155
	Df(3L)ZN47			5096316	6696471
	Df(3L)W5.4			5919622	7029849
	Df(3L)BSC411	5969060	6618726		
	Df(3L)Exel6109	6736213	6936639		
	Df(3L)BSC224	6957557	7150109		
	Df(3L)BSC374	6957558	7032145		

	Df(3L)RM5-1			6999777	7287396
	Df(3L)Exel6110	7087906	7149284		
	Df(3L)BSC117	7242575	7328086		
	Df(3L)pbl-X1			7349893	8129687
	Df(3L)Exel8104	7353086	7522363		
	Df(3L)ZP1			7889239	8254722
75A6-76D5	Df(3L)W10 (1.9%)			17867203	18202039
	Df(3L)fz2 (1.7%)			19148197	19226562
	Df(3L)BSC20 (0.8%)			19360266	19492579
	Df(3L)kto2 (3.5%)			19380732	19924632
	Df(3L)BSC8			17656096	18009745
	Df(3L)Cat			18056276	18834273
	Df(3L)ED4782	18988994	19163802		
	Df(3L)XS533			19481010	20314886
83B7-83D1	Df(3R)BSC47 (3.9%/1.7%)			1509535	1756808
	Df(3R)BSC464	1474083	2037668		

In cases where rescue occurred with both strains, percent rescue with *D. mauritiana w139* is presented first followed by percent rescue with *D. mauritiana* iso 105. Overlapping deficiencies that do not rescue are also listed.

[a] Breakpoints molecularly mapped.

[b] Breakpoints inferred based on cytology.

[c] *Df(3L)RM5-2* produced pharate hybrid males in initial crosses, but further testing failed to reproduce this result. See Table S1.

2.3.3 Characterization of 75A-76D region

The initial cross with *Df(3L)W10* produced 7.5% hybrid pharate males with *D. mauritiana* iso105, but additional replicates failed to produce rescue. Summing across replicates yields a final rescue of 1.9%. Two deficiencies, *Df(3L)BSC8* and *Df(3L)Cat*, which when combined encompass all of *Df(3L)W10*, did not rescue. However, other deficiencies adjacent to *Df(3L)W10* did rescue. *Df(3L)fz2*, which is ~80 kb and predicted to be ~1 Mb distal to *Df(3L)W10* rescued males at a similar low level (1.7%). Two additional deficiencies, *Df(3L)BSC20* and *Df(3L)kto2*, ~130 kb distal to *Df(3L)fz2* gave weak rescue (0.8% and 3.5%, respectively). These results suggest the possibility of multiple minor-effect genes in the 75A-76D region, but the low and variable level of rescue precluded further mapping.

2.3.4 Characterization of the 65D-E region

The *Df(3L)BSC27* and *Df(3L)BSC33* deletions both rescued pharate males greater than 10%. They are predicted to be 213 kb and 96 kb, respectively, and to be separated by 122 kb (Figure 2.2). When crossed together, we found trans- heterozygotes are viable, which is consistent with non-overlapping deficiencies. These results suggest either that there are 2 hybrid lethality loci in the region, or that they affect a single gene that is between them.

None of 13 additional deficiencies in this region produced pharate males (Figure 2.2, Table 2.3). The non-rescuing deficiencies include *Df(3L)BSC224*, which deletes all but 22 kb of *Df(3L)BSC27*, and *Df(3L)BSC117*, which encompasses the rescuing

Df(3L)BSC33. One possibility is that the rescuing deletion stocks contain second-site mutations that are responsible for the rescue. However, considering our results above showing variable rescue with *Lhr*$^{-}$ deletions, we conclude that *Df(3L)BSC27* and *Df(3L)BSC33* are identifying two candidate regions but that our power to further map the hybrid rescue gene(s) is severely limited by false negatives.

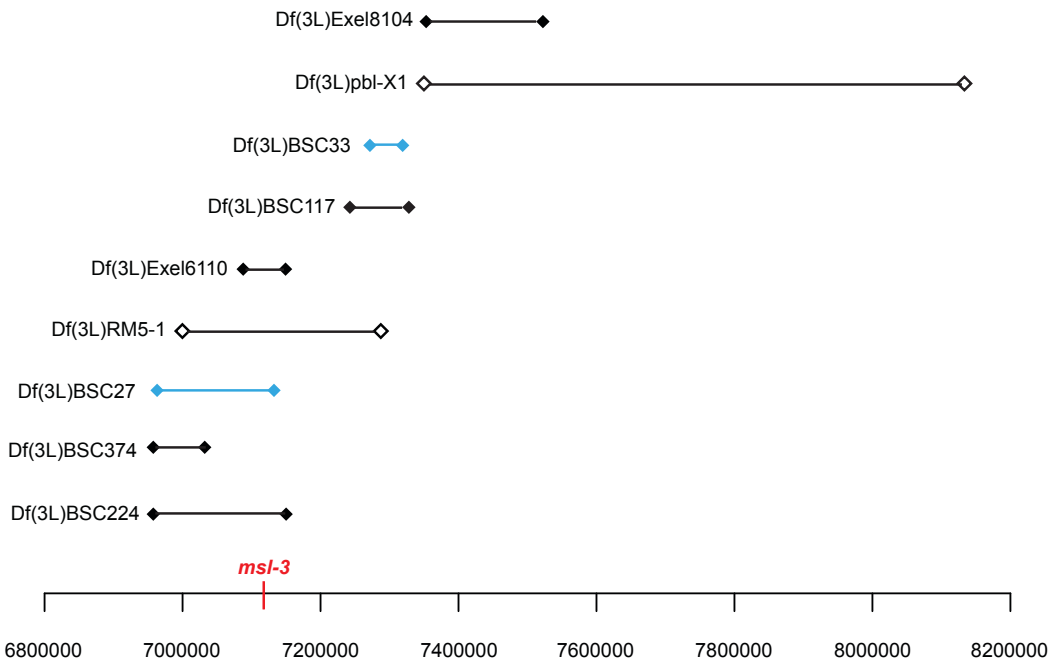


Figure 2.2. Deficiencies screened spanning 64C-66A. *Df(3L)BSC27* and *Df(3L)BSC33* spanning 65D4-65F6 produced rescued hybrid male pharates when crossed to *D. mauritiana* (blue bars). Seven deficiencies spanning this region that did not rescue are also shown. Filled endpoints denote molecularly defined deletions, whereas open endpoints indicate estimated breakpoints. Deficiencies *Df(3L)Exel6110* and *Df(3L)BSC27* were tested for complementation with *msl-3*^{1}; neither complemented *msl-3*^{1}, consistent with their molecularly mapped breakpoints. Complementation results are presented in Table S2.

2.3.5 Characterization of molecular breakpoints by NGS of rescuing 3L deficiencies

We attempted to confirm the breakpoints of the rescuing deletions by short-read sequencing. We identified the breakpoints of *Df(3L)BSC27* using both Hydra and Delly, which utilize paired-end read analysis (Quinlan *et al.* 2010; Rausch *et al.* 2012). The breakpoints were identified to be within 200 bp of those reported previously (Table 2.3). Pindel, which utilizes split-read analysis, did not identify this deletion, likely due to the 50 bp read length; however, our analysis of read-depth further supported this deletion (Figure 2.3). Neither paired-end nor split-read analysis determined the correct deletion in *Df(3L)BSC33*. Pindel identified a deletion partially overlapping the predicted region and matching the predicted size but it was not supported by read-depth analysis (Figure 2.3). However, read-depth analysis did identify a deletion that corresponds well to the predicted molecular breakpoints previously reported (Table 2.3). We conclude that the *Df(3L)BSC27* and *Df(3L)BSC33* deletion stocks are correct.

Sixty-three genes are annotated within *Df(3L)BSC27* and *Df(3L)BSC33* (Table 2.4). We prioritized candidates based on shared characteristics with *Hmr* and *Lhr*, such as encoding proteins that are nuclear, rapidly evolving, highly expressed in ovaries, chromatin-binding, have MADF and/or BESS domains, and are heterochromatic. Six genes (*tow, msl-3, Mis12, cdc27, bin, MED4,* and *mei-P22*) encode nuclear proteins involved in processes including dosage compensation, mitosis, transcription, and meiotic recombination. The msl-3 protein contains a chromatin-binding chromo domain (Koonin *et al.* 1995) as well as an MRG domain that is implicated in chromatin

remodeling (Bertram and Pereira-Smith 2001), while bin has a forkhead DNA-binding domain (Pérez Sánchez *et al.* 2002). Mis12 localizes specifically to the kinetochore and functions in mitotic spindle formation (Goshima *et al.* 2007). CG9948 is largely uncharacterized, but is of interest because it contains a MADF domain, similar to Hmr. Five of the candidate genes encode proteins highly expressed in the ovaries (*RpL18, CG9953, Cdc27, Galphai,* and *sgl)* with functions including translation, proteolysis, mitosis, receptor binding and oxidoreductase activity.

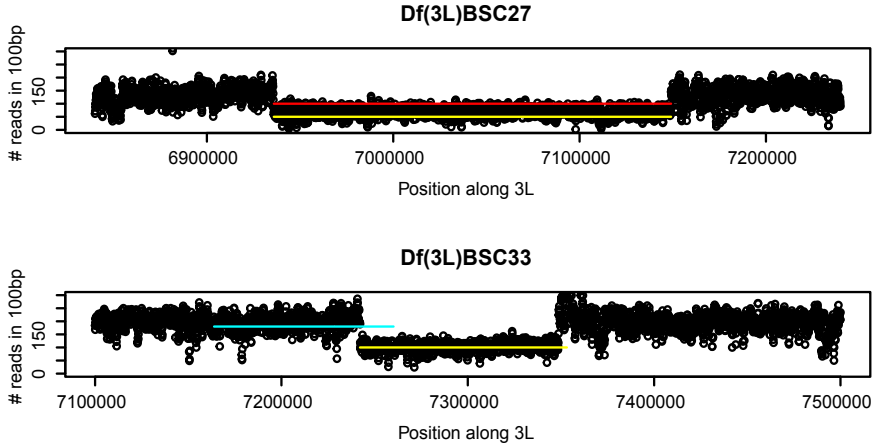


Figure 2.3. Coverage plots of deficiency stocks based on sequencing data. Coverage analysis supports the predicted deletions for *Df(3L)BSC27* and *Df(3L)BSC33.* The number of reads mapping within 100 bp intervals is plotted against the corresponding position on 3L for the sequenced deficiency lines. Segments in yellow represent the predicted location of deletions based on previous molecular or cytological estimates. Segments in red and cyan represent deletions predicted by paired-end approaches (Hydra and/or Delly) and Pindel, respectively.

Table 4. Genes mapped within rescuing deficiencies

Deficiency	No. of genes (high priority candidates)	Genes[1]
Df(3L)BSC27	44 (9)	***tow***, ***msl-3***, ***Cpr65Ec***, *CG17744, CG10077, Surf1, corn, form3, Cpr65Eb, mp,* ***Mis12***, *melt,* ***Galphai***, ***CG9953***, ***CG9948***, *CR32385, CG10063, CG34030, CG43439,* ***bin***, *CG14823, CG10075, CG32391, CG8629, Cpr65Ea, CG15829, CG8641, CG8628, CG32388, BBS1, Dbi, CG10064,* ***sgl***, *Prat2, ms(3)04202, mp, Me, Vn, CS3-1, rip, dv, E(Ubx)3L, jv*
Df(3L)BSC33	21 (4)	***MED4***, *unc-13-4A, mRpL50, CG14830, Neos, CG14829, CR43470,* ***mei-P22***, *Dscam2,* ***RpL18***, *CG14826, BHD,* ***Cdc27***, *ms(3)04202, CS3-1, anon-65Ea, CG8628, Dscam2, corn, E(Ubx)3L, form3*

[1] High priority candidates are indicated in bold, using criteria described in Results.

We focused on the gene *male specific lethal-3* (*msl-3*) in part because, like *Hmr* and *Lhr,* it encodes an adaptively evolving chromatin-binding protein (Rodriguez *et al.* 2007). The msl-3 protein is part of the dosage compensation complex (DCC) that binds to the X chromosome in males to mediate hyper-transcription (Gorman *et al.* 1995). A dosage compensation defect does not appear to be the direct cause of hybrid lethality because X chromosome transcripts are not preferentially affected in lethal hybrids (Wei *et al.* 2014). However, several observations suggest that dosage compensation genes may interact with hybrid lethality. Components of the dosage compensation complex (DCC) fail to localize to the X chromosome and H4K16Ac is not enriched on the X chromosome in hybrid males compared to pure species males (Pal Bhadra *et al.* 2006), although a later study did detect Msl-2 protein on the X (Thomae *et al.* 2013).

Additionally, mutations in *D. melanogaster* DCC genes including *msl-3* mildly enhance hybrid male viability when partially rescued by *Lhr* alleles (Barbash 2010a).

2.3.6 siRNA was ineffective in silencing in *D. simulans msl-3*

We first tested *D. melanogaster msl-3* as a candidate responsible for rescue by crossing *msl-3*^{1} females to *D. mauritiana*. No live or pharate hybrid males were observed (270 and 84 *msl-3*^{1}/+ hybrid females recovered in crosses with *D. mauritiana* iso105 and W139, respectively). We next attempted to test *D. simulans msl-3* (*sim-msl-3*) by knocking down its expression in hybrids using RNA interference (RNAi). The experimental challenge was to design an RNAi construct that would target the *D. simulans* ortholog but not the *D. melanogaster* ortholog, since removing both copies of *msl-3* would result in male lethality. The short hairpin microRNA (shmiR) method seemed ideal because it can achieve potent knockdown by expressing a single 21 bp siRNA from a modified miRNA-based vector (Haley *et al.* 2008). We used the Valium 20 transformation vector which expresses under the control of UAS sequences when crossed to a strain expressing the Gal4 activator protein (Ni *et al.* 2011). Our strategy was to transform these RNAi constructs into the *attP2* site on *D. melanogaster* chromosome 3, and then cross transformed stocks to a strain containing *actin-Gal4* on chromosome 2. These flies will express ubiquitously the shmiR, and when females are crossed to *D. simulans* males, one-quarter of the F1 hybrid progeny will inherit both the *Gal4* driver and the *UAS*-driven shmiR.

As a positive control, we first designed a vector that targets an insertion that is specific

to *sim-Lhr* (Tables 2.1, 2.5). Expression of this construct in *D. melanogaster* produced no phenotype. All four expected genotypes of F1 hybrid females were recovered. RT-PCR analysis showed that only *mel-Lhr* is expressed in *act-Gal4/+; UAS-shmIR-sim-Lhr/+* while both orthologs are expressed in control hybrid females (*+/+; UAS-shmIR-sim-Lhr/+*), demonstrating that the construct specifically knocks down *sim-Lhr* expression (Figure 2.4B). Interestingly, at 25°C females expressing the shmiR against *sim-Lhr* had the highest viability, which is consistent with *sim-Lhr* having a dominant effect on hybrid female viability at higher temperatures (Barbash *et al.* 2000). As predicted, knockdown of *sim-Lhr* rescued hybrid males, and RT-PCR analysis again demonstrated specific knockdown of the *sim-Lhr* otholog (Figure 2.4).

We designed 7 different shmiRs against *sim-msl-3* that each had mismatches to *mel-msl-3* (Table 2.1). Some were designed with modifications from published schemes (see Materials and Methods). Three of the constructs were lethal or semi-lethal to both sexes when expressed in *D. melanogaster* (Table 2.1), which must be due to off-target effects because *msl-3* is only required in males. A fourth construct (*msl-3-shRNA-600*) was lethal only to *D. melanogaster* males. Because *msl-3* is expressed in females, we could assay the effect of this shmiR, and found by RT-PCR that it does not reduce *msl-3* expression (Figure 2.4C). Therefore, this construct likely has a post-transcriptional effect on *mel-msl-3*.

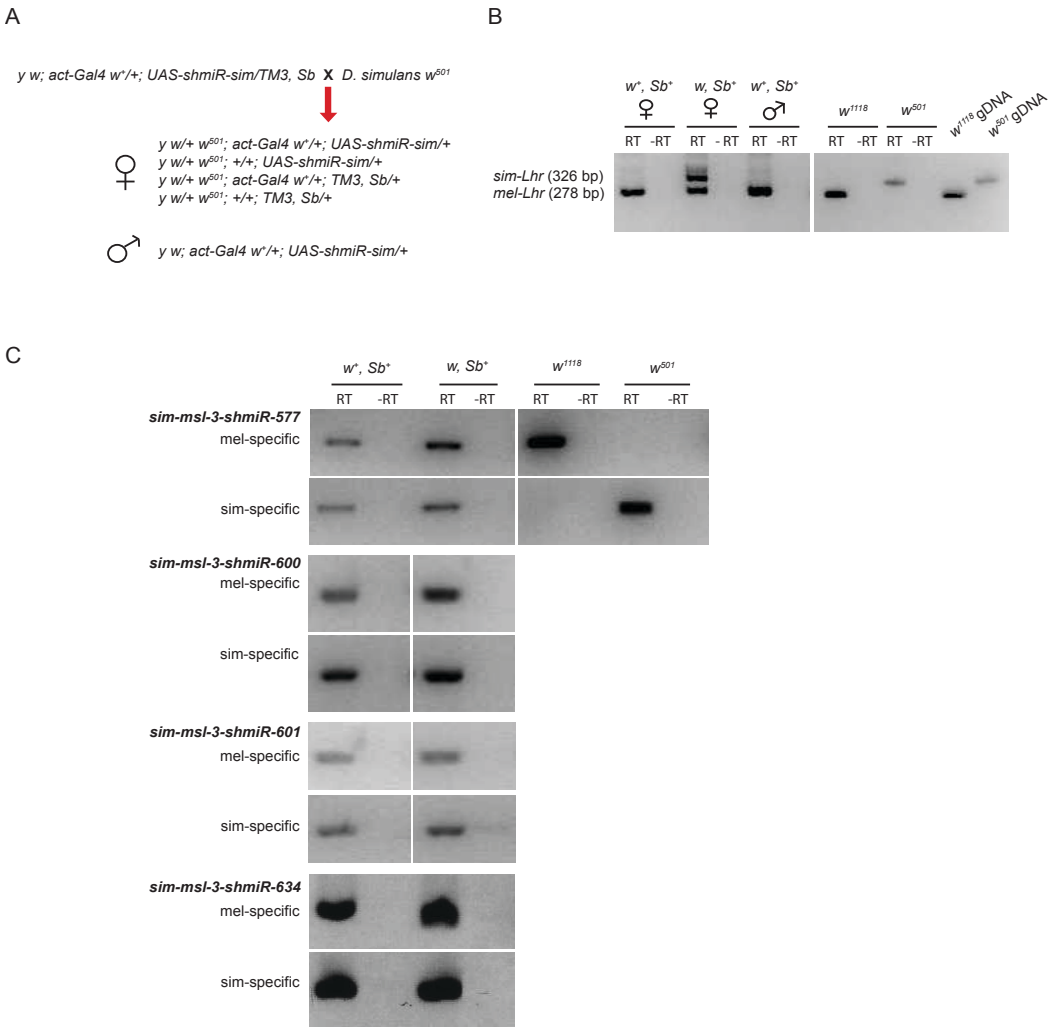


Figure 2.4. RT-PCR tests of shmiR knock-down of *D. simulans Lhr* and *msl-3*. A) *D. melanogaster* females expressing either *Lhr* or *msl-3* siRNAs targeting the *D. simulans* orthologs (abbreviated as *UAS-shmiR-sim*) were crossed to *D. simulans* males. w^+, Sb^+ hybrid progeny (1/4 of the total) inherit both the *Gal4* driver and *UAS-shmiR*. The males will survive if the *UAS-shmiR* construct knocks down expression of a hybrid lethality gene. B and C) RT-PCR tests of knockdown. B) Hybrid male and female progeny carrying both *act-Gal4* and *UAS-shmiR-sim-Lhr* (w^+, Sb^+) express *D. melanogaster mel-Lhr* (278 bp) but not *sim-Lhr* (326 bp), assayed using a single primer pair that detects an insertion in *sim-Lhr*. Hybrid females carrying only *UAS-shmiR-sim-Lhr* (*w*, Sb^+) were used as a control and express both orthologs. Right panel is RT-PCR and genomic DNA (gDNA) controls from *D. melanogaster* w^{1118} and *D. simulans* w^{501}. C) None of four tested *sim-msl3-shmiR* constructs silence *sim-msl-3* expression in hybrid female progeny. Separate PCR reactions were performed using primer pairs specific to either *mel-msl-3* or *sim-msl-3*, as confirmed using controls as above. Hybrid females carrying both *act-Gal4* and *UAS-shmiR-sim-msl-3* (w^+, Sb^+) expressed both *msl-3* orthologs. As a control, progeny only inheriting *UAS-shmiR-sim-msl-3* (*w*, Sb^+) were also assayed (except for *sim-msl-3-shmiR-577* where both *w*, Sb^+ and *w, Sb* animals were pooled), and expressed both orthologs as expected.

Table 2.5. Suppression of hybrid lethality by *UAS-shmIR-sim-Lhr*

		Number of progeny of indicated genotype (phenotype)								
Temp.		Sex of progeny		*GAL4/+; UAS/+* (w^+ Sb^+)		*GAL4/+;* *+/+* (w^+ Sb)		*+/+; UAS/+* (w Sb^+)		*+/+; +/+* (w Sb)
25°		Female		217		32		165		47
		Male		144		0		0		0
18°		Female		62		79		58		57
		Male		58		0		1		0

y w; P{w[+mC]=Act5C-GAL4}25FO1 /+; ϕ{UAS-shmIR-sim-Lhr}attP2, v^+ y^+/TM3,Sb D. melanogaster females were crossed to w^{501} *D. simulans* males. The transgenes are abbreviated as "*GAL4*" and "*UAS*" in the table headings.

The remaining 3 constructs were viable within *D. melanogaster* and thus could be crossed to *D. simulans* as we did for the *Lhr* control. None produced any hybrid males, but RT-PCR demonstrated that none silenced *sim-msl-3* expression. We were thus unable to test whether *sim-msl-3* affects hybrid male viability.

2.3.7 *mel-Hmr* does not cause lethality to X_{sim} hybrid males.

Our screen was limited to the autosomes, but we wished to test whether *mel-Hmr* can account entirely for the lethal effect of X_{mel} in hybrid males. X_{sim} (and X_{mau}) hybrid males are viable (Sturtevant 1920; Hutter *et al.* 1990), and a simple prediction is that the presence of *mel-Hmr* will kill X_{sim} hybrid males if *mel-Hmr* is the sole X-linked difference between these species involved in hybrid lethality. Hybrid sons carrying the paternal X

chromosome can be generated by crossing *D. simulans* males to compound-X *D. melanogaster* females. We tested the role of *mel-Hmr* using a *mel-Hmr-HA* transgene (Satyaki *et al.* 2014), and found that it had no effect on X_{sim} hybrid male viability (Table 2.6). We conclude that additional genes and/or sequences on X_{mel} are required for the fully penetrant lethality of X_{mel} hybrid males.

Table 2.6. A *mel-Hmr* transgene does not reduce X_{sim} hybrid male viability

Hybrid females		Hybrid males		
C(1)DX; +/+	*C(1)DX; ø{mel-Hmr-HA}/+*	*w/Y; +/+*	*w/Y; ø{mel-Hmr-HA}/+*	Relative viability
0	0	132	108	81.8% [1]

C(1)DX, y w f/Y; ø{mel-Hmr-HA}/+ D. melanogaster females were crossed to w^{501}*/Y D. simulans* males at 22-23°.
[1] Not significant by Chi-squared test ($p > 0.05$).

2.4 Discussion

2.4.1 Major-effect versus minor-effect HI genes

We used the Bloomington Deficiency Kit to screen for dominant suppressors of lethality in interspecific hybrids between *D. melanogaster* and *D. simulans*. Our screen is different from previous screens (Coyne *et al.* 1998; Presgraves 2003; Matute *et al.* 2010) in that it is designed to identify putative HI loci which cause dominant lethality, like *Hmr* and *Lhr*. The removal of either *Hmr* or *Lhr* suppresses hybrid male lethality, which classifies them as major-effect HI genes. Since we only screened for hybrid rescue with *D. mauritiana*, our screen is also predicated on the assumption that additional HI loci will be similar to *Hmr* and cause lethality with all 3 sibling species of *D. melanogaster*.

Our failure to identify any adult males suggests that the *D. melanogaster* genome does not harbor additional major-effect HI loci within the regions screened, but this does not exclude the possibility that there are additional factors of minor effect that contribute to hybrid lethality. We identified four regions, in addition to the region on 2R containing *Lhr*, which when deleted rescue hybrid male lethality to the pharate stage, suggesting the presence of minor-effect HI loci. However, the weak nature of these rescuing effects makes them susceptible to suppression by background effects. Without the ability to confirm them with multiple overlapping deficiencies, we conclude that the rescuing deficiencies identify regions that can be tentatively considered to contain minor-effect hybrid lethality loci.

The regions deleted by *Df(3L)BSC27* and *Df(3L)BSC33* gave the strongest level of rescue. Surprisingly these two deletions are distinct and do not overlap, as predicted by estimated cytological breakpoints and confirmed by complementation crosses and sequencing. It is possible, however, that both deletions affect a single gene. There are ~93 kb between *Df(3L)BSC27* and *Df(3L)BSC33*. Most enhancers in Drosophila are within 10 kb of their target sequence, but longer range interactions are known (Berman *et al.* 2004). For example, the *cut* gene is regulated by an enhancer 80 kb upstream of its promoter (Jack and DeLotto 1995). A recent genome-wide study estimated that ~28% of enhancers are >20 kb from their targets and can be over 100 kb away (Kvon *et al.* 2014).

Pontecorvo's experiments (1943) suggest that gene(s) on the *D. simulans* 3rd chromosome contribute to hybrid lethality. Based on previous findings that *D. melanogaster Lhr* has a weak effect on hybrid lethality, we reasoned that our screen has the ability to identify major-effect genes in *D. simulans* by detecting weak effects of deleting the *D. melanogaster* ortholog. We attempted to test *msl-3* as one such candidate by knocking down expression of *D. simulans msl-3* using shmiRs but unfortunately failed to do so.

2.4.2 Challenges with using shmiRs to knockdown gene expression.

The shmiR system is attractive because the expression of a single siRNA allows the design of siRNAs that target only one of the two orthologs in hybrids. We were successful in specifically targeting *sim-Lhr* by designing a shmiR targeting a species-

specific insertion in the *sim-Lhr* coding sequence. However, we were unable to achieve specific knockdown of *sim-msl-3*. Three constructs caused lethality to both sexes within *D. melanogaster* (Table 2.1), which must be off-target effects as *msl-3* is only required in males. One shmiR (*msl-3-sim-shmiR-600*) was lethal in *D. melanogaster* males, even though it has a mis-match to the *D. melanogaster msl-3* ortholog near its center. We found that *msl-3* mRNA expression is not reduced in the lethal males, suggesting that *msl-3* translation is likely being blocked via a microRNA-like effect of the seed sequence. Because as little as 6 bp of sequence homology near the 5' end of a microRNA can be sufficient for its activity (Lewis *et al.* 2005), this hypothesis would explain why the shmiR designed to target *D. simulans msl-3* would be able to target the *D. melanogaster* ortholog. This type of off-target effect will be challenging to predict since it requires limited homology, and suggests the need for caution when expressing shmiRs at high level. CRISPR-based methods may prove an effective alternative in the future for targeting mutations to *D. simulans* candidate HI genes

2.4.3 The potential role of X chromosome genes

The X chromosome contains 15% of *D. melanogaster* genes (Adams 2000) but we could not screen it because deletions of the X are lethal in males. We showed here that *mel-Hmr* is insufficient to explain the X-linked portion of hybrid male lethality, because a *mel-Hmr* transgene does not induce lethality of X_{sim} hybrid males (Table 2.6). This result is consistent with previous findings that a similar transgene significantly reduces viability of X_{mel}/X_{sim} hybrid females but does not induce full lethality (Barbash *et al.* 2003), in contrast to the invariant lethality of X_{mel}/X_{mel} hybrid females (Hutter *et al.* 1990).

Df(1)307-1-2 identifies a candidate region in 9D that is adjacent to but non-overlapping with *Hmr* (Barbash *et al.* 2003), but the causal gene remains unidentified. A recent screen examined *Y*-linked duplications of X_{mel} regions for lethality in X_{sim} hybrid males and identified 2 candidate regions (Matute and Gavin-Smyth 2014). One of these regions (9C-10B) is not a new discovery, as $Dp(1;2)v^{+75d}$ covering 9A2-10C2 was previously shown to reduce X_{mau}, X_{sec}, and X_{sim} hybrid male viability (Barbash *et al.* 2000; Orr and Irving 2000). The duplication also significantly reduces X_{mel}/X_{mau} and X_{mel}/X_{sim} hybrid female viability (Barbash *et al.* 2000), presumably due at least in part to the aforementioned effect of *Hmr*. This effect in females demonstrates that the 9C-10B duplication causes dominant lethality by interacting with *D. simulans* and *D. mauritiana* genes that are also dominantly acting and could be autosomal or X-linked, in contrast to the conclusion that it represents a dominant-recessive interaction between the X chromosomes (Matute and Gavin-Smyth 2014).

We have shown that *Hmr* cannot be the sole cause of the duplication-induced hybrid male lethality (Table 2.6), although it may be contributing. We suggest that lethality may result from the cumulative dosage increase of multiple duplicated genes, because well-characterized HI genes act as gain-of-function alleles in a hybrid background (Maheshwari and Barbash 2011). These putative dosage effects can only be tested for in a hybrid background. We reiterate here that testing fitness effects of chromosome aberrations within *D. melanogaster* is a useful general control but has no bearing on whether or not duplications or deficiencies are responsible for dosage effects in hybrids,

because HI genes by definition have distinct (and often opposite) properties in hybrid versus pure species backgrounds (Maheshwari and Barbash 2011).

Acknowledgments

We thank Dr. Margarida Moreira for help with identifying deficiency breakpoints, Kevin Wei, Michael McGurk, and Daniel Zinshteyn for helpful comments on the manuscript, and the Bloomington Drosophila Stock Center for the many deficiency stocks. Supported by NIH GM074737 to D.A.B.

CHAPTER 3

THE *HMR* AND *LHR* HYBRID INCOMPATIBILITY GENES SUPPRESS A BROAD RANGE OF HETEROCHROMATIC REPEATS[2]

3.1 Introduction

As populations diverge, their ability to reproduce with each other diminishes. Hybrid incompatibility (HI), the reduced viability and fertility of interspecific hybrids, is a major cause of reproductive isolation between nascent species and thus an important contributor to speciation. Many of the genes causing HI show evidence of adaptive evolution, typically manifest as excessive numbers of amino-acid-changing mutations compared to neutral expectations (Presgraves 2010; Maheshwari & Barbash 2011). These data do not, however, imply that natural selection acts directly on HI phenotypes. Rather, the prevailing model of HI formulated by Dobzhansky and Muller (D-M) emphasizes that incompatibilities evolve in two distinct steps. First, two or more loci diverge independently in two nascent species. Then, if these species later interbreed, these diverged genes may interact to cause deleterious HI phenotypes. The key insight of the D-M model is that hybrid lethality and sterility evolve as byproducts of intraspecific divergence (Maheshwari and Barbash 2011).

[2] This is a manuscript published as Satyaki, P. R. V., T. N. Cuykendall, K. H.-C. Wei, N. J. Brideau, H. Kwak *et al.*, (2014) *PloS Genet.* 10(3): e1004240. Author contributions are as follows: Conceived and designed the experiments: PRVS TNC KHCW DAB. Performed the experiments: PRVS TNC KHCW NJB HK SJ. Analyzed the data: PRVS TNC KHCW. Contributed reagents/materials/analysis tools: SA PMF. Wrote the paper: PRVS TNC KHCW DAB. Supplemental tables available online. doi:10.1371/journal.pgen.1004240.

Adaptive evolution therefore does ultimately lead to HI, but if we wish to identify the evolutionary forces that drive the divergence of HI genes, then we need to understand the function of these genes within species. The mechanisms by which HI genes cause sterility or lethality are important but separate issues. In fact, it remains uncertain whether the wild type functions of HI genes are generally predictive of the deleterious phenotypes that they cause within hybrids.

Pinpointing the function of HI genes and the causes of their adaptive evolution is a challenging goal. For example, the *Hybrid male rescue* (*Hmr*) gene causes large reductions in hybrid fitness (Barbash *et al.* 2003). Loss-of-function mutations in *D. melanogaster*, however, have only moderate effects on fertility and provide few insights into mechanistic underpinnings (Aruna *et al.* 2009). The nucleoporins provide an intriguing counterexample. Several have been implicated in hybrid lethality and found to evolve under adaptive evolution (Tang and Presgraves 2009). Mutations in nucleoporin subunits are lethal in *D. melanogaster*, but the genes have many pleiotropic functions and the challenge is to pinpoint which one(s) are driving evolutionary divergence.

Here we investigate two hybrid lethality genes, *Lethal hybrid rescue* (*Lhr*) and *Hmr*, which interact to cause F1 hybrid male lethality between *D. melanogaster* and *D. simulans* (Brideau *et al.* 2006). Both genes show extensive divergence in their coding sequences that is consistent with positive selection (Barbash *et al.* 2004; Brideau *et al.* 2006). For *Hmr* this sequence divergence appears to be required for hybrid lethality

because the *D. melanogaster* ortholog of *Hmr* causes hybrid lethality but the *D. simulans* ortholog does not (Barbash *et al.* 2004). For *Lhr*, however, both orthologs have hybrid lethal activity, with *D. simulans Lhr* having greater activity due to its higher expression level in hybrids (Maheshwari and Barbash 2012). That study left open the possibility that *Lhr* coding sequence divergence makes some contribution to hybrid lethality. Furthermore we found that *Lhr* from the more diverged species *D. virilis* has no hybrid lethal activity, suggesting that more extensive coding sequence divergence does have substantial functional consequences (Brideau and Barbash 2011).

These previous studies leave unanswered the fundamental question of what evolutionary force is driving adaptive sequence change, and necessitate a detailed understanding of *Hmr* and *Lhr* function within each of the hybridizing species. Loss of function alleles of *Hmr* and *Lhr* are strong suppressors of hybrid lethality, but are largely viable within *D. melanogaster* and *D. simulans*, respectively (Watanabe 1979; Hutter and Ashburner 1987).

Lhr (also known as HP3) protein localizes to heterochromatin (Brideau *et al.* 2006; Greil *et al.* 2007). Several other Drosophila HIs also involve heterochromatin or heterochromatin proteins, which is intriguing because genome size varies widely among Drosophila, largely as a consequence of variation in repetitive DNAs that make up the heterochromatin (Lohe and Roberts 1988; Bosco *et al.* 2007). Heterochromatin may have a much wider role in incompatibility because repetitive DNA variation is the major cause of the ~1000-fold variation in genome size among multicellular eukaryotes

(Gregory 2005). These DNAs can increase in copy number by general host processes such as unequal crossing over and duplication (Charlesworth *et al.* 1994). Alternatively, they may increase copy number by selfish properties such as transposition for TEs (Hickey 1982) and meiotic drive for satellite DNAs (Walker 1971). In either case, over-proliferation can be deleterious to their host species by causing genome instability, leading to the evolution of host defense mechanisms (Blumenstiel 2011). For example, one major mechanism is the piRNA pathway, where small (23-30 nt) RNAs derived from TE sequences are used to silence TE activity (Khurana and Theurkauf 2010). There are also hints that the piRNA pathway may regulate satellite DNAs (Usakin *et al.* 2007). Interestingly, piRNA regulatory genes often show signatures of adaptive evolution among Drosophila species (Lee and Langley 2012).

Genetic conflicts with selfish DNAs have been proposed as an important driver of HI (Presgraves 2010; Johnson 2010; Maheshwari and Barbash 2011), but little is known about what specific sequences are interacting with HI genes. *D. simulans* and *D. melanogaster* have great potential for addressing this question because they differ substantially from each other in genome size (Bosco *et al.* 2007), satellite DNA content (Lohe and Roberts 1988; Bosco *et al.* 2007), and in both the types and number of TEs that they harbour (Lerat *et al.* 2011). Here we report that *Hmr* and *Lhr* are required to repress transcription from both TEs and satellite DNAs. *Hmr* and *Lhr* also regulate telomeres, a third specialized type of heterochromatic sequence that serves to protect the ends of linear chromosomes (Andreyeva *et al.* 2005) and is composed of rapidly evolving DNA and proteins (Mefford and Trask 2002; Anderson *et al.* 2009; Raffa *et al.*

2011). Telomere variation can affect host fitness and genome stability, and has been proposed as another potential source of meiotic drive (Zwick *et al.* 1999; Anderson *et al.* 2009). We used a *D. simulans* mutation in *Lhr*, comparative cytology, and interspecific complementation with *Hmr* transgenes to identify classes of TEs and satellites that are regulated differentially between the species. We conclude that *Hmr* and *Lhr* provide an adaptive defense against multiple classes of repetitive DNA sequences that change rapidly in evolutionary time, can reduce host fitness, and have high potential to provoke genetic conflict.

3.2 Materials and Methods

3.2.1 Construction of the *Lhr^{KO}* mutant

We used the pW25 donor vector and ends-out homologous recombination method to make an *Lhr* mutant allele (Gong and Golic 2004). The donor vector was designed to recombine a *w$^+$* marker into *Lhr* and simultaneously remove 26 bp of the coding region. iProof (Biorad) was used to PCR amplify two genomic fragments from *y; cn bw sp* (*D. melanogaster*) genomic DNA. The 3768 bp *Lhr* upstream fragment, including 128 bp of the coding region of *Lhr,* was amplified with primers LUF-Fwd: 5'-ttggcgcgccAACAGGGTCGGCTGTCACATTT and LUF-Rev: 5'-ttggcgcgccGCGAGCATCTCCATGAGCAG (Tm=63°C) and cloned into the *Asc*I site of pW25 using the underlined sequences. The 3935bp *Lhr* downstream fragment that includes 806 bp of the *Lhr* coding region was amplified with primers LDF-Fwd: 5'-AAGCGGCCGCAGGTGGAGCCCAAAATGGACG and LDF-Rev: 5'-AAGCGGCCGCCACACATTGCGAATGCA G AAA (Tm=65°C) and cloned into the *Not*I site using the underlined sequences. Restriction digestion was used to pick a clone in which the 2 inserts and the *mini-white* gene were in the same orientation.

The construct was injected into a strain of *w^{1118}* (Genetic Services) and a transgenic line, *P{w$^+$, Lhr-KO}5-1*, with a lethal insertion on the X chromosome was obtained. *P{w$^+$, Lhr-KO}5-1/FM6* females were crossed to *y w; P{ry$^+$, hs-flpase}, P{v$^+$ hs-I-Sce} /TM6, Ubx* males. Two to three day-old larvae were heat shocked and *P{w$^+$, Lhr-KO}5-1/y w P{ry$^+$,*

hs-flpase}, P{v^{+} *hs-I-Sce}*/+ female progeny were crossed to w^{1118} males. Rare w^{+} sons were screened for homologous recombination events by PCR. Primer pairs Lhr-f1 5'-TTCGCACGTTGTGTTCAAGTAA-3', Lhr-r1 5'-GTAGC TTTCTCTTGGCGCTCTT -3' and Lhr-f2 5'- AACGTGCTCGTAGCTTTGGT-3'/, Lhr-r2 5'-TCGCGAAAATACTTCCGTCT-3' (Tm=58°C) produce no amplicons in the presence of the *white* insertion. Attempts to remove the w^{+} marker by *Cre* recombination were unsuccessful and the w^{+}-disrupted *Lhr* locus was designated as Lhr^{KO}.

To test the genetic effects of this mutation, we took advantage of a recent observation that a deficiency chromosome, which deletes *D. melanogaster Lhr* can weakly rescue *D. melanogaster-D. mauritiana* hybrid males to the pharate adult stage (Maheshwari and Barbash 2012). When we crossed Lhr^{KO} homozygous females to *D. mauritiana* males at 18°, we obtained 10.6% rescue of live males (17 males and 161 females). The stronger rescue observed here may be due to the fact that the mothers of the cross were homozygous for the Lhr^{KO} allele, since *Lhr* likely has strong maternal expression based on its high protein abundance in early embryos (Maheshwari and Barbash 2012).

3.2.2 *Hmr* transgenes

A *D. melanogaster Hmr-FLAG* transgene was made by inserting a 3X FLAG tag sequence (Hernan *et al.* 2000) immediately upstream of the stop codon of *Hmr* using fusion PCR into plasmid p72, which is a pCaSpeR2 vector containing a ~9.7kb fragment of the *Hmr* region (Barbash *et al.* 2003). Two *Hmr* fragments (L-arm and R-arm) were amplified from p72 with iProof polymerase by using primer pairs 739/738 and 736/740,

respectively. The primers 738 and 736 contain sequence encoding the FLAG tag and partially overlap to allow fusion in the subsequent stage. The primers 739 and 740 were combined with L-arm and R-arm products to produce a fused partial fragment of *Hmr* containing the 3X FLAG sequence. This fragment was cloned into the pCR-BluntII-Topo vector (Invitrogen) and sequenced completely between the *Avr*II and *Kpn*I restriction sites. The *Avr*II/*Kpn*I fragment was then cloned into the corresponding sites of the p72 plasmid. A 300 bp fragment containing the *attB* site was then PCR amplified from plasmid *pTA-attB* (gift from Dr. Michele Calos) using primers 502 and 503 and cloned into the *Not*I site. This fragment was digested with *Not*I (on the ends of 502 and 503), gel purified, and inserted into the *Not*I site of the plasmid containing *Hmr-FLAG*. We refer to this transgene as *mel-Hmr-FLAG*.

A *D. melanogaster Hmr-HA* transgene was made by inserting a 3XHA epitope tag between codons 466 and 467 of *Hmr*. Primers 215/1246 and 1247/495 were used to amplify 573 and 316 bp fragments, respectively. Primers 1246 and 1247 overlap and encode the HA tag. Fusion PCR containing these 2 products and primers 215/495 was performed. The PCR product was cloned into pCR-Blunt II-TOPO, and the insert was checked by sequencing. The insert was then cloned using *Spe*I and *BsiW*I back into a modified p72 containing an *attB* site inserted into the *Not*I site. The orientation and presence of the HA tag were checked by double digests and PCR. We refer to this transgene as *mel-Hmr-HA*.

A *D. simulans Hmr-FLAG* transgene was made by inserting the 3X FLAG tag sequence upstream of the stop codon in p89, a pBluescript II KS(+) plasmid containing the *D. simulans Hmr* insert that was used for the p92 transformation construct in (Barbash *et al.* 2004). Primers 751/753 and 750/752 were used to amplify 1.3kb and 1.8kb fragments of the insert, respectively, which were then joined by fusion PCR using primers 750/751. The fusion PCR product was cloned into pCR-Blunt II-TOPO and confirmed by sequencing. The insert was designed to have an *Hpa*I site near one end and a *Not*I site near the other. The *Not*I site was destroyed during cloning; however, the pCR-Blunt II-TOPO vector contains a *Not*I site within 40bp of the destroyed sequence. The insert was then cloned back into p89 using *Hpa*I and *Not*I. The orientation of the insert, as well as the addition of the FLAG tag, was checked by double digest with *Cla*I and *Hpa*I. The *D. simulans Hmr-FLAG* insert was then removed as a *Sac*II fragment. Klenow enzyme was used to fill-in the ends to allow cloning into the *Stu*I site of pCaSpeR2 containing an *attB* site inserted at its *Not*I site. We refer to this transgene as *sim-Hmr-FLAG.*

The *D. simulans Hmr-HA* transgene was made from plasmid p89 by inserting the HA tag at the region orthologous to *mel-Hmr-HA* (Barbash *et al.* 2004). Primers 135/1365 and 1247/1364 were used to amplify 861 bp and 827 bp fragments, respectively, from the p89 template, and were fused together using primers 1364/135. The fusion PCR product was then cloned into pCR-Blunt II-TOPO and the entire insert was checked by sequencing. The insert was then cloned back into p89 using *Spe*I and *Blp*I. Blunt end ligation, used for *sim-Hmr-FLAG* above, proved inefficient for transferring the insert into

the transformation vector. Therefore an *Xba*I site was added to the 3' end of *Hmr*-HA by amplifying the entire insert using primers 1402/1403. The PCR product was then gel purified and cloned back into pCR-Blunt II-TOPO. The polylinker contains an *Xba*I site 5' to the insert, allowing us to clone the entire insert into the *Xba*I site of pCaSpeR2 containing an *attB* site inserted at its *Not*I site. We refer to this transgene as *sim-Hmr-HA*.

Oligonucleotides for *Hmr* transgenes (all written 5'-3'). 739: AGCCAAATTGCCGACAGTAGCCAAG; 738: ATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCAGGCGGTGGCG GATTGACCTTG; 736: GACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTAGCT CTCGAAACTTTTGGCACACGTAG; 740: TTGTACTGCCATTAGGTATAGCTAACCATCC; 502: AAACCCGCGGCCGCATGCCCGCCGTGACCGTC; 503: AAACCCGCGGCCGCGATGTAGGTCACGGTCTCG; 152: TCTTCTTAGACTGCGGGTTG; 215: CAGCGCATGCGCGGCACCGTAT; 1246: ATAGTCCGGGACGTCATAGGGATAGCCCGCATAGTCAGGAACATCGTATGGGTAC ATTGCACTGTTGGTCATGCTCGT; 1247: TCCCTATGACGTCCCGGACTATGCAGGATCCTATCCATATGACGTTCCAGATTAC;G CTAGCACTGCCACAAGCATTGG; 495: GACACGCCCGTTCCCATAGT; 751: ACAGCGATTTGCGCAAGCCG; 753: TCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCAGGCGGTGGCGG

ATTTGCCTTCTTGGCGTATTTAGA; 750: GTGAATTGTAATACGACTCACTATAGGGCG; 752: GACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTAGCT CTCGAATCATTGGCACACG; 135: GAGGAGGACCCCACCTATAACTAC; 1365: ATAGTCCGGGACGTCATAGGGATAGCCCGCATAGTCAGGAACATCGTATGGGTAT GCACTGTTAGAAATGCTTGTGCTG; 1364: GCTGGCAATTTGGACTTTGT; 1402: GCGGGCGGTCATTATTAA; 1403: TATCTAGAGCGGCCGCGAGCTCTAATA.

3.2.3 Transgenic Fly Lines

φC31-mediated transgenesis was performed by Genetic Services using the *P{CaryP}attP2* integration site at cytological position 68A4 (Groth *et al.* 2004). Site specificity of integration was checked by PCR assays described in references (Venken *et al.* 2006; Maheshwari and Barbash 2012). *D. melanogaster* transformants were crossed to a *y w* strain. Wild type activity of the *Hmr-HA* transgene was tested for complementation of an *Hmr* rescue mutation in hybrids as done previously for Hmr^{+} transgenes (Barbash *et al.* 2003; Barbash *et al.* 2004). Here we crossed $Df(1)Hmr^{-}$*/FM6; ø{mel-Hmr-HA}* females to *D. simulans* w^{501} males. We recovered 193 w^{501}*/Y; +/+* hybrid males but only 1 w^{501}*/Y; ø{mel-Hmr-HA}/+* hybrid male, demonstrating that the transgene is Hmr^{+}.

3.2.4 Drosophila strains

Lhr^{KO} was outcrossed to *w*^{1118} for six generations. Sibling crosses were then used to generate a homozygous *w*^{1118}; *Lhr*^{KO}/*Lhr*^{KO} (abbreviated as *Lhr*^{KO}), a heterozygous *Lhr*^{KO}/+, and a wildtype *w*^{1118}; *Lhr*$^{+}$/*Lhr*$^{+}$ line (abbreviated as *Lhr*$^{+}$). All experiments with *Lhr* in this paper use these matched mutant and sibling controls unless otherwise specified. The *D. simulans Lhr*^{1} allele is caused by an insertion in the 5' UTR and appears to make no transcript by RT-PCR (Brideau *et al.* 2006). *Lhr*^{1} was outcrossed to the inbred wild-type line *w*^{501} for 3 generations to generate the stock *w*^{501}*; Lhr*^{1} (abbreviated as *Lhr*^{1}) and *w*^{501}*, Lhr*$^{+}$ (abbreviated as *Lhr*$^{+}$). *Lhr-HA* transgenes were described previously (Maheshwari and Barbash 2012). *y w* F10 was created by single-pair matings between siblings for 10 generations.

We refer to the *P{EPgy2}Hmr*^{3} allele that is marked with *y*$^{+}$ and *w*$^{+}$ described in (Aruna *et al.* 2009) as *Hmr*^{3}. *Df(1)Hmr*$^{-}$*, y w v*, abbreviated as *Df(1)Hmr*$^{-}$, is described in (Barbash and Lorigan 2007). In order to match backgrounds for the *Hmr* RNA-Seq experiments, the *Hmr*^{3} stock and the transgenic lines (*mel-Hmr-FLAG* and *sim-Hmr-FLAG*) were outcrossed to *y w* F10 for 6 generations and then made homozygous.

3.2.5 Fertility assays

Individual 1-2 day old virgin *Lhr*^{KO} and *Lhr*^{KO}/+ sibling females, obtained from crosses of *Lhr*^{KO}/+ at 27°C, were crossed to two *w*^{1118} males. Flies were transferred to a fresh vial every 5 days for 15 days. Vials in which either the female or both males were missing or dead were not scored or transferred. Total progeny from each remaining vial were

counted over a period of xx days. To create the heteroallelic siblings *Lhr^{KO}/Df(2R)BSC44* and *Lhr^{KO}/SM6a*, *Lhr^{KO}/Lhr^{KO}* were crossed to the *Lhr^{-}* deletion stock *Df(2R)BSC44/SM6a* (Brideau *et al*. 2006) The fertility assay was carried out as above except vials were flipped every 4-5 days.

3.2.6 Hatch rate assays

Lhr^{KO}/+ or *Lhr^{KO}/Lhr^{KO}* females were crossed to *w^{1118}* males at 27°C. Egg lays were carried out on grape juice/agar plates for 3 hour periods at either 2-3 days, 5-6 days or 10-11 days after eclosion of the female parents. The plates were maintained at 27°C and monitored over the next 24-36 hours for hatched eggs.

3.2.7 Crosses for generating *Hmr* genotypes for RNA-Seq of ovarian mRNA

y w Hmr^{3}; +/+ females were crossed to *y w; ø{mel-Hmr-FLAG}/ø{mel-Hmr-FLAG}* males. F1 males were crossed to *Df(1)Hmr^{-}/FM6*; *+/+* females to generate both *y w Hmr^{3}/Df(1)Hmr^{-}*; *ø{mel-Hmr-FLAG}/+* and *y w Hmr^{3}/Df(1)Hmr^{-}*; *+/+*. Similarly, *y w Hmr^{3}; +/+* females were crossed to *y w*; *ø{sim-Hmr-FLAG}/ø{sim-Hmr-FLAG}* males. F1 males were crossed to *Df(1)Hmr^{-}/FM6*; *+/+* females to generate *y w Hmr^{3}/Df(1)Hmr^{-}*; *ø{sim-Hmr-FLAG}/+*. Lastly, *y w*; *+/+* females were crossed to *y w*; *ø{mel-Hmr-FLAG}/ø{mel-Hmr-FLAG}* males. F1 males were crossed to *Df(1)Hmr^{-}/FM6*; *+/+* females to generate the heterozygous wildtype control, *y w/Df(1)Hmr^{-}*; *+/+*. These crosses were done at 27ºC and in triplicate to generate 3 biological replicates.

3.2.8 Crosses for generating pure-species and hybrid samples for RNA-Seq of larvae

The *Df(1)Hmr⁻, y w v/FM7i, P{w+ mC =ActGFP }JMR* stock (abbreviated as *Df(1)Hmr⁻ /FM7i, GFP*) was described previously (Barbash and Lorigan 2007). A stock with the matching *Hmr⁺* genotype*, y w v/FM7i, P{w+ mC =ActGFP}JMR* (abbreviated as *Hmr⁺/FM7i, GFP*) was created by crossing *y w v/Y* males with *Df(1)Hmr⁻/FM7i, GFP* females. *FM7i, GFP /Y* males from this *Hmr⁺* stock were then crossed to *Df(1)Hmr⁻ /FM7i, GFP* females for 10 generations in order to make the autosomal backgrounds comparable between the two stocks.

To generate hybrids, *Df(1)Hmr⁻/FM7i, GFP* or *Hmr⁺/FM7i, GFP* were crossed to *v/Y D. simulans* males. For each cross, 6 replicates were made each containing 25 0-12 hour-old virgin females and 50 4-6 day-old virgin males. Hybrid larval sons not carrying the balancer were selected by their *y⁻* mouth hook and *GFP⁻* body phenotypes. Additionally, some crosses were allowed to develop to ensure that only *Df(1)Hmr⁻* crosses produced hybrid sons. To generate *D. melanogaster* samples, 3 replicates of 10 *Df(1)Hmr⁻/FM7i, GFP* or *Hmr⁺/FM7i, GFP* virgin females were crossed to 15 *FM7i, GFP/Y* males. Larval sons not carrying the balancer were selected by *y⁻* and *GFP⁻* phenotypes. To generate *D. simulans* samples, 3 replicates of 10 *y w D. simulans* virgin females were crossed to 15 *v/Y D. simulans* males. Larval sons were selected by *y⁻*.

3.2.9 Preparation of protein lysates for semi-quantitative Western blots.

50 mg of 1-17 hr embryo collections were dounced 30 times with a tight pestle in 500ul buffer A1 (15mM HEPES, pH=7.5; 15mM NaCl; 60mM 1M KCl; 4mM $MgCl_2$; 0.5% TritonX-100; 0.5mM DDT) and then centrifuged for 5 minutes at 4°C. The pellet was washed with 500µl buffer A1 and centrifuged. This process was repeated another two times. The pellet was lysed by douncing in 200 µl SDS lysis buffer (500µl 10% SDS, 200µl 1M Tris, pH=8.0, 40µl 0.5M EDTA, 100µl 100X protease inhibitor, 10µl 0.5M EGTA, 50µl 100mM PMSF, 9.1ml water). The lysate was allowed to rotate at 4°C for 20 minutes and then centrifuged. The supernatant was removed, quantitated using the Bradford assay and was run on an SDS-PAGE gel.

3.2.10 Anti-Lhr antibodies and Western blots

An Lhr cDNA was cloned into pDEST17 (Invitrogen). The expressed protein from *E. coli* was purified using Ni-Ag beads under denaturing conditions (8M urea), dialyzed down to 2M urea and injected into rabbits (Cocalico). The antisera was then purified by coupling purified His-Lhr to CnBr-activated Sepharose beads in the presence of 1% Triton-X and removing urea by dialysis. Antisera was eluted in 0.2 M glycine, pH2.8 and then neutralized with 1M Tris, pH8.5. The antibody failed to detect Lhr in immunofluorescent experiments but was used for Western blots in Figure 3.4 at 1:4000 in 5% milk-TBST and HRP conjugated anti-rabbit secondary antibody at 1:2000 dilution. HA-tagged Lhr was detected with 1:1000 dilution of rat anti-HA (Roche, 3F10) and HP1a was detected with a 1:700 dilution of mouse monoclonal supernatant (C1A9, DSHB).

3.2.11 Co-immunoprecipitation

0~16 hour-old embryos were collected, dechorionated and snap frozen in liquid nitrogen. Embryos were then resuspended to 10x embryo volume of Buffer A (10mM Tris-Cl pH 8.0, 300mM sucrose, 3mM $CaCl_2$, 2mM Mg $acetate_2$, 0.1% Triton X-100, 0.5mM DTT, 0.5mM PMSF) and homogenized with a dounce homogenizer. The homogenized lysate was centrifuged at 700g for 10 minutes at 4° to pellet the nuclei. The supernatant was removed, the pelleted washed once in Buffer A, the nuclei centrifuged again and then resuspended in 1x embryo volume of Buffer MN (15mM Tris-Cl pH7.4, 250mM sucrose, 60mM KCl, 1.0mM $CaCl_2$, 0.5mM DTT, 1x protease inhibitor cocktail). The nuclear lysate was sonicated briefly, micrococcal nuclease added to a concentration of 500 units/ml, and the chromatin digested for 1 hour at 4° with gentle agitation. EDTA and Triton X-100 were then added to a concentration of 5mM and 0.1% respectively, to inactivate nuclease activity and solubilize the proteins, followed by incubation at 4° for 1 hour. After a second brief sonication, the digest was centrifuged at 12,000g for 10 min at 4° and the supernatant was collected. 50µl of the chromatin digest was diluted in IP Wash Buffer (50mM Tris-Cl pH7.4, 100mM NaCl, 0.1% Triton X-100) with 1x protease inhibitor cocktail to a final volume of 125µl per co-immunoprecipitation mixture. 15 µl of protein G-conjugated magnetic beads and 2-5µl of antibody were added followed by incubation for 4 hours at 4° with gentle agitation. The beads were washed 3 times in IP Wash Buffer. The immunoprecipitated proteins were then eluted by boiling the beads in 1x Laemmli sample buffer for 5 minutes and analyzed by immunoblotting.

3.2.12 RT-PCR and qRT-PCR assays

RNA extraction, cDNA synthesis and qRT-PCR assays were performed as in reference (Maheshwari and Barbash 2012), using 2-5µg of RNA. qRT-PCR experiments included three technical replicates of three separate biological replicates. Primers included: Lhr-f1 5'caccATGAGTACCGACAGCGCCGAGGAA, Lhr-r1 5' ACACTTGGTTTTCGGCACATC CGC, Lhr-f2 5' GTAGCTTTCTCTTGGCGCTCTT, Lhr-r2 5' GTAAGTGAACTGAAGCTGC GTTGG, EDTP-F 5'GCTGGCAGGTGG TTACCGACA, EDTP-R 5'CGTGGCCAGGTTCA TGGATGA, Bap55-F 5' CCGAGAGTC TCTTTGACAATGCA, and Bap55-R 5'GCCTCTT CGTACTCCTGCGA. Hmr-f1 5' TAAGTTCGCCTTCCGCACATACC and Hmr-r1 5' GACCAGAAACCTGAGTTGCTCCA. *HeT-A* and *RpL32* (also known as *Rp49*) transcript levels were measured with primers from reference (Pane *et al.* 2007).

3.2.13 qPCR of *HeT-A* DNA copy number

The Invitrogen DNEasy kit was used to make genomic DNA from Lhr^{KO} and Lhr^{+} female carcasses that were free of ovarian tissue. Primers Het-s2 and Het-as2 amplify from the coding sequence of *HeT-A* (Klenov *et al.* 2007). *HeT-A* copy number was normalized to *RpL32* (also known as *Rp49*) copy number using primers from reference (Pane *et al.* 2007).

3.2.14 RNA-Seq samples

For samples from ovaries, flies were kept at 27°C for several generations prior to and during the experiment. Freshly eclosed females were collected and aged 2-3 days and then transferred to fresh food with yeast paste for another 2-3 days. RNA was extracted, from ovaries dissected in chilled 1X PBS, using Trizol. Ovarian mRNA-Seq libraries were constructed at the Epigenomics Core Facility at Weill Cornell Medical College using the poly(A) enrichment method. Libraries were sequenced using the Illumina HiSeq2000 platform to produce 50bp single reads which were then trimmed for quality and filtered to remove rRNA reads. One biological sample each from Lhr^{KO} and Lhr^{+} was duplexed and run in a single lane. 51,193,832 filtered reads were obtained for Lhr^{+} and 41,688,028 reads for Lhr^{KO}. Three biological replicates each of *D. simulans* w^{501} and Lhr^{1} ovarian mRNA libraries were run on a single lane and the number of filtered reads ranged from 36,472,726 to 43,449,879. For experiments with *Hmr*, two biological replicates were included for each genotype and all 8 samples were multiplexed in a single lane. The number of filtered reads for each sample ranged from 23,863,381 to 27,490,644. For larval samples, around 30 larvae were collected for each genotype and flash frozen in liquid N_2. RNA was extracted from 2 biological replicates of each genotype using Trizol. Larval RNAseq libraries were generated and bar-coded using the TruSeq kit, and run in one lane of an Illumina HiSeq 2000 100bp yielding 13,707,247 to 20,373,267 filtered reads per sample, except for one library which produced only 7,840,004 reads.

3.2.15 RNA-Seq analysis

Reads mapping to either rRNA or repetitive DNA were filtered out using Bowtie (Langmead *et al.* 2009)and the filtered reads were mapped to the unmasked *D. melanogaster* genome using Tophat (Trapnell *et al.* 2009). The BAM file outputs were used by Cuffdiff with the -b option (Trapnell *et al.* 2010). All *.fasta and *.gtf files were based on the release 5.68 of the *D. melanogaster* genome from ENSEMBL. To find differentially expressed genes in *D. simulans*, we aligned reads to the *D. melanogaster* genome with Tophat, allowing two mis-matches. While this approach could potentially reduce mapping ability for diverged genes, it allowed us to take advantage of the better assembly and annotation of the *D. melanogaster* genome.

To maximize the TEs considered in our analyses, we mapped reads to two different databases using Bowtie. First, reads were uniquely mapped to a database consisting of all the annotated TE insertions in the *D. melanogaster* and *D. simulans* genomes (Kelleher *et al.* 2012); we refer to this as the individual-insertion database. While this database likely represents most TE families present in our stocks, some TEs may either be absent from the assembled genome or be represented by copies that are sufficiently diverged such that they impact our ability to correctly assess transcript levels. These elements include the telomeric element *TAHRE*, which has only a few insertions in the genome and is known to be absent from the reference genome since only two telomeres are included in the assembly (George *et al.* 2006). Therefore we also mapped reads, allowing for either 0 mismatches when aligning reads from *D. melanogaster* or 3 mismatches when aligning reads from *D. simulans* or hybrids, to a

database consisting of the consensus sequences of the annotated TEs and repeats found in Repbase as well as *de novo* predicted TEs generated by piler-DF using the 12 *Drosophila* genomes (Kelleher *et al.* 2012); we refer to this as the consensus-sequence database. Only reads that mapped uniquely within the same family were included in the subsequent analyses of differential expression. Mismatches allowed for each alignment are mentioned in figure legends. Statistical significance of differential expression among TEs was calculated with F.E.T. in the DEG-seq package (Wang *et al.* 2010).

To analyze reads mapping to satellite DNAs, we built a database using a curated file from the Berkeley Drosophila Genome Project (http://www.fruitfly.org/sequence/sequence_db/na_re.dros) which itself was constructed from GenBank sequences. This file includes some mis-annotated TEs and non-satellite sequences. We counted reads that mapped to these repeats without any mismatches and calculated statistical significance of differential expression among satellites with F.E.T. in the DEG-seq package.

3.2.16 Small RNA sequencing and analysis

Libraries were prepared as described but no oxidation was carried out (Li *et al.* 2009). Briefly, total RNA was extracted from 5-6 day old Lhr^{KO} and Lhr^{+} ovaries using the mirVANA kit (Invitrogen). Total RNA was size fractionated on a 15% Urea-PAGE gel to enrich for 18-29 nt small RNA, excised and eluted and then subjected to 2S rRNA depletion. This small RNA was ligated to a 3' RNA adapter, gel purified, and then ligated to a 5' DNA adapter. The adapter-ligated small RNAs were reverse transcribed and

PCR amplified. The amplified PCR products were gel purified, quantified and sequenced in two lanes of a HISeq 2000 machine.

Only reads with a 3' adapter were kept, which was then removed using a custom script (Kelleher *et al.* 2012). These reads were binned by size as either miRNA/siRNA (17-22 nt) or piRNA (23-30 nt). rRNA, tRNA and snoRNA sequences were filtered from these reads and the remaining reads were further filtered to keep only those reads that mapped to either the unmasked genome,or the satellite DNA database described above, or Repbase consensus sequences (Jurka *et al.* 2005). These filtered reads included 89,953,149 piRNA reads and 40,859,119 siRNA reads in Lhr^{KO}, and 120,143,855 piRNA reads and 36,388,192 siRNA reads in Lhr^{+}.

piRNA reads were mapped uniquely to all *D. melanogaster* sequences from Repbase using Bowtie, allowing for one mismatch. Ping-Pong scores were calculated using reads mapped with up to 1 mismatch, as described in reference (Kelleher *et al.* 2012). For mapping to piRNA clusters, we built an index using sequences extracted from the Release 5 DM3 genome on the UCSC genome database and GenBank with coordinates of individual piRNA clusters obtained from reference (Brennecke *et al.* 2007). piRNA reads were uniquely mapped to piRNA clusters with zero mismatches and significance for differential expression was calculated using F.E.T implemented in DEG-seq. siRNA reads were mapped uniquely to all *D. melanogaster* sequences from Repbase with Bowtie, without allowing for any mismatches.

3.2.17 Immuno-fluorescence and Immuno-FISH

Immunofluorescence and FISH were performed on embryos and ovaries as described in references (Aruna *et al.* 2009; Ferree and Barbash 2009). Polytene chromosomes were dissected in 0.7% NaCl, squashed, and fixed in 1.8% PFA, 45% acetic acid for 17 minutes. They were then washed in 1% Triton X in PBS for 10 minutes, then washed in 5% milk in PBS for 1 hour, incubated with primary antibody overnight at 4°C, washed in 5% milk in PBS for 10 minutes, incubated with secondary antibody for 1 hour at room temperature, and then washed for 10 minutes in buffer A (0.15M NaCl, 0.2% NP40 substitute, 0.2%Tween 20) followed by 10 minutes in buffer B (0.20M NaCl, 0.2% NP40 substitute, 0.2%Tween 20).

Rat anti-HA antibody (Roche, 3F10) was used at 1:100, rat anti-Vasa (DSHB) was used at 1:25, Fibrillarin (Abcam, Ab5281) was used at 1:100, anti-HP1a antibody (C1A9, DSHB) was used at 1:100. Alexa fluorophore-conjugated secondary antibodies were used to detect the primary antibody. Fluorescently labeled probes against GA-rich satellites, AACAC, 2L3L, 359 bp and Dodeca were obtained from Sigma with sequences described in references (Ferree and Barbash 2009; Dernburg 2011; Maheshwari and Barbash 2012). Imaging was carried out using a Zeiss 710 confocal microscope at Cornell University's Microscopy and Imaging Facility.

3.2.18 Yeast two-hybrid assays

A full-length coding-sequence plasmid of *D. melanogaster Hmr* was made by correcting 3 frame-shift errors in the RE54143 cDNA (Barbash *et al.* 2003). Two errors in exon 5

were replaced by ligating in a ~1.6 kb *Xba*I-*Hind*III fragment from the LD22117 cDNA, followed by replacement of a 2172 bp *Nde*I-*Zra*I fragment from the p83 genomic clone (Barbash *et al.* 2003). The coding sequence was then PCRd out and cloned into pENTR/D-TOPO. The *D. simulans Hmr* CDS was PCRd out of cDNA and cloned into pENTR/D-TOPO. The *Lhr* plasmids and yeast two-hybrid destination vectors and assays are described in reference (Brideau *et al.* 2006).

3.2.19 Data Access

Illumina sequence data from this study are available from the NCBI website under BioProject number PRJNA236022.

3.3 Results

3.3.1 Lhr and Hmr form a complex with HP1a

Lhr protein localizes to a subdomain of pericentric heterochromatin in early embryos (Maheshwari and Barbash 2012). To explore possible similarities with Hmr, we examined the localization of Hmr with a 3X-HA epitope-tagged *Hmr* transgene (see Materials and Methods). mel-Hmr-HA colocalizes with HP1a and H3K9me2 at heterochromatin in nuclear cycle 14 embryos (Figure 3.1A). We then used Immuno-FISH to determine its localization relative to specific heterochromatic satellite DNA sequences. mel-Hmr-HA does not overlap with the X-linked 359-bp satellite but colocalizes with dodeca, a GC-rich pericentromeric satellite on chromosome 3. This pattern mimics that seen previously with Lhr (Maheshwari and Barbash 2012). Additionally, mel-Hmr-HA colocalizes with GA-rich repeats and the 2L3L satellite in embryos (Figure 3.1B). Colocalization between mel-Hmr-HA with both dodeca and GA-rich repeats is also observed in ovarian nurse cells from *Hmr*^{3}*; mel-Hmr-HA* females, indicating localization is not a consequence of overexpression (Figures 3.2B, C). Unlike Lhr (Maheshwari and Barbash 2012), mel-Hmr-HA localizes to the nucleolus in early embryos (Figure 3.1C), suggesting that Hmr may have some distinct functions from Lhr.

The largely similar localization patterns of Hmr and Lhr raise the possibility that they physically interact. We performed co-immunoprecipitation (co-IP) studies from embryo extracts and found that mel-Lhr-HA and mel-Hmr-FLAG co-IP (Figure 3.1D). *mel-Lhr-HA* was previously shown to express at wild type levels (Maheshwari and Barbash

2012), and *mel-Hmr-FLAG* is expressed significantly lower than wild type levels (Figure 3.3), demonstrating that these results are not due to overexpression. Lhr was previously shown to bind to, co-localize with, and be dependent on HP1a for correct heterochromatic localization (Giot *et al.* 2003; Brideau *et al.* 2006; Greil *et al.* 2007; Brideau and Barbash 2011). We therefore tested if HP1a also associates with Hmr. IPs with HP1a pulled down mel-Lhr-HA and mel-Hmr-FLAG, but the reciprocal IPs failed to pull down detectable HP1a (Figure 3.1E).

Yeast two-hybrid assays show that Hmr and Lhr from *D. melanogaster* interact, suggesting that the co-IP reflects a direct interaction between the proteins (Figure 3.1F). This interaction is likely mediated via the BESS domains within Lhr and Hmr (Brideau et al 2006), a 40 amino-acid motif found in 19 proteins in *D. melanogaster* that has been implicated in protein-protein interactions and homo-oligomerization (Bhaskar and Courey 2002). We also found that the *D. simulans* orthologs interact, as do the heterospecific combinations; the strength of interactions varied widely but exploring the potential significance of this result will require a more quantitative assay.

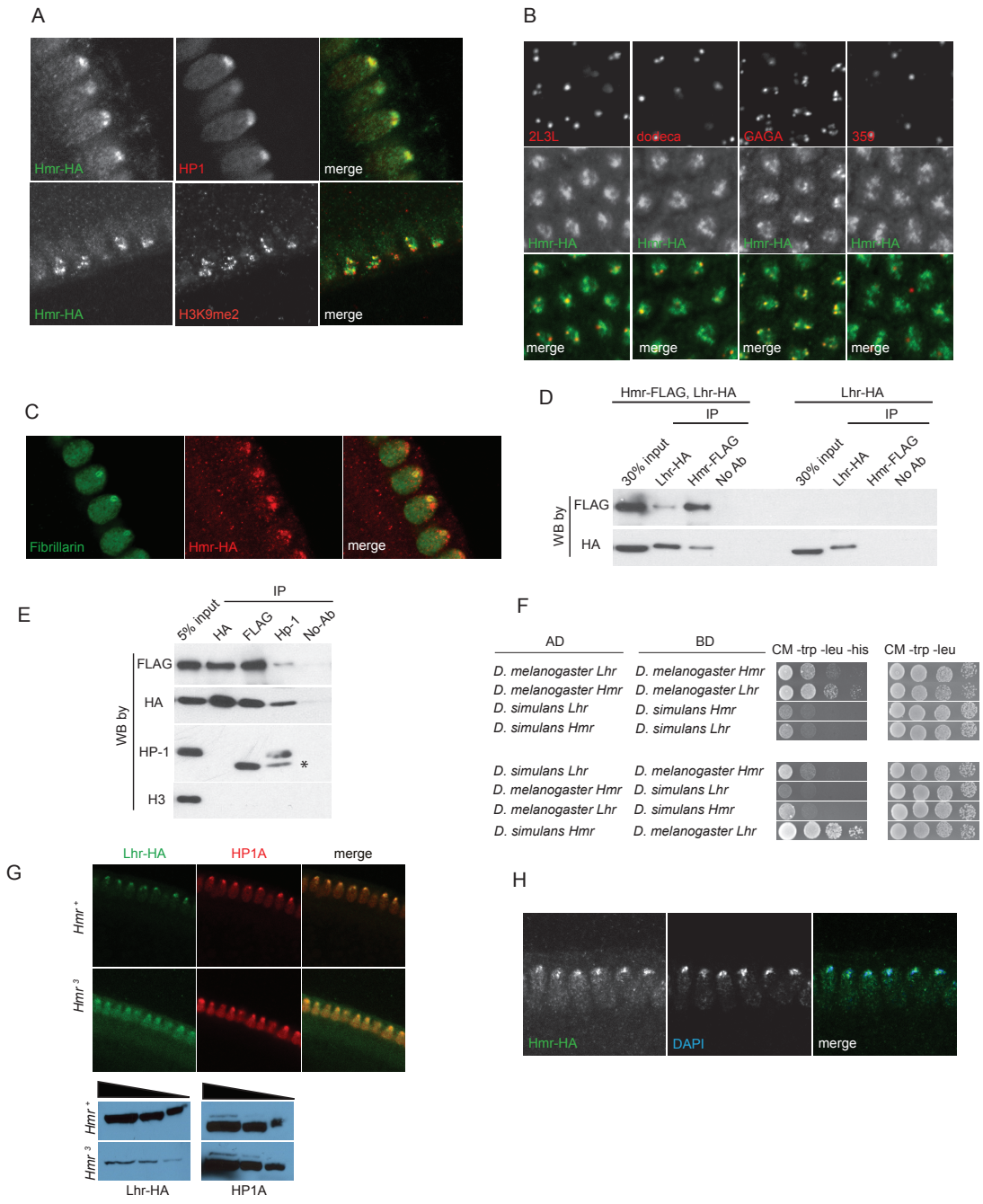


Figure 3.1. Hmr forms a complex with Lhr and HP1a and is required to stabilize Lhr. (A) mel-Hmr-HA (green) colocalizes with both HP1a (top) and H3k9me2 (middle; both red) in nuclear cycle 14 embryos. The HP1a costain is in a *mel-Hmr-HA* backgrouond, while the H3K9me2 is in a *Hmr*^{3}*; mel-Hmr-HA* backround. A negative control shows no HA signal in *w*^{1118} embryos lacking the *mel-Hmr-HA* transgene (bottom). Scale bars represent 10 μm. (B) mel-Hmr-HA (green) colocalizes with 2L3L, dodeca and GA-rich satellites but not with the 359 bp repeat satellite repeat in *mel-Hmr-HA* (all FISH probes red). Scale bars represent 5 μm. (C) mel HMR-HA (red) colocalizes with the nucleolar marker Fibrillarin (green) in early embryos. Scale

bars represent 10 μm. (D) mel-Lhr-HA and mel-Hmr-FLAG co-immunoprecipitate from *D. melanogaster* embryo extracts derived from flies expressing both transgenes (left 4 lanes) but not from flies expressing only Lhr-HA (right 4 lanes). Extracts were IP'd with the indicated antibodies, and then probed by Western Blots (WB) with the same or different antibodies. (E) Lhr-HA, Hmr-FLAG and HP1a co-immunoprecipitation from embryo extracts. Specificity is indicated by lack of immunoprecipitation of histone H3. Asterisk indicates the antibody light chain. (F) Lhr and Hmr interact in a yeast-two hybrid assay. Interactions were detected by growth on complete media (CM) lacking histidine (his); growth controls were performed on CM lacking tryptophan (trp) and leucine (leu). The top 4 panels test for interactions between orthologs from the same species; the bottom 4 between heterospecific orthologs. AD, activation domain; BD, DNA binding domain. (G) Lhr-HA is detectable in Hmr^{3} and localizes to heterochromatin, as indicated by co-localization with HP1a. Note that a higher gain was used in the Hmr^{3} panels compared to the Hmr^{+} panels in order to detect Lhr-HA, and is reflected in the higher background. Western blots confirm that Lhr-HA levels are reduced in Hmr^{3}. HP1a is used as a loading control. (H) Hmr-HA maintains its localization to DAPI-dense heterochromatin in Lhr^{KO}; *Hmr-HA* embryos. Scale bars represent 10 μm.

We next examined protein localization in mutant backgrounds to test the potential mutual dependence of Lhr and Hmr for their localization to heterochromatin. We made a *D. melanogaster Lhr* mutation by recombining a *mini-white* gene into the *Lhr* locus to create the Lhr^{KO} allele (Figure 3.4A). In Lhr^{KO}, transcription from *Lhr* but not flanking genes is greatly reduced, and no Lhr protein is detectable (Figure 3.4B, C). These results demonstrate that Lhr^{KO} is a strong loss of function allele, which we confirmed in hybrid rescue crosses (see Materials and Methods).

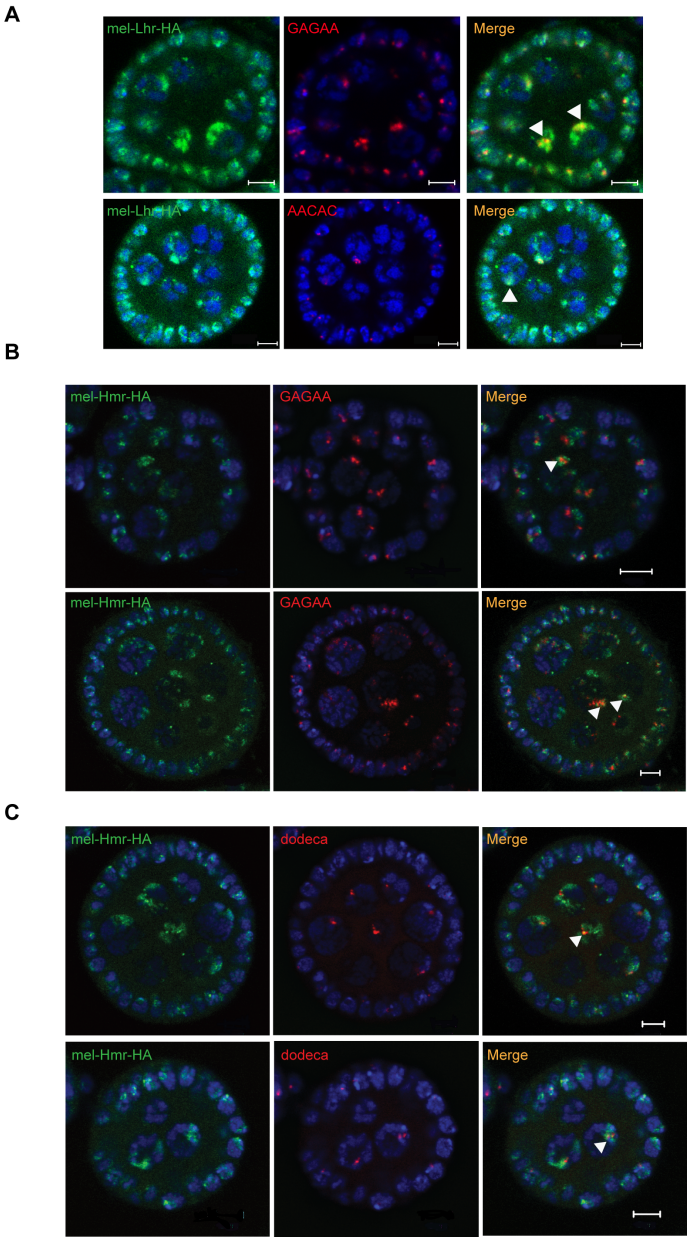


Figure 3.2. Lhr and Hmr colocalize with specific satellite sequences in ovaries. Nurse cell nuclei (blue) are stained with DAPI in all panels. Scale bars represent 5 μm. (A) mel-Lhr-HA (green) colocalizes with GAGAA(red, top panel) and AACAC (red, bottom panel) in the nurse cells of Lhr^{KO}/+; *LhrHA/*+ovaries. Arrows point to overlaps between bright FISH and HA-staining foci. (B) mel-Hmr-HA (green) colocalizes with GAGAA (red) and (C) dodeca (red) in nurse cells of Hmr^{3}; *mel-Hmr-HA/mel-Hmr-HA* ovaries in a subset of nuclei. Arrows point to overlaps between FISH signals and the brightly staining foci of mel-Hmr-HA. Two different egg chambers are shown for both dodeca and GAGAA.

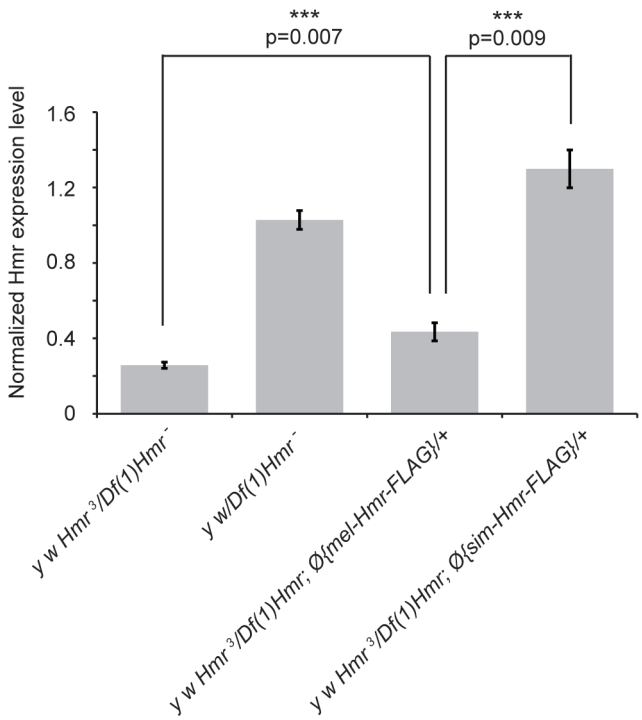


Figure 3.3. qRT-PCR analysis of *Hmr-FLAG* transgenes. *Hmr* transcript levels in transgenic lines were compared to the host strain (Hmr^{-}) and also to $Hmr^{+/-}$. The transgenes are heterozygous, therefore both the transgenic lines and $Hmr^{+/-}$ carry one copy of Hmr^{+}. RNA was isolated from ovaries and *Hmr*expression levels were normalized relative to *RpL32*. Error bars represent standard error within 3 biological replicates. The difference in the expression level of *mel-Hmr-FLAG* and *sim-Hmr-FLAG* is significant ($p=0.009$, two-tailed *t*-test with equal variance). Additionally, the expression of *mel-Hmr-FLAG* is significantly different than an endogenous copy of *Hmr* ($p=0.007$, two-tailed *t*-test with equal variance).

Lhr-HA levels are greatly reduced in Hmr^{3} mutant embryos but when examined at high gain a small amount of Lhr-HA is detectable in heterochromatin (Figure 3.1G). This result suggests that Hmr is not absolutely required to localize Lhr to heterochromatin, though it remains possible that some Hmr protein is made in the Hmr^{3} mutant. In a reciprocal experiment, Hmr-HA localization appears normal in Lhr^{KO} (Figure 3.1H). In combination with previous results, our data suggest that Lhr localization to heterochromatin depends on HP1a, and that Hmr stabilizes Lhr.

3.3.2 *Lhr* is required for female fertility

Lhr^{KO} flies are almost fully viable (22.25% compared to the expected 25% in crosses between heterozygotes at 27°; $p<0.05$ by Chi-squared; N=2813 total flies scored). However, comparison of Lhr^{KO} with a background-matched Lhr^{+} control (see Materials and Methods) showed that Lhr^{KO} females have substantially lower fertility, particularly at higher temperatures. One to five day old Lhr^{KO} females display only a fraction of the fertility of Lhr^{KO}/+ and later become sterile (Figure 3.5A). We confirmed this in a different Lhr^{-} background where a similar reduction in fertility occurs at later stages (Figure 3.5B). In a separate experiment we found that the hatch rate of the eggs laid by Lhr^{KO}/Lhr^{KO} mothers is low and declines with increasing maternal age (Table S1). This Lhr^{KO} female fertility phenotype is strikingly similar to that of *Hmr* mutants (Aruna *et al.* 2009), suggesting that *Hmr* and *Lhr* may function in a common regulatory pathway.

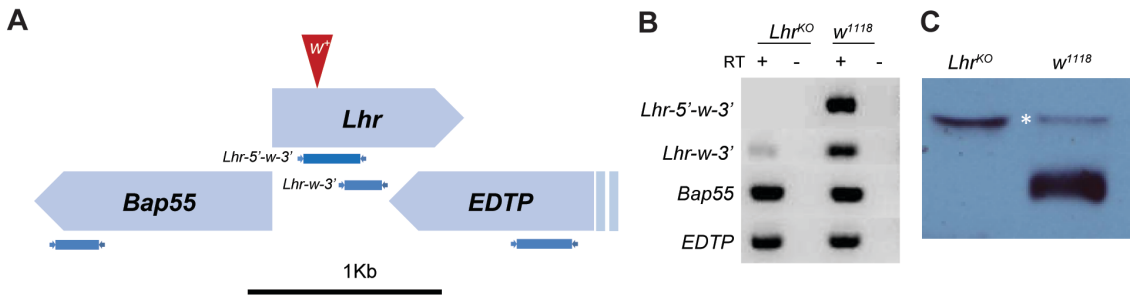


Figure 3.4. The *D. melanogaster* Lhr^{KO} allele generated by homologous recombination. (A) *Lhr* and flanking genes are shown, the red triangle labeled w^+ indicates the site of the insertion in the Lhr^{KO} allele, which is predicted to be ~4.7 kb based on the structure of the targeting vector. Products used in RT-PCR reactions in (B) are shown below the genes. *EDTP* gene is partial; *w+* insertion not to scale. (B) RT-PCR from adult females shows no *Lhr* transcript spanning the w^+ insertion (*Lhr-5′ -w-3′*) in Lhr^{KO}. A highly reduced amount of *Lhr* transcript is detected 3′ to the *w+* insertion (*Lhr-w-3′*). The flanking genes *Bap55* and *EDTP* are not affected. w^{1118} was used as a Lhr^+ control. +, − indicates presence or absence of reverse transcriptase (RT). (C) Western analysis shows that Lhr^{KO} produces no protein. A non-specific band indicated by the asterisk is used a loading control.

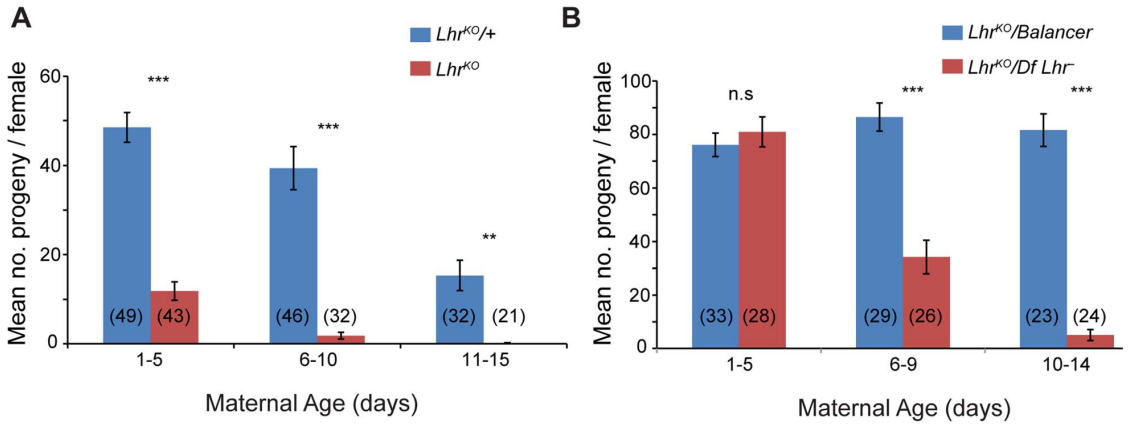


Figure 3.5. *Lhr* mutant females have reduced fertility. Total adult progeny from single Lhr^{KO}/Lhr^{KO} (A) or Lhr^{KO}/*Df(2R)BSC44,* Lhr^- (B) females were compared at 27° to heterozygous female siblings (Lhr^{KO}/+ for (A); Lhr^{KO}/*SM6a* for (B)). The difference between the fertility of genotypes was tested by a two-tailed *t*-test. n.s= not significant, **p<0.01,***p<0.001. The number of individuals tested for each experiment is shown at the bottom of the bars. The error bars represent S.E.M. Crosses were performed at 27°.

3.3.3 *Lhr* and *Hmr* are required to repress transposable elements

We performed an RNA-Seq comparison of ovaries from Lhr^{KO} and Lhr^{+} to investigate the cause of this fertility reduction and discovered a widespread increase in transposable element (TE) transcripts. Using two different TE mapping methods (see Materials and Methods) we found that transcripts from 99 families were at least 2-fold upregulated, with 38 elements being at least 10 fold upregulated (Figure 3.6A; Table S2). Mis-regulated TEs include elements with germline expression, such as the telomeric non-LTR retrotransposons *HeT-A* (350.7 fold) and *TART* (51.76 fold), the LTR retrotransposon *copia* (19.8 fold), and the DNA transposon *bari-1* (44.7 fold). TEs expressed only in the somatic follicle cells, such as *Gypsy* (3.8 fold) and *Zam* (7 fold) were also upregulated. In addition, qRT-PCR in two different genetic backgrounds confirmed the massive increase in *HeT-A* transcript levels (185-846-fold; Figure 3.7). These results demonstrate that the telomeric TEs are especially sensitive to *Lhr* regulation.

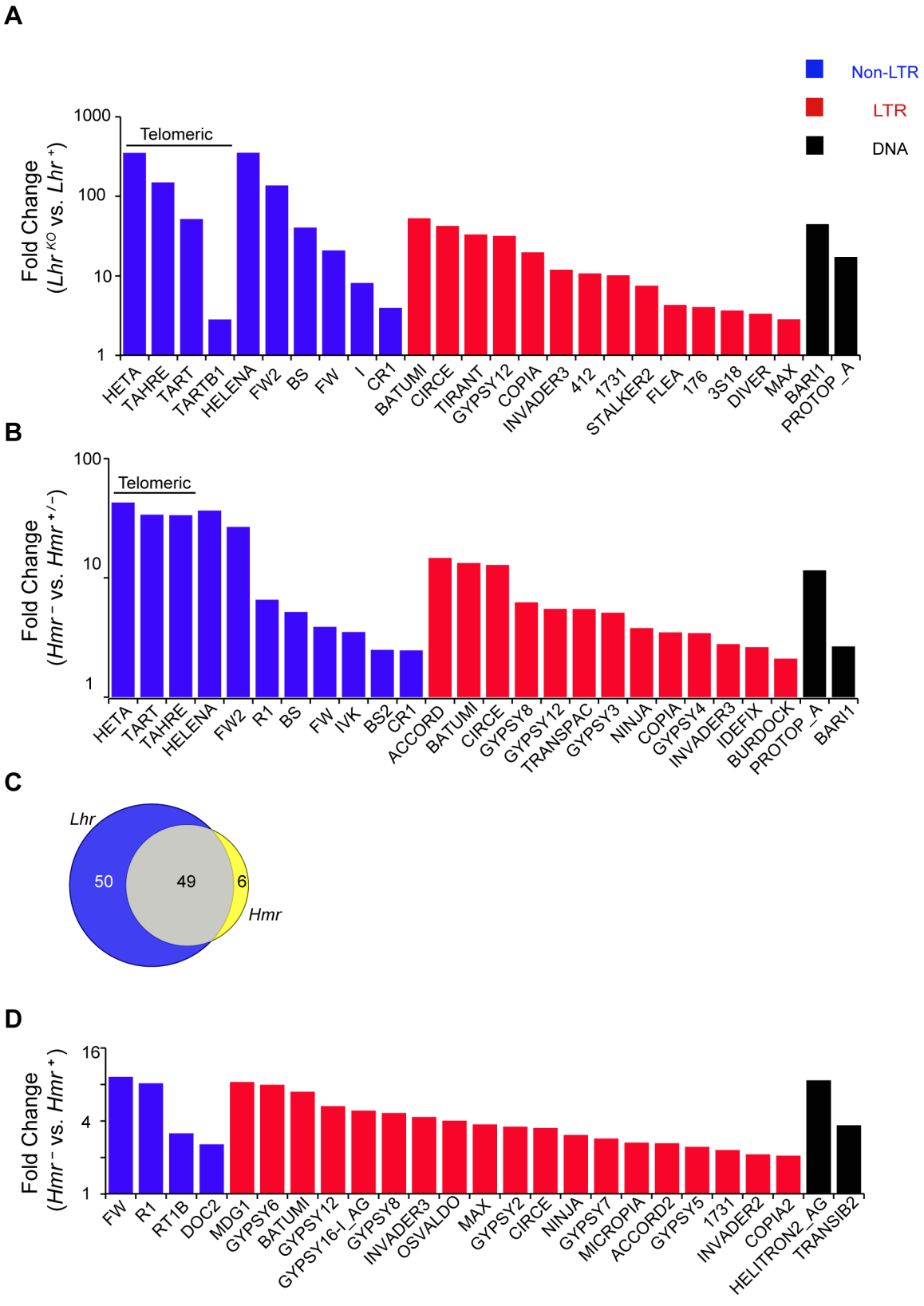

A
Non-LTR
LTR
DNA
Telomeric
Fold Change
(Lhr KO vs. Lhr +)
1000
100
10
1
HETA
TAHRE
TART
TARTB1
HELENA
FW2
BS
FW
I
CR1
BATUMI
CIRCE
TIRANT
GYPSY12
COPIA
INVADER3
412
1731
STALKER2
FLEA
176
3S18
DIVER
MAX
BARI1
PROTOP_A
B
Telomeric
Fold Change
(Hmr− vs. Hmr +/−)
100
10
1
HETA
TART
TAHRE
HELENA
FW2
R1
BS
FW
IVK
BS2
CR1
ACCORD
BATUMI
CIRCE
GYPSY8
GYPSY12
TRANSPAC
GYPSY3
NINJA
COPIA
GYPSY4
INVADER3
IDEFIX
BURDOCK
PROTOP_A
BARI1
C
Lhr
50
49
6
Hmr
D
Fold Change
(Hmr− vs. Hmr +)
16
4
1
FW
R1
RT1B
DOC2
MDG1
GYPSY6
BATUMI
GYPSY12
GYPSY16-I_AG
GYPSY8
INVADER3
OSVALDO
MAX
GYPSY2
CIRCE
NINJA
GYPSY7
MICROPIA
ACCORD2
GYPSY5
1731
INVADER2
COPIA2
HELITRON2_AG
TRANSIB2

Figure 3.6. TE misregulation in *Lhr* and *Hmr* mutants. (A and B) Analysis of Lhr^{KO} (A) and Hmr^{-} (B) ovaries. Reads with zero mismatches were mapped separately to the individual-insertion or consensus-sequence TE databases. A subset of TEs that are significantly different between genotypes are shown and include those with the 25 lowest p-values obtained from individual-insertion mapping analysis, but excluding all *centroid* repeats (Smith *et al.* 2007). Additionally shown are *TAHRE*, which is only found in the consensus-sequence database, as well as *TARTB1* for Lhr^{KO}, which is significant but not among the 25 top hits in the Lhr^{KO} individual-insertion analysis. (C) 49 TEs are upregulated at least 2 fold in both Lhr^{KO} and Hmr^{-}. TE families include those resulting from mapping reads to the insertion database, as well as families found only when reads were mapped to the consensus database. (D) Reads from *Hmr* mutant or wildtype male larvae with up to three mismatches were mapped to the individual-insertion or consensus-sequence TE databases. All TE families, excluding *centroids*, that were significantly upregulated in the insertion sequence based analysis are shown here. Note the different Y-axis scales in A, B and D. Classification of DNA, LTR and non-LTR elements is from reference (Kaminker *et al.* 2002).

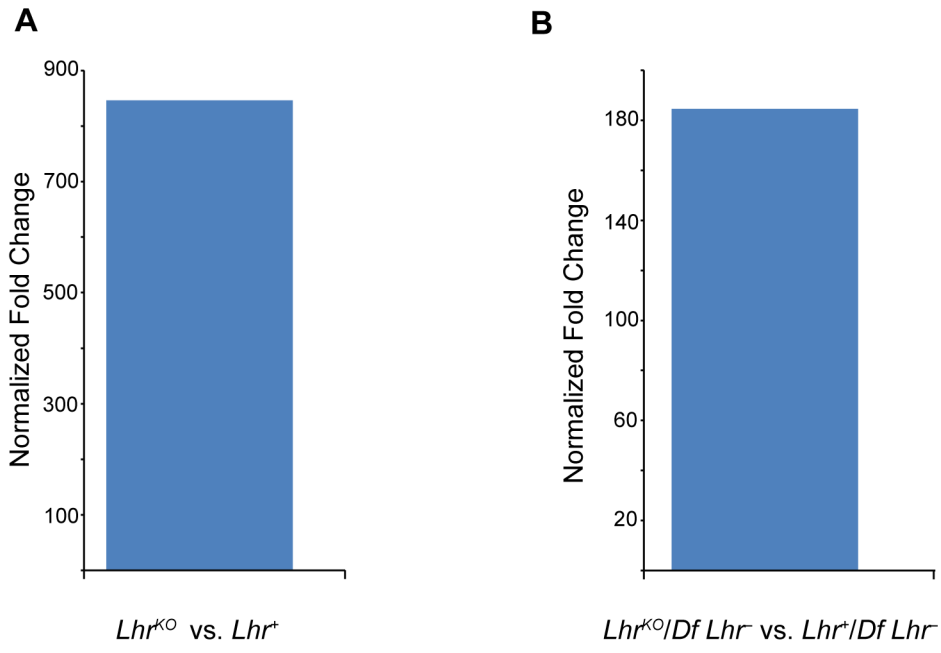


Figure 3.7. qRT-PCR analysis shows elevated *HeT-A* levels in *Lhr* mutants. qPCR was used to estimate the transcript levels of *HeT-A* relative to the gene *RpL32* in poly-A primed cDNA samples obtained from ovarian RNA from two different Lhr^{-} backgrounds and matching controls. (A) Ratio of *HeT-A*/*RpL32*in Lhr^{KO} vs. Lhr^{+}, showing mean from 3 biological replicates. Significance of fold change was calculated using Welch's one-tailed *t*-test; $p<0.05$. (B) Ratio of *HeT-A*/*RpL32* in Lhr^{KO}/*Df(2R)BSC44* vs. Lhr^{+}/ *Df(2R)BSC44*, showing mean from 4 biological replicates. Significance of fold change was calculated using the one tailed Wilcoxon rank sum test; $p<0.05$.

We also performed RNA-Seq analysis of an *Hmr* mutant (*Df(1)Hmr^{-}/Hmr^{3}*, abbreviated below as *Hmr^{-}*). We compared it to a heterozygous control (*Df(1)Hmr^{-}/y w Hmr^{+}*, abbreviated below as *Hmr^{-}/Hmr^{+}*) because it closely matches the genetic background of the mutant genotype, and also serves as a control for *Hmr* transgenic genotypes that are described below. We found that 55 different TE families are upregulated at least 2 fold in *Hmr* mutants, with 14 being upregulated at least 10 fold (Figure 3.6B; Table S3). Notably, the telomeric retrotransposons *HeT-A* and *TART* are again among the most highly upregulated. Strikingly, the TEs affected by *Hmr* are largely a subset of *Lhr*-regulated TEs, suggesting that they act together to regulate multiple TE families (Figure 3.6C). The smaller number of mis-regulated families in *Hmr^{-}* likely reflects the fact that we are comparing *Hmr^{-}* mutants to heterozygotes, but *Lhr* mutants to wild type.

Since some germline TE repressor genes also regulate somatic TE expression (Perrat *et al.* 2013), we performed RNA-Seq to compare TE expression between 72-76 hour-old *Df(1)Hmr^{-}/Y* and *Hmr^{+}/Y D. melanogaster* male larvae. This also served as a control for experiments described below to address whether TE mis-expression may be contributing to hybrid lethality. We found that 31 TEs exhibit a statistically significant ≥2 fold upregulation (Figure 3.6D; Table S4), but there are two striking differences compared to *Hmr* mutant ovaries. First, different TEs are affected, with the telomeric retrotransposons in particular not upregulated in the larvae. Second, the magnitude of TE derepression is lower in larvae.

3.3.4 *Lhr* and *Hmr* affect expression of heterochromatic genes

We next examined potential effects on protein-coding genes. Remarkably few genes (11 in Hmr^{-}; 0 in Lhr^{KO}) show a statistically significant misregulation in either *Lhr* or *Hmr* mutants (FDR 0.05; Tables S5, S6). However, a comparison of fold change in the expression of all heterochromatic versus all euchromatic genes found that heterochromatic genes are downregulated to a greater extent for both mutants, although the effect is stronger in Lhr^{KO} (Figure 3.8). Lhr preferentially associates with heterochromatic genes in an embryonic cell culture line (Greil *et al.* 2007); our results suggest that Lhr and Hmr have a small positive effect on expression of some heterochromatic genes.

3.3.5 *Lhr* and *Hmr* mutants have long telomeres

Drosophilidae have lost the telomerase-based mechanism of telomere elongation and instead use the regulated transposition of the *HeT-A*, *TART* and *TAHRE* retrotransposons (Pardue and Debaryshe 2011). Strikingly, these were among the 3 most strongly affected TEs in Lhr^{KO} and Hmr^{-} ovaries (Figure 3.6). We therefore investigated in more detail the localization of Lhr and Hmr proteins to the telomere (Brideau *et al.* 2006). Cytological markers on polytene chromosomes have been used to describe three distinct regions in the telomere, with HP1a localizing exclusively to the "cap," a proeinaceous structure at the most distal end of telomeres (Andreyeva *et al.* 2005; Raffa *et al.* 2011).

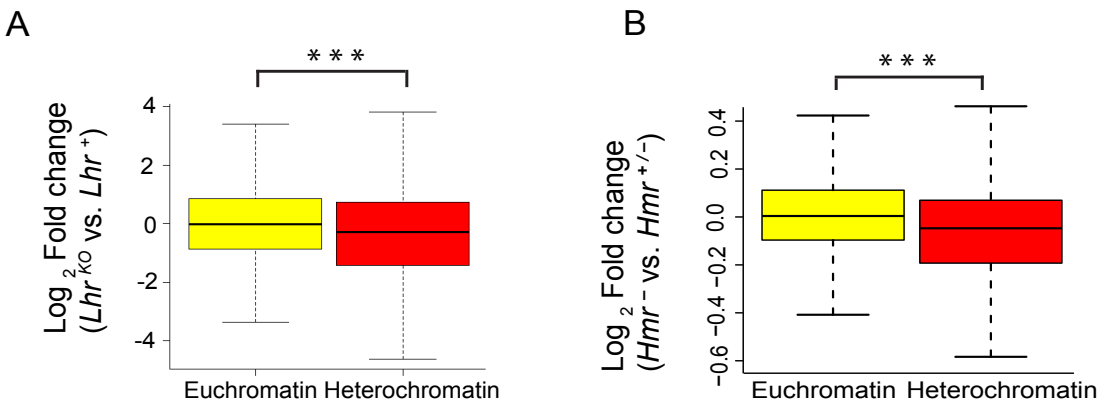


Figure 3.8. Reduced expression of heterochromatic genes in *Lhr* and *Hmr* mutants.
Loss of *Lhr* (A) and *Hmr* (B) leads to a statistically significant reduction in the expression of heterochromatic genes. Significance of difference was calculated using the Wilcoxon rank sum test with continuity correction (for (A) p = 3.549e-05, for (B) p =1.461e-09). Box plots show $\log_2$ fold change of 7838 euchromatic and 370 heterochromatic genes for (A) and 7451 euchromatic and 344 heterochromatic genes for (B). The definition of the euchromatin-heterochromatin boundary for all chromosomes comes from experiments done in S2 tissue culture cells, except for 3R, which comes from the cytogenomic border (Riddle *et al.* 2011).

mel-Lhr-HA and mel-Hmr-HA overlap with HP1a, showing that Lhr and Hmr localize to the cap but not to more proximal regions (Figure 3.9A, B). Localization is not due to the doubling of dosage of these proteins in the transgenic lines because it also occurs in mutant *Hmr*^{3}*; Hmr-HA/Hmr-HA* and *Lhr*^{KO}*/+; Lhr-HA/+* genotypes (Figure 3.10). The localization of Lhr and Hmr to the cap, the primacy of the cap in the regulation of telomeric length, and the increase in the transcript levels of telomeric retrotransposons in *Lhr* and *Hmr* mutants led us to ask if these mutations cause long telomeres. We quantitated *HeT-A* DNA copy number by qPCR in *Lhr*^{KO} flies maintained at 27°C separately from its matched wild-type control strain for ~40 generations. We found that *HeT-A* copy number increased approximately 6 fold in *Lhr*^{KO} (Figure 3.9C). We also examined *HeT-A* DNA copy number in an *Hmr*^{3} mutant stock, and found ~4-16 fold higher abundance than in the *Hmr*$^{+}$ stocks *y w* and Canton-S (Figure 3.9D).

3.3.6 Satellite DNA transcripts are upregulated in *Lhr* and *Hmr* mutants.

Hmr and Lhr both localize to pericentric heterochromatin, which is largely composed of TEs and satellite DNAs. The potential effects of heterochromatin proteins on the levels of transcripts from satellites have not been widely explored. We therefore used our RNA-Seq data to examine transcript levels from 143 repeats in a repeat-sequence database (see Materials and Methods). Transcripts from most repeats are found at low abundance in *Lhr*$^{+}$ with only 17 producing more than 10 reads (Table S7). Four different satellite classes are significantly higher in *Lhr*^{KO} versus *Lhr*$^{+}$ ovaries, including three that collectively make up more than 8% of the *D. melanogaster* genome (Lohe and Roberts

1988): AAGAC, AACAC, and the GA-rich satellites (Figure 3.11A). The GAGAA satellite showed the strongest effect, with an approximately 30-fold increase.

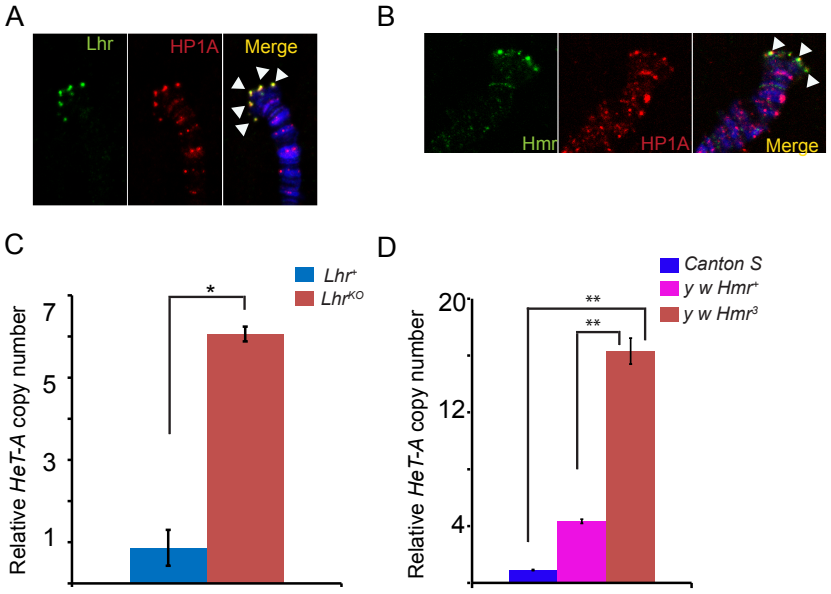


Figure 3.9. Lhr and Hmr are telomere cap proteins required for regulating telomere length. Lhr-HA (A) and Hmr-HA (B) localize to telomeres. Co-immunostaining with anti-HA and anti-HP1a shows that both proteins colocalize at the cap (arrowheads). The merged images include DAPI to stain DNA, shown in blue. Lhr^{KO} (C) and Hmr^{3} (D) have increased *HeT-A* copy number. qPCR was used to estimate the abundance of *HeT-A* and *rp49* from Lhr^{KO}, Lhr^{+}, *y w* Hmr^{3}, a matched *y w* Hmr^{+} control, and the wild-type Canton S strain. Genomic DNA was isolated from carcasses of females whose ovaries were removed in order to minimize the amount of polytenized DNA present. Relative *HeT-A* copy number is the ratio of *HeT-A/rp49.* The error bars represent S.E.M for three replicates. The significance of the differences between the genotypes was calculated using two tailed *t*-test; *= $p<0.05$; ** = $p< 0.01$. Scale bars = 5μm.

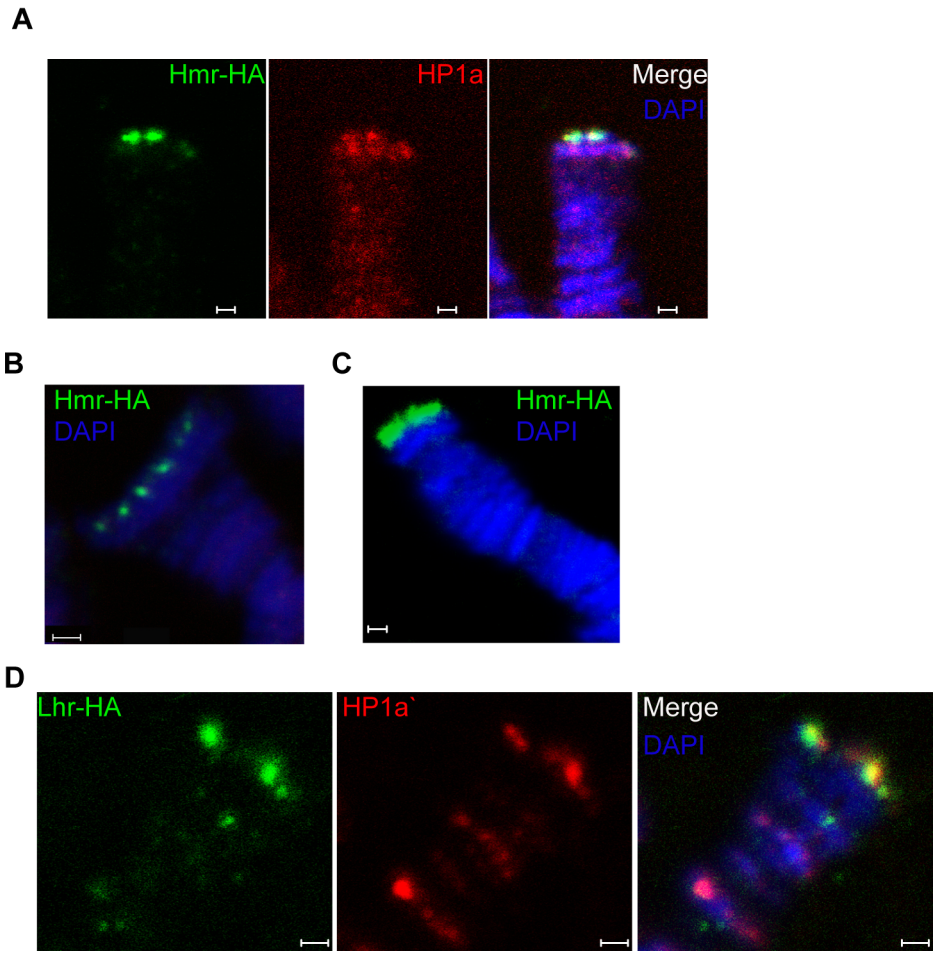


Figure 3.10. Localization of Hmr-HA and Lhr-HA to the telomeres is independent of dosage of endogenous copies. mel-Hmr-HA (green) in *Hmr*^{3}; *Hmr-HA* (A–C) and mel-Lhr-HA (green) in *Lhr*$^{KO/}$+; *Lhr-HA/*+ (D) colocalize with HP1A (red) at the telomere cap on polytene chromosomes. mel-Hmr-HA shows a range of distributions at the telomere, including punctate (B) and continuous across the chromosome terminus (C). Scale bar is 1 μm.

These results raise the question of whether transcriptional regulation of specific satellite DNAs reflects a direct association with Lhr. Lhr was not previously tested for association with either GA-rich satellites, which are found on all chromosomes in *D. melanogaster* (Lohe *et al.* 1993), or with the AACAC satellite found on chromosomes 2 and Y (Platero

et al. 1998). We found that Lhr-HA colocalizes extensively with the GA-rich and AACAC satellites in the nurse cell nuclei of early stage egg chambers (Figure 3.11B, 3.2A).

In our *Hmr* RNA-Seq data, the number of reads mapping to each repeat family was generally very small, but 3 satellite families are significantly derepressed by at least 4 fold in Hmr^{-} (Figure 3.11C; Table S8), including GAGAA which has a 19 fold increase in expression. This finding is consistent with the localization of mel-Hmr-HA to GA-rich satellites above (Figure 3.1B). Additionally, the satellite Z37541, which binds nuclear lamins, is upregulated 5 fold in Hmr^{-} (Baricheva *et al.* 1996).

Although Lhr-HA localizes to the dodeca satellite (Maheshwari and Barbash 2012); we detected very few reads in either our Lhr^{+} or Lhr^{KO} samples; likewise we did not find upregulation of dodeca in our *Hmr* RNA-Seq data. We conclude that Hmr and Lhr proteins are required to regulate transcript levels of a subset of satellites to which they localize.

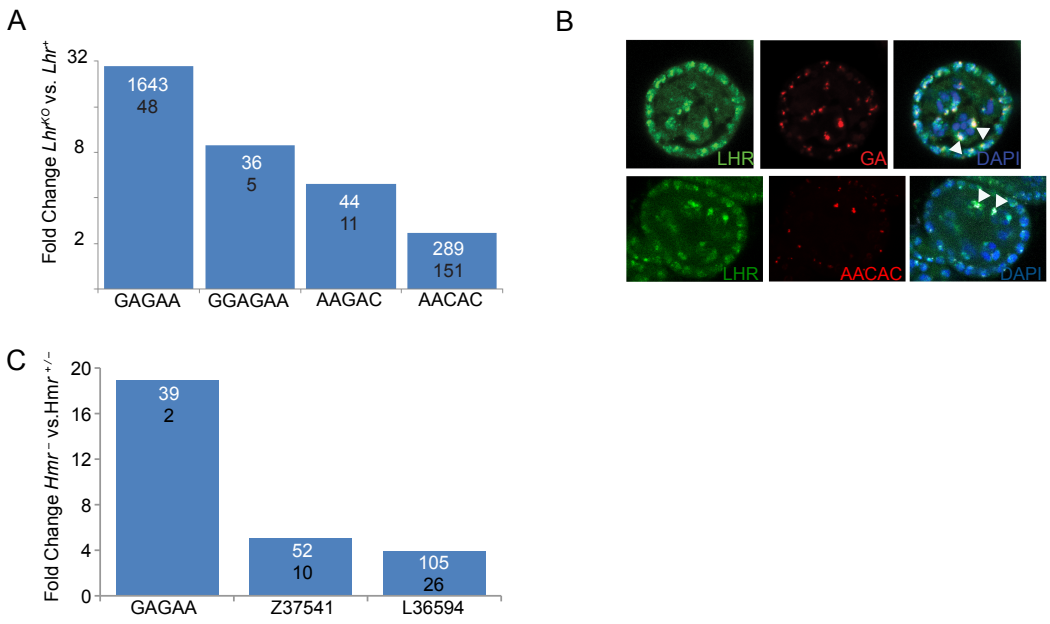


Figure 3.11. *Lhr* and *Hmr* repress satellite DNA transcription (A) Fold increase in satellite transcripts of Lhr^{KO} versus Lhr^{+}. Numbers within the bars show normalized reads mapping to each satellite, the numerator from Lhr^{KO} and the denominator from Lhr^{+}. All differences have $p<0.01$ by F.E.T. test. (B) Lhr-HA (green) colocalizes with GA-rich and AACAC satellites (red) in ovarian nurse cell nuclei (arrowheads) DAPI is shown in the merged images in blue. Scale bar = 10μm. (C) Fold increase in satellite transcripts in Hmr^{-} versus $Hmr^{+/-}$. Numbers within the bars show normalized reads mapping to each satellite, the numerator from Hmr^{-} and the denominator from $Hmr^{+/-}$. All differences have $p<0.001$ by F.E.T. test.

3.3.7 siRNA and piRNA patterns are largely normal in *Lhr*^{KO}

The wide spectrum of TEs derepressed in *Lhr* and *Hmr* mutants is similar to mutations in piRNA regulatory genes such as *AGO3* and *aub* that post-transcriptionally regulate TEs via small-RNA-mediated silencing (Vagin *et al.* 2006; Li *et al.* 2009). We therefore investigated a range of phenotypes that are associated with defects in the piRNA pathway. *AGO3* and *aub* mutants disrupt Vasa localization to the peri-nuclear small-RNA processing center, the nuage, and exhibit drastic reductions in the piRNA fraction (23-30nt) (Li *et al.* 2009; Malone *et al.* 2009). We found, however, that Vasa localizes normally in *Lhr*^{KO} (Figure 3.12A). We then sequenced the small RNA pool in *Lhr*^{KO} and found that the piRNA level is comparable to *Lhr*$^{+}$ with only a minor reduction in longer piRNAs (Figure 3.12B). This pattern contrasts with mutants such as *aub* and *spn-E* that show a severe loss of piRNAs (Malone *et al.* 2009). We looked more closely for TE-specific defects and found that piRNAs mapping to most individual TE families are comparable between *Lhr*$^{+}$ and *Lhr*^{KO} (Figure 3.12C; Table S9). We also examined "ping-pong" processing, which produces piRNAs from opposing strands with a characteristic 10 nucleotide overlap (Li *et al.* 2009; Malone *et al.* 2009). Ping-pong scores are generally higher in *Lhr*$^{+}$ (Figure 3.12D; Table S10) but several points argue against there being a significant defect in ping-pong or piRNA processing in *Lhr*^{KO}. First, the magnitude of the difference between genotypes is low, with the ping-pong score being >=2-fold higher in *Lhr*$^{+}$ for only 26/140 TEs. Furthermore, half of these 26 have ping-pong scores <0.01 in *Lhr*$^{+}$ (Table S10), suggesting that those TE families are not significantly processed by ping-pong in wild type flies. Second, these differences in

ping-pong scores between *Lhr*$^{+}$ and *Lhr*^{KO} are much milder compared to mutations in genes such as *spn-E* (Malone *et al.* 2009). Third, many of the TEs showing differences in ping-pong scores are not strongly derepressed in *Lhr*^{KO}. Conversely, many TEs that are strongly derepressed in *Lhr*^{KO}, including *HeT-A*, have ping-pong scores that are comparable to wild-type. Fourth, some TEs with elevated mRNA levels also show increased ping-pong signatures, probably because of increased processing through a functional ping-pong pathway. We suggest therefore that the moderate trend towards reduced ping-pong scores in *Lhr*^{KO} does not refelct a failure in the ping-pong cycle. Instead, it may result from a skew in the ratio of sense: antisense piRNAs, because *Lhr*^{KO} flies have high levels of TE transcripts that can be processed into sense piRNAs. An analogous argument has been made for mutations in the Drosophila *Gtsf1/asterix* gene, which derepress TEs and give an altered ratio of sense and antisense piRNAs but appear to do so downstream of piRNA biogenesis (Dönertas *et al.* 2013).

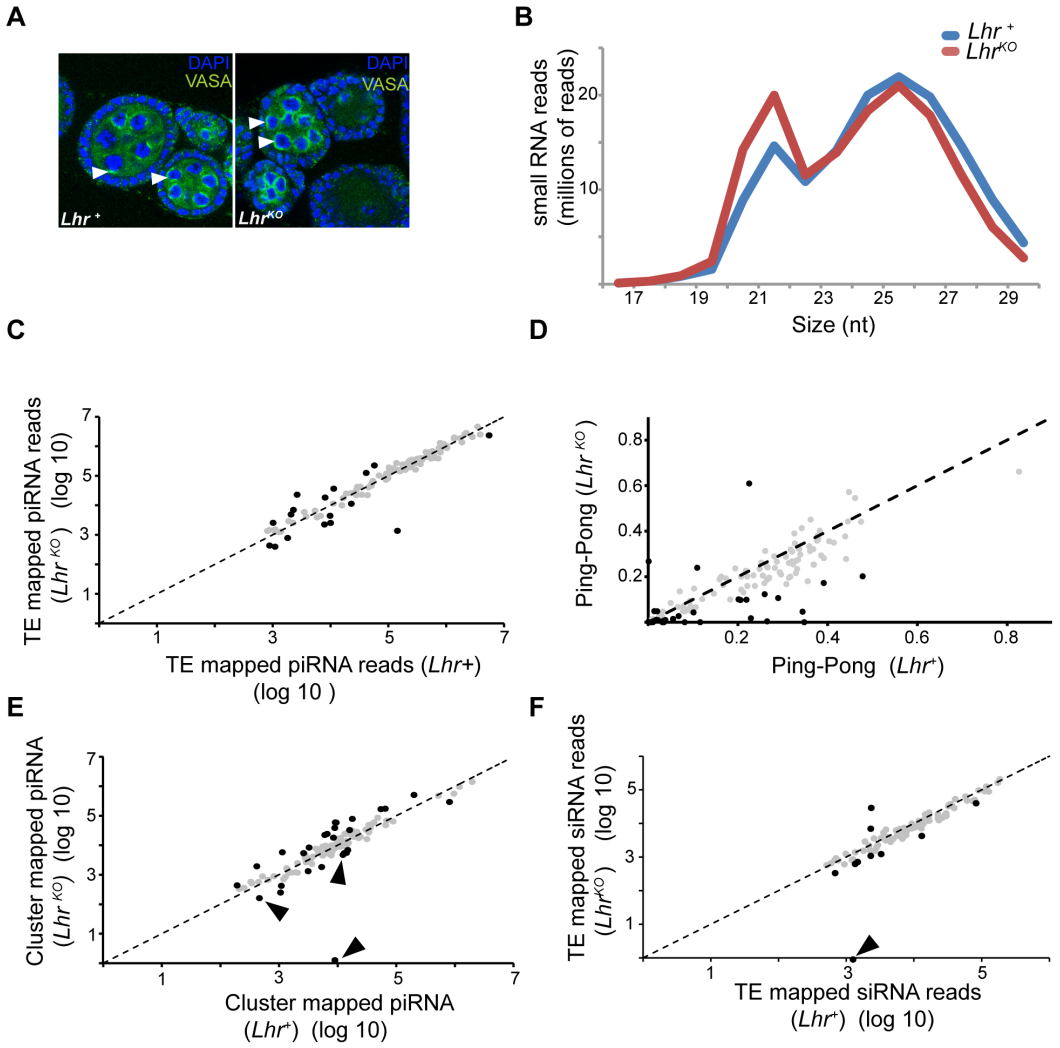


Figure 3.12. Small RNA patterns are largely unaffected in *Lhr*^{KO} (A) VASA (green) marks the peri-nuclear nuage (white arrowheads) and shows no difference in localization between *Lhr*$^{+}$ and *Lhr*^{KO} ovaries. (B) siRNA (17-22 nt) without mismatches and piRNA (23-30 nt) with up to one mismatch were mapped to a reference sequence set containing the *D. melanogaster* r5.68 genome, *D. melanogaster* sequences from Repbase and the repeat-sequence database. The number of mapped *Lhr*^{KO} reads was normalized to the total number of mapped *Lhr*$^{+}$ reads. (C) Filtered piRNA reads were mapped uniquely to the Repbase TE consensus sequences with one allowed mismatch. 121 TE families producing >=1000 reads summed over both genotypes are shown. Black circles represent TE families whose fold change between *Lhr*^{KO} and *Lhr*$^{+}$ is greater than 2 fold ($p < 0.001$). (D) Ping-pong scores of TE families in *Lhr*^{KO} and *Lhr*$^{+}$. Black circles represent TE families whose fold change in ping-pong score between *Lhr*^{KO} and *Lhr*$^{+}$ is greater than 2 fold (Table S10). (E) Plot shows the number of unique piRNAs mapped to piRNA clusters, with one allowed mismatch and normalized between genotypes. piRNA clusters with >=500 reads summed over both genotypes are shown. Black arrowheads point to sub-telomeric piRNA clusters. Black circles indicate clusters whose fold change between *Lhr*^{KO} and *Lhr*$^{+}$ is greater

than 2-fold (*p < 0.001)*. (F) Unique siRNA (17-22 nt) were mapped as in (C), except no mismatches were allowed. 96 TE families are plotted that have >=1000 reads summed over both genotypes. Black circles represent TEs whose siRNA levels changed by >2 fold. siRNA mapping to the TAS repeat HETRP are almost completely lost (arrow). For (C, D, F) significance values were calculated using F.E.T., implemented in DEG-seq.

We searched further for possible defects in piRNA production by examining piRNAs that map to 122 primary-piRNA-generating heterochromatic clusters (Brennecke *et al.* 2007). piRNAs originating from most of the major clusters are not significantly affected in Lhr^{KO}, but clusters 16 and 11 of the 122 clusters are at least two-fold higher or lower, respectively in Lhr^{KO} (Figure 3.12E; Table S11). Some of the most strongly affected clusters are associated with telomeres. Cluster 3 consists entirely of telomeric retrotransposons and is upregulated 4.3 fold in Lhr^{KO}. Sub-telomeric cluster 11 shows a complete loss of unique piRNAs, while clusters 33 and 4 are 2.6 and 2.9 fold downregulated, respectively. These 3 clusters consist mainly of HETRP telomere-associated (TAS) repeats and are therefore not expected to contribute to TE repression; their misregulation instead suggests that *Lhr* is required for regulating chromatin states at telomeres.

The siRNA pathway has also been implicated in repressing TEs in the ovary (Blumenstiel and Hartl 2005; Czech *et al.* 2008; Rozhkov *et al.* 2010). We found that siRNAs mapping to the vast majority of TE families, including those mapping to *HeT-A*, are not significantly different between Lhr^{KO} and Lhr^{+}, suggesting that *Lhr* is not generally required for siRNA biogenesis (Figure 3.12F; Table S12). Taken together, our results indicate that defects in small RNA synthesis are not the cause of TE

derepression in *Lhr*^{KO}. An intriguing possibility is that Lhr is a piRNA-dependent effector of TE silencing.

3.3.8 Comparing *Lhr* function in *D. simulans* and *D. melanogaster*

We propose that the dynamic sequence turnover of repetitive DNAs is the selective pressure driving the adaptive sequence divergence of *Lhr* and *Hmr*. This hypothesis implies that the localization and/or function of the Lhr protein have changed between species, due to co-evolution with species-specific repetitive DNAs. The *Lhr*^{1} allele in *D. simulans* (Watanabe 1979) presents a rare opportunity to compare the function of a rapidly evolving heterochromatin protein between sibling species. We performed RNA-Seq from ovaries of *Lhr*^{1} females and a matched *Lhr*$^{+}$ control (see Materials and Methods). We found essentially no *Lhr* transcript reads in the *Lhr*^{1} mutant strain (Table S13), strongly suggesting that this allele is null.

D. simulans has many of the same satellites as *D. melanogaster* but they are generally of lower abundance (Lohe and Roberts 1988). We therefore first examined satellite DNA expression in the *Lhr*^{1} and *Lhr*$^{+}$ (control) RNA-Seq data. Unlike in *D. melanogaster* *Lhr*^{KO}, we found few reads mapping to repeats in either genotype and no significant differences between them. We conclude that *Lhr* has a unique role in *D. melanogaster* to repress satellite DNA transcription. The AACAC satellite that co-localizes with Lhr in *D. melanogaster* (Figure 3.11B) is absent in *D. simulans* (Platero *et al.* 1998). The GAGAA satellite is also drastically different in *D. simulans*, being eight-fold less abundant and found only on the sex chromsomes (Lohe and Roberts 1988; Platero *et al.*

1998) . To determine if this interspecific difference in satellite content reflects divergent localization patterns of Lhr orthologs, we examined *D. simulans* ovaries expressing a previously characterized *sim-Lhr-HA* transgene (Maheshwari and Barbash 2012). While Lhr-HA is juxtaposed to dodeca in both species, as previously described (Maheshwari and Barbash 2012), the strongest foci in *D. simulans* do not overlap with GAGAA (Figure 3.13A). These results demonstrate that Lhr has evolved distinct localization patterns to at least two satellites between *D. melanogaster* and *D. simulans.*

We next examined TE expression and discovered a broad spectrum of TEs derepressed in *D. simulans* Lhr^{1}, with 80 TE families showing a greater than two-fold up-regulation (Figure 3.13B; Table S14). Upregulated TEs again include the telomeric transposable elements *HeT-A, TART,* and *TAHRE,* other germline elements such as *Nomad,* and somatic TEs such as *Zam* and *Gypsy 5*. 53 transposable elements were commonly mis-regulated in both *D. melanogaster* and *D. simulans,* showing that the function of *Lhr* in repressing TEs is broadly conserved between species (Figure 3.13C). However, the fold increases of most individual TE families are lower than seen in *D. melanogaster* Lhr^{KO}. For example, *HeT-A* is 352 fold upregulated in Lhr^{KO} but only 23.8 fold upregulated in Lhr^{1}.

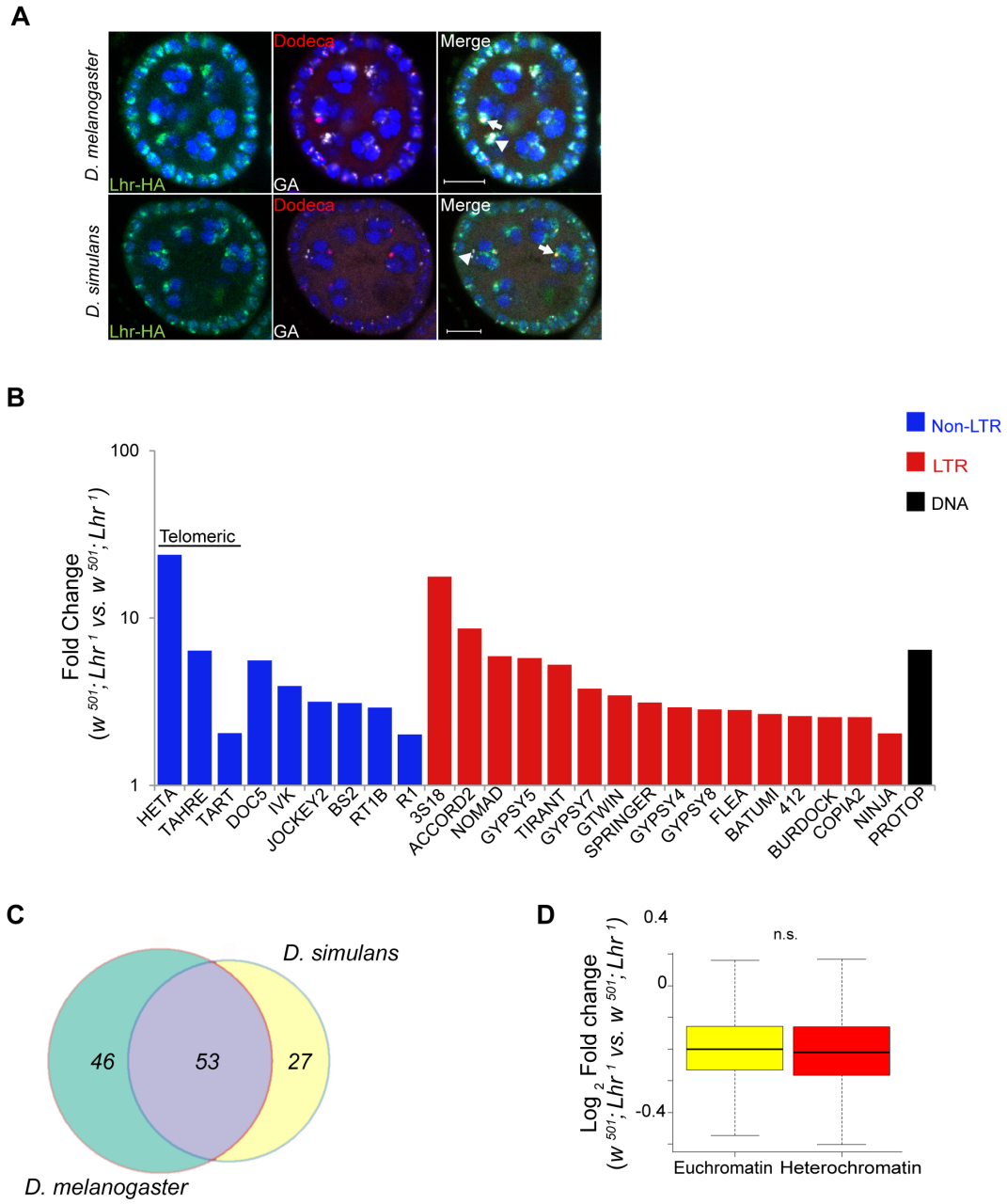


Figure 3.13. Analysis of *Lhr* function in *D. simulans* (A) Immuno-FISH experiment shows that the brightest mel-Lhr foci colocalize with Dodeca (red, arrow) and GA satellites (white, arrowhead) in *D. melanogaster* (upper panel). The brightest sim-Lhr foci either colocalize or are juxtaposed with Dodeca (arrow) but are not associated with GA-rich satellites (arrowhead). All panels contain DAPI shown in blue. Scale bar = 10μm. (B) Fold changes in TE expression between w^{501}; Lhr^{1} and w^{501}; Lhr^{+} were calculated for uniquely mapping reads with zero mismatches to the individual-insertion database and with three mismatches to the consensus-sequence database. Three mismatches are required to account for the divergence of TE insertions in *D. simulans* from the consensus sequences, which are largely defined from *D. melanogaster* TEs. The 25 most significantly derepressed TE families in the individual-insertion sequence based analysis are shown here (excluding *centroids*), as well as *TAHRE*, which is

found only in the consensus-sequence database. Classification of DNA, LTR and non-LTR elements is from reference (Kaminker *et al.* 2002). (C) Comparison of TE misregulation between *D. melanogaster* and *D. simulans Lhr* mutations. The diagram includes all TE families that were upregulated at least two fold, including those in individual-insertion database analysis as well as those that are only represented in the consensus-sequence database analysis. (D) Comparison of euchromatic and heterochromatic gene expression in *D. simulans* w^{501}*;* Lhr^{1}, as described in Figure 3.8. The euchromatin-heterochromatin border has not been experimentally determined in *D. simulans* and was defined from *D. melanogaster.* Analysis includes 7479 euchromatic and 350 heterochromatic genes (p = 0.12, Wilcoxon rank sum test with continuity correction).

We further discovered that *Lhr* loss in *D. simulans* does not significantly affect the expression of heterochromatic genes (Figure 3.13D, Table S13), in contrast with our similar analysis of Lhr^{KO} in *D. melanogaster* (Figure 3.8A). This result suggests that pericentric genes in *D. melanogaster* are more sensitive to changes in heterochromatin state than in *D. simulans.* Overall, our results demonstrate that *Lhr* function correlates with the increased repeat content and larger amount of heterochromatin found in *D. melanogaster*.

3.3.9 Comparison of *Hmr* ortholog function

To examine the functional consequences of *Hmr* divergence, we took an alternative approach of transforming *sim-Hmr* transgenes into *D. melanogaster*. We found that sim-Hmr-HA, like mel-Hmr-HA, localizes to heterochromatin in *D. melanogaster* (Figure 3.14A).

To examine potential differences in TE and satellite regulation, we used parallel *mel-Hmr-FLAG* and *sim-Hmr-FLAG* transgenes, crossed them into an Hmr^{-} background ($Df(1)Hmr^{-}/Hmr^{3}$), and performed RNA-Seq on ovarian mRNA. Our expectation was

that divergence of *Hmr* between the orthologs might manifest as the failure of *sim-Hmr-FLAG* to complement the derepression of TEs in *Hmr*$^{-}$.

As a control for the function of the transgenes, we compared the heterozygous wild type *Hmr*$^{-}$/*Hmr*$^{+}$ to *Hmr*$^{-}$*; ø{mel-Hmr-FLAG}/+*, as each genotype has one wild type copy of *Hmr*$^{+}$. The majority of the upregulated TEs in *Hmr*$^{-}$ (Figure 3.6B) are suppressed by the *mel-Hmr-FLAG* transgene; however, 9 out of 182 families ranged from 2 to 9 times more highly expressed in *Hmr*$^{-}$*; ø{mel-Hmr-FLAG}/+* than *Hmr*$^{-}$/*Hmr*$^{+}$ (Figure 3.14B). This result suggests that *mel-Hmr-FLAG* does not fully complement the *Hmr* mutant phenotype, which may reflect its decreased expression compared to a wild type allele (Figure 3.3), though it is also possible that some differences may result from TE polymorphisms that remain between the strains. qRT-PCR also demonstrated that *sim-Hmr-FLAG* expresses in *D. melanogaster* at ~3x the level of *mel-Hmr-FLAG* (Figure 3.3), a difference previously seen with *Lhr* transgenes (Maheshwari and Barbash 2012). Because *Hmr* is a negative regulator of TE expression, we suggest that this expression difference will not bias our goal of identifying TEs that are not fully repressed by *sim-Hmr-FLAG*.

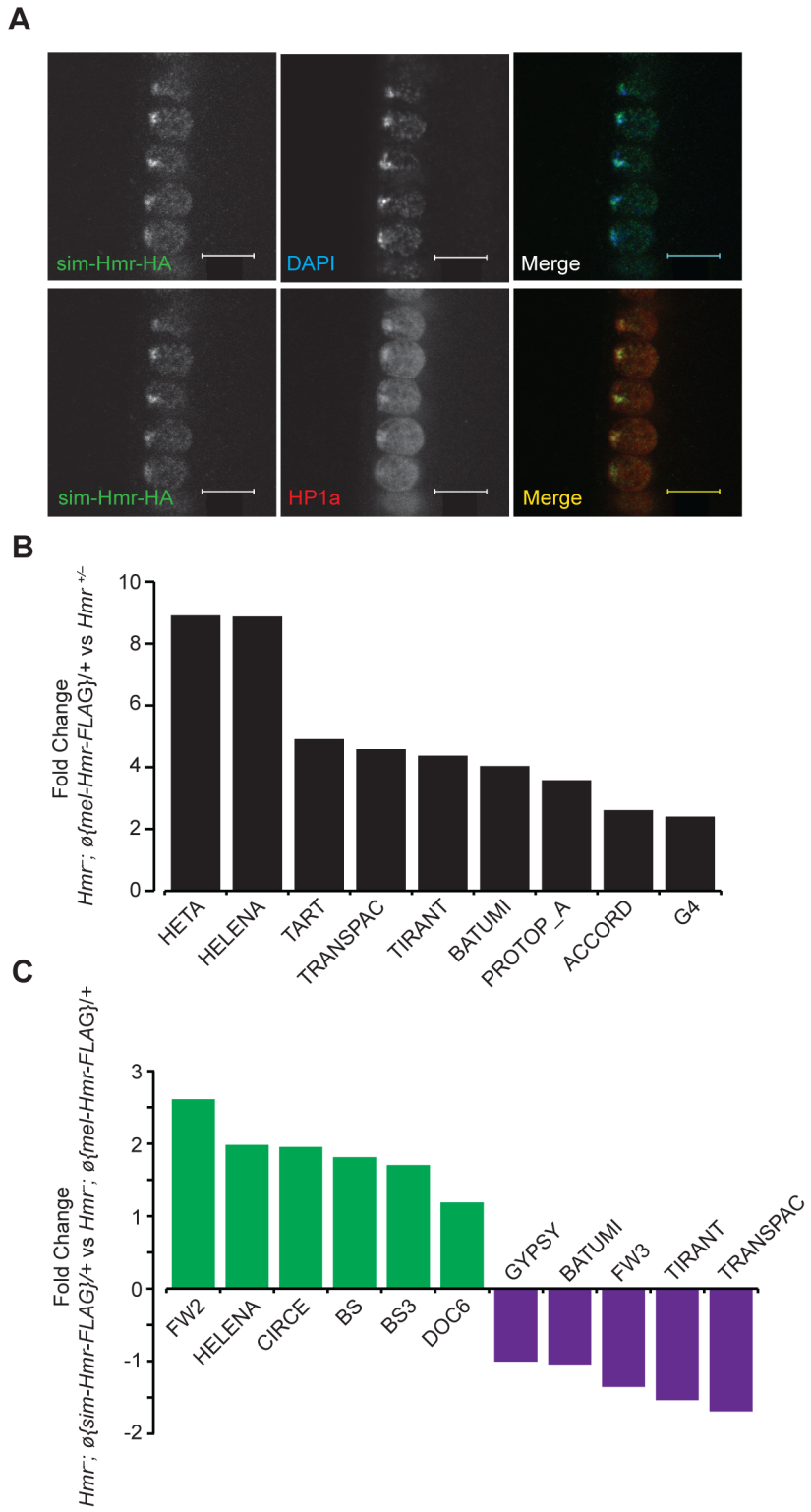


Figure 3.14. *Hmr* orthologs have diverged in their effects on a small subset of TEs. (A) sim-Hmr-HA colocalizes with HP1a (red) in nuclear cycle 14 *D. melanogaster* Hmr^{3}*; sim-Hmr-HA* embryos. The *sim-Hmr-HA* transgene was transformed into *D. melanogaster* at the identical

attP2 site used for *mel-Hmr-HA* above (Figure 3.1). (B) *mel-Hmr-FLAG* does not fully complement TE derepression in Hmr^{-}. 9 TE families are 2-9x more highly expressed in Hmr^{-}; *ø{mel-Hmr-FLAG}/+* compared to $Hmr^{+/-}$. (C) Comparison of TE expression in Hmr^{-}*; ø{mel-Hmr-FLAG}/+* and Hmr^{-}*; ø{sim-Hmr-FLAG}/+*. For B and C, reads were mapped to the individual-insertion database. TEs are considered differentially expressed in the pairwise comparisons if there was at least a 2x fold change and $p<0.001$.

We did not find any difference in satellite DNA expression; however, we found 11 TE families that are differentially expressed between the transgenic genotypes (Figure 3.14C). Five are more highly expressed in Hmr^{-}*; ø{mel-Hmr-FLAG}/+* with fold changes ranging from 2-3, of which 3 are incompletely repressed by *mel-Hmr-FLAG* in the control cross described above (*Transpac*, *Tirant*, and *Batumi*). The differential expression of these 5 families likely reflects the inability of *mel-Hmr-FLAG* to fully complement Hmr^{-} and the higher expression level of *sim-Hmr-FLAG.*

More intriguing are 6 TE families that are 2-6x more highly expressed in Hmr^{-}*; ø{sim-Hmr-FLAG}/+* than in Hmr^{-}*; ø{mel-Hmr-FLAG}/+*, implying that *sim-Hmr-FLAG* is unable to fully complement the derepression of these elements. *BS* and *Doc6* (also known as *Juan*) elements are present at a mean frequency of about 0.1 in a population of Portuguese *D. melanogaster* (Kofler *et al.* 2012) and have low pairwise identity in the reference genome (Bergman and Bensasson 2007), suggesting that they are likely active. The mean population frequencies of 4 of the other families (*BS3*, *Circe*, *Helena*, and *FW2*) are near 1, suggesting that these TEs are fixed and therefore currently inactive in *D. melanogaster*. *Helena,* though, appears to have been active more recently within *D. simulans* (Rebollo *et al.* 2008). We suggest that *BS, Doc6* and *Helena* are

candidates for future investigation of co-evolution with *Hmr* in either *D. melanogaster* or *D. simulans*.

3.3.10 Transposable elements are upregulated in hybrids

In light of our discovery that *Lhr* and *Hmr* are required for TE repression within *D. melanogaster* and *D. simulans*, we investigated TE activity in lethal (Hmr^{+}) hybrid male larvae. Because most TEs have different expression levels between *D. melanogaster* and *D. simulans*, we defined mis-regulated TEs as being at least two-fold higher than both parental species, as done in a previous analysis (Kelleher *et al.* 2012). We found that 42 LTR and non-LTR elements are significantly upregulated in lethal (Hmr^{+}) hybrid male larvae with 2 others being downregulated (Figure 3.15A; Table S15).

We next examined whether TE misregulation correlates with hybrid lethality by comparing the lethal Hmr^{+} hybrid males to viable Hmr^{-} hybrid males (Figure 3.15B, Table S16). The expression of 29 TEs is significantly lower in viable Hmr^{-} hybrids. Because *Hmr* functions as a repressor of TEs in *D. melanogaster* male larvae (Figure 3.6C), these differences may reflect a general difference between lethal and viable hybrids rather than the presence or absence of *Hmr* activity. In fact, only 4 of the 29 TEs downregulated in Hmr^{-} hybrid male larvae are upregulated in Hmr^{-} *D. melanogaster* male larvae (Table S4).

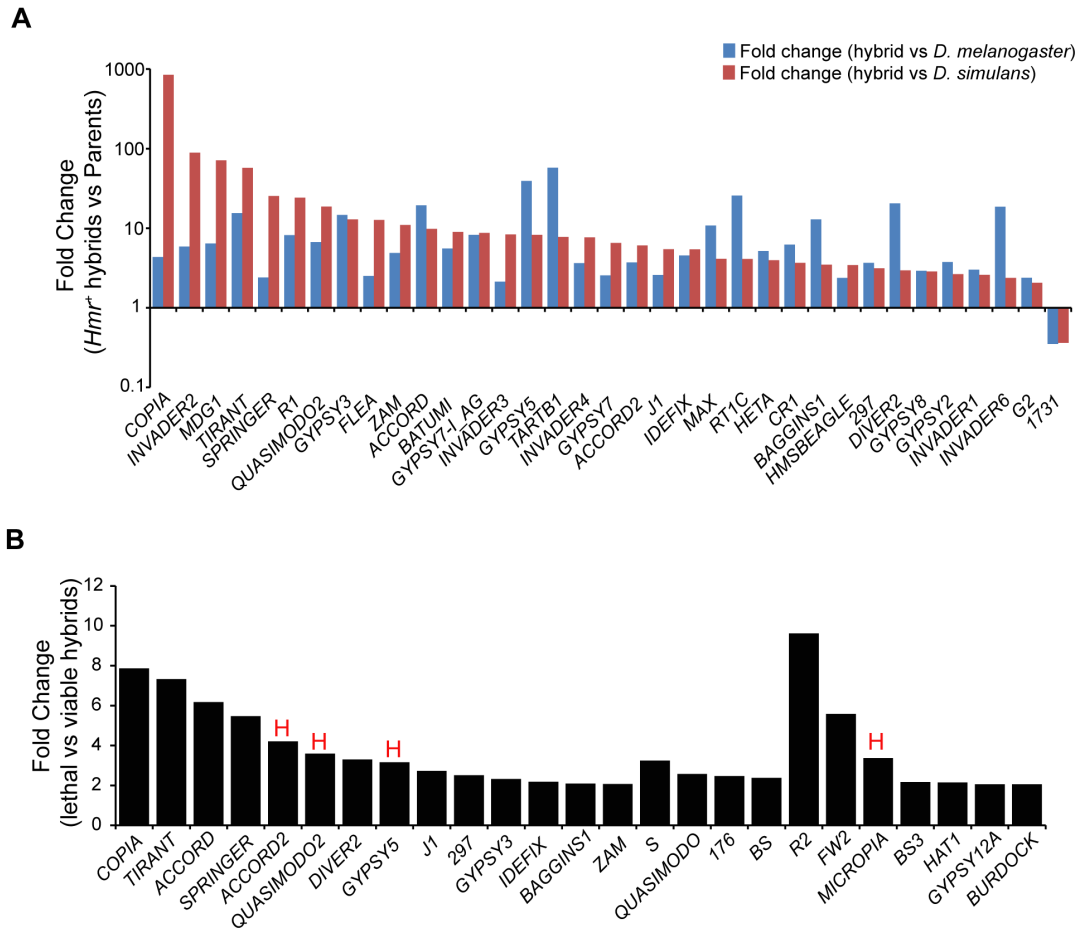


Figure 3.15. TE misregulation in hybrid males. (A) Fold change of TEs up- or downregulated ≥2-fold in *Hmr*$^{+}$ hybrid male larvae relative to both *D. melanogaster* and *D. simulans* male larvae. Uncharacterized *centroids* are not shown. (B) Fold change of TEs with significantly higher expression in lethal *Hmr*$^{+}$ versus viable *Hmr*$^{-}$ hybrid male larvae. "H" indicates TEs that are significantly upregulated in *Hmr*$^{-}$ *D. melanogaster* male larvae compared to *Hmr*$^{+}$ *D. melanogaster* male larvae from Figure 3.6D. Note the different Y axis scales between panels A and B. TE families include those resulting from mapping reads to the individual-insertion database, as well as families found only when reads were mapped to the consensus-sequence database. Reads unique to each TE class were mapped allowing for up to 3 mismatches.

In addition, we found modest increases (2-4 fold) in the activity of 5 TE families in living hybrids. None are significantly upregulated in *Hmr⁻ D. melanogaster* male larvae (Table S4). They include *TAHRE* and may reflect higher levels of cell proliferation in viable hybrids. Taken together our results suggest that TE overexpression is unlikely to be causing hybrid lethality.

3.4 Discussion

3.4.1 Lhr and Hmr interact with HP1a

We and others previously reported that Lhr (also known as HP3) interacts with HP1a (Giot *et al.* 2003; Brideau *et al.* 2006; Greil *et al.* 2007; Brideau and Barbash 2011). Here we report that Hmr also interacts with Lhr, and both are present in a complex together with HP1a. Consistent with this interaction, many of the roles we report here for Lhr and Hmr have been described for HP1a, including localizing to heterochromatin, regulating TE and pericentric gene expression, and controlling telomere length (Eissenberg and Elgin 2000; Savitsky *et al.* 2002; Wang and Elgin 2011). However, unlike mutations in *Su(var)205* which enodes HP1a (Eissenberg *et al.* 1990), mutations in *Hmr* and *Lhr* are viable. Furthermore, Hmr and Lhr do not localize to the 359 bp satellite which forms a substantial fraction of X-linked pericentric heterochromatin (Figure 3.1) (Maheshwari and Barbash 2012). These findings suggest that Hmr and Lhr are not ubiquitous heterochromatin proteins, leaving open the intriguing question of what guides their localization specificity.

The interaction of Hmr and Lhr with HP1a, as well as their effects on TEs in somatic cells have recently been independently reported (Thomae *et al.* 2013; AA Alekseyenko and M. Kuroda, personal communication). Thomae *et al.* 2013 also report other findings similar to ours here including repressive effects of Hmr and Lhr on TEs in somatic tissues and their localization to telomeres. Several conclusions are similar between the two studies and with previously published conclusions. Thomae *et al.* 2013 observe upregulation of TEs in hybrids but conclude that they are unlikely to be the direct cause

of hybrid lethality, a conclusion we reach below using different methods. Their conclusion that hybrids are highly sensitive to *Hmr* dosage is in concordance with previous studies, such as the previous observation that a ~9.7 kb Hmr^{+} transgene causes dosage-dependent lethality to hybrid females (Barbash *et al.* 2003). This conclusion also fits well with the discovery that hybrids are highly sensitive to *Lhr* dosage (Maheshwari and Barbash 2012).

One area of possible discrepancy is the viability effects and cellular phenotypes associated with *Hmr* and *Lhr* mutants versus RNAi knockdown. Thomae *et al* report a high rate of mitotic defects in *Lhr* RNAi knockdown tissue culture cells, yet we found that Lhr^{KO} flies are almost fully viable (see Results), as are *Lhr* RNAi knockdown animals (Thomae *et al.* 2013). We also have not observed the lethality or morphological defects in *Hmr* mutants that are reported for Hmr RNAi cells and animals (Thomae *et al.* 2013). For example, Aruna et al. (Aruna *et al.* 2009) found reduced longevity but no effect on viability up to eclosion of flies carrying the $Df(1)Hmr^{-}$ allele, a deletion of the 5' end of *Hmr*. Further work is necessary to determine if these discrepancies reflect phenotypes associated with the use of RNA interference or differences between assaying whole animals versus tissue-culture cells, such as the aneuploid state of cultured cell lines (Zhang *et al.* 2010).

3.4.2 Rapidly evolving heterochromatin proteins and repetitive DNA variation

Several HIs involve heterochromatin proteins or heterochromatic sequences, leading to the suggestion that genetic conflicts between selfish DNAs and host fitness are an important force that is driving the evolution of HI (Presgraves 2010; Johnson 2010; Maheshwari and Barbash 2011; Sawamura 2012).

TE and satellite abundance varies widely among species and is a major contributor to genome-size variation. The evolutionary causes of this variation have been widely debated for many years (Doolittle 2013). When considering genetic conflict theories, it is important to first exclude alternative evolutionary causes of repetitive DNA variation. One explanation is neutrality, with repeat variation governed by mutational processes, in particular the balance between insertions and deletions (Petrov 2002). Insertion/deletion models are particularly appropriate for inactive and degenerate TEs, and perhaps also for certain classes of satellites that are no longer homogenized by concerted evolution (Shepelev *et al.* 2009).

Selectionist models fit better for active repeats, and must be invoked if the adaptive evolution of heterochromatin proteins is proposed to reflect co-evolution with repetitive DNA. One model is that some repeats are co-opted for host functions. Drosophila's telomeric retrotransposons are a relevant example that is discussed below. We also consider three, non-mutually exclusive selective costs associated with repetitive DNA when discussing the evolution of *Hmr* and *Lhr.*

One potential cost arises from the overall load of repetitive DNAs, including increased genome size and instability. A second is direct genetic conflict. We define genetic conflict here to refer to fitness costs imposed by selfish DNAs that have evolved specific mechanisms to increase their transmission (Werren 2011). Such conflicts could be caused by highly active individual repeats, for example during hybrid dysgenesis caused by introduction of a TE family into naive strains (Bregliano and Kidwell 1983). Finally, genetic conflicts can have indirect costs, such as pleiotropic fertility defects caused by repeat expansions involved in meiotic drive (Fishman and Saunders 2008).

3.4.3 *Hmr* and *Lhr* repress transposable elements

TEs define selfish DNA (Doolittle 2013). They infect most genomes, can self-mobilize and increase their copy number, and destabilize genomes via spontaneous mutations, ectopic recombination, and deleterious increases in genome size (Lee and Langley 2010; González and Petrov 2012). Adaptive evolution of TE-defense genes can therefore be readily interpreted as the host species responding to the fitness cost of TEs (Blumenstiel 2011).

Like *Hmr* and *Lhr*, many piRNA pathway genes are also evolving under positive selection (Lee and Langley 2012). This raises the possibility that *Lhr* and *Hmr* are co-evolving with the piRNA pathway proteins. However, the lack of major perturbations in the piRNA pool in Lhr^{KO} suggests that Lhr and Hmr function downstream or independently of piRNA biogenesis. Piwi, guided by piRNA, has been proposed to recruit repressive heterochromatin components including HP1a and histone methyl

transferases to transposable elements (Klenov *et al.* 2011; Wang and Elgin 2011). One possibility is that Lhr and Hmr function downstream of Hp1a to repress TEs via RNA degradation machinery such as the nuclear exosome (Yamanaka *et al.* 2012).

We note that *AGO3* is moderately down-regulated in both Lhr^{KO} (3.4 fold) and Hmr^{-} (~2 fold) (Tables S5, S6), likely because the gene is pericentromeric. Two results demonstrate that this modest reduction in *AGO3* cannot explain the broad effects on TEs in *Hmr* and *Lhr* mutants. First, *AGO3* expression is unaffected in *D. simulans* Lhr^{1}, which also shows widespread TE derepression. Second, *AGO3* mutants have major disturbances to their piRNA pool (Li *et al.* 2009), which we did not observe in Lhr^{KO} (Figure 3.12).

3.4.4 *Hmr* and *Lhr* regulate telomeres

While TE repression is typically viewed in terms of genetic conflicts, the relationship between *Lhr*, *Hmr* and the telomeric TEs resembles symbiosis. These TEs have been domesticated by Drosophila species for tens of millions of years to serve a vital host function, and thus are not considered selfish DNA (Villasante *et al.* 2007; Pardue and Debaryshe 2011). The telomeric TEs were among the most strongly derepressed in *Hmr* and *Lhr* mutants, in some cases more than 100 fold. We also observed increases in *HeT-A* DNA copy number in *Hmr* and *Lhr* stocks. Increased telomeric TE expression does not necessarily increase *HeT-A* DNA copy number and cause longer telomeres, suggesting that multiple factors control telomere length (Shpiz and Kalmykova 2012). If so, then *Lhr* and *Hmr* must control multiple processes at the telomere. This is supported

by the localization of both proteins to the telomere cap, a protective structure that prevents telomere fusions (Raffa *et al.* 2011). The strong reduction in Lhr^{KO} of piRNAs from three TAS-repeat containing sub-telomeric piRNA clusters is particularly intriguing. piRNA production from clusters is dependent on them maintaining a heterochromatic state (Rangan *et al.* 2011), which could explain why *Lhr* is required for TAS piRNA expression while it acts as a repressor in most other circumstances.

3.4.5 *Hmr* and *Lhr* regulate species-specific satellite DNAs

We discovered several striking examples that suggest species-specific co-evolution of *Hmr* and *Lhr* with satellite DNAs. We found that *D. melanogaster* Hmr and Lhr proteins localize to and repress transcripts from GA-rich satellites. GA-rich satellites are ~8 fold less abundant in *D. simulans* (Lohe and Roberts 1988) but are cytologically detectable; nevertheless we find that sim-Lhr does not localize to them. GA-rich satellites also have low abundance in the outgroup species *D. erecta* (Lohe and Roberts 1988), implying that the differential abundance with *D. simulans* reflects an increase in *D. melanogaster*. Similarly we discovered that mel-Lhr-HA localizes to AACAC in *D. melanogaster*, a repeat that is absent in *D. simulans* (Sage and Csink 2003) (Sage & Csink 2003). Furthermore, we detected moderate up-regulation of several other satellite transcripts only in *D. melanogaster*. Our results suggest that *Lhr* and *Hmr* may have evolved in *D. melanogaster* to mitigate the deleterious consequences of satellite expansion, which can include ectopic recombination, increased genome size, and destabilized chromosome segregation (Charlesworth *et al.* 1994; Ferree and Prasad 2012).

Satellite transcripts have been reported from various tissues in wild type *D. melanogaster* (Bonaccorsi *et al.* 1990; He *et al.* 2012), but little is known about their production. They could be products of either non-specific transcription or read-through from adjacent TEs. Increased levels of satellite transcripts are observed in *D. melanogaster spn-E* mutants, suggesting that RNA interference or piRNA pathways control satellite transcript levels (Usakin *et al.* 2007).

3.4.6 Is the adaptive evolution of *Hmr* and *Lhr* driven by diverging heterochromatic repeats?

We find that at a broad scale, *Lhr* and *Hmr* from both *D. melanogaster* and *D. simulans* regulate heterochromatic repetitive DNAs but very few genes. This finding is consistent with previous analyses demonstrating that some functions of these genes are conserved between species (Barbash *et al.* 2004; Aruna *et al.* 2009; Brideau and Barbash 2011; Maheshwari and Barbash 2012). But many of the repeats regulated by *Lhr* and *Hmr* are rapidly evolving, raising the question of whether specific repetitive DNAs are directly driving the adaptive evolution of the *Lhr* and *Hmr* coding sequences between species. A simple prediction is that *D. simulans* orthologs should fail to fully repress such repeats when placed into *D. melanogaster*, a prediction that we tested for *Hmr*.

The *BS* non-LTR retrotransposon is significantly derepressed in *D. melanogaster* Hmr^{-} and Lhr^{KO} and in *D. simulans* Lhr^{1} mutants. Interestingly, *BS* appears to be transpositionally active in *D. melanogaster* but inactive in *D. simulans* (Granzotto *et al.*

2011). One interpretation is that *BS* was active in the common ancestor and regulated by *Hmr* and *Lhr*. The genes would continue to co-evolve with *BS* in *D. melanogaster*, making the *sim-Hmr* ortholog less effective at repressing *BS* elements in *D. melanogaster*. In this scenario, *Hmr* and *Lhr* are engaged in a recurrent genetic conflict with *BS* elements that leads to their sequence divergence. Consistent with this prediction, we found significantly higher expression in *Hmr⁻; ø{sim-Hmr-FLAG}/+* compared to *Hmr⁻; ø{mel-Hmr-FLAG}/+*.

Copia shows a different pattern, with ~20-fold up-regulation in Lhr^{KO} but only ~2-fold in Lhr^{1} (and only when mapping to the consensus-sequence database), as well as significant derepression in *Hmr⁻*. *Copia* expression level can be high in *D. melanogaster* but is variable among populations. In contrast, *copia* elements in *D. simulans* typically contain deletions in regulatory elements required for expression, and transcripts are undetectable by Northern blot analysis (Csink and McDonald 1995). These results suggest that *Hmr* and *Lhr* could be *D. melanogaster* host factors that defend against a TE that is currently active within the species. However, *copia* was fully repressed in *Hmr⁻; ø{sim-Hmr-FLAG}/+*, demonstrating that adaptive divergence of *Hmr* by itself does not affect *copia* regulation.

Overall, we found surprisingly few cases of overexpression associated with *Hmr* divergence, including no effects on satellite DNAs (Figure 3.14). We also note that most of the TEs identified other than *BS* and *Doc6* are likely transpositionally inactive in

D. melanogaster (Kofler *et al.* 2012), which makes it more challenging to fit a scenario of direct and recurrent evolution between *Hmr* and specific TEs.

We suggest several possible interpretations of these results. One is that *Hmr* and *Lhr* adaptive divergence is in fact driven largely or solely by *BS* and/or *Doc6*, a hypothesis that will require understanding the mechanism by which *Hmr* and *Lhr* affect expression of these TEs. Second is that *Hmr* and *Lhr* may be co-evolving with other genes, and that multiple diverged genes need to be replaced simultaneously in order to detect their effects on other TEs and satellite DNAs. Third is that more sensitive assays are needed, for example monitoring TE transposition rates over multiple generations. A fourth possibility is an alternative to genetic conflict scenarios that arises from population-genetic models. These models suggest that the fitness costs of individual TE families are likely extremely weak under most circumstances. The adaptive evolution of repressor proteins may therefore reflect the cumulative load of repeats within a genome (Lee and Langley 2012). This alternative view could be applicable to *Hmr* and *Lhr* since they repress a large number of TEs and satellites. Finally, *Hmr* and *Lhr* may have additional unidentified phenotypes that are the direct targets of adaptive evolution.

3.4.7 Repeat load, adaptation and hybrid incompatibilities

D. simulans has a smaller genome with ~4-fold less satellite DNA (Lohe and Roberts 1988; Bosco *et al.* 2007) and significantly fewer TEs (Dowsett and Young 1982; Lerat *et al.* 2011) compared to *D. melanogaster*. This large difference in repeat content between *D. melanogaster* and *D. simulans* may have wider consequences. We found reduced

expression from pericentric heterochromatin genes in *Hmr* and *Lhr* mutants in *D. melanogaster*. This reduction may reflect the fact that pericentric genes have evolved to use heterochromatin proteins such as Lhr and Hmr to maintain gene expression in a repeat-rich environment (Yasuhara and Wakimoto 2006) (Yasuhara & Wakimoto 2006). Pericentric genes in species with fewer repeats would presumably not require these proteins. Consistent with this model, we found that *Lhr* loss in *D. simulans* has a negligible impact on pericentric gene expression. This finding suggests that *Lhr* and *Hmr* have an adaptive role in blocking effects on gene expression arising from increasing repetitive DNA copy number.

If each genome is uniquely adapted to its repetitive DNA content, then the shock of hybridization may lead to misregulation of TEs and satellites. TEs are activated in various animal and plant hybrids but the consequences, if any, for hybrid fitness are largely unclear (Castillo and Moyle 2012). We found substantial TE misregulation in hybrid male larvae (Figure 3.15A). Since these hybrids are agametic (Kerkis 1933), this TE expression comes from somatic tissues. The fitness cost of this upregulation is unclear as somatic TE overexpression is not necessarily lethal within *D. melanogaster* (Ghildiyal *et al.* 2008; Kawamura *et al.* 2008) (Ghildiyal et al 2008; Kawamura et al 2008). Comparison of lethal Hmr^{+} and viable Hmr^{-} hybrid males demonstrates that lethal hybrids have more TE expression (Figure 3.15B) than the viable hybrids, which in turn have more TE expression than either of its parents. However, this TE misregulation seems unconnected with *Hmr* as the TEs differentially expressed between Hmr^{+} and Hmr^{-} hybrid male larvae are largely distinct from those between Hmr^{+} and Hmr^{-} *D.*

melanogaster male larvae. Further, while *Hmr*$^{-}$ causes rampant TE over-expression within *D. melanogaster*, it is associated with reduced TE levels in hybrids. These observations argue that the TE derepression in hybrids is unrelated to the pure species function of *Hmr*. This finding is consistent with previous genetic studies which demonstrate that the wild type *Hmr*$^{+}$ allele causes hybrid lethality and thus behaves as a gain-of-function allele in hybrids (Barbash *et al.* 2000; Orr and Irving 2000). More generally it underscores the unique nature of the hybrid genetic background (Maheshwari & Barbash 2011). Somatic TE overexpression may result from breakdown in the siRNA or piRNA pathways due to incompatibilities among multiple rapidly evolving TE regulators.

One clear example is known where a species-specific difference in a satellite DNA causes incompatibility between Drosophila species (Ferree and Barbash 2009). But the toll caused by heterochromatic differences may more commonly be indirect, as heterochromatin proteins diverge in response to changes in heterochromatic DNA repeats. Recent work suggests that hybrid female sterility may be caused by incompatibilities among rapidly evolving piRNA proteins rather than by species-specific differences in TEs (Kelleher *et al.* 2012). We suggest that the role of *Hmr* and *Lhr* in regulating the activity of three highly dynamic classes of heterochromatin has led to their recurrent adaptive evolution, and secondarily, to their involvement in interspecific hybrid lethality.

Acknowledgements

We would like to thank S. Parhad and Dr. William Theurkauf (U. Mass Medical Center) for being very generous with their expertise, time, and resources in allowing us to construct small RNA libraries at the Theurkauf lab. We thank Amanda Manfredo for assistance with construction of the larval hybrid RNA-seq libraries, Dr. Erin Kelleher and Dr. Ali Raza Awan for advice on analysis of RNA-seq data, and Carol Bayles for imaging assistance at the Cornell Microscope and Imaging Facility. We would also like to thank Dr. Heather Flores and Dr. Shamoni Maheshwari for ideas and helpful discussions, and Dr. Erin Kelleher and the anonymous reviewers for helpful comments on the manuscript.

CHAPTER 4

ANALYSIS OF THE MECHANISM OF HMR TE REPRESSION

4.1 Introduction

Transposable elements (TEs) have invaded the genomes of nearly all eukaryotes. In *Drosophila*, TE content accounts for 20% of the genome (Barrón *et al.* 2014), while the human genome is comprised of nearly 50% TE sequence (Lander *et al.* 2001). TEs are generally considered to be selfish genetic elements that strain the host genome and are rarely beneficial. Despite their highly mutagenic capability, TEs also provide a rich source of variation, thereby influencing genome evolution. Nonetheless, active TEs threaten genomic integrity due to their ability to disrupt coding sequence and promote ectopic recombination. Eukaryotic genomes have evolved small RNA-mediated pathways to control TE mobility, such as the PIWI-interacting (piRNA) pathway. The piRNA pathway is a well studied TE silencing mechanism that functions distinctly in the germline and the soma (Li *et al.* 2009; Malone *et al.* 2009).

The piRNA pathway exploits 26-30 nucleotide small RNAs (piRNAs), which are processed from piRNA clusters, to target TE transcripts for degradation. piRNA clusters are found primarily in heterochromatic regions of the genome and contain mostly degenerate TE sequences representing the majority of annotated elements (Brennecke *et al.* 2007). *D.melanogaster* contains 3 PIWI proteins (Piwi, Argonaute3 (AGO3), and Aubergine (Aub)) capable of dsRNA-dependent endonucleolytic activity (Saito *et al.* 2006; Gunawardane *et al.* 2007). While *piwi* is expressed in both the soma and

germline, *AGO3* and *aub* expression is restricted to the germline (Malone *et al.* 2009). Piwi and Aub bind antisense piRNAs to target and subsequently cleave TE transcripts, thereby generating sense piRNAs which are then bound by AGO3 and direct cleavage of piRNA cluster transcripts in a cycle known as the 'ping pong' amplification loop which is unique to the germline. Both *AGO3* and *aub* mutants exhibit decreased piRNA pools (Li *et al.* 2009; Malone *et al.* 2009). Germline *piwi* knockdown ovaries also exhibit a general reduction in the piRNA pool (Rozhkov *et al.* 2013).

While the 'ping pong' amplification loop exhibits post-transcriptional regulation of TEs, there is evidence for additional piRNA-mediated silencing at the transcriptional level. Nuclear localization of Piwi is required for TE silencing (Klenov *et al.* 2011) suggesting a possible role in chromatin modification. Several studies have implicated Piwi in heterochromatin-dependent silencing at TEs (Klenov *et al.* 2011) in a process that requires HP1 (Wang and Elgin 2011). Recent genome-wide studies have also supported a role for Piwi in establishing a repressive chromatin environment at TEs (Sienski *et al.* 2012; Le Thomas *et al.* 2013; Rozhkov *et al.* 2013), though there is some disagreement among studies. Le Thomas *et al*. (2013) showed that in *piwi* GLKD (germline knock down) ovaries, derepression of TE transcripts correlates with increased RNA Polymerase II (Pol II) occupancy at TE promoters. The authors argue that the observed Piwi-mediated transcriptional repression is likely attributed to the role of Piwi in directing histone H3 lysine 9 trimethylation (H3K9me3) (Le Thomas *et al.* 2013). Rozhkov *et al.* (2013) observed a similar correlation between nascent TE transcripts and transposon mRNAs in somatic *piwi* knockdown, but observed only a modest

increase in nascent transcripts relative to TE mRNAs in germline *piwi* knockdown ovaries, concluding that Piwi exerts multiple levels of regulation in the germline (Rozhkov *et al.* 2013).

A caveat of germline knockdown studies is that the RNAi response is not active in early-stage egg chambers (Rozhkov *et al.* 2013), however, *piwi* null mutants exhibit atrophied ovaries, which complicate interpretation. Klenov et al. (2014) exploited a *piwi* mutation that inhibits nuclear localization ($piwi^{Nt}$). The $piwi^{Nt}$ flies have morphologically normal ovaries while losing the ability to silence both somatic and germline TEs (Klenov *et al.* 2011). Klenov *et al.* (2014) profiled H3K4me2 occupancy, a mark of active transcription, in $piwi^{Nt}$ mutants and found a significant increase at several TE families, including the telomeric TEs, *HeT-A* and *TAHRE*. However, H3K9me3 and HP1a were only reduced in a subset of TE families, indicating that while the majority of TEs maintain their heterochromatic state in $piwi^{Nt}$ ovaries, nuclear piwi is required for complete repression (Klenov *et al.* 2014).

piRNA biogenesis and processing is a critical component of piRNA-mediated silencing. Several proteins have been implicated in the transcriptional regulation of dual-strand piRNA clusters, which are transcribed from both genomic strands and expressed primarily in the germline (Siomi *et al.* 2011). SETDB1/Eggless is a histone methyl-transferase that catalyzes methylation at H3K9 and is required for cluster transcription (Rangan *et al.* 2011). The Cutoff (Cuff) protein, which physically interacts with the HP1 homolog, Rhino (Rhi), is also required for piRNA cluster transcription (Pane *et al.* 2011).

In the absence of Rhi, only transcription from dual-strand clusters is abolished (Klattenhoff *et al.* 2009). A recent genome-wide study demonstrated that Rhi functions in a complex with Cuff and Deadlock (Del) (RDC complex) that is required for dual-strand piRNA production (Mohn *et al.* 2014). While Piwi does not mediate H3K9 methylation at the major dual-strand clusters, such as *cluster 42AB* (Klenov *et al.* 2014; Mohn *et al.* 2014), it is required to recruit the RDC to individual TE insertions which function as piRNA source loci (Mohn *et al.* 2014). In agreement, Klenov *et al.* (2014) did not find a significant reduction in H3K9me3 occupancy at the majority of analyzed clusters (including *cluster 42AB*) in $piwi^{Nt}$ ovaries, except for a modest reduction at clusters on chromosome 4.

Similar to *piwi* and other piRNA pathway genes, the hybrid incompatibility (HI) genes, *Hybrid male rescue (Hmr)* and *Lethal hybrid rescue (Lhr)*, regulate TE transcript levels in the female germline (Satyaki *et al.* 2014). *Hmr* interacts genetically with *Lhr* to cause lethality of *D. melanogaster/D. simulans* interspecific hybrid sons (Brideau *et al.* 2006). Both *Hmr* and *Lhr* encode heterochromatin-binding proteins and regulate the expression of a broad range of TE families. Notably, among the most highly derepressed TEs in both *Hmr* and *Lhr* mutants are the telomeric TEs, *HeT-A*, *TAHRE*, and *TART*. Interestingly, these TEs are also significantly upregulated in *piwi* mutants (Le Thomas *et al.* 2013; Rozhkov *et al.* 2013) leading us to propose a role for Hmr and Lhr in piRNA-mediated silencing.

Satyaki *et al.* 2014 probed the mechanism through which Lhr regulates TEs by assaying the small RNA pool in *Lhr* mutants but found no appreciable effect on piRNA levels compared to Lhr^{+}. This result is in stark contrast with the dramatic reduction in piRNA levels observed in both *aub* and *spn-E* mutants (Malone *et al.* 2009). Because Hmr and Lhr physically interact in *D. melanogaster* (Satyaki *et al.* 2014), we expect that Hmr and Lhr participate in a shared mechanism to regulate TEs and therefore consider it unlikely that Hmr functions in piRNA production. Here I test whether Hmr regulates TEs at the transcriptional level by comparing H3K9me3 enrichment and RNA Polyermase II (Pol II) occupancy genome-wide between wild-type and Hmr^{3} mutant ovaries (Figure 4.1). I find that Hmr does not have a significant effect on H3K9me3 deposition or RNA Pol II occupancy at TEs and conclude that the major mode of TE regulation elicited by Hmr is post-transcriptional. I analyze existing ChIP-seq data sets of piRNA pathway mutants that are known to transcriptionally regulate TEs to confirm my conclusion. Finally, I find few significant differences in the genome-wide pattern of H3K9me3 enrichment, with the striking exception of the pericentromeric region of the X chromosome, which exhibits decreased H3K9me3 in *Hmr* mutants.

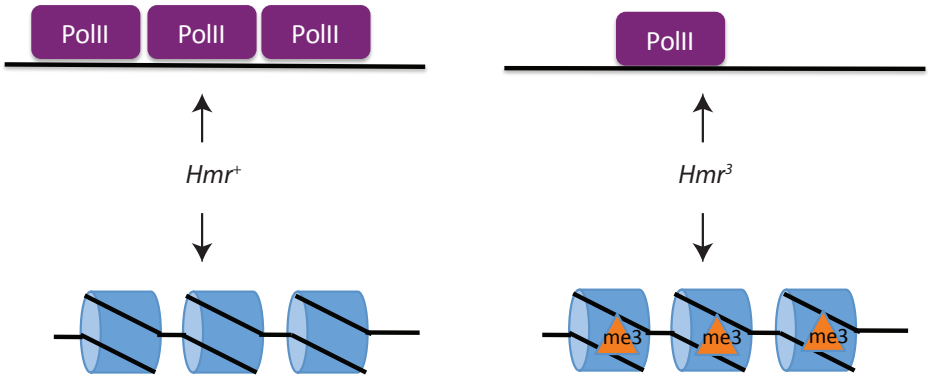


Figure 4.1. Schematic of expected result if Hmr silences TEs transcriptionally. If Hmr acts transcriptionally to silence TEs, we would expect to see increased occupancy in the *Hmr*^{3} background of Pol II at those TEs that are significantly derepressed in *Hmr* mutants. Likewise, if Hmr acts to affect chromatin structure, we expect a decrease of H3K9me3 deposition in the *Hmr*^{3} background at Hmr-regulated TEs.

4.2 Materials and Methods

4.2.1 Fly stocks

Stock *y[1] w[67c23] P{w[+mC] y[+mDint2]=EPgy2}EY12237* known as *Hmr*^{3}, was outcrossed to *y w F10* for 6 generations in April 2012. Sibling crosses were done to generate a homozygous matched line called *y w Hmr*^{3}. Two biological replicates of each ChIP-seq were performed.

4.2.2 ChIP

One hundred ovary pairs were dissected in ice-cold 1X PBS from flies aged 5-6 days at room temperature, with the final 2 days being supplemented with yeast paste. Ovaries were cross-linked in 1.8% paraformaldehyde/1X PBS for 10 min at room temperature, quenched with 125 mM glycine for 5 min, and washed three times in 1X PBS. Ovaries were lysed in sonication buffer (20 mM Tris pH8.0, 2 mM EDTA, 0.5 mM EGTA, 1X Roche Complete P.I. Tab, 0.5% SDS) using a motor-driven pestle. A Bioruptor (Diagenode) was used to shear the DNA to the desired 300-500 bp range using the following settings: 2 cycles at 15 min each at “High” with 20 sec ON/1 min OFF pulses. The resulting extract was spun at 13,000 rpm for 10 min at 4°C and then diluted with IP buffer (0.5% Trition X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol) to 0.05% SDS. For each ChIP, 200 μl of Protein A agarose beads (Millipore 16-125) were washed three times in IP buffer. The equilibrated beads were then added

to the diluted extract (100 μl beads + 900 μl chromatin) and incubated for 2 hours at 4°C. The beads were spun for 1 min at 1000 rpm and the supernatant transferred to a new tube. For the IP western and ChIP-qpcr, anti-Rabbit H3K9me3 (abcam ab1220, 1:100) and a 1:10 dilution of anti-Rabbit Rpb3 (gift from John T. Lis lab) were used. For the ChIP-seq experiment, 5 μg of anti-Rabbit H3K9me3 antibody (abcam ab8898) or 10 μl anti-Rabbit Rpb3 serum (gift from John T. Lis lab) were added to the respective samples and incubated over night at 4°C. We switched to the anti-Rabbit H3K9me3 antibody (abcam ab889) because it was previously shown to be used successfully for ChIP (Rozhkov *et al.* 2013). The chromatin-immunoprecipitated samples were subsequently washed in a series of ice-cold buffers: low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl), high salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 500 mM NaCl), LiCl (10 mM Tris pH 8.0, 0.25 M LiCl, 1% NP40, 1% NaDeoxycholate), and then rinsed two times in TE buffer (10 mM Tris pH8.0, 1 mM EDTA), followed by treatment with 50 μg RNase A. The DNA-protein-antibody complexes were eluted from the beads by incubating two times for 15 min in elution buffer (1% SDS, 0.1 M $NaHCO_3$) at room temperature. Cross-links were reversed by heating at 65°C. The samples were treated with 40 μg Proteinase K for 2 hours at 37 °C.

4.2.3 IP western

Following the pull-down, the DNA-protein-antibody complexes were washed and 0.5X volume of 2X SDS/PAGE sample buffer was added. The samples were boiled for 5m

and approximately ~3.5 ovary pairs were run of each sample. Primary antibodies: monoclonal rabbit anti-H3K9me3 (abcam ab1220; 1:1000), polyclonal rabbit anti-Pol II Rpb3 (gift from John T. Lis lab, 1:1000). Secondary antibody: HRP conjugated goat anti-rabbit secondary antibody (1:20,000). Detection was achieved with the ECL Western blotting substrate (Pierce).

4.2.4 ChIP-qPCR

5 μl of each ChIP DNA and input (1:100 dilution) were used as templates. *HeT-A* was amplified using Het-s2 and Het-as2 which anneal to the CDS (Klenov *et al.* 2007) and normalized to the 60D region amplified by the primers used in Klenov *et al.* 2007. *Rp49* was amplified using primers rp49 s2 and rp49 as2 (Klenov *et al.* 2007) and normalized to 60D.

4.2.5 ChIP-Seq

ChIP-DNA was quantified using the QuBit. The Pol II Rpb3 ChIP DNA yields ranged from 6.6 ng-14 ng, the H3K9me3 ChIP DNA yields ranged from 11 ng-19 ng, and the input DNA yielded ~3.4 μg for each genotype. Libraries were subsequently prepared in house using the Illumina TruSeq Nano DNA Sample prep kit according to the protocol with the following revisions. Approximately 1 ng of ChIP DNA was used from each sample to prepare libraries while 100 ng of each input sample was used. After adapter ligation, one half of the supernatant was used in the PCR reaction to enrich for adapter-ligated molecules. Eight cycles were used during the enrichment step for all samples in

order to obtain the required amount for sequencing (1 ng/μl). In the case of the Pol II Rpb3 ChIP DNA and one replicate of *Hmr*^{3} H3K9me3 ChIP DNA, the remaining half supernatant was used in a subsequent PCR with 12 cycles and the resulting library was then combined with the previous one. Sequencing was performed by Cornell University Institute of Biotechnology using a single lane of the HiSeq 2000 and generating 100 bp single end reads.

4.2.6 Sequencing data preprocessing

Read quality was assessed using FastQC (http://www.bioinformatics.bbsrc .ac.uk/projects/fastqc/). The number of reads obtained per sample ranged from 9,046,326 to 22,264,529 for the input, with an average of 16,572,672 reads per sample. The SLIDINGWINDOW function of Trimmomatic 0.32 (Bolger *et al.* 2014) was used to trim reads at a base quality threshold of 20 and then filter out reads less than 40 bp (Table 4.1).

4.2.7 ChIP-seq data analysis

TE analysis

For all data sets, reads were mapped to a database containing TE consensus sequences downloaded from RepBase and then modified in house by Michael McGurk (file Repbase_19.06_DM_Curated_7-14-14). Detailed notes of the modifications are listed in the accompanying ReadMe.txt. Alignments were performed using Bowtie2 (2.2.3) (Langmead and Salzberg 2012). All reads that did not map to the TE consensus database (Table 4.1) were then mapped to Dmel Release 6.01 downloaded from the UCSC genome browser (Table 4.1). TE alignments with mapping quality less than 20 were removed using Samtools. Coverage plots of read depth along TE length were obtained using a custom python script written by Michael McGurk. For the Pol II ChIPs, read counts to genes were calculated using HT-Seq (Anders *et al.* 2015) and the reflat annotation file downloaded from the UCSC table browser. For the H3K9me3 ChIPs, the genome was binned into 1000 bp windows and the corresponding read counts were calculated for each window. This is a more accurate normalization for H3K9me3 because, unlike RNA Pol II, it is not enriched at genes.

The two genotypes (*y w* Hmr^{3} and *y w* Hmr^{+}) were outcrossed for six generations in 2012, and the experiments presented here were performed in the summer of 2014. Therefore, we had reason to expect some amount of divergence in TE copy number. In order to take this into account, we normalized read counts by TE copy number. We used reads from the input Hmr^{+} and input Hmr^{3} samples to estimate copy number of both TEs and genes using the following formula:

$$\text{Copy Number} = \frac{\text{Number of reads mapping to TE family or gene or bin}}{\text{Average input read depth} \times \text{TE length or gene length or bin size}}$$

For the Pol II Rpb3 ChIP samples, copy number of TE families and genes were calculated for each genotype, *Hmr*$^{+}$ and *Hmr*^{3}. Total copy number for each genotype was obtained by combining the respective files. For the H3K9me3 ChIP samples, copy number of TE families and number of reads mapping to 1000 bp bins were calculated for each genotype and total copy number was then generated by combining the respective files for each genotype. The read counts for each ChIP DNA sample were then normalized by the copy number of each TE family or gene or bin from the corresponding genotype. The resulting normalized read counts were entered into DESeq 1.16.0(Anders and Huber 2010) and size factors were estimated. Differential enrichment of TEs was determined based on adjusted p-values < 0.05 and fold change greater than 1.5.

Table 4.1. Read and mapping information.

Sample	Total # raw reads	# reads > 40 bp	# reads mapping to TE db 1 time	# reads mapping to TE db > 1 times	% reads mapping to TEs	% TE alignments with MQ > 20
Input wt	30,532,218	22,319,773	1,679,085	148,793	8.19%	80.39%
Input *Hmr*^{3}	30,748,832	21,989,841	1,749,340	161,175	8.69%	80.54%
H3K9me3 wt R1	12,433,741	9,046,326	2,437,294	180,131	28.93%	80.83%
H3K9me3 *Hmr*^{3} R1	24,271,000	17,192,678	4,816,207	371,316	30.17%	80.76%
H3K9me3 wt R3	14,779,885	10,754,383	3,260,699	231,748	32.47%	81.18%
H3K9me3 *Hmr*^{3} R3	31,678,014	22,264,529	6,744,944	496,534	32.52%	81.34%
Pol II wt R1	28,685,162	19,887,111	1,174,093	109,921	6.46%	80.75%
Pol II *Hmr*^{3} R3	29,786,980	21,180,342	1,353,748	161,352	7.15%	78.99%
Pol II wt R1	24,599,721	16,458,265	933,255	96,262	6.26%	80.08%
Pol II *Hmr*^{3} R3	22,715,517	15,797,743	906,585	116,875	6.48%	78.52%

Genome analysis

Genome alignments (described above) were filtered at a mapping quality threshold of 20. The remaining alignments were used as input for MACS2 2.1.0 (Zhang *et al.* 2008). A minimum FDR cutoff of 0.01 was employed for all datasets. For H3K9me3, the broad peak option was used. To identify differentially bound peaks, I used the R package DiffBind (3.1) (Ross-Innes *et al.* 2012). Peaks were classified as differentially bound if the absolute fold difference was greater than 1.5 at FDR < 0.1. Peaks were classified as heterochromatic or euchromatic by translating the coordinates from (Riddle *et al.* 2011) to Dmel Release 6.01 using the coordinates converter on FlyBase.

4.2.8 RNA-seq data re-analysis

RNA-seq data from *Hmr* transheterozygous mutants and *Hmr*$^{+}$ from Satyaki *et al.* (2014) was re-analyzed to allow for comparison between datasets. Raw reads were processed as described in (Satyaki *et al.* 2014). Reads were aligned to TE consensus sequences using Bowtie2 (Langmead and Salzberg 2012). Reads with mapping quality < 20 were discarded. Unmapped reads were then aligned to Dmel Release 6.01 using Tophat (Trapnell *et al.* 2009). Gene counts were obtained using HT-seq as described above (Anders *et al.* 2015). Count files for each sample containing both TE and gene counts were used as input in DESeq (Anders and Huber 2010) to determine differential expression.

4.3 Results

4.3.1 Pol II Rpb3 and H3K9me3 pull down efficiencies

To verify that the antibodies pulled down their respective targets, I performed IP-westerns. There was a distinct band at the expected size of 33 kD for Pol II Rpb3 IP and the input, which served as a positive control (Figure 4.2A). There was a faint band in the H3K9me3 IP compared to input (Figure 4.2B) at the expected size of 17 kD. I subsequently switched to the abcam H3K9me3 ChIP-grade antibody for the ChIP-seq. These tests confirmed that the immunoprecipitation and pull down conditions were working. In order to test Pol II and H3K9me3 IP efficiency, I performed ChIP-qpcr. I used *rp49* as a positive control for the Pol II ChIP because it is ubiquitously and highly expressed. Pol II Rpb3 efficiency was determined by normalizing the amount of *rp49* in the IP samples to the amount in the input, representing the amount of *rp49* pulled down (Figure 4.3A). Pull-down efficiency between the samples ranged from 0.4% to 5%. Previous Pol II Rpb3 efficiencies obtained in our lab were ~1%. For comparison, the pull down efficiency of an intergenic spacer, known as 60D, was also calculated (Figure 4.3A). Finally, *rp49* enrichment was calculated by normalizing over 60D to control for noise (Figure 4.3B). The amount of enrichment ranged from 12-47, indicating substantial variability among samples. There is no significant difference in fold enrichment between Hmr^{+} and Hmr^{3} (two-tailed *t*-test, not significant), however, this represents a single biological replicate. I used *HeT-A* as a positive control for H3K9me3 ChIP because it is enriched for this mark. H3K9me3 pull-down efficiency was stronger

than what I observed for Pol II, however, because of a stronger 60D signal, overall enrichment of *HeT-A* was comparatively lower (Figure 4.3C,D). The enrichment for H3K9me3 relative to 60D is comparable to the enrichment seen in (Klenov *et al.* 2007).

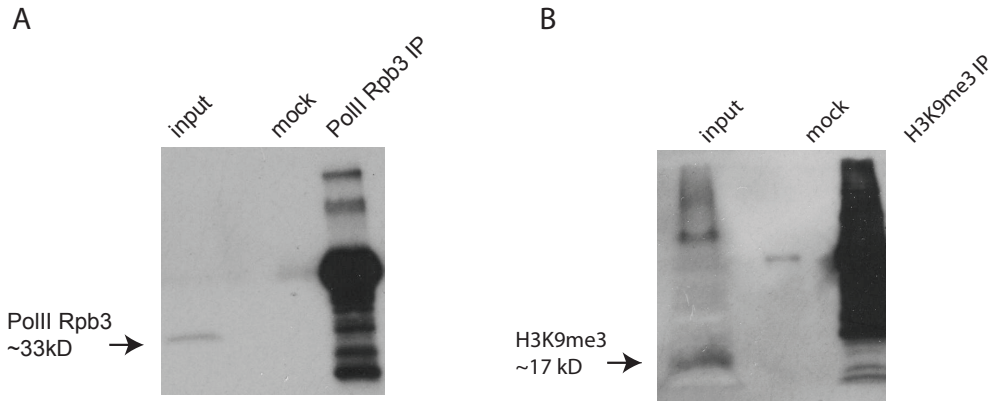


Figure 4.2. IP westerns showing Pol II Rpb3 and H3K9me3 antibody specificity to targets. Approximately 4 ovary pairs dissected from *y w Hmr*$^{+}$ are loaded. The mock is shown as a negative control for each IP. The overexposed bands are the heavy (50 kD) and light (23 kD) chains of the rabbit IgG antibody. (A) Pol II Rpb3 IP western shows a clear band in the IP sample corresponding to the observed band in the input control. (B) H3K9me3 IP western shows a faint band in the IP sample corresponding to the input control.

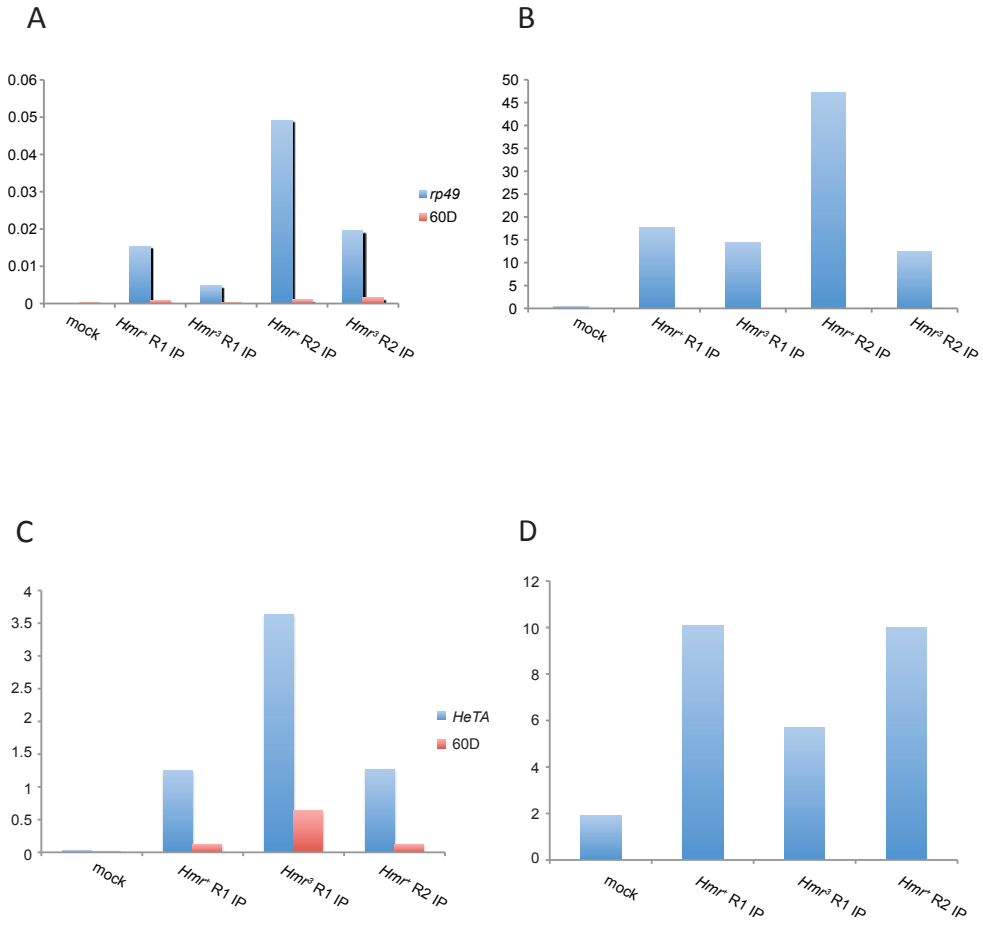


Figure 4.3. Pol II ChIP-qpcr to determine pull-down efficiency. Two biological replicates of each genotype were performed. (A, B) Pol II ChIP-qpcr. (C, D) H3K9me3 ChIP-qpcr. Pull-down efficiency of *rp49* (blue) was determined by normalizing the amount of *rp49* in the IP samples to input. The 60D intergenic spacer was included as a control for noise and its pull-down efficiency is also shown (red) (A). The mock was included as a negative control. *rp49* enrichment in each IP sample was determined by normalizing the amount of *rp49* in each IP sample to the amount of 60D in the corresponding sample (B). Pull-down efficiency of *HeT-A* (blue) was determined by normalizing the amount of *HeT-A* in the IP samples to input. The 60D intergenic spacer was included as a control for background and the mock as a negative control (C). *HeT-A* enrichment in each IP sample was determined by normalizing the amount of *HeT-A* in each IP sample to the amount of 60D in the corresponding sample (D).

4.3.2 ChIP-seq data quality

Read reproducibility

Two biological replicates of each ChIP-seq experiment were performed. I first examined the reproducibility of the reads between replicates. The genome was divided in 500 bp bins and coverage of each bin was calculated for all samples. The Pearson correlation coefficient (PCC) was computed to determine the correlation of mapped read counts between replicates. I found PCC values > 0.9 across all replicates, indicating high read reproducibility (Figure 4.4). I also determined correlations between non-replicates (data not shown). The correlation of mapping reads between the H3K9me3 and Pol II ChIP experiments ranged from 0.0387434 to 0.09442091 for all pairwise combinations between samples, indicating that the pull-downs were specific. For the H3K9me3 ChIP, correlations between genotypes (Hmr^{+} vs. Hmr^{3}) ranged from 0.7550587 to 0.8843369, suggesting that Hmr has some effect on H3K9me3 deposition. Finally, for the Pol II ChIP, correlations between genotypes were comparable to that observed between replicates (0.8886091-0.955921) (data not shown).

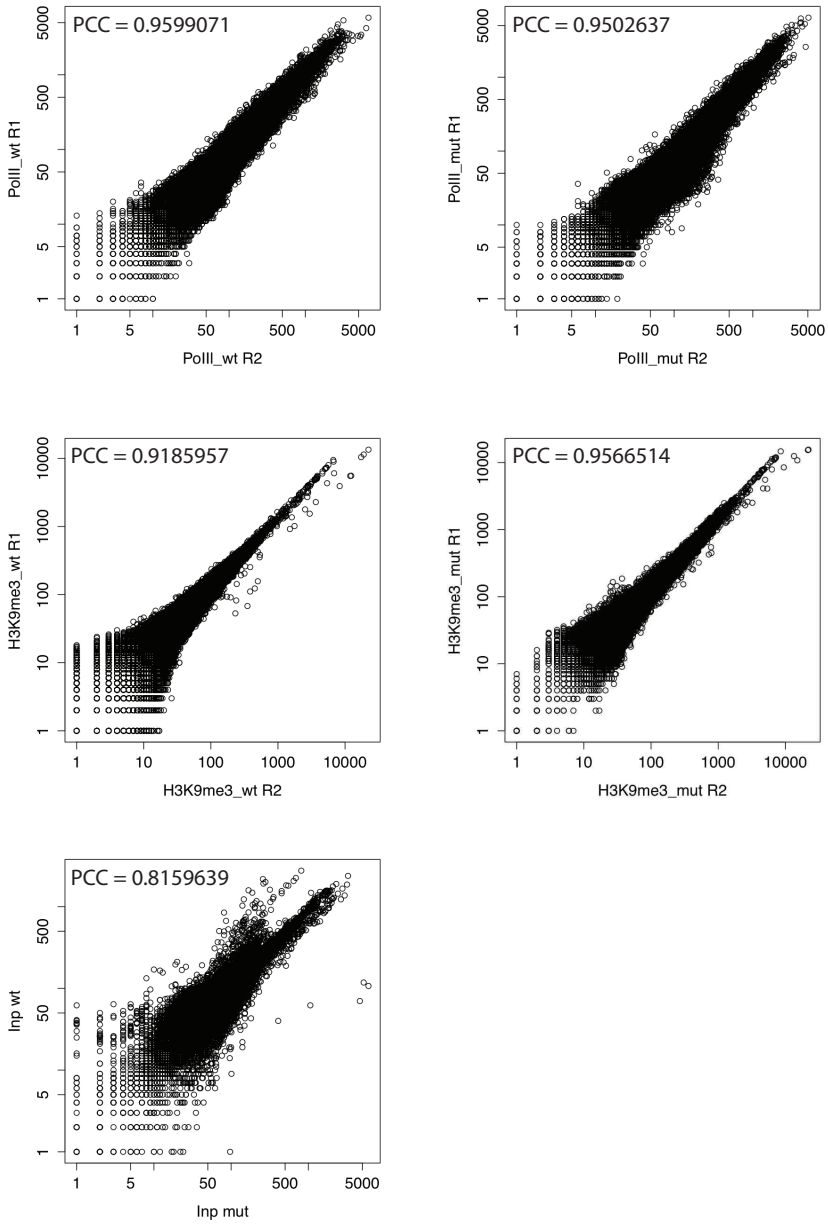


Figure 4.4. Reads between replicates are highly correlated. Pearson's correlation coefficient of mapped read counts at 500 bp bins was calculated between replicate samples. All ChIP samples showed high correlation with PCC > 0.9, indicating high read reproducibility. PCC between the *Hmr*$^{+}$ and *Hmr*^{3} input samples was also computed. As expected, this correlation is lower because of the divergence in the genomic backgrounds.

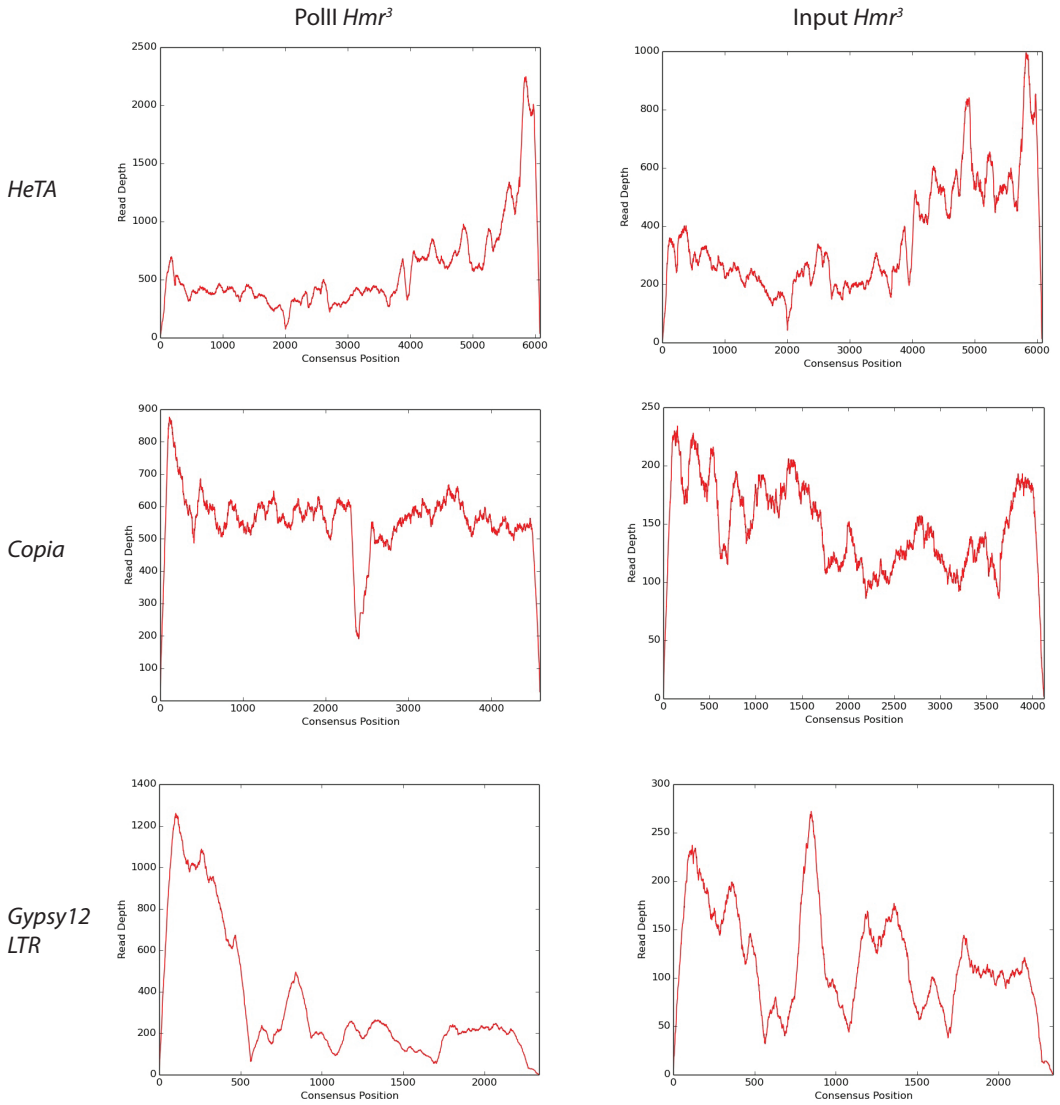


Figure 4.5. Coverage plots demonstrate reads map across TE lengths. Shown are representative TE coverage plots with read depth on the y-axis and consensus TE length on the x-axis. The majority of TEs display reads mapping throughout the TE body, indicating the presence of full-length elements, in contrast to fragments.

Coverage plots

Read coverage along TE lengths was assessed in order to verify that reads mapping to TEs are not primarily mapping to TE fragments. With very few exceptions, read depth was consistent across TE lengths in all our samples. Figure 4.5 shows example coverage plots for one biological replicate of *Hmr3* Pol II Rpb3 ChIP DNA and the corresponding input.

H3K9me3 enrichment at TEs

In order to confirm successful pull-down of H3K9me3, I tested for enrichment at TEs. I mapped ChIP-seq DNA reads and the corresponding input control to a database containing TE consensus sequences in *D. melanogaster* and then calculated RPM for each sample. H3K9me3 is clearly enriched at TEs in contrast to Pol II ChIP-seq DNA, which does not show enrichment at transposons, conforming to expectation (Figure 4.6A). I examined all transposons for an inverse relationship between H3K9me3 and Pol II fold change due to Hmr^3, but did not observe one (Figure 4.6B).

4.3.3 Hmr^3 and Hmr^+ backgrounds have diverged

We mapped input control reads from Hmr^3 and Hmr^+ to the TE consensus database and used DESeq to assess whether the two backgrounds have diverged in TE copy number. DESeq found a small number of TE families that were overrepresented in Hmr^3 compared to Hmr^+ (Figure 4.7A), suggesting that despite having been outcrossed for 6 generations, these strains have indeed diverged in copy number with respect to a

specific subset of TEs. Notably, the telomeric TEs *HeT-A, TART-A*, and *TAHRE* are present at higher copy numbers in the *Hmr* mutant.

This result led us to refine our differential enrichment analysis at TEs. To take TE copy number variation into account, we normalized the ChIP-seq read counts to TEs by TE copy number. TE copy number between the two genotypes is highly correlated (r_s = 0.9821426) with a few notable exceptions mentioned above (Figure 4.7B). We also calculated gene copy number in each genotype and found it to be highly correlated as well, though surprisingly less so than TE copy number (Figure 4.7C). The higher correlation for TE copy number is likely due to the lower copy number of genes (i.e. 2), which causes stronger sampling effects.

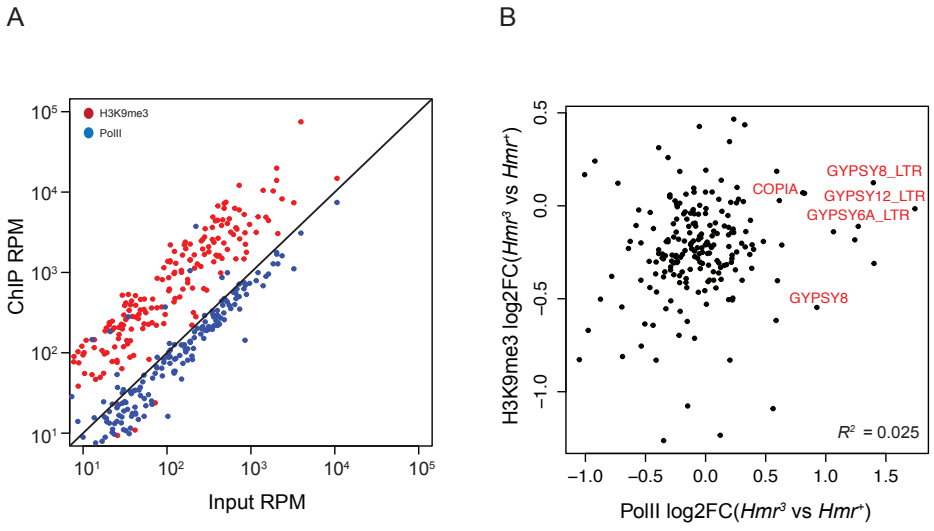


Figure 4.6. High enrichment of H3K9me3 at TEs in *Hmr*$^+$. (A) Each dot corresponds to a TE consensus sequence. H3K9me3 shows clear enrichment at TEs (red) compared to Pol II (blue). ChIP RPM calculated from respective *Hmr*$^+$ samples for each ChIP. Input RPM reflects *Hmr*$^+$ reads. (B) There is no inverse relationship between the differential fold enrichment of Pol II and H3K9me3 due to the *Hmr* mutation (R^2 = 0.025). TEs identified in red exhibit a significant increase in Pol II occupancy in the *Hmr* mutant.

4.3.4 The major mode of Hmr TE regulation is not transcriptional

I found only a handful of TEs with a significant change in Pol II occupancy due to the *Hmr* mutation including *gypsy 8*, *copia*, the telomeric TEs *TART-A* and *HeT-A*, and the *gypsy 12* and *gypsy 6* LTRs (Figure 4.8A). I re-analyzed the RNA-seq data from Satyaki *et. al* 2014 using the updated TE consensus database and refined the normalization procedure to better account for library size (see RNA-seq data analysis). My results largely agreed with the previous analysis and confirmed significant changes in fold expression for *copia* and the telomeric TEs *HeT-A* and *TART-A*, and the *gyspy 12* LTR in *Hmr* mutants relative to wildtype controls (Satyaki *et al.* 2014). Surprisingly, three TE families exhibited a decrease in Pol II occupancy in *Hmr* mutants (*GTWIN*, *R2*, and *gypsy 5*). None of these TEs were detected as having a significant change in steady-state RNA levels. I found only a very weak correlation (Figure 4.9A) between the change in TE transcript abundance and the change in Pol II occupancy in *Hmr* mutants (ρ = 0.1844975, p-value = 0.01326).

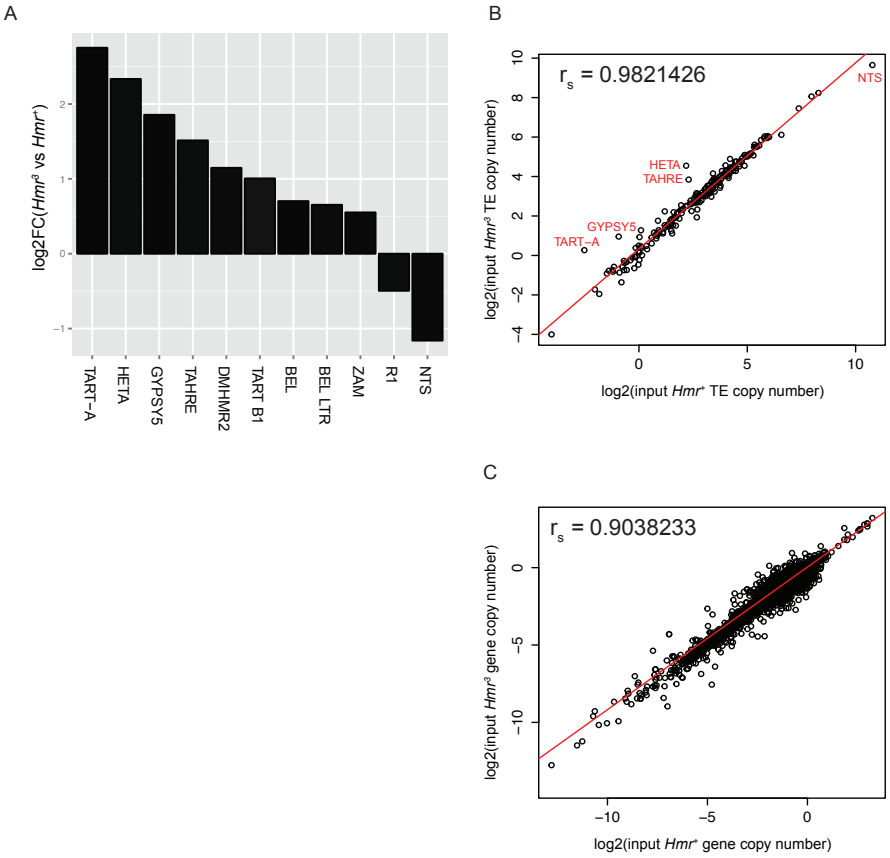


Figure 4.7. TE and gene copy numbers are highly correlated between genotypes. (A) To determine whether there was divergence in TE copy number between *Hmr*^{3} and *Hmr*$^{+}$, I used DESeq to evaluate whether there is a significant difference in the number of reads mapping to TEs between the two input samples. A handful of TEs are present in *Hmr*^{3} at higher copy number than in *Hmr*$^{+}$, including *HeT-A, TART-A,* and *TAHRE.* (B) Spearman's correlation coefficient shows that with very few exceptions, notably the telomeric TEs, TE copy number is very similar between *Hmr*^{3} and *Hmr*$^{+}$. (C) Spearman's correlation shows that gene copy number is similar in both genotypes, as expected.

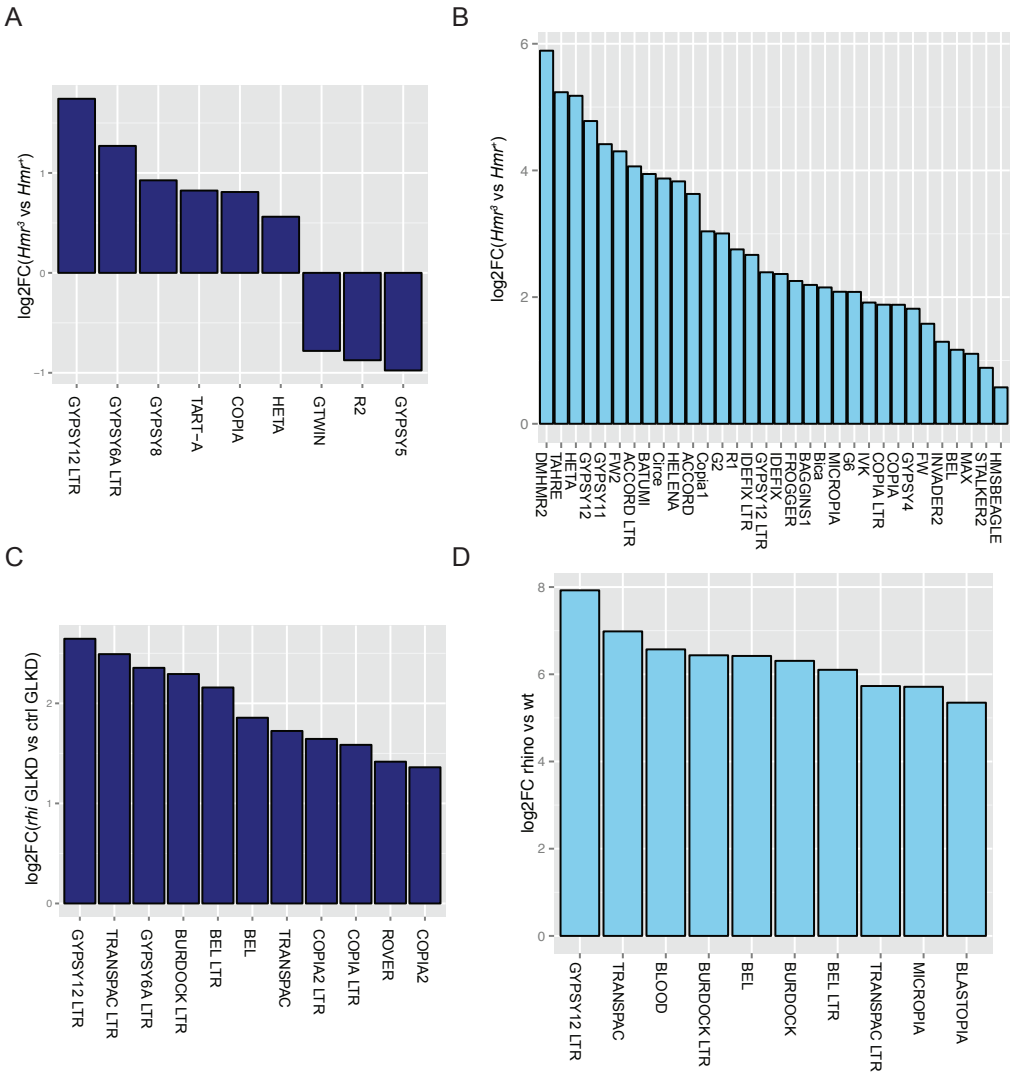


Figure 4.8. Hmr does not transcriptionally regulate the majority of TEs. Reads were mapped to the TE consensus database. The number of reads mapping to each family was calculated and transformed to account for copy number differences. Enrichment was calculated by normalizing to input. Differential enrichment was analyzed using DESeq. (A) A subset of TEs showed increased occupancy of Pol II in the *Hmr* mutant relative to the wild-type control. TEs were considered significant if adjusted p-value < 0.01. (B) Several TE families are upregulated in transheterozygous *Hmr* mutants relative to a heterozygous control. Shown are TEs upregulated at least 1.5 fold. (C) and (D) Data from (Mohn *et al.* 2014) (C) A subset of TE families display increased Pol II occupancy in *rhi* mutants relative to controls. (D) *gypsy 12 LTR*, *transpac*, *burdock LTR*, *bel*, and the *bel LTR* exhibit increased expression as well as increased Pol II occupancy in *rhi* mutants.

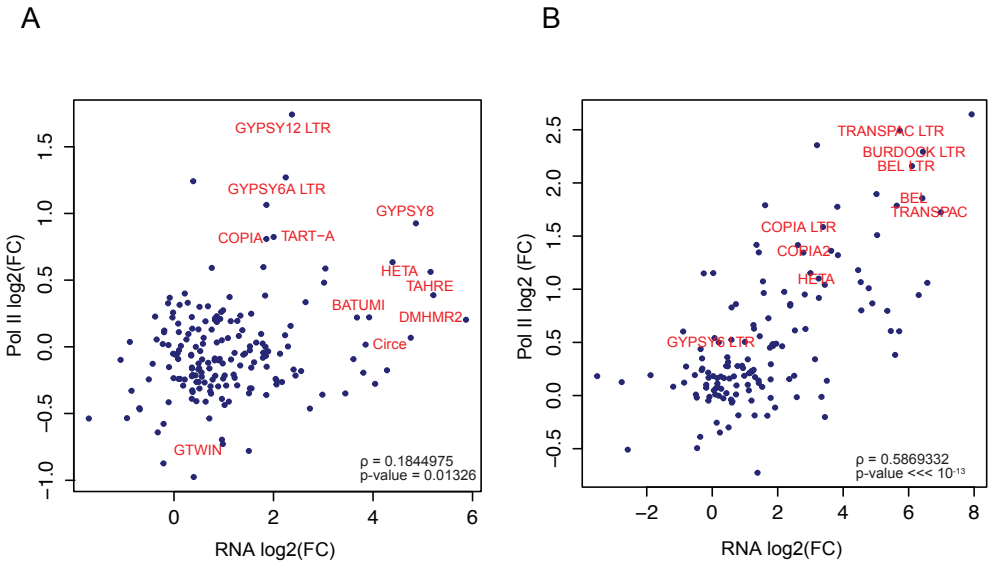


Figure 4.9. Correlation between Pol II occupancy and RNA expression. (A) There is no correlation between change in TE transcript abundance and change in Pol II occupancy in *Hmr* mutants relative to wild-type control. (B) There is a strong positive correlation between change in TE transcript abundance and change in Pol II occupancy in *rhi* mutants, indicative of TGS.

The most definitive way to determine transpositional activity is to observe new insertions, however, this is challenging to measure. One approach is to estimate the population frequencies of individual insertions (González *et al.* 2008; Petrov *et al.* 2011; Kofler *et al.* 2012). The mean population frequency of individual elements of an active family is expected to be low because recent insertions will be rare. The c*opia* element is known to be transpositionally active and is present at a low population frequency (less than 0.2) in a Portuguese population of *D. melanogaster* characterized in (Kofler *et al.* 2012). Both *gypsy 8* and *gypsy* 12 are near fixation in the Portuguese population (Kofler *et al.* 2012), suggesting they are no longer transpositionally active.

In order to better understand the role of Hmr in TE transcriptional silencing, I re-analyzed available Pol II ChIP-seq data sets from piRNA mutants. Depletion of *rhino* (*rhi*) causes loss of piRNAs originating from dual-strand clusters and an increase in TE steady-state RNA levels (Klattenhoff *et al.* 2009). Recently Mohn *et al.* (2014) showed that Rhi functions in a complex with Cutoff (Cuff) and Deadlock (Del) to promote transcription from dual-strand piRNA clusters. Given this function of Rhi, I would expect increased occupancy of Pol II at TEs upon *rhi* depletion at TEs that are in Rhi-dependent clusters. I observed an increase in RNA Pol II at 8 TE families, the majority of which are LTRs (Figure 4.8C). In contrast to *Hmr*, the telomeric TEs did not exhibit a significant fold change in RNA Pol II occupancy in *rhi* GLKD ovaries. The *gypsy 12* and *gypsy 6* LTRs displayed the greatest fold change between *rhi* GLKD and the control. Because these elements are likely inactive, this result is perplexing. However, Mohn *et al.* (2014) found that *rhi* depletion also resulted in the loss of *gypsy 12* piRNAs,

indicating that this element is still being targeted by piRNAs. Two different families of *copia* also exhibited significant fold changes. *Transpac*, *burdock*, and *rover* have population frequencies less than 10% (Kofler *et al.* 2012; Kelleher and Barbash 2013) and are therefore also likely transpositionally active. This is consistent with the finding that the majority of the derepressed TEs in *rhi* mutants are present at low frequency (Kelleher and Barbash 2013). A subset of these TEs is significantly derepressed upon *rhi* depletion (Figure 4.8D) and there is a strong positive correlation (ρ = 0.5869332, p-value < 10^{-13}) between the change in transcript abundance and RNA Pol II occupancy, confirming that *rhi* acts at the transcriptional level (Figure 4.9B). I also performed my analysis on data from germline *piwi* knockdown ovaries presented in Le Thomas *et al.* (2013), but found no significant changes in Pol II occupancy. Le Thomas *et al.* (2013) reported modest fold changes in a number of TEs, including *HeT-A*, *TART*, and the *gypsy 12 LTR*. This discrepancy is likely a result of different analysis pipelines. They normalized for library size by calculating RPM and RPKM values and accounted for potential differences in pull-down efficiency between samples by comparing RNA Pol II levels at TSS between the samples to derive a normalization factor.

4.3.5 Hmr does not modify chromatin structure at TEs

Only three TE families displayed decreased H3K9me3 enrichment in *Hmr* mutant ovaries: *gypsy 2, HetRP,* and *HeT-A* (Figure 4.10A). Reads did not map across the whole length of *HetRP*, suggesting that this signal does not arise from full-length elements. *Gypsy 2* has a mean population frequency less than 0.2 (Kofler *et al.* 2012), indicating it is possibly active. There was no significant anti-correlation (ρ = -0.04039713,

p-value = 0.591) between steady-state RNA levels and H3K9me3 enrichment in *Hmr* mutants (Figure 4.10C).

I re-analyzed H3K9me3 ChIP-seq data from *piwi*^{Nt} ovaries from Klenov *et al.* (2014) and found several TE families exhibiting significant changes in H3K9me3 enrichment (Figure 4.10C). The majority displayed a decrease of H3K9me3 enrichment in *piwi*^{Nt}. The DNA transposon *Helitron,* which replicates using a distinctive rolling circle mechanism (Kapitonov and Jurka 2001), exhibited the largest fold change with an 8 fold reduction in H3K9me3 in *piwi* mutants. *HeT-A*, *gypsy 2*, and the *gypsy 2* LTR had a similar fold change to what was observed in *Hmr* mutants. *Tirant, mdg3*, *copia*, *blood*, and *412* also had significant fold changes and are all potentially active based on population frequencies < 10% (Kofler *et al.* 2012; Kelleher and Barbash 2013). Seven TE families had an increase in H3K9me3 enrichment due to loss of nuclear piwi (Figure 4.10C). Klenov *et al.* (2014) reported a number of TEs with modest differences in H3K9me3 enrichment, including *gypsy*, *412, HeT-A*, and *TAHRE,* which showed increased enrichment in *piwi* mutants. However, it is unclear how they determined significance. We also re-analyzed H3K9me3 ChIP-seq data from *piwi* GLKD ovaries from Mohn *et al.* (2014), but found no signficant fold changes compared to the control KD. It has been noted that the RNAi response does not turn on at the onset of oogenesis and that residual Piwi may be present in early stage egg chambers (Rozhkov *et al.* 2013). Rozhkov *et al.* (2013) also performed H3K9me3 ChIP-seq on *piwi* GLKD ovaries and found significant changes in the telomeric TEs as well as *jockey.* Mohn *et al.* (2014) do

not specify the stock number for the shRNA *piwi* line, but it is presumably the same as used in Rozkhov *et al.* (2013) (VDRC 101658).

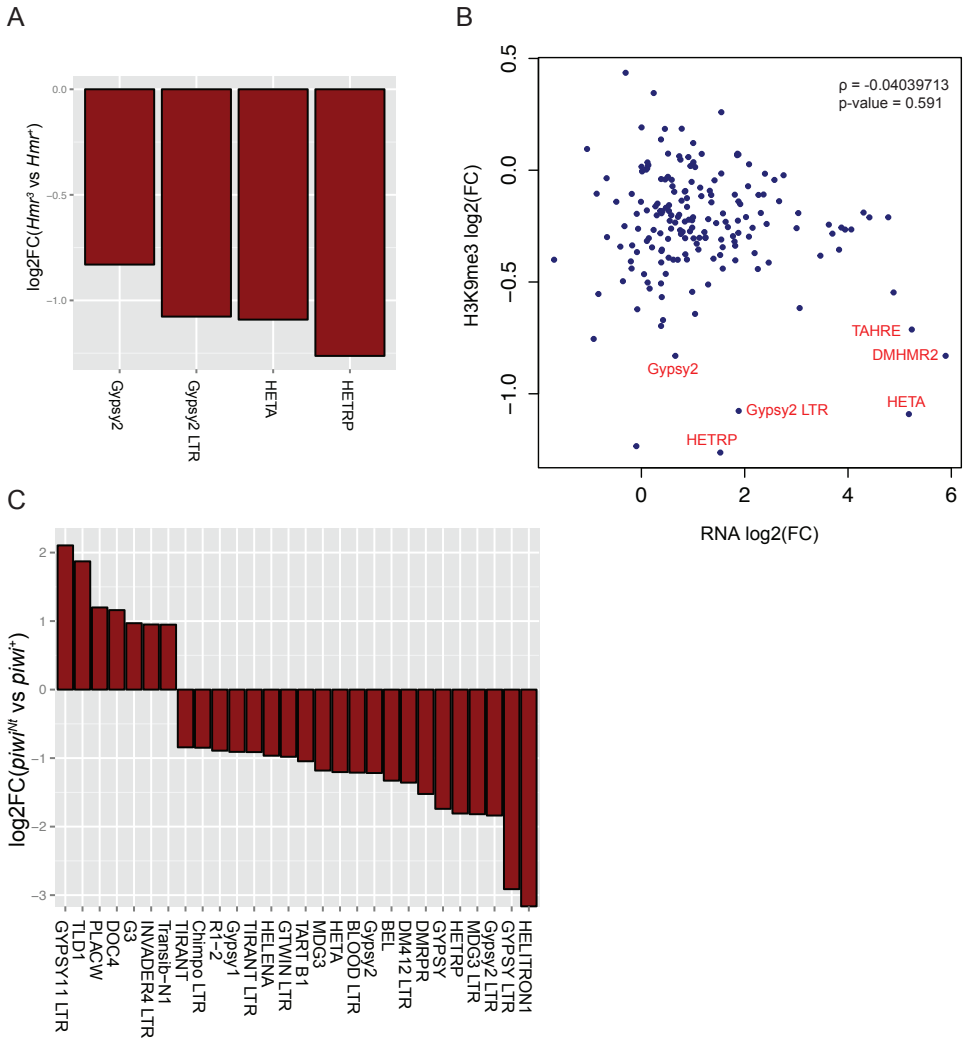


Figure 4.10. Hmr does not modify chromatin structure at most TEs. (A) Few TEs showed differential enrichment due to loss of *Hmr*. TEs were considered significant if adjusted p-value < 0.01. (B) There is no correlation between change in TE transcript abundance and change in fold enrichment due to Hmr. Telomeric TEs as well as TEs showing H3K9me3 depletion are labelled in red. (C) Data from (Klenov *et al.* 2014) Loss of nuclear *piwi* causes depletion of H3K9me3 at numerous TEs. TEs were considered significant if adjusted p-value < 0.01.

4.3.6 Hmr regulates H3K9me3 deposition on the X chromosome

To analyze whether Hmr affects H3K9me3 deposition genome-wide, I identified peaks of enrichment using MACS2 (Zhang *et al.* 2008) and then determined differential binding using the R package DiffBind (Ross-Innes *et al.* 2012). Nearly 500 peaks were identified as differentially enriched for H3K9me3 throughout the genome (Figure 4.11A). The peak width ranged from 100 bp – 44,500 bp, but the vast majority were less than 5 kb. Strikingly, nearly half of these peaks are on the X chromosome. Both the 2nd and 3rd chromosomes displayed very similar patterns. The peaks were distributed throughout the arms and did not show enrichment in heterochromatin. Surprisingly, the majority of differential peaks were due to increased H3K9me3 deposition in the mutant.

The X chromosome displayed a unique pattern of H3K9me3 deposition compared to the autosomes. Approximately 1/3 of the differential peaks are within heterochromatin and comprise more than 500,000 bp, which is nearly half the sum of base pairs displaying differential H3K9me3 genome-wide. There are two prominent clusters of differential peak enrichment in the pericentromeric region of the X. At the heterochromatin-euchromatin border, there is a sharp increase in H3K9me3 in the mutant. However, there is a large cluster of peaks downstream in which H3K9me3 is depleted in the mutant (Figure 4.11B). Intriguingly, the *flamenco* piRNA cluster resides within this pericentric region. The *flamenco* cluster functions in somatic support cells in the ovary and regulates somatic elements including *gypsy*, *idefix*, and *ZAM* (Malone *et al.* 2009).

In order to determine the functional consequences of this effect on the X chromosome, I

analyzed the expression of the annotated features spanning region 20,971,415 – 23,530,193. Very few genes are misregulated in *Hmr* mutants (Satyaki *et al.* 2014) and there is no enrichment of these genes on the X. In addition, there are also several transposon insertions within this region. However, the RNA-seq dataset does not have the power to identify individual insertions that are misregulated. I did have power to map reads to 6 individual elements, 3 of which are located on the X (2 *HeT-A* insertions and 1 *TART* insertion), however, these insertions are all located at the distal telomere.

It is possible that the striking pattern observed on the X chromosome is an artifact and reflects differences in the assembly of heterochromatin on the X versus the autosomes in release 6. However, there is no evidence for this in the assembly notes as the size of unmapped regions is approximately equivalent across the chromosomes (Hoskins *et al.* 2015). The X chromosome contains a slightly higher density of TEs than the autosomes (Kofler *et al.* 2012; Cridland *et al.* 2013), but this is unlikely to account for the significant excess of differential peaks observed on the X. Overall, our results suggest that Hmr has a complex role in the regulation of heterochromatin on the X chromosome.

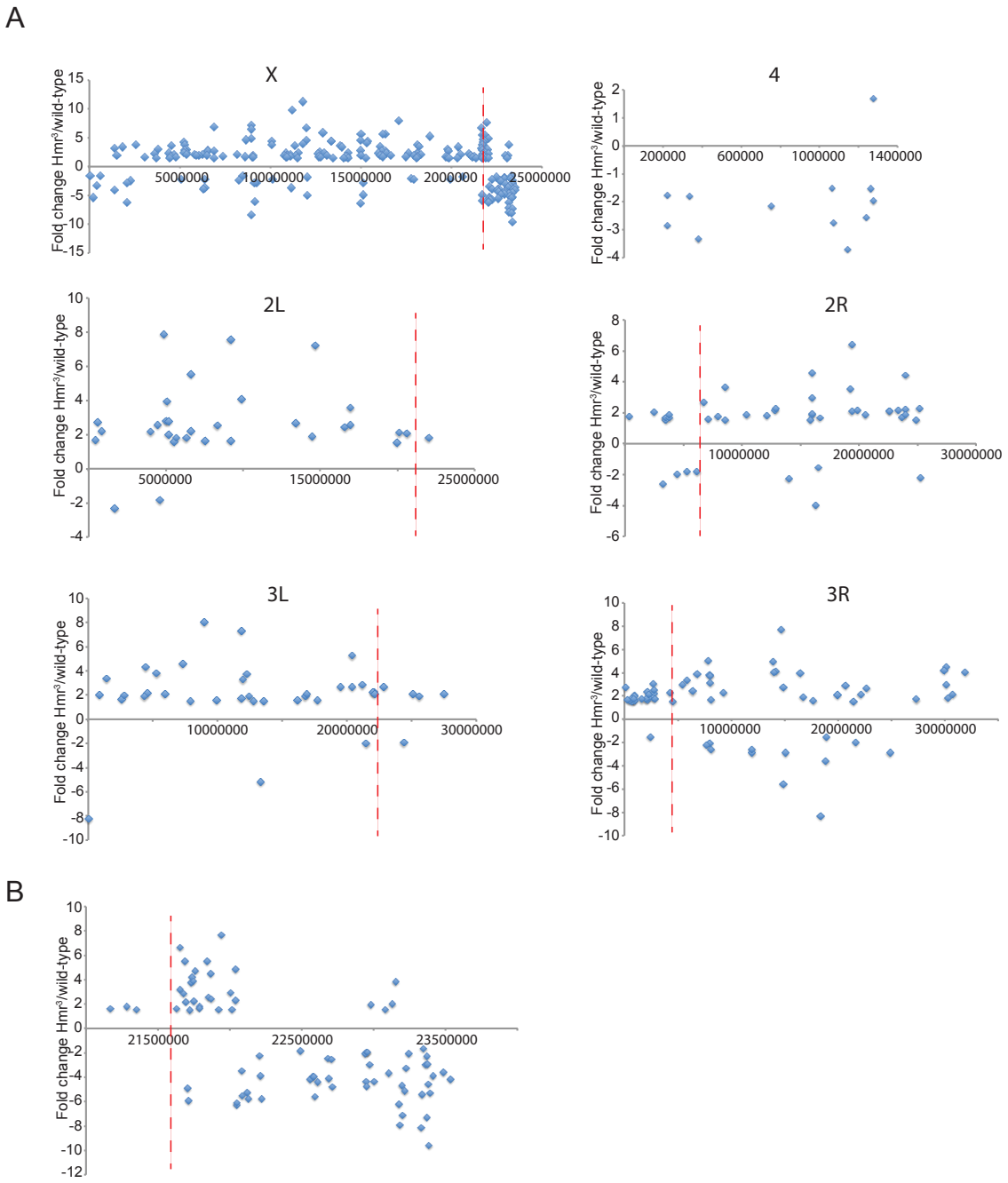


Figure 4.11. Hmr regulates X heterochomatin. Peaks were called as significant at a FDR of 0.01. The y-axis is the fold change between *Hmr*^{3} and *Hmr*$^{+}$. Fold change < 0 indicates decreased H3K9me3 deposition in the mutant while fold change > 0 indicates increased H3K9me3 in the mutant relative to wild type. The dashed red lines mark heterochromatin-euchromatin boundaries. (A) H3K9me3 peaks displaying significant differential enrichment between *Hmr*$^{+}$ and *Hmr*^{3} are plotted along the length of each chromosome. (B) Magnification of X chromosome region spanning 21171054 – 23531495. H3K9me3 is increased in the mutant near the heterochromatin-euchromatin boundary, but decreased within pericentric heterochromatin.

4.4 Discussion

4.4.1 Analysis pipelines lead to distinct conclusions

This study highlights the importance of rigorously assessing analysis methods when mapping to repetitive sequences. The analysis of repetitive sequences is not trivial and can be confounded by multi-mapping reads and differences in repeat copy number. Here I aligned reads to a database containing TE consensus sequences, which precludes issues arising due to multi-mapping reads and is the standard practice. However, different pipelines can lead to distinct results. If the discrepancies are due to TE copy number divergence, there are two possible causes: 1) the strain backgrounds are different (i.e. RNAi experiments) or 2) the mutants cause increased TE transposition (i.e. *Hmr*).

Klenov *et al.* (2014) utilized the *piwi*^{Nt} mutant to analyze the effect of *piwi* on H3K9me3 enrichment. They reported a number of TEs that exhibited a decrease in H3K9me3 upon *piwi* loss. I also found a large of TEs with significantly reduced H3K9me3, however our results did not completely overlap. I determined whether differences in TE copy number between the strains used in Klenov *et al.* (2014) could account for these differences. I mapped input reads from both genotypes (*piwi*^{Nt} and wild-type) and used DESeq to determine differential enrichment, but did not find any significant differences in TE copy number, suggesting that the observed discrepancies are due to distinct analysis pipelines.

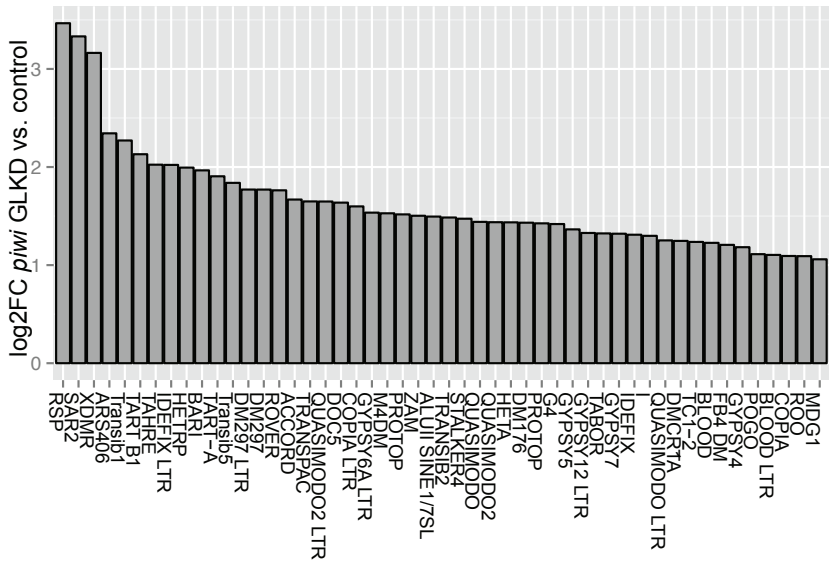


Figure 4.12. TE copy number in *piwi* and *white* GLKD strains. Divergence in TE copy number between RNAi strains used in LeThomas *et al.* (2013) may explain some of the discrepancies in results presented in LeThomas *et al.* (2013) and those presented here.

Le Thomas *et al.* (2013) used RNAi to knockdown *piwi* in ovaries and found an increase in Pol II occupancy at a number of TEs. However, using my pipeline, no TEs displayed a significant increase in Pol II upon *piwi* depletion. I mapped input reads from *piwi* GLKD as well as *white* GLKD (control) and used DESeq to determine whether there was a significant difference in the number of reads mapping to TEs. I found that 50 TE families have diverged in copy number between these strains (Figure 4.12). However, one caveat of this analysis is that typically only one input sample is sequenced per genotype. Thus, this analysis lacks the statistical power of replicates. Nonetheless, it is possible that because my analysis accounts for copy number differences, I did not observe the same results as presented in Le Thomas *et al.* (2014).

This finding highlights the importance of accounting for copy number differences when analyzing TEs. In RNAi studies, the strains being compared are derived from lines carrying distinct RNAi constructs and are thus subject to genetic variability. TEs are highly susceptible to this variation due to their dynamic nature. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system elegantly eliminates background variability, at least initially, and therefore provides a more accurate depiction of gene function. However, copy number variation can occur as lines diverge, therefore, appropriate normalization techniques must be considered.

4.4.2 Telomeric TE copy number increases in *Hmr* mutants

Despite being outcrossed for multiple generations, the strains used here, Hmr^{+} and Hmr^{3}, have diverged in TE copy number for a small number of families. The telomeric TEs were among those that have increased significantly in copy number. Satyaki *et al.* (2014) previously showed that both *Lhr* and *Hmr* mutants exhibit increases in *HeT-A* copy number ranging from 6-14 fold. Here, by quantifying ChIP input data, I observed that *TART* and *TAHRE* also display similar copy number increases in the *Hmr* mutant ranging from 2-6 fold. *DMHMR2* expression was highly upregulated in *Hmr* mutants and has doubled in copy number. There is little in the literature regarding this repeat, except that it resides within heterochromatin. *Gypsy 5* also increased in copy number in the *Hmr* mutant. However, transcript abundance of this element was not significantly affected. *Gypsy 5* is expressed exclusively in somatic cells and is putatively active based on the presence of full-length copies in the genome (Bartolomé *et al.* 2009). One

caveat here is that the strains were unlikely to be fully matched after 6 generations and therefore, the copy number divergence in *DMHMR2* and *gypsy 5* might be due to strain differences, rather than *Hmr.* Nonetheless, the vast majority of TEs are present at equivalent copy numbers between the strains, indicating that the increased RNA transcript abundance in *Hmr* mutants is not due to differences in copy number.

TE expression is typically used as a proxy for TE activity, but the exact relationship between them is unknown. Importantly, expression does not necessarily imply activity as several inactive TEs are derepressed in *Hmr* mutants, for example *gypsy 12* and *Helena* (Kofler *et al.* 2012; Satyaki *et al.* 2014). The majority of TEs in *D. melanogaster* are retrotransposons, which transpose through an RNA intermediate and therefore, increased transcription is likely correlated with the rate of transposition. However, many TEs are regulated post-transcriptionally by the piRNA pathway, which targets and degrades transposon mRNAs. The increase in copy number of the teomeric TEs in *Hmr* mutants suggests that their derepression causes increased activity. Hmr localizes to telomere ends (Satyaki *et al.* 2014) and could be a component of the telomere cap. Alternatively, it is possible that this phenotype is not a direct consequence of derepression, but instead caused by an additional unknown function of Hmr.

4.4.3 Mechanism of Hmr TE repression

The results shown here do not support a role for Hmr in global TGS of TEs because the amount of RNA Pol II bound at the majority of TEs was not significantly different

between *Hmr*^{3} and *Hmr*$^{+}$. RNA Pol II occupancy did increase significantly at a small subset of TEs upon *Hmr* depletion, including the telomeric TEs *HeT-A* and *TART*, further supporting a role for Hmr in telomere function (discussed below). Curiously, three families (*GTWIN, R2, Gypsy 5*) displayed a significant reduction in the amount of bound RNA Pol II upon loss of *Hmr*, however none of these elements are regulated by Hmr. Therefore, this result is not in direct conflict with previous results.

In addition, there was little effect on H3K9me3 deposition at TE sequences in *Hmr* mutants, indicating that Hmr does not function in the modification of chromatin at TEs. The conclusion that Hmr does not act at the transcriptional level is surprising because Hmr is nuclear and interacts with HP1 (Thomae *et al.* 2013; Satyaki *et al.* 2014). Nascent RNA transcription can be measured using GRO-seq (Core *et al.* 2008). This is a more sensitive assay than Pol II ChIP-seq to measure transcriptional activity due to incomplete knowledge of the precise relationship between the amount of bound Pol II and level of expression. Rozhkov *et al*. 2013 performed GRO-seq in both somatic and germline tissues in *piwi* KD ovaries. In the soma, nascent TE transcription was strongly positively correlated with steady state transcription levels upon loss of Piwi, supporting its role in TGS (Rozhkov *et al.* 2013). In order to fully examine a potential transcriptional role for Hmr at TEs, it is necessary to measure nascent transcription in *Hmr* mutants. Lack of a positive correlation between nascent transcription and steady-state RNA levels would confirm the findings presented here.

Satyaki *et al.* (2014) suggest that Hmr and Lhr could participate in RNA degradation via the nuclear exosome (Yamanaka *et al.* 2012). The exosome is a multi-subunit complex possessing 3' to 5' RNA endonucleolytic activity and is found in both the cytoplasm and nucleus. The nuclear exosome functions in both the processing of precursor RNAs, as well as degradation of aberrant transcripts. In addition, it has been implicated in the regulation of mRNA transcript levels (Roth *et al.* 2005; Lee *et al.* 2005; Kuai *et al.* 2005) and has been shown to associate with actively transcribed regions (Andrulis *et al.* 2002; Hessle *et al.* 2009, 2012). I suggest that Hmr could act as a cofactor of this complex to elicit TE silencing in the nucleus. It is unknown whether the nuclear exosome targets TE transcripts for degradation. However, it has been implicated in the degradation of gene transcripts containing TE sequences at the 5' end resulting from inefficient splicing (Kuan *et al.* 2009). None of the core exosome components, which include several of the ribosomal RNA-processing proteins, are known interactors of Hmr, but Hmr does localize to the nucleolus (Thomae *et al.* 2013; Satyaki *et al.* 2014). In addition, the genomic binding profile of exosome components closely parallels that of BEAF-32 (Lim *et al.* 2013), which physically interacts with Hmr (Thomae *et al.* 2013).

4.4.4 Does Hmr function within the piRNA pathway?

Hmr and *Lhr* are rapidly evolving as are several piRNA pathway genes (Lee and Langley 2012), suggesting that divergence at *Hmr* and *Lhr* could be caused by co-evolution with the piRNA pathway (Satyaki *et al.* 2014). However, whether Hmr and Lhr function within the piRNA pathway is unknown. Lhr does not affect piRNA abundance nor cause abnormal ping pong signatures, so it is unlikely to function in piRNA

biogenesis (Satyaki *et al.* 2014). Furthermore, Vasa, which is required for secondary piRNA processing, maintains localization to the nuage in *Lhr* mutants (Satyaki *et al.* 2014), but this remains untested for Hmr. Co-immunostaining of Hmr and Lhr show that these two proteins do not completely overlap (S. Prasad, unpublished data), suggesting additional unique roles for each. Therefore, it is worth assaying the piRNA pool in *Hmr* mutants. Knowledge of the binding sites of Hmr will also shed light on whether Hmr is involved in piRNA biogenesis.

The results for Hmr presented here do not specifically mirror the effects of known piRNA protein mutants analyzed here. For example, *Hmr* does not have a strong effect on H3K9me3 enrichment as observed in *piwi* mutants. Furthermore, there is no correlation between expression and RNA Pol II occupancy of Hmr-regulated TEs in *Hmr* mutants, as observed for *rhi*. However, this does not exclude the possibility that Hmr functions within the piRNA pathway. Thomae *et al.* (2013) used tandem-affinity purification in *D. melanogaster* Schneider cells (S2) to identify protein interactors with Hmr and Lhr. While none of the identified factors are known piRNA-effector proteins, *Stonewall* (*Stwl*), *cenp-c*, and *mod(mdg4)* were shown to interact with Hmr (Thomae *et al.* 2013). Notably, all three of these proteins displayed significant TE derepression in the germline screen by Czech *et al.* (2013) to identify novel piRNA pathway genes. Alternatively, Hmr may function independently of the piRNA pathway.

4.4.5 Hmr functions at the telomere

There is substantial evidence linking Hmr and Lhr to telomeric processes. Satyaki *et al.* (2014) showed that the telomeric TEs were among the highest derepressed elements upon loss of *Lhr* or *Hmr*. Loss of *Lhr* was also shown to affect the abundance of piRNAs mapping to clusters containing the HETRP telomere-associated sequence (TAS) repeat, suggesting that Lhr is required for expression of telomere-associated clusters (Satyaki *et al.* 2014). Among the few TEs exhibiting differential Pol II occupancy were both *HeT-A* and *TART*. Furthermore, H3K9me3 signal at *HeT-A* was depleted. These observations suggest that Hmr may mediate TGS at *HeT-A* by establishing a repressive state. Because of the critical function of the telomeric TEs, precise regulation is essential to maintain genome integrity. Intriguingly, the piRNA pathway also functions in telomere regulation. Depletion of Piwi causes massive upregulation of the telomeric TEs (Le Thomas *et al.* 2013) as well as decreased piRNAs mapping to *HeT-A*, *TAHRE*, and *TART* (Rozhkov *et al.* 2013). Like Lhr, Piwi appears to be required for expression from telomeric piRNA clusters. H3K9me3 is required for piRNA cluster transcription (Rangan *et al.* 2011), therefore, the decrease of H3K9me3 at the telomeric TEs predicts a concomitant loss of telomere-associated piRNAs in *Hmr* mutants, but this remains to be tested. Khurana *et al.* 2010 showed that the piRNA pathway genes *aubergine (aub)* and *armitage (armi)* function in telomere protection in a process mediated by telomeric piRNAs (Khurana *et al.* 2010). Therefore, it is likely that both Hmr and Lhr function in telomere protection and are possibly components of the cap, which is supported by their co-localization with HP1a at telomere ends.

The telomeric retrotransposons are present in all known Drosophila species. However, despite their critical role in telomere maintainence, both *HeT-A* and *TART* display elevated rates of sequence divergence compared to euchromatic genes and other retrotransposons (Casacuberta and Pardue 2005). In fact, even within a single stock, distinct *HeT-A* subfamilies have been detected (Danilevskaya *et al.* 1998; Abad 2004). The majority of these families are actively transcribed, though at varying levels. Interestingly, an analysis between *HeT-A* sequence variability between the GIII strain (which is known to have extended telomeres) and Oregon-R (which has normal telomere length) showed that, despite the increased telomere length in GIII, the same level of sequence diversity between subfamilies was observed in each strain (Piñeyro *et al.* 2011). The authors conclude that this is likely due to shared constraints (Piñeyro *et al.* 2011). The question then arises of what is causing the rapid evolution of the telomeric sequences. Piñeyro *et al.* 2011 propose two possibilities: 1) the observed sequence divergence is a result of a balance between the the inherent nature of TEs to escape host regulation and its function as a domesticated element in the host or 2) that only a few regions of *HeT-A* sequence are under purifying selection and function is maintained despite extensive divergence outside those regions. Petit *et al.* 2012 identified a small conserved region within the 3' UTR which corresponds to a piRNA target sequence, however, whether this sequence is required for *HeT-A* function is unknown (Petit *et al.* 2012). Because of the dynamic nature of the telomeric retrotransposons, it is possible that the telomeric function of Hmr is responsible for its rapid evolution.

4.4.6 Does the effect on the X chromosome indirectly regulate TE sequences genome-wide?

While *Hmr* mutants did not exhibit abnormal H3K9me3 at TE sequences, the genome-wide analysis of this mark yielded a striking pattern on the X chromosome. Hmr appears to maintain H3K9me3 within the pericentromeric region of the X, but represses its accumulation at the heterochromatin-euchromatin border, suggesting that Hmr has a complex role in X heterochromatin regulation. It is possible that this function has downstream indirect effects on the regulation of TEs genome-wide. For example, with the exception of the 4th chromosome, the autosomes displayed increased H3K9me3 in *Hmr* mutants relative to wild-type. It is possible that loss of H3K9me3 in X pericentric heterchromatin results in the redistribution of the mark to the other chromosomes. The 4th chromosome contained very few peaks of differential enrichment, however, nearly all of these regions displayed decreased H3K9me3 deposition in the mutant. Additional characteriziation of the chromatin structure of the X in *Hmr* mutants is required for a more complete understanding of Hmr's role in modulating X chromosome heterochromatin and to determine whether this function mediates TE regulation.

CHAPTER 5

HMR CHROMATIN IMMUNOPRECIPITATION

5.1 Introduction

Hybrid incompatibility (HI) is an example of post-zygotic reproductive isolation and contributes to speciation by decreasing the fitness of interspecific hybrids. Genes causing incompatibility are collectively called HI genes and are typically thought to play important roles in speciation. The Dobzhansky-Muller (D-M) model asserts that HI is the result of lineage-specific divergence and that the incompatibility arises due to the negative epistatic interaction of the two derived alleles in the hybrid background. F1 hybrid sons from crosses between *Drosophila melanogaster* females and *D. simulans* males are larval lethal. Lethality is caused by the epistatic interaction of *D. melanogaster Hybrid male rescue* (*Hmr*) and *D. simulans Lethal hybrid rescue* (*Lhr*) in the hybrid background (Brideau *et al.* 2006).

Loss-of-function mutations in either *D. melanogaster Hmr* or *D. simulans Lhr* rescue hybrid males to the adult stage (Watanabe 1979; Hutter and Ashburner 1987). Deletions spanning *D. melanogaster Lhr* rescue hybrid males to the pharate stage indicating that the hybrid lethal function of *Lhr* has diverged between the two species (Maheshwari and Barbash 2012, Cuykendall *et al.* 2014). The hybrid lethal function of *Hmr* has also diverged. Patroclinous exceptional hybrid males inheriting the *D. simulans* X are fully viable. Intriguingly, both *Hmr* and *Lhr* are rapidly evolving due to positive selection (Barbash *et al.* 2004; Brideau *et al.* 2006). In order to elucidate what is driving the

divergence of *Hmr* and *Lhr*, it is imperative to first determine the intraspecific functions of these genes. This chapter focuses on probing the molecular function of the Hmr protein in *D. melanogaster*.

Hmr encodes a 1,430 amino acid chromatin-binding protein containing four MADF domains (Barbash *et al.* 2003; Maheshwari *et al.* 2008). The MADF domain is homologous to the DNA-binding domain in the Adf1 transcription factor (England *et al.* 1992), suggesting a possible role in gene regulation for Hmr (Maheshwari *et al.* 2008). Fluorescent in situ hybridization (FISH) experiments on embryos and ovaries showed that Hmr localizes to heterochromatin and associates with a subset of satellite sequences, including the pericentromeric satellite dodeca (AATAACATAG) on 2L and 3L (Abad *et al.* 1992), GA-rich repeats, and the 2L3L satellite (Satyaki *et al.* 2014). Further, Hmr regulates the abundance of TE transcripts by exerting a repressive effect (Satyaki *et al.* 2014). TEs are derepressed up to 40 fold in *Hmr* transheterozygous mutants when compared to the heterozygous Hmr^{+} controls, with the most striking effect at the telomeric TEs, *HeT-A*, *TAHRE*, and *TART*.

Hmr colocalizes with HP1a at the telomeric cap, which is a protein-rich structure at the distal end of the telomere (Andreyeva *et al.* 2005; Raffa *et al.* 2011; Satyaki *et al.* 2014). In addition, there is some evidence for a direct physical interaction between Hmr and HP1a (Thomae *et al.* 2013; Satyaki *et al.* 2014). HP1a is enriched in heterochromatin (James and Elgin 1986; James *et al.* 1989) and is required for silencing a subset of TEs

(Wang and Elgin 2011). ChIP-chip in BG3 and S2 cells shows HP1 enrichment at TEs as well as at active genes located within heterochromatin (Riddle *et al.* 2011).

The major derepression of TEs in *Hmr* mutants suggests that Hmr may localize to specific TEs to exert its repressive effect. I aim to identify the binding sites of Hmr throughout the genome using chromatin immunoprecipitation followed by whole genome sequencing (ChIP-seq). Several factors complicate this endeavor. First, there is not a working ChIP grade antibody to Hmr. Therefore, I am using transgenic lines expressing *Hmr-HA* transgenes (Satyaki *et al.* 2014) in the Hmr^{3} mutant background to pull-down Hmr-HA using a rabbit anti-HA antibody. Secondly, there is not a positive control for the ChIP because the binding sites of Hmr are unknown. Third, Hmr is a large protein and I have observed that it is prone to degradation.

In addition, I am interested in the divergence of *Hmr* between *D. melanogaster* and *D. simulans*. *D. simulans* Hmr maintains heterochromatic localization in a *D. melanogaster* background (Satyaki *et al.* 2014). Further, only a handful of TEs are differentially regulated between the orthologs (Satyaki *et al.* 2014). However, the sequence divergence between the orthologs is substantial and could indicate binding site divergence at a resolution higher than is detectable by immunofluorescence. Therefore, I will also assay *D. simulans* Hmr by using a *D. melanogaster* Hmr^{3} transgenic line expressing *D. simulans Hmr-HA.*

5.2 Methods

5.2.1 Fly stocks

Stock 3-73 (*y[1] w[67c23] P{w[+mC] y[+mDint2]=EPgy2}EY12237*) known as *Hmr*^{3}, was outcrossed to *y w F10* for 6 generations in April 2012. Sibling crosses were done to generate a homozygous line called *y w Hmr*^{3}. Females from this stock were crossed to *y w*; *TM3 Sb/TM6 Hu* males to obtain *Hmr*^{3}; *TM3 Sb/+* males which were then backcrossed to *y w Hmr*^{3} females to generate *Hmr*^{3}*; TM3 Sb/+* females.

Males from the transgenic lines (*y w ; ø{mel-Hmr-HA}* and *y w ; ø{sim-Hmr-HA})* (described in Satyaki *et al.* 2014) carrying either *mel-Hmr-HA* or *sim-Hmr-HA* transgene were crossed to *y w Hmr*^{3} females. Male progeny (*y w Hmr*^{3}; *{ø}/+)* were then crossed to *Hmr*^{3}*; TM3 Sb/+* females. Male and female progeny heterozygous for *TM3 Sb* and the transgene were selected and crossed to obtain stocks homozygous for the transgenes, *y w Hmr*^{3}; *ø{mel-Hmr-HA}* and *y w Hmr*^{3}; *ø{sim-Hmr-HA}*.

5.2.2 qRT-PCR

RNA extraction, cDNA synthesis, and qRT-PCR were performed as described in (Maheshwari and Barbash 2012). For the qRT-PCR, three technical replicates were performed. Primers: Hmr cDNA F (1428): TAAGTTCGCCTTCCGCACATACC Hmr cDNA R (1429) TGACCAGAAACCTGAGTTGCTCCA. *Rp49* was amplified using primers from reference (Fiumera 2004). *Hmr* levels were normalized relative to *rp49*.

HeT-A and 60D were amplified using primers from (Klenov *et al.* 2007) and *HeT-A* was normalized to 60D.

5.2.3 ChIP

One hundred ovary pairs were dissected in ice cold 1X PBS from flies aged 5-6 days at room temperature, with the final 2 days being supplemented with yeast paste. Ovaries were cross-linked in 1.8% paraformaldehyde/1X PBS for 10 min at room temperature, quenched with 125 mM glycine for 5 min, and washed three times in 1X PBS. Ovaries were lysed in sonication buffer (20 mM Tris pH8.0, 2 mM EDTA, 0.5 mM EGTA, 1X Roche Complete P.I. Tab, 0.5% SDS) using a motor-driven pestle. A Bioruptor (Diagenode) was used to shear the DNA to the desired 300-500 bp range using the following settings: 2 cycles at 15 min each at "High" with 20 sec ON/1 min OFF pulses. The resulting extract was spun at 13,000 rpm for 10 min at 4°C and then diluted with IP buffer (0.5% Trition X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol) to 0.05% SDS. For each ChIP, 200 μl of Protein A/G agarose beads (Millipore 16-125) were washed three times in IP buffer. The equilibrated beads were then added to the dilution extract (100 μl beads + 900 μl chromatin) and incubated for 2 hours at 4°C. The beads were spun for 1 min at 1000 rpm and the supernatant transferred to a new tube. Initially, the IP was done using a 1:10 dilution of polyclonal rat anti-HA (Roche 3F10). However, I switched to a ChIP grade polyclonal rabbit anti-HA antibody (ab9110, 1:10) for better yield. The chromatin-immunoprecipitated samples were subsequently washed in a series of ice-cold buffers: low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl), high salt (0.1% SDS, 1% Triton X-100, 2

mM EDTA, 20 mM Tris pH 8.0, 500 mM NaCl), LiCl (10 mM Tris pH 8.0, 0.25 M LiCl, 1% NP40, 1% NaDeoxycholate), and then rinsed two times in TE buffer (10 mM Tris pH8.0, 1 mM EDTA), followed by treatment with 50 μg RNase A. The DNA-protein-antibody complexes were eluted from the beads by incubating two times for 15 min in elution buffer (1% SDS, 0.1 M $NaHCO_3$) at room temperature. Cross-links were reversed by heating at 65°C. The samples were treated with 40 μg Proteinase K for 2 hours at 37 °C.

5.2.4 Western blot and IP western

For the IP western, the DNA-protein-antibody complexes were washed and 0.5X volume of 2X SDS/PAGE sample buffer was added following pull-down. The samples were boiled for 5m and approximately ~3.5 ovary pairs were run of each sample. Primary antibody: polyclonal rat anti-HA (Roche 3F10, 1:10 dilution). Secondary antibody: HRP conjugated goat anti-rat secondary antibody (1:20,000). Detection was achieved with the ECL Western blotting substrate (Pierce). For western blots to detect endogenous Hmr, primary antibodies (monoclonal rat anti-Hmr 2C10 and 3D8) were a gift from Axel Imhof (Thomae *et al.* 2013), N-terminal and C-terminal antibodies were made and purified by Shuqing Ji in house; all were used at 1:50 in 5% milk-TBST. HRP-conjugated anti-rat secondary antibody was used at 1:2000 dilution but failed to detect a band at 170 kD.

5.2.5 ChIP-qPCR

5 μl of each ChIP DNA and input (1:100 dilution) were used as templates. *HeT-A* was amplified using Het-s2 and Het-as2 which anneal to the CDS (Klenov *et al.* 2007) and normalized to the 60D region amplified by the primers used in Klenov *et al.* 2007.

5.3 Results and Discussion

5.3.1 *Hmr-HA* transgenes are expressed

We used qRT-PCR to assay whether the *Hmr* transgenes are expressed. Both *mel-Hmr-HA* and *sim-Hmr-HA* are expressed at significantly higher levels when compared to the negative control, *y w* Hmr^3 (Figure 5.1). While *sim-Hmr-HA* is expressed at a higher level compared to the *mel-Hmr-HA*, this difference is not significant (p=0.059, two-tailed *t*-test with equal variance). However, this pattern mirrors the relative expression levels of the *Hmr-FLAG* transgenes in (Satyaki *et al.* 2014).

5.3.2 Anti-Hmr antibodies

I tested two endogenous monoclonal Hmr antibodies raised in rat from (Thomae *et al.* 2013) on ovaries dissected from Hmr^+ and Hmr^3 flies for specificity. I tried concentrations of 1:500, 1:100, and 1:50, but failed to see a band for Hmr (Figure 5.2). In addition, I also tested two purified antibodies made in-house (by Shuqing Ji) recognizing the N-terminus and C-terminus of the protein. While the N-terminus antibody produced a band in *D. melanogaster*, it did not work in *D. simulans* (Figure 5.2B). Therefore, I proceeded with using the rabbit anti-HA antibody to pull-down HA-tagged Hmr.

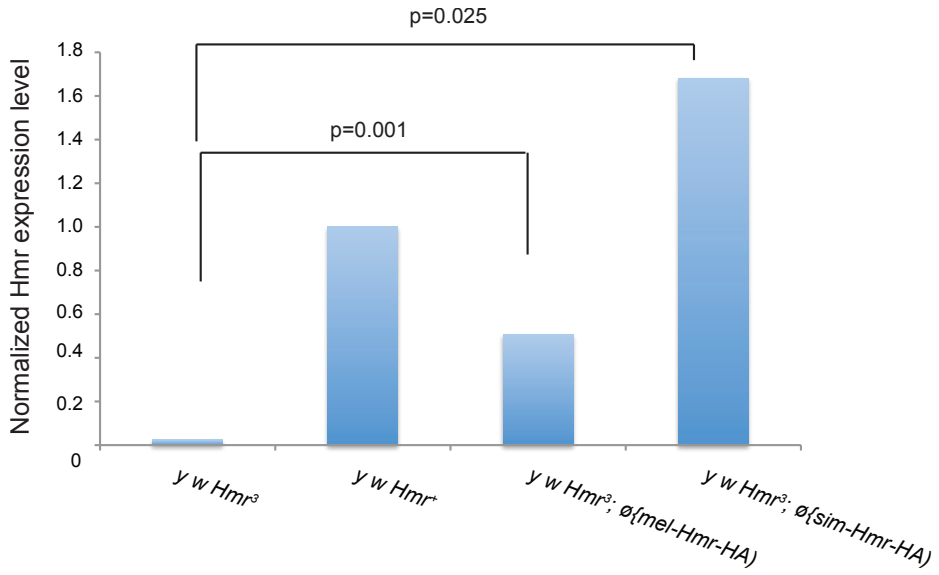


Figure 5.1. *Hmr-HA* transgenes are expressed. qRT-PCR analysis of *Hmr-HA* transgenes. *Hmr* transcript levels in transgenic lines were compared to the host strain *y w Hmr*^{3} and a *y w Hmr*$^{+}$ control. The transgenes are homozygous. Expression level was normalized relative to *rp49*. The difference in expression level of *mel-Hmr-HA* and *sim-Hmr-HA* is not significant, but the differences between *mel-Hmr-HA* and *Hmr3* and *sim-Hmr-HA* and *Hmr3* are significant (p=0.001, p=0.025, respectively, two-tailed t-test).

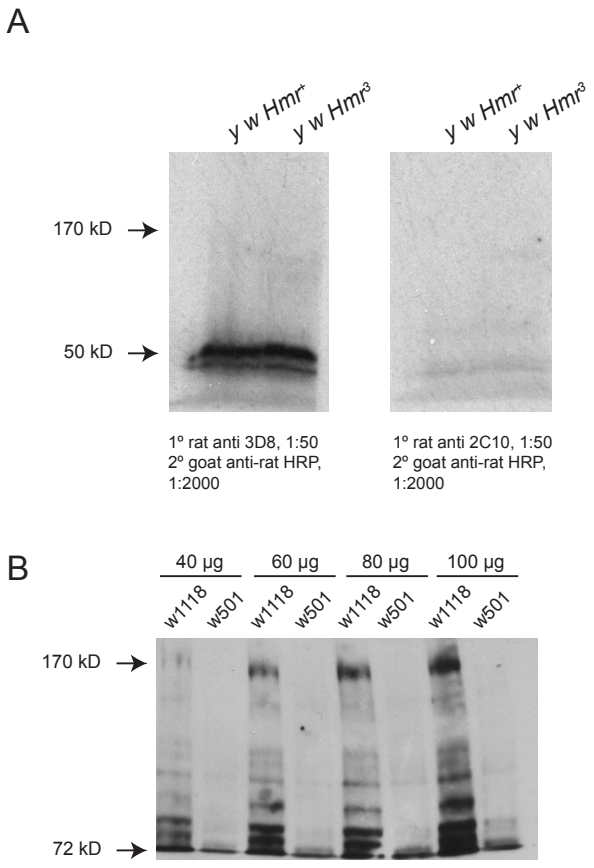


Figure 5.2. Anti-Hmr antibodies (A) I tested monoclonal rat antibodies to Hmr from (Thomae *et al.* 2013) on ovaries dissected from *Hmr+* and *Hmr3* lines. Primary antibodies were used at 1:50 in 5% milk-TBST and detected using HRP-conjugated anti-rat secondary antibody at a dilution of 1:2000. A band for Hmr was not observed. The bands at 50 kD are likely the heavy chains of the IgGa (B) I collected embryos 5-6 h after egg laying and isolated the protein fraction. I tested an Hmr antibody (1:1000) designed to recognize the N-terminal portion of the protein. The antibody recognized *D. melanogaster* (w1118), but not *D. simulans* (w501) Hmr. HRP-conjugated goat anti-guinea pig was used at 1:2000. Exposure time was 2 min.

5.3.3. Hmr is prone to degradation and pull-down is weak

I initially used a rat anti-HA antibody to pull-down Hmr-HA. The lab has used this antibody successfully for westerns and immunofluorescence. However, this does not necessarily indicate that the antibody is suitable for ChIP. To test this, I performed an IP western. I used *y w Hmr*^{3}; *ø{mel-Hmr-HA}* ovaries for all IP optimization steps. Using the antibody at very high concentrations of 1:10 and 1:20 for the IP produced bands at the expected size for Hmr (Figure 5.3A). However, in addition to the band at 170 kD, I noticed several additional smaller bands, indicative of degradation products. Furthermore, in order for the bands at 170 kD to be visible, I had to do an over night exposure. I included a positive control consisting of 10 freshly dissected ovaries from *y w Hmr*^{3}; *ø{mel-Hmr-HA}*, as well as an input control. Curiously, I also observed a band in the negative control, *Hmr*^{3} input, which is likely due to spill over from the positive control. In order to increase pull-down efficiency, I switched to a ChIP grade polyclonal rabbit anti-HA antibody for all subsequent pull-downs. ChIP grade antibodies are pre-tested for ChIP efficiency and background levels and are therefore recommended. I observed bands for full size Hmr for IPs containing a lower concentration of antibody than used previously for the rat anti-HA antibody (Figure 5.3B). Furthermore, bands were visible after only a 35 min exposure. However, degradation products were still present.

I also tested whether pull-down was more efficient using solely Protein G agarose beads or a mixture of Protein A and Protein G agarose beads to recognize the rabbit IgG antibody. I did not observe a strong difference and for subsequent tests used a

mixture to maximize antibody recognition. Despite attempts to load ~10 ovary pairs for each sample, there was substantially more protein in the *Hmr*^{3}*; ϕ{mel-Hmr-HA}* sample than the corresponding input. This likely results from the cross-linking step. Over fixation can result in Hmr being bound to insoluble material, resulting in decreased protein yields.

In order to decrease Hmr degradation, I performed a time course experiment to determine the step at which degradation was occurring. I first tested whether degradation is due to the dissection process. Ovaries are dissected in ice-cold 1X PBS and then immediately collected and put on ice. Because the incubation on ice can vary due to the number of ovaries dissected at one time, I tested whether the length of incubation prior to lysis contributed to degradation. I dissected ovaries in groups of 50 pairs. For one group, I immediately lysed in 2X SDS/PAGE sample buffer. For the other group, I incubated on ice for 2 h prior to lysis. I did not observe a difference in the amount of degradation products and concluded that degradation is not occurring as a result of the dissection process (Figure 5.3C).

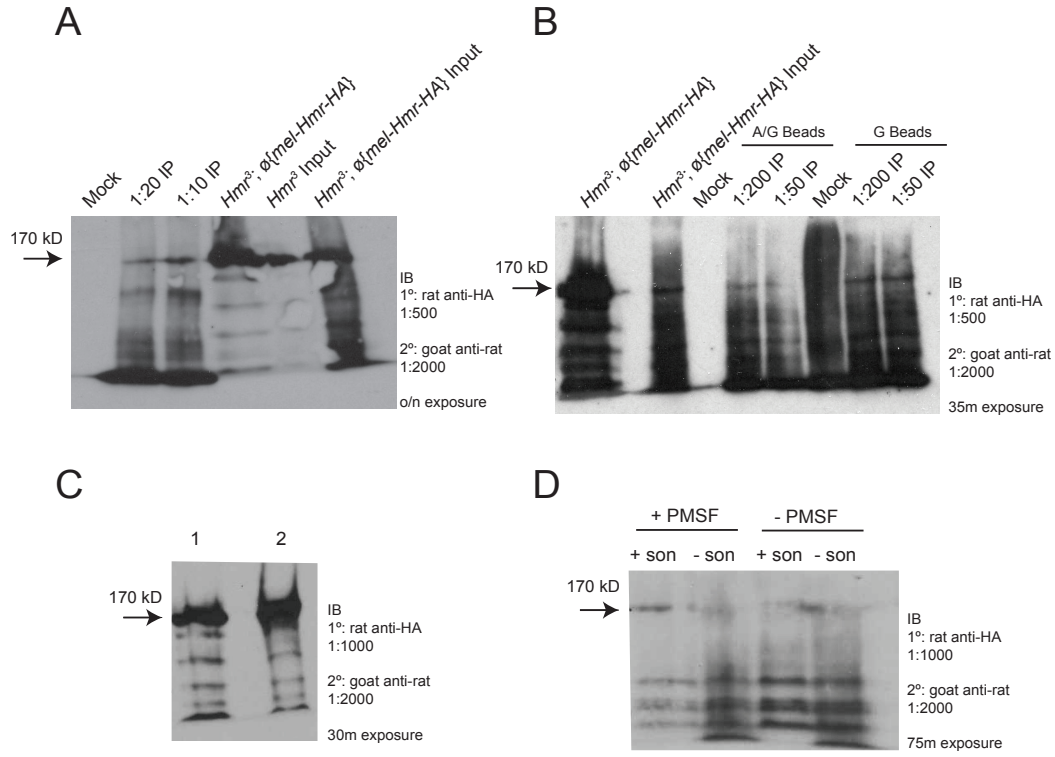


Figure 5.3. Hmr is prone to degradation resulting from sonication. (A) In our initial attempts at Hmr-HA ChIP, we used a polyclonal rat anti-HA antibody for the IP at concentrations of 1:10 and 1:20 on ovaries dissected from *Hmr^3; ϕ{mel-Hmr-HA}* and observed a band corresponding to the expected size of Hmr at 170 kD. The *Hmr^3; ϕ{mel-Hmr-HA}* lane contains protein isolated from 10 ovary pairs dissected freshly and serves as an additional positive control. Rat anti-HA antibody was used at 1:500 dilution and goat anti-rat secondary antibody was used at a dilution of 1:2000. Approximately 4 ovary pairs of each sample were loaded unless stated otherwise. (B) We switched to a ChIP-grade rabbit anti-HA antibody to gain higher pull-down efficiency. Approximately 10 ovary pairs of each sample were loaded. The *Hmr^3; ϕ{mel-Hmr-HA}* contains protein isolated from 10 ovary pairs freshly dissected. Exposure time was significantly less than in (A) and IP antibody concentrations were lower, indicating the rabbit anti-HA has higher affinity for the HA tag than the rat anti-HA antibody. The IP was performed on ovaries dissected from *Hmr^3; ϕ{mel-Hmr-HA}* and the inputs were siphoned off prior. A mixture of Protein A and G agarose beads was compared to Protein G agarose beads to determine whether a mixture would produce higher yield. The IB was done at concentrations described in (A). (C) We performed a time course experiment to determine whether degradation was occurring as a result of the dissection process. Approximately 4 ovary pairs per sample were loaded. Lane 1: ovaries were lysed immediately following dissection. Lane 2: ovaries were incubated for 2 h prior to lysis. Degradation was not increased by the 2 h incubation prior to lysis. (D) We tested whether sonication was contributing to the observed degradation of Hmr. Approximately 4 ovary pairs dissected from *Hmr^3; ϕ{mel-Hmr-HA}* were loaded for each sample. **+ PMSF** indicates ovaries were treated with the protease inhibitor, PMSF, prior to fixation. **+ son** and **– son** indicate sonication treatment.

Next, I investigated whether degradation is due to sonication and whether adding the protease inhibitor PMSF, prior to fixation, was able to mitigate the level of observed degradation. I dissected 200 ovary pairs and added 1 μl PMSF to 100 of the ovary pairs prior to fixation in 1.8% PFA. I then split each group (+ PMSF, - PMSF) in half. Approximately 50 ovary pairs from each treatment were sonicated for 30 min. The remaining half was not sonicated. Western blot analysis showed that sonication does not have a large effect on degradation and addition of PMSF prior to fixation appeared to stabilize full-length protein recovered after sonication (Figure 5.3D). However, western blots are qualitative, and it is possible that sonication contributes to degradation.

Micrococcal nuclease (MNase) is sometimes used as an alternative to sonication. Because MNase is a nuclease, it is potentially prone to sequence bias, while sonication is expected to be more random. I treated 50 fixed ovary pairs with 1 μl MNase for 5 min, 10 min, or 20 min. I followed MNase treatment by 0, 5, 10, or 20 min of sonication. A subset of these experiments is shown in Figure 5.4. I observed that some amount of sonication is optimal to more fully solubilize the chromatin. I chose to continue MNase optimization followed by 10 min of sonication. I next tested the DNA fragment size range obtained from varying incubation times with MNase and observed fragment ranges from just below 100-200 bp resulting from 5 and 10 min MNase incubations. For ChIP-seq, a slightly higher range (100-500 bp) is preferable for library construction, so I held MNase incubation time constant and instead diluted MNase 1:10, 1:50, 1:100, 1:500, 1:1000 (Figure 5.6A). Because the gel was difficult to interpret, I also measured fragment size using the Bioanalyzer (Biotechnology Resource Center, Cornell University). The

fragment analysis showed a 200-960 bp fragment range for the 1:20 dilution (5 U) for 5 min at 37°C (Figure 5.5A) followed by 10 min of sonication. Following these modifications, I repeated the ChIP and obtained ChIP DNA at concentrations comparable to what I obtained for the H3K9me3 ChIPs performed in Chapter 4 in the range of 5-10ng. However, I could not replicate the previously observed fragment range and instead observed that the chromatin had not been sheared effectively (Figure 5.5B).

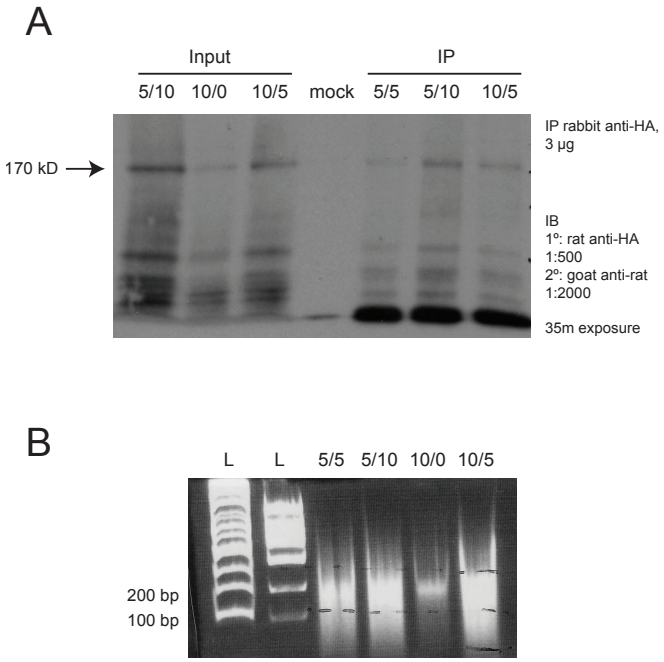
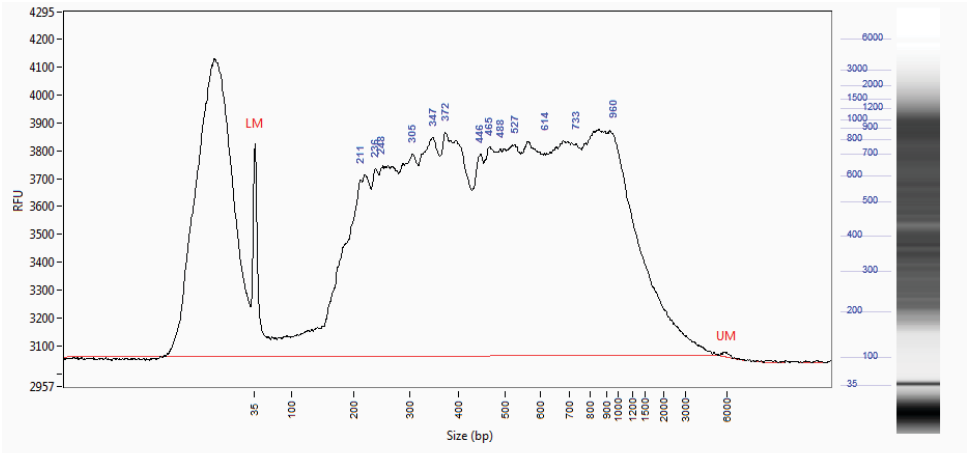


Figure 5.4. Decreased sonication time yields increased full-length Hmr. (A) Approximately 5 ovary pairs dissected from *Hmr*^{3}*; ϕ{mel-Hmr-HA}* were loaded. The first number represents the time in minutes of MNase incubation and the second number represents the length of sonication. 10 min of sonication produces the highest yield of full-length Hmr while minimizing degradation. (B) DNA was isolated and run on a 1.5% agarose gel. The first number represents the amount of time in min for the MNase incubation and the second number represents the amount of time in min for sonication.

I reasoned that MNase may be vulnerable to variability and re-focused my attempts to achieve optimal shearing. I weighed whether I would obtain higher yields using the MNase method which yields more full length Hmr but bound a wider range of fragment sizes, thereby decreasing sensitivity, or my previous method which yields less full size Hmr bound to more uniform fragment sizes. I reverted to the original method using 30 min of sonication. However, when I tested the fragment size range, I observed a range well below 100 bp, which is too small to make DNA sequencing libraries. I inferred this could actually be RNase contamination and treated the samples with an RNase cocktail, containing both RNase A and RNase T1, which should eradicate RNA that is observable on a gel (Figure 5.6B). While the RNA appeared more degraded, the DNA was still not sheared to the proper size range expected based on the sonication conditions (Figure 5.6C). I also tested whether amount of time between fixation and reversing cross-links affected the ChIP DNA yield (Figure 5.6B, C) and observed higher yield for samples which were de-cross linked immediately. I made fresh glycine solution to ensure proper quenching of the formaldehyde.

A



B

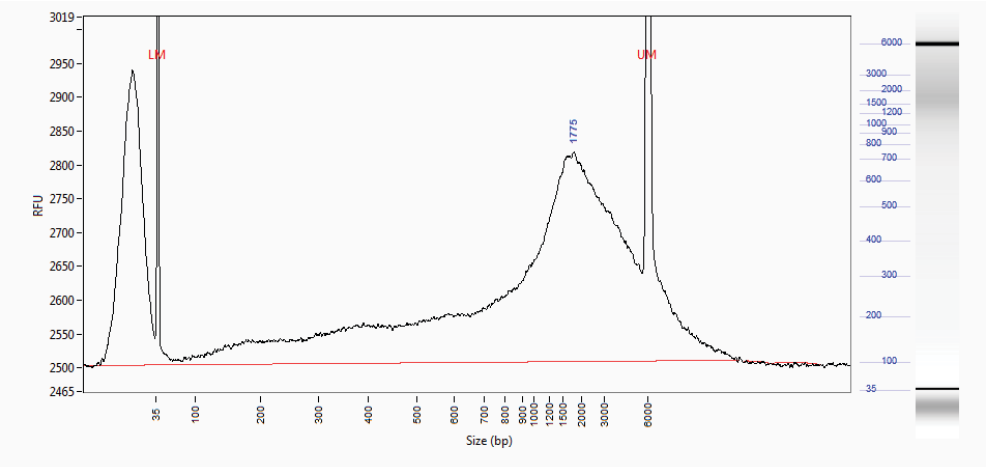


Figure 5.5. MNase dilution of 1:20 produces suitable fragment size range. (A) Dissected ovaries were fixed and then treated for 5 min with a 1:20 dilution of MNase and then subjected to 10 min of sonication. Cross-links were reversed and DNA was isolated and submitted for fragment analysis using the Bioanalyzer. (B) We were unable to replicate the results in (A) using the same conditions for the MNase treatment and sonication, indicating that MNase is prone to variability.

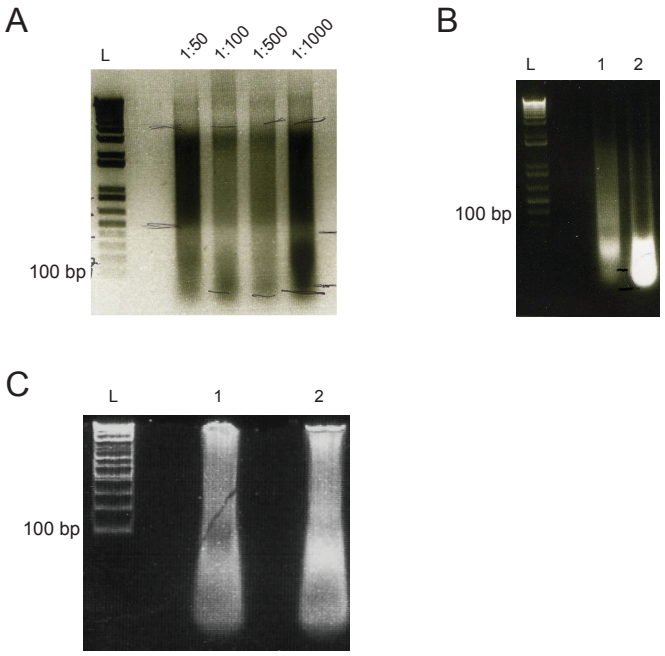


Figure 5.6. MNase prohibits accurate fragment length distributions. (A) We dissected ovaries from *Hmr*^{3}*; ϕ{mel-Hmr-HA}.* We sheared chromatin using a combination of MNase incubation followed by 10 min sonication. MNase incubation was held constant at 5 min, but we tried a dilution series of MNase in order to increase the fragment range distribution. The MNase concentrations used are listed above the wells. After shearing, cross-links were reversed and DNA was extracted and run on an agarose gel. (B) and (C) Lane 1 is chromatin that was incubated at 4°C overnight prior to reversing cross-links. Lane 2 is chromatin that was reverse cross-linked immediately following sonication. (B) RNA contamination is abundant. (C) Samples were treated with RNase cocktail after sonication. RNA contamination was decreased, but the desired fragment size range was still not observed.

5.3.4 Absence of Hmr enrichment at TEs

While bona fide binding sites of Hmr are unknown a priori, I propose that it possibly binds near TE sequences due to its role in TE regulation. I performed ChIP followed by qpcr and tested three TE families (*HeT-A*, *Copia*, and *Helena*) that were massively derepressed in *Hmr* mutants, for enrichment (Figure 5.7). I did not observe enrichment at any of the tested TE families. This is particularly surprising for *HeT-A* and its promoter because Hmr localizes to telomeres (Satyaki *et al.* 2014). However, it is possible that Hmr binds via HP1a at the telomeres and is therefore not close enough to the sequence to be effectively cross-linked. Alternatively, Hmr may bind TE sequences not tested here.

5.3.5 Conclusions

The experiments described here are ongoing. The problems I have encountered reaffirm the necessity of a bona fide binding site of Hmr in order to effectively optimize the ChIP protocol. In future ChIP optimizations for proteins with less well known binding sites, it is advisable to optimize conditions such as antibody concentration, fixation times, and shearing conditions using cell culture.

It is possible that Hmr is not close enough to DNA to observe actual enrichment. Previous attempts at Lhr ChIP were performed in the lab, however, enrichment of

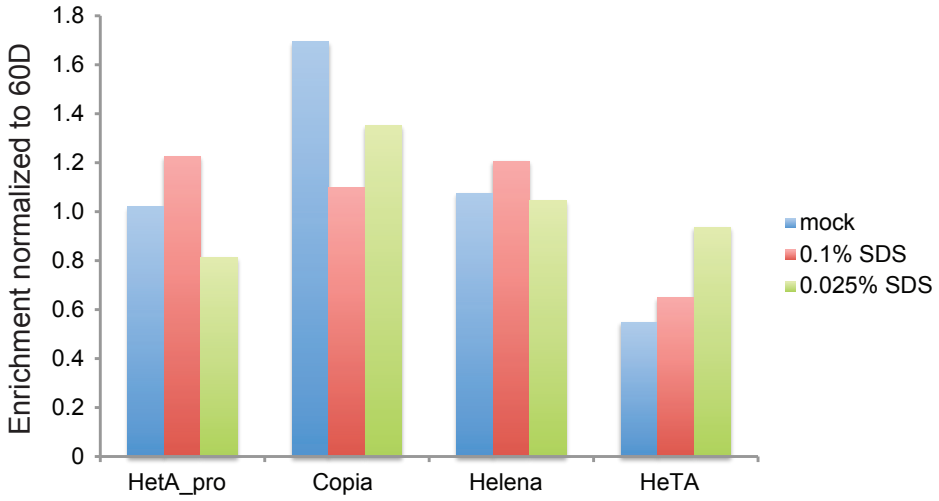


Figure 5.7. TEs are not enriched for Hmr. We used ChIP-qpcr to test for enrichment of Hmr-HA at a subset of TEs that were massively upregulated in *Hmr* mutants. The ChIP was performed as described in section 5.2.4 using the Rat anti-HA antibody (Roche 3F10, 1:10 dilution). Ovaries were dissected from *Hmr*^{3}*; ϕ{mel-Hmr-HA}* and tested for two different concentrations of SDS dilution prior to the IP. SDS is necessary for efficient sonication but needs to be diluted prior to the IP so as not to interfere with antibody binding. We did not observe noticeable differences between 0.1% and 0.025% SDS dilutions. None of the TEs tested here showed enrichment for Hmr-HA.

Lhr in the pull-downs (assayed by western), were not observed (experiments performed by P. Satyaki), indicating Lhr may not contact DNA directly. While Lhr contains a BESS domain, which is a putative protein binding domain, no DNA-binding domains have been identified. Because Hmr and Lhr physically interact, it may be that the Hmr-Lhr complex is too far from the DNA to be effectively cross-linked. However, the presence of four MADF domains strongly suggests that Hmr contacts chromatin in close enough proximity for ChIP.

Identification of Hmr binding sites would further clarify the molecular function of Hmr. While it is known that Hmr represses TEs, its mechanism is unknown. Derepression of TEs is a hallmark of piRNA pathway mutants. The piRNA pathway protects the germline from unsolicited TE activity in eukaryotes. While the post-transcriptional gene silencing (PTGS) function of the piRNA pathway is the most well characterized, there is increasing evidence for piRNA-mediated transcriptional gene silencing (TGS). Several recent genome-wide studies have suggested a role for Piwi in establishing a repressive chromatin environment at TEs due to enrichment of heterochromatic marks, such as HP1 and H3K9me3, upon depletion of Piwi (Sienski *et al.* 2012; Le Thomas *et al.* 2013; Rozhkov *et al.* 2013). The binding profile of Piwi remains unknown (Marinov *et al.* 2015), however, the piRNA-effector protein Rhino (Rhi) was recently shown to be enriched specifically at dual-strand piRNA clusters (Mohn *et al.* 2014). It is unknown whether Hmr functions in the piRNA pathway. It seems likely that Hmr binds to TEs and other repetitive sequences in heterochromatin, though I did not observe enrichment at a set of TEs highly derepressed in *Hmr* mutants. Therefore, Hmr may bind a set of TEs not tested or may not bind TEs at all. Enrichment at TEs or piRNA clusters would indicate a role for Hmr in chromatin modification or TGS. Further studies, including assessing piRNA pools in *Hmr* mutants, are required for a complete understanding of how Hmr suppresses heterochromatic repeats.

BIBLIOGRAPHY

Abad, J. P., 2004 Genomic Analysis of Drosophila melanogaster Telomeres: Full-length Copies of HeT-A and TART Elements at Telomeres. Molecular Biology and Evolution 21: 1613–1619.

Abad, J. P., M. Carmena, S. Baars, R. D. Saunders, D. M. Glover *et al.*, 1992 Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A 89: 4663–4667.

Adams, M. D., 2000 The Genome Sequence of Drosophila melanogaster. Science 287: 2185–2195.

Anders, S., and W. Huber, 2010 Differential expression analysis for sequence count data. Genome Biol. 11: R106.

Anderson, J. A., W. D. Gilliland, and C. H. Langley, 2009 Molecular population genetics and evolution of Drosophila meiosis genes. Genetics 181: 177–185.

Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166–169.

Andreyeva, E. N., E. S. Belyaeva, V. F. Semeshin, G. V. Pokholkova, and I. F. Zhimulev, 2005 Three distinct chromatin domains in telomere ends of polytene chromosomes in Drosophila melanogaster Tel mutants. J. Cell. Sci. 118: 5465–5477.

Andrulis, E. D., J. Werner, A. Nazarian, H. Erdjument-Bromage, P. Tempst *et al.*, 2002 The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature 420: 837–841.

Aruna, S., H. A. Flores, and D. A. Barbash, 2009 Reduced fertility of Drosophila melanogaster hybrid male rescue (Hmr) mutant females is partially complemented by Hmr orthologs from sibling species. Genetics 181: 1437–1450.

Barbash, D. A., 2011 Comment on “A Test of the Snowball Theory for the Rate of Evolution of Hybrid Incompatibilities.” Science 333: 1576–1576.

Barbash, D. A., 2010a Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. Genetics 184: 313–316.

Barbash, D. A., 2010b Ninety years of Drosophila melanogaster hybrids. Genetics 186: 1–8.

Barbash, D. A., 2007 Nup96-dependent hybrid lethality occurs in a subset of species from the simulans clade of Drosophila. Genetics 176: 543–552.

Barbash, D. A., P. Awadalla, and A. M. Tarone, 2004 Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. PLoS Biol 2: e142.

Barbash, D. A., and J. G. Lorigan, 2007 Lethality in Drosophila melanogaster/Drosophila simulans species hybrids is not associated with substantial transcriptional misregulation. J. Exp. Zool. B Mol. Dev. Evol 308: 74–84.

Barbash, D. A., J. Roote, and M. Ashburner, 2000 The Drosophila melanogaster hybrid male rescue gene causes inviability in male and female species hybrids. Genetics 154: 1747–1771.

Barbash, D. A., D. F. Siino, A. M. Tarone, and J. Roote, 2003 A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc. Natl. Acad. Sci. U.S.A 100: 5302–5307.

Baricheva, E. A., M. Berrios, S. S. Bogachev, I. V. Borisevich, E. R. Lapik *et al.*, 1996 DNA from Drosophila melanogaster beta-heterochromatin binds specifically to nuclear lamins in vitro and the nuclear envelope in situ. Gene 171: 171–176.

Barrón, M. G., A.-S. Fiston-Lavier, D. A. Petrov, and J. González, 2014 Population Genomics of Transposable Elements in *Drosophila*. Annual Review of Genetics 48: 561–581.

Bartolomé, C., X. Bello, and X. Maside, 2009 Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. Genome Biology 10: R22.

Bayes, J. J., and H. S. Malik, 2009 Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science 326: 1538–1541.

Bergman, C. M., and D. Bensasson, 2007 Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 104: 11340–11345.

Berman, B. P., B. D. Pfeiffer, T. R. Laverty, S. L. Salzberg, G. M. Rubin *et al.*, 2004 Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biol. 5: R61.

Bertram, M. J., and O. M. Pereira-Smith, 2001 Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. Gene 266: 111–121.

Bhaskar, V., and A. J. Courey, 2002 The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. Gene 299: 173–184.

Blumenstiel, J. P., 2011 Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet. 27: 23–31.

Blumenstiel, J. P., and D. L. Hartl, 2005 Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. Proc. Natl. Acad. Sci. U.S.A. 102: 15965–15970.

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

Bonaccorsi, S., M. Gatti, C. Pisano, and A. Lohe, 1990 Transcription of a satellite DNA on two Y chromosome loops of Drosophila melanogaster. Chromosoma 99: 260–266.

Bosco, G., P. Campbell, J. T. Leiva-Neto, and T. A. Markow, 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.

Bregliano, J.-C., and M. Kidwell, 1983 Hybrid dysgenesis determinants, in *Mob Genet Elements*, Academic Press, New York.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell 128: 1089–1103.

Brideau, N. J., and D. A. Barbash, 2011 Functional conservation of the Drosophila hybrid incompatibility gene Lhr. BMC Evol. Biol 11: 57.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang *et al.*, 2006 Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science 314: 1292–1295.

Casacuberta, E., and M.-L. Pardue, 2005 HeT-A and TART, two Drosophila retrotransposons with a bona fide role in chromosome structure for more than 60 million years. Cytogenetic and Genome Research 110: 152–159.

Castillo, D. M., and L. C. Moyle, 2012 Evolutionary Implications of Mechanistic Models of TE-Mediated Hybrid Incompatibility. International Journal of Evolutionary Biology 2012: 1–12.

Cattani, M. V., and D. C. Presgraves, 2012 Incompatibility Between X Chromosome Factor and Pericentric Heterochromatic Region Causes Lethality in Hybrids Between Drosophila melanogaster and Its Sibling Species. Genetics 191: 549–559.

Charlesworth, B., P. Sniegowski, and W. Stephan, 1994 The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215–220.

Core, L. J., J. J. Waterfall, and J. T. Lis, 2008 Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845–1848.

Coyne, J. A., S. Simeonidis, and P. Rooney, 1998 Relative paucity of genes causing inviability in hybrids between Drosophila melanogaster and D. simulans. Genetics 150: 1091–1103.

Cridland, J. M., S. J. Macdonald, A. D. Long, and K. R. Thornton, 2013 Abundance and Distribution of Transposable Elements in Two Drosophila QTL Mapping Resources. Molecular Biology and Evolution 30: 2311–2327.

Csink, A. K., and J. F. McDonald, 1995 Analysis of copia sequence variation within and between Drosophila species. Mol. Biol. Evol. 12: 83–93.

Cuykendall, T. N., P. Satyaki, S. Ji, D. M. Clay, N. B. Edelman *et al.*, 2014 A Screen for F1 Hybrid Male Rescue Reveals No Major-Effect Hybrid Lethality Loci in the Drosophila melanogaster Autosomal Genome. G3: Genes|Genomes|Genetics 4: 2451–2460.

Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde *et al.*, 2008 An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798–802.

Czech, B., J. B. Preall, J. McGinn, and G. J. Hannon, 2013 A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. Molecular Cell 50: 749–761.

Danilevskaya, O. N., K. Lowenhaupt, and M. L. Pardue, 1998 Conserved subfamilies of the Drosophila HeT-A telomere-specific retrotransposon. Genetics 148: 233–242.

Dernburg, A. F., 2011 In situ hybridization to somatic chromosomes in Drosophila. Cold Spring Harb Protoc 2011:

Dobzhansky, T., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York.

Dönertas, D., G. Sienski, and J. Brennecke, 2013 Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev. 27: 1693–1705.

Doolittle, W. F., 2013 Is junk DNA bunk? A critique of ENCODE. Proc. Natl. Acad. Sci. U.S.A. 110: 5294–5300.

Dowsett, A. P., and M. W. Young, 1982 Differing levels of dispersed repetitive DNA among closely related species of Drosophila. Proc. Natl. Acad. Sci. U.S.A. 79: 4570–4574.

Eissenberg, J. C., and S. C. Elgin, 2000 The HP1 protein family: getting a grip on chromatin. Curr. Opin. Genet. Dev. 10: 204–210.

Eissenberg, J. C., T. C. James, D. M. Foster-Hartnett, T. Hartnett, V. Ngan *et al.*, 1990 Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 87: 9923–9927.

England, B. P., A. Admon, and R. Tjian, 1992 Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc. Natl. Acad. Sci. U.S.A 89: 683–687.

Ferree, P. M., and D. A. Barbash, 2009 Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol 7: e1000234.

Ferree, P. M., and S. Prasad, 2012 How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. Genet Res Int 2012: 430136.

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

Fiumera, A. C., 2004 Sperm Competitive Ability in Drosophila melanogaster Associated With Variation in Male Reproductive Proteins. Genetics 169: 243–257.

Gao, G., J.-C. Walser, M. L. Beaucher, P. Morciano, N. Wesolowska *et al.*, 2010 HipHop interacts with HOAP and HP1 to protect Drosophila telomeres in a sequence-independent manner. The EMBO Journal 29: 819–829.

George, J. A., P. G. DeBaryshe, K. L. Traverse, S. E. Celniker, and M.-L. Pardue, 2006 Genomic organization of the Drosophila telomere retrotransposable elements. Genome Res. 16: 1231–1240.

Ghildiyal, M., H. Seitz, M. D. Horwich, C. Li, T. Du *et al.*, 2008 Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320: 1077–1081.

Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang *et al.*, 2003 A protein interaction map of Drosophila melanogaster. Science 302: 1727–1736.

Gong, W. J., and K. G. Golic, 2004 Genomic deletions of the Drosophila melanogaster Hsp70 genes. Genetics 168: 1467–1476.

González, J., K. Lenkov, M. Lipatov, J. M. Macpherson, and D. A. Petrov, 2008 High Rate of Recent Transposable Element–Induced Adaptation in Drosophila melanogaster (M. A. F. Noor, Ed.). PLoS Biology 6: e251.

González, J., and D. A. Petrov, 2012 Evolution of genome content: population dynamics of transposable elements in flies and humans. Methods Mol. Biol. 855: 361–383.

Gorman, M., A. Franke, and B. S. Baker, 1995 Molecular characterization of the male-specific lethal-3 gene and investigations of the regulation of dosage compensation in Drosophila. Development 121: 463–475.

Goshima, G., R. Wollman, S. S. Goodwin, N. Zhang, J. M. Scholey *et al.*, 2007 Genes Required for Mitotic Spindle Assembly in Drosophila S2 Cells. Science 316: 417–421.

Granzotto, A., F. R. Lopes, C. Vieira, and C. M. A. Carareto, 2011 Vertical inheritance and bursts of transposition have shaped the evolution of the BS non-LTR retrotransposon in Drosophila. Mol. Genet. Genomics 286: 57–66.

Gregory, T. R., 2005 Synergy between sequence and size in large-scale genomics. Nat. Rev. Genet. 6: 699–708.

Greil, F., E. de Wit, H. J. Bussemaker, and B. van Steensel, 2007 HP1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. EMBO J. 26: 741–751.

Groth, A. C., M. Fish, R. Nusse, and M. P. Calos, 2004 Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166: 1775–1782.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura *et al.*, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5’ end formation in Drosophila. Science 315: 1587–1590.

Haley, B., D. Hendrix, V. Trang, and M. Levine, 2008 A simplified miRNA-based gene silencing method for Drosophila melanogaster. Dev. Biol. 321: 482–490.

He, B., A. Caudy, L. Parsons, A. Rosebrock, A. Pane *et al.*, 2012 Mapping the pericentric heterochromatin by comparative genomic hybridization analysis and chromosome deletions in Drosophila melanogaster. Genome Res. 22: 2507–2519.

Hernan, R., K. Heuermann, and B. Brizzard, 2000 Multiple epitope tagging of expressed proteins for enhanced detection. BioTechniques 28: 789–793.

Hessle, V., P. Bjork, M. Sokolowski, E. Gonzalez de Valdivia, R. Silverstein *et al.*, 2009 The Exosome Associates Cotranscriptionally with the Nascent Pre-mRNP through Interactions with Heterogeneous Nuclear Ribonucleoproteins. Molecular Biology of the Cell 20: 3459–3470.

Hessle, V., A. von Euler, E. Gonzalez de Valdivia, and N. Visa, 2012 Rrp6 is recruited to transcribed genes and accompanies the spliced mRNA to the nuclear pore. RNA 18: 1466–1474.

Hickey, D. A., 1982 Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101: 519–531.

Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6 reference sequence of the *Drosophila melanogaster* genome. Genome Research 25: 445–458.

Hutter, P., and M. Ashburner, 1987 Genetic rescue of inviable hybrids between Drosophila melanogaster and its sibling species. Nature 327: 331–333.

Hutter, P., J. Roote, and M. Ashburner, 1990 A genetic basis for the inviability of hybrids between sibling species of Drosophila. Genetics 124: 909–920.

Jack, J., and Y. DeLotto, 1995 Structure and regulation of a complex locus: the cut gene of Drosophila. Genetics 139: 1689–1700.

James, T. C., J. C. Eissenberg, C. Craig, V. Dietrich, A. Hobson *et al.*, 1989 Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. Eur. J. Cell Biol 50: 170–180.

James, T. C., and S. C. Elgin, 1986 Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol. Cell. Biol. 6: 3862–3872.

Johnson, N. A., 2010 Hybrid incompatibility genes: remnants of a genomic battlefield? Trends Genet. 26: 317–325.

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany *et al.*, 2005 Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110: 462–467.

Kaminker, J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas *et al.*, 2002 The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. 3: RESEARCH0084.

Kapitonov, V. V., and J. Jurka, 2001 Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences 98: 8714–8719.

Kawamura, Y., K. Saito, T. Kin, Y. Ono, K. Asai *et al.*, 2008 Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature 453: 793–797.

Kelleher, E. S., and D. A. Barbash, 2013 Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. Mol. Biol. Evol. 30: 1816–1829.

Kelleher, E. S., N. B. Edelman, and D. A. Barbash, 2012 Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 10: e1001428.

Kerkis, J., 1933 Development of gonads in hybrids between Drosophila melanogaster an Drosophila simulans. J Exper Zool 66: 477–509.

Khurana, J. S., and W. Theurkauf, 2010 piRNAs, transposon silencing, and Drosophila germline development. J. Cell Biol. 191: 905–913.

Khurana, J. S., J. Xu, Z. Weng, and W. E. Theurkauf, 2010 Distinct Functions for the Drosophila piRNA Pathway in Genome Maintenance and Telomere Protection (R. S. Hawley, Ed.). PLoS Genetics 6: e1001246.

Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu *et al.*, 2009 The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. Cell 138: 1137–1149.

Klenov, M. S., S. A. Lavrov, A. P. Korbut, A. D. Stolyarenko, E. Y. Yakushev *et al.*, 2014 Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res. 42: 6208–6218.

Klenov, M. S., S. A. Lavrov, A. D. Stolyarenko, S. S. Ryazansky, A. A. Aravin *et al.*, 2007 Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res. 35: 5430–5438.

Klenov, M. S., O. A. Sokolova, E. Y. Yakushev, A. D. Stolyarenko, E. A. Mikhaleva *et al.*, 2011 Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc. Natl. Acad. Sci. U.S.A. 108: 18760–18765.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics 156: 1913–1931.

Kofler, R., A. J. Betancourt, and C. Schlötterer, 2012 Sequencing of Pooled DNA Samples (Pool-Seq) Uncovers Complex Dynamics of Transposable Element Insertions in Drosophila melanogaster (D. J. Begun, Ed.). PLoS Genetics 8: e1002487.

Koonin, E. V., S. Zhou, and J. C. Lucchesi, 1995 The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Res. 23: 4229–4233.

Kuai, L., B. Das, and F. Sherman, 2005 A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. 102: 13962–13967.

Kuan, Y.-S., P. Brewer-Jensen, W.-L. Bai, C. Hunter, C. B. Wilson *et al.*, 2009 Drosophila Suppressor of Sable Protein [Su(s)] Promotes Degradation of

Aberrant and Transposon-Derived RNAs. Molecular and Cellular Biology 29: 5590–5603.

Kvon, E. Z., T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani *et al.*, 2014 Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, International Human Genome Sequencing Consortium, 2001 Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Lee, A., A. K. Henras, and G. Chanfreau, 2005 Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae. Mol. Cell 19: 39–51.

Lee, Y. C. G., and C. H. Langley, 2012 Long-term and short-term evolutionary impacts of transposable elements on Drosophila. Genetics 192: 1411–1432.

Lee, Y. C. G., and C. H. Langley, 2010 Transposable elements in natural populations of Drosophila melanogaster. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365: 1219–1228.

Lerat, E., N. Burlet, C. Biémont, and C. Vieira, 2011 Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473: 100–109.

Lewis, B. P., C. B. Burge, and D. P. Bartel, 2005 Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20.

Lim, S. J., P. J. Boyle, M. Chinen, R. K. Dale, and E. P. Lei, 2013 Genome-wide localization of exosome components to active promoters and chromatin insulators in Drosophila. Nucleic Acids Research 41: 2963–2980.

Li, C., V. V. Vagin, S. Lee, J. Xu, S. Ma *et al.*, 2009 Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell 137: 509–521.

Lohe, A. R., A. J. Hilliker, and P. A. Roberts, 1993 Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics 134: 1149–1174.

Lohe, A., and P. Roberts, 1988 Evolution of satellite DNA sequences in Drosophila., pp. 148–186 in Cambridge Univ. Press.

Maheshwari, S., and D. A. Barbash, 2012 Cis-by-Trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. PLoS Genet. 8: e1002597.

Maheshwari, S., and D. A. Barbash, 2011 The genetics of hybrid incompatibilities. Annu. Rev. Genet. 45: 331–355.

Maheshwari, S., J. Wang, and D. A. Barbash, 2008 Recurrent positive selection of the Drosophila hybrid incompatibility gene Hmr. Mol. Biol. Evol 25: 2421–2430.

Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie *et al.*, 2009 Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 137: 522–535.

Marinov, G. K., J. Wang, D. Handler, B. J. Wold, Z. Weng *et al.*, 2015 Pitfalls of Mapping High-Throughput Sequencing Data to Repetitive Sequences: Piwi's Genomic Targets Still Not Identified. Dev. Cell 32: 765–771.

Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne, 2010 A Test of the Snowball Theory for the Rate of Evolution of Hybrid Incompatibilities. Science 329: 1518–1521.

Matute, D. R., and J. Gavin-Smyth, 2014 Fine mapping of dominant X-linked incompatibility alleles in Drosophila hybrids. PLoS Genet. 10: e1004270.

McDermott, S. R., and R. M. Kliman, 2008 Estimation of isolation times of the island species in the Drosophila simulans complex from multilocus DNA sequence data. PLoS ONE 3: e2442.

Mefford, H. C., and B. J. Trask, 2002 The complex structure and dynamic evolution of human subtelomeres. Nat. Rev. Genet. 3: 91–102.

Mohn, F., G. Sienski, D. Handler, and J. Brennecke, 2014 The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell 157: 1364–1379.

Muller, H., 1940 *The New Systematics, ed Husley JS*. The Clarendon Press, Oxford.

Muller, H. J., and G. Pontecorvo, 1940 Recombinants between Drosophila Species the F1 Hybrids of which are Sterile. Nature 146: 199–200.

Ni, J.-Q., R. Zhou, B. Czech, L.-P. Liu, L. Holderbaum *et al.*, 2011 A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nature Methods 8: 405–407.

Orr, H. A., and S. Irving, 2000 Genetic analysis of the hybrid male rescue locus of Drosophila. Genetics 155: 225–231.

Orsi, G. A., E. F. Joyce, P. Couble, K. S. McKim, and B. Loppin, 2010 Drosophila I-R hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. J. Cell. Sci. 123: 3515–3524.

Pal Bhadra, M., U. Bhadra, and J. A. Birchler, 2006 Misregulation of sex-lethal and disruption of male-specific lethal complex localization in Drosophila species hybrids. Genetics 174: 1151–1159.

Pane, A., P. Jiang, D. Y. Zhao, M. Singh, and T. Schüpbach, 2011 The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J. 30: 4601–4615.

Pane, A., K. Wehr, and T. Schüpbach, 2007 zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev. Cell 12: 851–862.

Pardue, M.-L., and P. Debaryshe, 2011 Adapting to life at the end of the line: How Drosophila telomeric retrotransposons cope with their job. Mob Genet Elements 1: 128–134.

Pérez Sánchez, C., S. Casas-Tintó, L. Sánchez, J. Rey-Campos, and B. Granadino, 2002 DmFoxF, a novel Drosophila fork head factor expressed in visceral mesoderm. Mechanisms of Development 111: 163–166.

Perrat, P. N., S. DasGupta, J. Wang, W. Theurkauf, Z. Weng *et al.*, 2013 Transposition-driven genomic heterogeneity in the Drosophila brain. Science 340: 91–95.

Petit, N., D. Piñeyro, E. López-Panadès, E. Casacuberta, and A. Navarro, 2012 HeT-A_pi1, a piRNA Target Sequence in the Drosophila Telomeric Retrotransposon HeT-A, Is Extremely Conserved across Copies and Species (C. Lalueza-Fox, Ed.). PLoS ONE 7: e37405.

Petrov, D. A., 2002 Mutational equilibrium model of genome size evolution. Theor Popul Biol 61: 531–544.

Petrov, D. A., A.-S. Fiston-Lavier, M. Lipatov, K. Lenkov, and J. Gonzalez, 2011 Population Genomics of Transposable Elements in Drosophila melanogaster. Molecular Biology and Evolution 28: 1633–1644.

Piñeyro, D., E. López-Panadès, M. Lucena-Pérez, and E. Casacuberta, 2011 Transcriptional analysis of the HeT-A retrotransposon in mutant and wild type stocks reveals high sequence variability at Drosophila telomeres and other unusual features. BMC Genomics 12: 573.

Platero, J. S., A. K. Csink, A. Quintanilla, and S. Henikoff, 1998 Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in Drosophila. J. Cell Biol. 140: 1297–1306.

Pontecorvo G, 1943 Viability interactions between chromosomes of Drosophila melanogaster and Drosophila simulans. Journal of Genetics 45: 51–66.

Presgraves, D. C., 2003 A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. Genetics 163: 955–972.

Presgraves, D. C., 2010 The molecular evolutionary basis of species formation. Nat. Rev. Genet 11: 175–180.

Quinlan, A. R., R. A. Clark, S. Sokolova, M. L. Leibowitz, Y. Zhang *et al.*, 2010 Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. 20: 623–635.

Raffa, G. D., L. Ciapponi, G. Cenci, and M. Gatti, 2011 Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. Nucleus 2: 383–391.

Rangan, P., C. D. Malone, C. Navarro, S. P. Newbold, P. S. Hayes *et al.*, 2011 piRNA production requires heterochromatin formation in Drosophila. Curr. Biol. 21: 1373–1379.

Rausch, T., T. Zichner, A. Schlattl, A. M. Stutz, V. Benes *et al.*, 2012 DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28: i333–i339.

Rebollo, R., E. Lerat, L. L. Kleine, C. Biémont, and C. Vieira, 2008 Losing helena: the extinction of a drosophila line-like element. BMC Genomics 9: 149.

Riddle, N. C., A. Minoda, P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz *et al.*, 2011 Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res. 21: 147–163.

Rodriguez, M. A., D. Vermaak, J. J. Bayes, and H. S. Malik, 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 104: 15412–15417.

Ross-Innes, C. S., R. Stark, A. E. Teschendorff, K. A. Holmes, H. R. Ali *et al.*, 2012 Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481: 389–393.

Roth, K. M., M. K. Wolf, M. Rossi, and J. S. Butler, 2005 The nuclear exosome contributes to autogenous control of NAB2 mRNA levels. Mol. Cell. Biol. 25: 1577–1585.

Rozhkov, N. V., A. A. Aravin, E. S. Zelentsova, N. G. Schostak, R. Sachidanandam *et al.*, 2010 Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. RNA 16: 1634–1645.

Rozhkov, N. V., M. Hammell, and G. J. Hannon, 2013 Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev. 27: 400–412.

Ruby, J. G., A. Stark, W. K. Johnston, M. Kellis, D. P. Bartel *et al.*, 2007 Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 17: 1850–1864.

Sage, B. T., and A. K. Csink, 2003 Heterochromatic self-association, a determinant of nuclear organization, does not require sequence homology in Drosophila. Genetics 165: 1183–1193.

Saito, K., K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi *et al.*, 2006 Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 20: 2214–2222.

Satyaki, P. R. V., T. N. Cuykendall, K. H.-C. Wei, N. J. Brideau, H. Kwak *et al.*, 2014 The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet. 10: e1004240.

Savitsky, M., O. Kravchuk, L. Melnikova, and P. Georgiev, 2002 Heterochromatin protein 1 is involved in control of telomere elongation in Drosophila melanogaster. Mol. Cell. Biol. 22: 3204–3218.

Sawamura, K., 2012 Chromatin evolution and molecular drive in speciation. Int J Evol Biol 2012: 301894.

Sawamura, K., M. T. Yamamoto, and T. K. Watanabe, 1993 Hybrid lethal systems in the Drosophila melanogaster species complex. II. The Zygotic hybrid rescue (Zhr) gene of D. melanogaster. Genetics 133: 307–313.

Shepelev, V. A., A. A. Alexandrov, Y. B. Yurov, and I. A. Alexandrov, 2009 The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet. 5: e1000641.

Shpiz, S., and A. Kalmykova, 2012 Control of Telomere Length in Drosophila., pp. 33–56 in *Li, B, editor. Review of Selected Topic of Telomere Biology.*, Rijeka: InTech.

Sienski, G., D. Dönertas, and J. Brennecke, 2012 Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964–980.

Siomi, M. C., K. Sato, D. Pezic, and A. A. Aravin, 2011 PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. 12: 246–258.

Smith, C. D., R. C. Edgar, M. D. Yandell, D. R. Smith, S. E. Celniker *et al.*, 2007 Improved repeat identification and masking in Dipterans. Gene 389: 1–9.

Sturtevant, A. H., 1920 Genetic Studies on DROSOPHILA SIMULANS. I. Introduction. Hybrids with DROSOPHILA MELANOGASTER. Genetics 5: 488–500.

Tang, S., and D. C. Presgraves, 2009 Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. Science 323: 779–782.

Thomae, A. W., G. O. M. Schade, J. Padeken, M. Borath, I. Vetter *et al.*, 2013 A Pair of Centromeric Proteins Mediates Reproductive Isolation in Drosophila Species. Developmental Cell 27: 412–424.

Le Thomas, A., A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao *et al.*, 2013 Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 27: 390–399.

Ting, C. T., S. C. Tsaur, M. L. Wu, and C. I. Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282: 1501–1504.

Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan *et al.*, 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511–515.

Usakin, L., J. Abad, V. V. Vagin, B. de Pablos, A. Villasante *et al.*, 2007 Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in Drosophila melanogaster ovaries. Genetics 176: 1343–1349.

Vagin, V. V., A. Sigova, C. Li, H. Seitz, V. Gvozdev *et al.*, 2006 A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313: 320–324.

Venken, K. J. T., Y. He, R. A. Hoskins, and H. J. Bellen, 2006 P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314: 1747–1751.

Villasante, A., J. P. Abad, R. Planelló, M. Méndez-Lago, S. E. Celniker *et al.*, 2007 Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res. 17: 1909–1918.

Walker, P. M., 1971 Origin of satellite DNA. Nature 229: 306–308.

Wang, S. H., and S. C. R. Elgin, 2011 Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. U.S.A. 108: 21164–21169.

Wang, L., Z. Feng, X. Wang, X. Wang, and X. Zhang, 2010 DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26: 136–138.

Watanabe, T. K., 1979 A gene that rescues the lethal hybrids between Drosophila melanogaster and Drosophila simulans. Jpn. J. Genet. 54: 325–331.

Wei, K. H.-C., A. G. Clark, and D. A. Barbash, 2014 Limited Gene Misregulation Is Exacerbated by Allele-Specific Upregulation in Lethal Hybrids between Drosophila melanogaster and Drosophila simulans. Molecular Biology and Evolution 31: 1767–1778.

Werren, J. H., 2011 Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc. Natl. Acad. Sci. U.S.A. 108 Suppl 2: 10863–10870.

Wu, H., H. Ma, C. Ye, D. Ramirez, S. Chen *et al.*, 2011 Improved siRNA/shRNA functionality by mismatched duplex. PLoS ONE 6: e28580.

Yamanaka, S., S. Mehta, F. E. Reyes-Turcu, F. Zhuang, R. T. Fuchs *et al.*, 2012 RNAi triggered by specialized machinery silences developmental genes and retrotransposons. Nature 493: 557–560.

Yasuhara, J. C., and B. T. Wakimoto, 2006 Oxymoron no more: the expanding world of heterochromatic genes. Trends Genet. 22: 330–338.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler, and Z. Ning, 2009 Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25: 2865–2871.

Zachar, Z., and P. M. Bingham, 1982 Regulation of white locus expression: the structure of mutant alleles at the white locus of Drosophila melanogaster. Cell 30: 529–541.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson *et al.*, 2008 Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9: R137.

Zhang, Y., J. H. Malone, S. K. Powell, V. Periwal, E. Spana *et al.*, 2010 Expression in aneuploid Drosophila S2 cells. PLoS Biol. 8: e1000320.

Zwick, M. E., J. L. Salstrom, and C. H. Langley, 1999 Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in Drosophila melanogaster. Genetics 152: 1605–1614.