2018-05-31

Adaptive Evolution of piRNA pathway in Drosophila

Swapnil S. Parhad
University of Massachusetts Medical School

ADAPTIVE EVOLUTION OF PIRNA PATHWAY IN *DROSOPHILA*

A Dissertation Presented

By

SWAPNIL S. PARHAD

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 31, 2018

Program in Molecular Medicine

ADAPTIVE EVOLUTION OF PIRNA PATHWAY IN *DROSOPHILA*

A Dissertation Presented

By

SWAPNIL S. PARHAD

This work was undertaken in the Graduate School of Biomedical Sciences

INTERDISCIPLINARY GRADUATE PROGRAM

Under the mentorship of

William E. Theurkauf, Ph.D., Thesis Advisor

Zhiping Weng, Ph.D., Member of Committee

Phillip D. Zamore, Ph.D., Member of Committee

Manuel Garber, Ph.D., Member of Committee

Cedric Feschotte, Ph.D., External Member of Committee

Jeremy Luban, M.D., Chair of Committee

Mary Ellen Lane, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

May 31, 2018

DEDICATION

To my family, friends and teachers

Many times, they play all the roles…

ACKNOWLEDGEMENTS

First of all, I would like to thank my thesis advisor Bill. He is an awesome mentor, great scientist and thinker and amazing person. It always felt like that I am working with him, not for him. He has created a great fun filled atmosphere for doing Science. He always amazes me with his great ideas. Despite being well established in the field, he remains down to earth person and is always receptive towards new ideas from anyone. He is always there to discuss about projects, to look under the microscope whenever we had new results. I always enjoyed chatting with him about projects and never felt intimidated to pitch models different than his. He taught me "to change my model based on data, not data based on model". It never felt stressful to get a particular result. All the experiments felt like a mystery to be unfolded, whatever the result may be. I am deeply indebted to Bill, for setting up a great example not only as a scientist, but also as amazing human being.

Second person I would like to thank is Zhiping. She is great scientist and a very nice person. She has always helped me with any question I have. It would not have been possible to do all the projects without her able guidance, especially for Bioinformatics analysis. I am very grateful for her efforts to make sure we have all the means necessary for our data analysis. I am astonished by her contagious enthusiasm for Science and ability to learn different topics. She always motivates me to learn more about Computation so that I can independently analyze my own data. I will always be grateful to Zhiping for her guidance. Bill reminds me of my dad and Zhiping of my mom. They

make the lab feel like home. Sometimes I feel like a pampered kid. When I was young, I used to ask for video games. Now I ask for sequencing…

I would like to thank my lab-mates for maintaining a fun filled environment to work in. First of all, I want to thank our lab manager Birgit. I cannot imagine how our lab would run without her. She is a great colleague and awesome lab manager. She is always there if you are busy and need some help. She always made sure we have all reagents necessary to do our experiments. She taught me how to use Confocal and image processing. I have learnt a lot about both Scientific and non-Scientific things from her. I want to thank Nadine for teaching me fly genetics and its tricks when I joined the lab. She showed how fly pushing can be fun. I want to thank all the past and current members of Theurkauf lab for teaching me various techniques along the way, for their insightful discussions and for being lifelong friends. Zhao for taught me various deep sequencing techniques. I miss playing tennis with Zhao. Fan helped with various molecular biology techniques. Alfred taught me Python and always encourages me to learn more about it. I want to thank Travis and Gen for being in the lab during late hours. It was always nice to talk to someone amongst the humming of instruments. I want to thank Nick and Sam. It is a lot of fun to talk to them about anything. They bring to light different perspectives. I want to thank Anetta, Subhanita and Anna for all the discussions.

I want to thank the past and present members of Zhiping's lab for their help in Bioinformatic analysis. All the work would not have been possible without their able support. Jessie introduced me to Bioinformatics. She laid the foundations for many of the data analysis pipelines. I want to thank Shikui and Tianxiong (Bear) for writing the

pipelines which we can use. They were and still promptly and happily respond to any questions about analysis. I would have been lost without their expert help and advice. I want to thank Jiali for his help with TEMP analysis. I am grateful to Hao, Michael and Arjan for taking care of the backup servers and genome browsers. They have always promptly responded and fixed any bugs. I appreciate help from Wei for Ping-Pong analysis, Xiao-Ou for splicing analysis and Thom and Sweta for Structural analysis.

I want to thank my thesis committee members Phil, Jeremy, Manuel and Ollie. They brought different views, asked great questions and guided me about my projects. I especially want to thank Phil and members from his lab Gwen, Chengjian, Zhao, Tim, Bo, Xuan, Stephan, Cindy, Daniel, and Tiffanie for their help with reagents, protocols and discussions about the projects. I want to thank Cedric Feschotte for being the external member of the examination committee. I want to thank Julius Brennecke, Kyoichi Sawamura, Trudi Schupbach, Dan Barbash, Jim Kadonaga, and Benjamin Loppin for generously sharing reagents and fly stocks. I appreciate the collaborative environment in the RNA biology community at UMass. It was always nice to see Victor in the audience. He would ask great questions even in a boring talk and make the topic more interesting. I am grateful to IGP coordinators Darla and Joan, GSBS office and UMass Immigration office for their administrative support. I want to thank John Leszyk for Mass spectrometry analysis and sequencing core for their assistance.

I want to thank my friends, who made UMass feel like home away from home. Mayuri, Ankita, Vahbiz, Mihir, Sneha, Varun, Siva, Sandhya, Karthik, Ankit, Pallavi & Pallavi, Sreya,were always there to hang out and for any help. I want to thank Ami,

Hemal, Sonal, Divya, Naveen, Paurav, Meetu, Tyler, Lucy, Dwijit, Gautam, Aparna, Satyajeet, Harshada, Sam, Seemin, Sri, Rehan, Rajesh, Rohit, Sudesh, Aditya, Arvind, Anubhab, Harish, Adi, Sungwook, Amena, Jogi, Eashwar, Djade, Melissa, Emiliano, Bill, Priyanka, Wazim, Sumeet, Shubham, Sanchaita, Srikanth, Vijay, Sivapriya, Sourav, Helen, Ei, Ben, Lin, Ratna, Nisha, James for keeping life entertained. I want to thank my TIFR friends Ashwin, Neha, Hari, Mugdha, Areen, Hema, Satyaki, Suman, Saumya, Ajeet, Moumita, Rubul, Saikesav, Dhanno, Arkarup, Sanchari, Angika, Vishal, Dida, Vasudha, Deepika, Tanay, Arun, Rakshita, Ashwin, Avinash for all the fun times.

I am always grateful to my parents, who always strived hard to give me best education opportunities, believed in me to pursue anything I want and their unconditional love, support and care. I would be nothing without their efforts and care. I want to thank my sister Sandhya and my cousins for their love. Last but not the least, I want to thank my wife Nishi for her love and support during the journey. I feel fortunate to have met her.

ABSTRACT

Major fraction of eukaryotic genomes is composed of transposons. Mobilization of these transposons leads to mutations and genomic instability. In animals, these selfish genetic elements are regulated by a class of small RNAs called PIWI interacting RNAs (piRNAs). Thus host piRNA pathway acts as a defense against pathogenic transposons. Many piRNA pathway genes are rapidly evolving indicating that they are involved in a host-pathogen arms race. In my thesis, I investigated the nature of this arms race by checking functional consequences of the sequence diversity in piRNA pathway genes.

In order to study the functional consequences of the divergence in piRNA pathway genes, we swapped piRNA pathway genes between two sibling Drosophila species, *Drosophila melanogaster* and *Drosophila simulans*. We focused on RDC complex, composed of Rhino, Deadlock and Cutoff, which specifies piRNA clusters and regulates transcription from clusters. None of the *D. simulans* RDC complex proteins function in *D. melanogaster*. Rhino and Deadlock interact and colocalize in *D. simulans* and *D. melanogaster*, but *D. simulans* Rhino does not bind *D. melanogaster* Deadlock, due to substitutions in the rapidly evolving Shadow domain. Cutoff from *D. simulans* stably binds and traps *D. melanogaster* Deadlock. Adaptive evolution has thus generated cross-species incompatibilities in the piRNA pathway which may contribute in reproductive isolation.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

COPYRIGHT NOTICE

Chapter I

Introduction

Summary

Transposons are major genome constituents that can mobilize and trigger mutations and DNA damage. Transposon silencing is particularly critical in the germline, which is dedicated to transmission of the inherited genome, and PIWI interacting RNAs (piRNAs) guide a host defense system that transcriptionally and post-transcriptionally silences transposons during germline development. Germline control of transposons by the piRNA pathway is conserved, but many piRNA pathway genes are evolving rapidly under positive selection, and the piRNA biogenesis machinery shows remarkable phylogenetic diversity. Conservation of core function combined with rapid gene evolution is characteristic of a host-pathogen arms race. We review evidence that transposons and the piRNA pathway are engaged in an evolutionary tug of war, and propose that unique features of this arms race may help drive speciation.

1.1. Brief Introduction

Single celled organisms to complex animals face the threat of pathogens, which are countered by powerful adaptive and innate immune systems. However, the targets of host defense systems can mutate to evade detection, and pathogens frequently express inhibitors that suppress the host immune response. Host-pathogen interactions thus lead to cycles of adaptive evolution, with positive selection acting on pathogen mutations that evade host defenses and allow propagation, leading to positive selection of host mutations that restore pathogen control. This drives a "Red Queen arms race" characterized by rapid co-evolution of interacting host and pathogen genes (Van Valen, 1973). Transposons are integral genome constituents that can mobilize and cause mutations and genomic instability, and piRNA pathway functions as the host defense against these pathogens. Many piRNA pathway genes show evidence of adaptive evolution (Blumenstiel et al., 2016), suggesting that they are engaged in an arms race with the transposons they control. Here we contrast the conserved mechanisms that drive transposon replication with the divergent processes that produce the piRNAs that silence them and speculate that this unending tug of war may have profound evolutionary consequences.

1.2. Transposons: Genomic pathogens

1.2.1. Diverse transposons, conserved transposition mechanisms

Transposons (aka transposable elements or jumping genes) were discovered by Barbara McClintock through cytogenetic analysis of mosaic pigmentation patterns in maize kernels (Fedoroff, 2012; McClintock, 1950). Since this initial finding, transposons have been found in essentially every organism (Bao et al., 2015; Canapa et al., 2015). They are also remarkably diverse. For example, there are over 120 transposon families in *Drosophila melanogaster* (*D. melanogaster*). However, these diverse mobile elements move by a very limited number of transposition mechanism (figure 1.1) (Curcio and Derbyshire, 2003; Huang et al., 2012).

DNA transposable elements move by a cut-paste mechanism, and encode a transposase that recognizes inverted terminal repeats, catalyzing excision of an existing element and integration of the excised double stranded DNA intermediate into a new site. Examples of DNA transposon are P-elements in *Drosophila* and Tc1 elements *Caenorhabditis elegans* (Bingham et al., 1982; Rosenzweig et al., 1983). Helitrons move through a single stranded DNA intermediate, which leaves the donor element intact. The helitron transposase nicks one end of a donor element and the target site, and the 3' end of the target is ligated to the 5' end of the donor element. Rolling circle replication from the 3' end of the donor nick displaces one strand of the transposon and generates a new second strand of the donor copy. The displaced transposon copy forms an acceptor heteroduplex with a loop containing the new helitron. Replication then resolves the

heteroduplex, generating a new copy of the element (Figure 1.1) (Feschotte and Wessler, 2001; Kapitonov and Jurka, 2001; Thomas and Pritham, 2015). Retrotransposons move by a copy-paste mechanism with an RNA intermediate. These elements are related to retroviruses and encode a reverse transcriptase that makes a DNA copy from a transposon transcript, which is integrated into a new site. A subset of retrotransposons encode gag, pol and env proteins, and can make viral particles. These endogenous retroviruses can move from cell to cell, or from animal to animal, leading to horizontal transfer. For example, *Drosophila gypsy* elements expressed in the somatic follicle cells of the ovary "infect" adjacent germline cells (Song et al., 1997).

Retrotransposons are further subdivided by structure and replication capacity. Elements that have long terminal repeats and encode reverse transcriptase are termed LTR retrotransposons, and include Ty elements in *Saccharomyces cerevisiae* and Burdock from *Drosophila* (Boeke et al., 1985; Ponomarenko et al., 1997). A subset of retrotransposons lack LTRs, and are classified as long interspersed elements (LINEs) or short interspersed elements (SINEs). LINEs are autonomous and encode a reverse transcriptase and endonuclease that mediates transposition, while SINEs are non-autonomous and use LINE encoded enzymes to move. Jockey from *D. melanogaster* and L1 in mammals are examples of non-LTR retrotransposons (Fanning and Singer, 1987; Priimagi et al., 1988). The diverse transposon families that populate the genomes of single celled organisms to complex animals thus move by a limited number of highly conserved mechanisms.

Figure 1.1

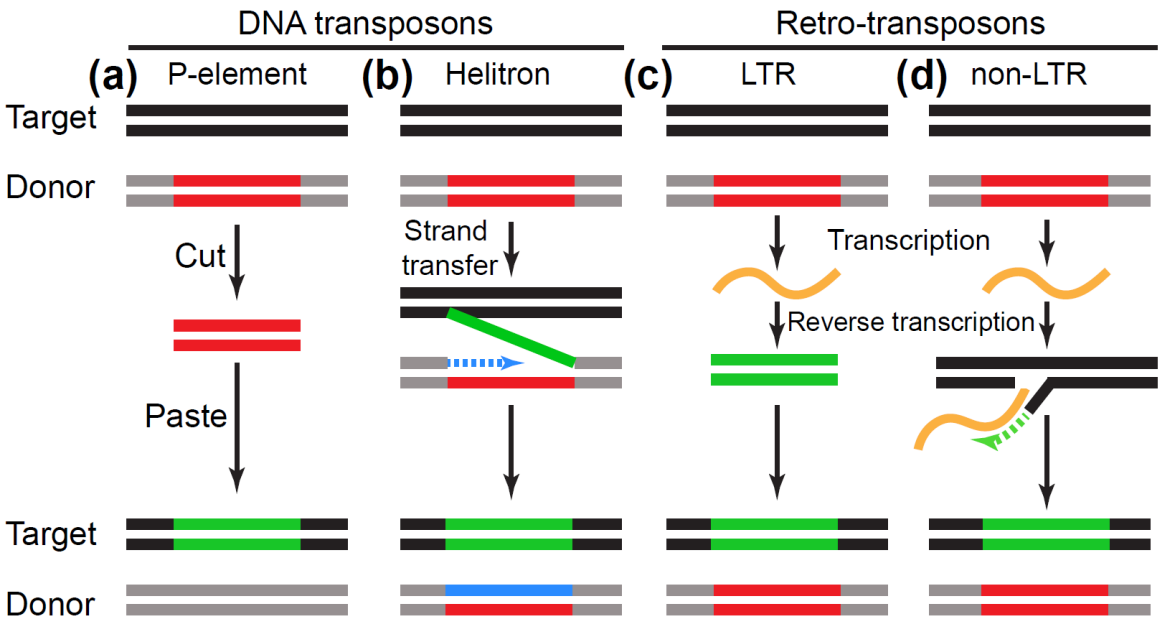

DNA transposons
Retro-transposons
(a) P-element
(b) Helitron
(c) LTR
(d) non-LTR
Target
Donor
Cut
Paste
Strand transfer
Transcription
Reverse transcription
Target
Donor

Figure 1.1. Mechanism of transposition for different transposons

Transposition mechanisms for major eukaryotic transposons. (a) and (b) are DNA transposons which transpose through a DNA intermediate. (c) and (d) are RNA or retro-transposons which transpose through a RNA intermediate. Target and donor sites are shown in black and grey respectively. Old and new TEs are shown in red and green respectively. Examples of such transposons are denoted. (a) DNA transposons, such as P-elements, excise from the donor and insert into a new site. (b) Helitrons transfer one DNA strand from the donor to the recipient site. The donor site synthesizes a new strand (shown in blue). The recipient site also synthesizes a new strand. (c) LTR retro-transposons transcribe into RNA. This RNA is reverse transcribed and inserted into a new site. (d) Non-LTR retro-transposons also transcribe into RNA. The RNA is reverse-transcribed at the insertion site. Thus, the original donor site is unaffected for retro-transposons.

1.2.2. Pathogenic characteristics of transposons

1.2.2.1. Transposons have deleterious effects on host

Transposons can disrupt the host genome by inducing insertional mutations, facilitating ectopic recombination and rearrangements, and triggering DNA breaks. Transposon insertions into exons can disrupt coding sequence leading to truncated proteins, while intron insertions can alter splicing and disrupt gene function, or generate novel and potentially deleterious fusion proteins. Promoter or enhancer insertions can change the expression pattern of genes whereas insertions in 5' or 3' UTRs can affect the gene regulation. Transposon induced mutations have been linked to cancer and other diseases (Ayarpadikannan and Kim, 2014; Belancio et al., 2008). Transposition also leads to DNA nicks and double strand breaks, and errors in repair of these lesions can lead to recombination between transposon repeats, triggering chromosomal duplications, deletions, translocations and inversions (Hedges and Deininger, 2007). Limiting the "pathogenic" consequences of transposition is therefore essential to maintaining genome integrity.

1.2.2.2. Transposons can spread between populations

The abundance of transposons in the genome reflects a combination of replication in germ cells, leading to vertical transmission of new insertions, and assembly of virus-like particles that infect new hosts, leading to horizontal transmission. Multiple instances of horizontal transfer of TEs have been observed between both closely related and distant species (Gilbert and Feschotte, 2018; Schaack et al., 2010). For example, P-elements are

DNA transposable elements that recently moved from *Drosophila willistoni* into *D. melanogaster*. These species are separated by approximately 50 MYA, but the sequence of the P-elements they harbor differ only by one nucleotide (Daniels et al., 1990). P-elements have swept through wild populations of *D. melanogaster* over the past few decades (Engels, 1992), and the same element is currently spreading through wild *Drosophila simulans* (*D. simulans*) populations (Kofler et al., 2015). Transposon family sequences in distinct *Drosophila* species generally show less sequence divergence than protein coding genes (Bartolome et al., 2009; Lerat et al., 2011; Sanchez-Gracia et al., 2005). For example, Piwi, which binds the piRNAs that silence P elements, shows 33% divergence between *D. melanogaster* and *D. willistoni*. Similar patterns are observed in animals and plants (reviewed in (Schaack et al., 2010)). SPIN family of transposons show horizontal transfer in mammals and tetrapods (Pace et al., 2008), and horizontal transfer of Tc1 like transposons is observed in fishes and frogs (Leaver, 2001). Horizontal TE transfer thus appears to be widespread, requiring a host defense system that can adapt to new invaders.

1.3. piRNA pathway: Host Immune response for transposons

Animals produce small PIWI interacting RNAs (piRNAs) to control transposons, which represent both endogenous genome pathogens and exogenous invaders (Ghildiyal and Zamore, 2009). With exogenous viruses or bacteria, the immune response is mounted after infection. The piRNA pathway, by contrast, must continuously suppress

TEs that are integral genome components, and respond to invasion of new pathogens. piRNA biogenesis and function have been extensively studied in flies (Huang et al., 2017), mice (Fu and Wang, 2014), worms (Kasper et al., 2014), but have also been characterized in planarians (Friedlander et al., 2009), fish (Houwing et al., 2007), chicken (Rengaraj et al., 2014) and humans (Ha et al., 2014). However, transposon control by piRNAs is best understood in model organisms.

piRNAs were identified in *Drosophila*, through an analysis of *Stellate* (*Ste*) silencing by the *Suppressor of Stellate (Su(Ste))* locus (Aravin et al., 2001). In this system, mutation in *Su(Ste)* leads to male sterility and over-expression of *Ste* protein, with assembles into prominent needle-like crystals in the testes (Bozzetti et al., 1995; Palumbo et al., 1994). Aravin et al showed that *Su(Ste)* encodes short RNAs that are complementary to *Ste,* and that mutations in *SpnE,* subsequently shown to be required for piRNA production, lead to over-expression of Ste and a subset of transposons. Subsequent analysis of the tissue distribution of short RNAs, performed by direct cloning and sequencing, identified a 23-30 nt long RNAs matching transposons in germline tissue. While miRNAs and siRNAs are produced from double stranded precursors by Dicer endonuclease cleavage, production of these germline enriched small RNAs is Dicer independent. Similar small RNAs were subsequently found in mouse testes, and shown to bind to the mouse homologs of Piwi, a *Drosophila* protein required for germline development (Cox et al., 1998; Lin and Spradling, 1997). Piwi is founding member of the PIWI clade of Argonaute proteins, and these novel small RNAs were therefore named Piwi-interacting RNAs (piRNAs).

Figure 1.2

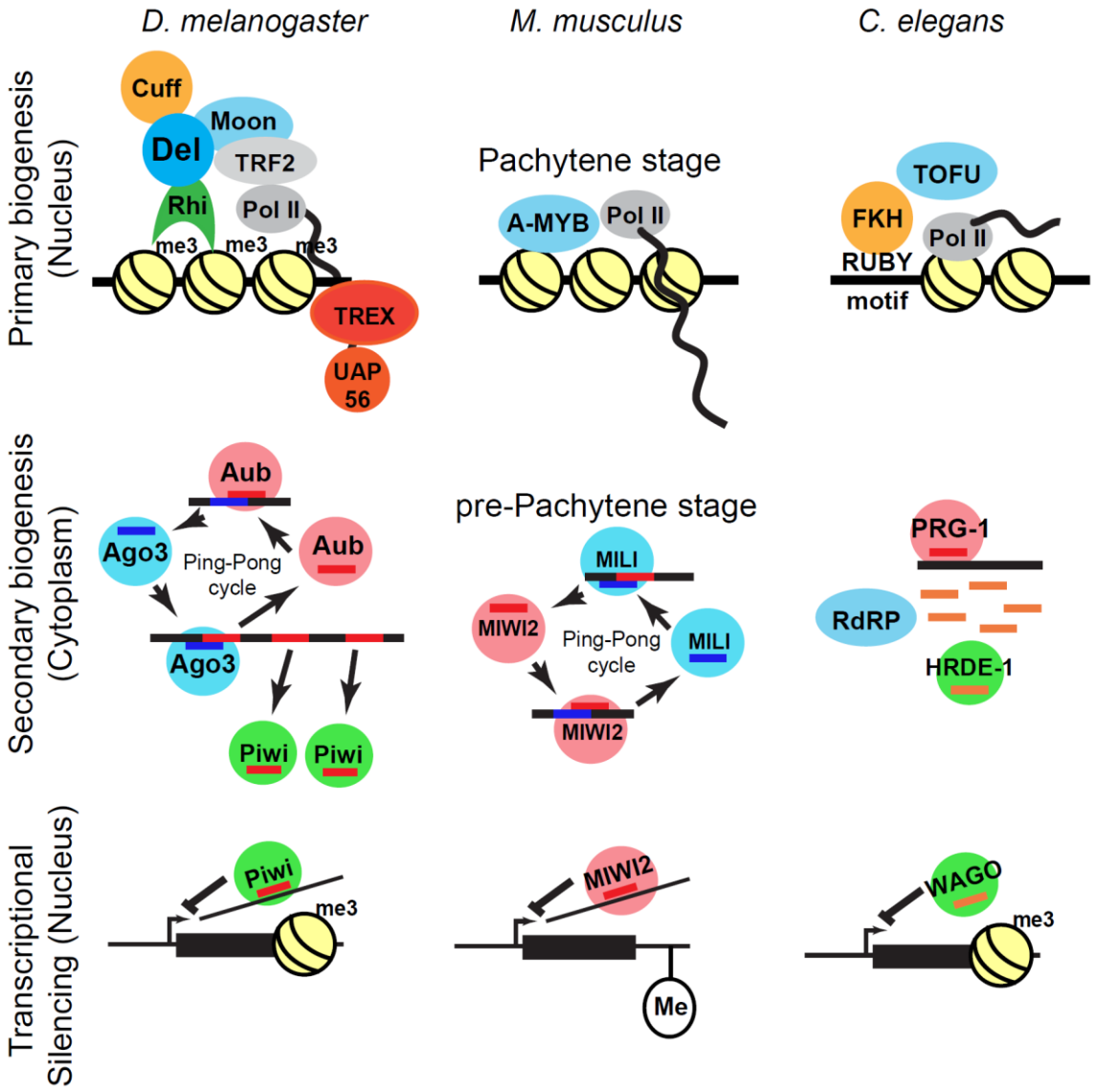

D. melanogaster
M. musculus
C. elegans
Primary biogenesis (Nucleus)
Cuff
Moon
Del
TRF2
Rhi
Pol II
me3
me3
me3
TREX
UAP 56
Pachytene stage
A-MYB
Pol II
TOFU
FKH
Pol II
RUBY motif
Secondary biogenesis (Cytoplasm)
Aub
Ago3
Ping-Pong cycle
Aub
Ago3
Piwi
Piwi
pre-Pachytene stage
MILI
MIWI2
Ping-Pong cycle
MILI
MIWI2
PRG-1
RdRP
HRDE-1
Transcriptional Silencing (Nucleus)
Piwi
me3
MIWI2
Me
WAGO
me3

Figure 1.2. piRNA biogenesis mechanisms in different model organisms

Simplified models for piRNA biogenesis is shown for *D. melanogaster* (flies), *M. musculus* (mice) and *C. elegans* (worms). Top row shows primary piRNA biogenesis mechanism at piRNA clusters inside nucleus, middle row shows secondary piRNA biogenesis in cytoplasm and the bottom row shows transcriptional silencing by piRNAs. In flies, piRNA clusters are marked by H3K9me3. The transcription by RNA Pol II is facilitated by RDC complex with Moonshiner and TRF2. The transcripts are bound by TREX and UAP56. In the cytoplasm, these transcripts are processed into piRNAs by Ping-Pong amplification cycle. The phased piRNAs are further bound by Piwi which can lead to transcriptional silencing by directing histone modification. In mice, A-MYB acts as transcription factor for pachytene piRNA cluster transcription by RNA Pol II. The Ping-pong cycle in the pre-pachytene stage leads to amplification of piRNAs. piRNA bound MIWI2 can silence transposons by directing DNA methylation. In worms, each piRNA cluster encodes for one piRNA and has its own promoter identified by Ruby motif. The RNA Pol II mediated transcription is directed by Forkhead and TOFU proteins. Recognition of target by piRNA bound PRG-1 leads to generation of 22G-RNAs. These secondary 22G-RNAs can mediated transcriptional silencing when in complex with worm specific Argonautes (WAGOs).

1.3.1. piRNA biogenesis in *Drosophila*

In flies, mutations that disrupt the piRNA transposon silencing system lead to female sterility and defects in embryonic patterning that can be easily quantified by visual inspection of the eggs produced by mutant females (Klattenhoff et al., 2007). At the time piRNAs were first described, maternal genetic control of embryonic patterning was a mature field, but the molecular function of many of the genes required for embryonic patterning were not well understood. However, germline genome instability and activation of damage signaling through ATR and Chk2 kinases, in both DNA repair and piRNA pathway mutations, were found to trigger these distinctive patterning defects (Klattenhoff et al., 2007). As a result, previously identified patterning genes (Schupbach and Wieschaus, 1989, 1991), and new mutations identified in genome-wide screens for mutations triggering patterning defects and transposon over-expression (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013), led to the rapid identification of numerous component of the *Drosophila* piRNA pathway. These studies, with advances in high-throughput sequencing of small RNAs and a high quality genome assembly, helped define the machinery that produces piRNA precursors, processes these long RNAs into mature piRNAs, and silences their targets.

1.3.1.1. Primary piRNA biogenesis

The primary piRNAs that initiate transposon silencing are derived from specific genomic loci, called piRNA clusters, composed of nested transposon insertions, which function as archive of transposon sequences (figure 1.3a) (Bergman et al., 2006;

Brennecke et al., 2007). *Drosophila* ovaries are composed of cysts containing the germline nurse cells and oocyte, surrounded by a monolayer of somatic follicle cells. In the germline, the dominant clusters contain random transposon arrays and produce piRNAs from both genomic strands. In the follicle cells, by contrast, clusters produce piRNAs from one strand, and transposon fragments are strongly biased in the anti-sense direction relative to transcription (Malone et al., 2009). Fly ovaries thus produce piRNAs targeting transposons by two distinct mechanisms.

In the *Drosophila* germline, the dominant piRNA clusters are bound by the HP1 homolog Rhino, which forms a complex with the linker Deadlock (Klattenhoff et al., 2009; Le Thomas et al., 2014; Mohn et al., 2014; Parhad et al., 2017; Zhang et al., 2014). Deadlock recruits Moonshiner and TRF2 (TATA box binding protein related factor 2), which promotes RNA polymerase II transcription from both strands of these transposon rich loci (Andersen et al., 2017). Rhino also interacts with the DXO homolog Cuff, which functions with Rhino to suppress splicing and poly-adenylation of piRNA precursor transcripts (Chen et al., 2016; Mohn et al., 2014; Zhang et al., 2014). This may help differentiate these piRNA precursors from gene transcripts. Unspliced cluster transcripts are bound by the DEAD box protein UAP56 and the THO complex, which may deliver these transcripts to nuclear pores for transport to the cytoplasm for processing (Hur et al., 2016; Zhang et al., 2012a).

Most of the cytoplasmic piRNA processing machinery, along with the piRNA binding Piwi proteins Aub and Ago3, localize to perinuclear nuage granules (Malone et al., 2009), but the endonuclease Zuc and a partner protein Papi localize to the

mitochondrial outer membranes, and the helicase Armi localizes to both nuage and mitochondria (Huang et al., 2014). Precursor processing may therefore require shuttling between nuage and mitochondria. Precursor cleavage by Ago3, localized to nuage, or by the mitochondrial nuclease Zuc, generates intermediates carrying the 5' end of mature piRNA (Han et al., 2015; Mohn et al., 2015). These intermediates bind to the Piwi proteins, and 3' trimming by the Nibbler exonuclease, or by direct cleavage by Ago3, generates mature piRNAs (Hayashi et al., 2016), which are 2'-O-methylated by Hen1, which stabilizes their 3' ends (Horwich et al., 2007).

Somatic follicle cells do not express the components of RDC (Rhino, Deadlock, and Cutoff) complex (Mohn et al., 2014). The uni-strand piRNA clusters have promoters and are transcribed like mRNAs with 5' cap, polyA tail and splicing signatures (Goriaux et al., 2014; Mohn et al., 2014; Zanni et al., 2013). In the cytoplasm, these transcripts are processed in cytoplasmic complex called Yb bodies, composed of Yb, Armi, Zuc and Vret, and loaded into Piwi (Handler et al., 2011; Murota et al., 2014; Qi et al., 2011; Saito et al., 2010; Zamparini et al., 2011). Zuc mediated cleavage generates the 5' ends of piRNAs (Han et al., 2015; Mohn et al., 2015). The 3' ends can be produced by Zuc mediated cleavage or trimming by Nibbler (Hayashi et al., 2016). These are also 2'-O-methylated by Hen1 at 3' ends (Horwich et al., 2007). The piRNA bound Piwi can enter the nucleus and transcriptionally silence TEs (Sienski et al., 2012).

Figure 1.3

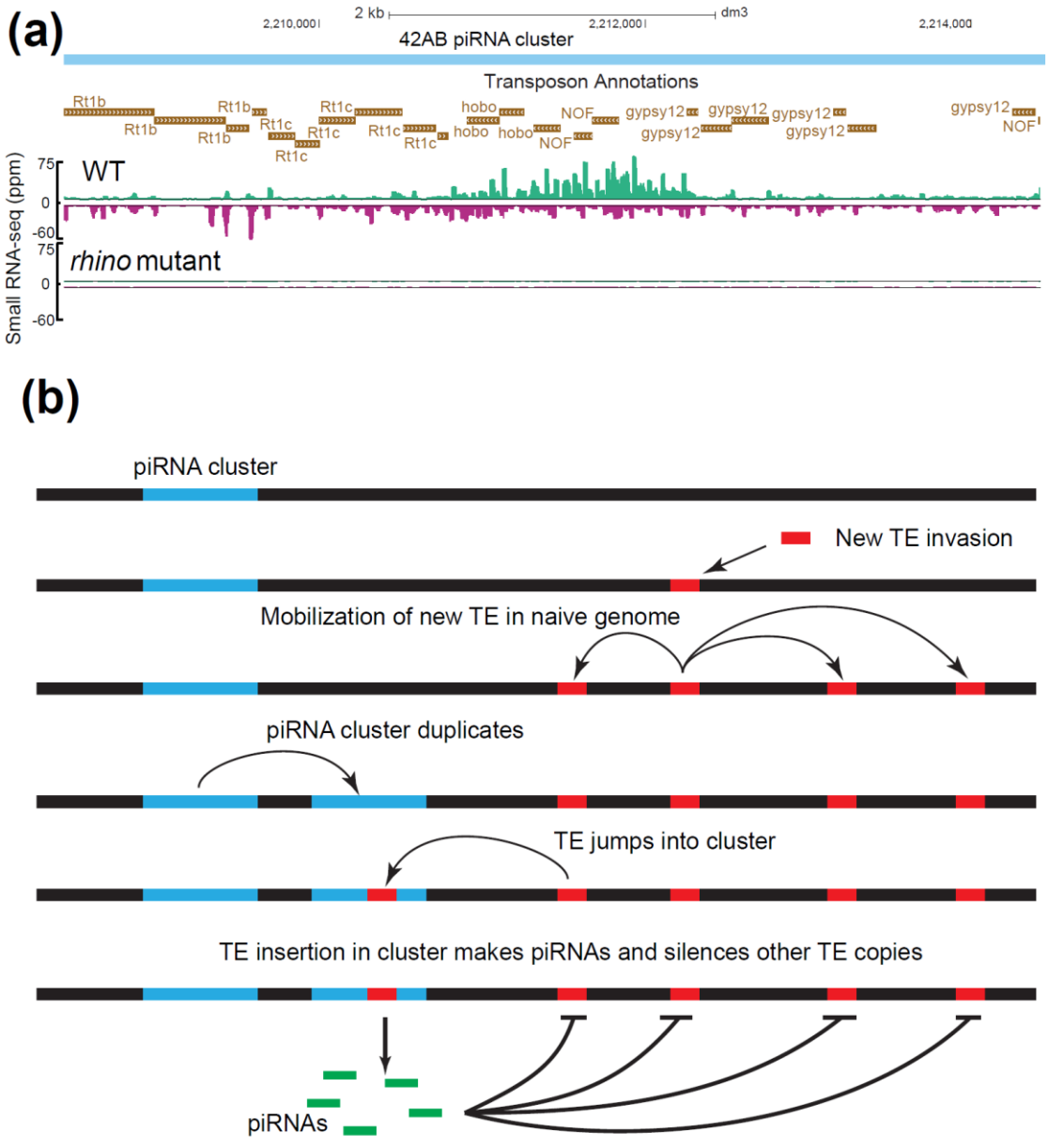

(a)
2 kb
dm3
2,210,000
2,212,000
2,214,000
42AB piRNA cluster
Transposon Annotations
Rt1b
Rt1c
hobo
NOF
gypsy12
WT
rhino mutant
Small RNA-seq (ppm)
75
0
-60
(b)
piRNA cluster
New TE invasion
Mobilization of new TE in naive genome
piRNA cluster duplicates
TE jumps into cluster
TE insertion in cluster makes piRNAs and silences other TE copies
piRNAs

Figure 1.3. Organization of piRNA clusters and possible conflicts with transposons

(a) Genome browser view of 42AB piRNA cluster in *D. melanogaster* showing transposon organization within the cluster and piRNA levels in wild type and *rhino* mutant ovaries. (b) Cluster conflicts with transposons. When a transposon jumps into a new species, it can transpose fast and make multiple copies. The piRNA clusters duplicate so as to increase the landing sites for the new TE to jump into. Once the new transposon jumps into the cluster, piRNAs are generated against the new transposon and these piRNAs can silence the new TE. Thus, it is advantageous for clusters to incorporate new TE, but is deleterious for the transposons.

1.3.1.2. Secondary piRNA biogenesis

In the germline, the Piwi protein Aub preferentially binds piRNAs that are anti-sense to transposon transcripts, and directs cleavage of these transcripts to produce the precursors of sense strand piRNAs bound to Ago3, which are subsequently trimmed and methylated. This Ago3 bound piRNA can then cleave the piRNA cluster transcript, producing precursors that are loaded into Aub. This secondary piRNA amplification loop is called the Ping-Pong cycle (Brennecke et al., 2007; Gunawardane et al., 2007), and is regulated by the DEAD box helicase Vasa and Tudor domain protein Qin (Xiol et al., 2014; Zhang et al., 2011). This Ping-Pong cycle is only present in the germline and not in the follicle cells, as these proteins are germline specific. Cleavage of precursor transcript by Ago3-piRNA complexes also initiates production of phased piRNAs that are bound by Piwi, the founding member of the Piwi clade (Han et al., 2015; Mohn et al., 2015).

Aub and Ago3 are active endonucleases that localize to nuage and the cytoplasm. In addition to driving piRNA biogenesis, these proteins post-transcriptionally silence transposons by cleaving homologous transcripts (Brennecke et al., 2007; Gunawardane et al., 2007). Piwi, by contrast, localizes to the nucleus, where it directs transcriptional silencing through Panoramix and Asterix, which direct repressive histone modification of Piwi targets (Le Thomas et al., 2013; Muerdter et al., 2013; Yu et al., 2015b). This appears to involve co-transcriptional recognition of nascent transposon transcripts, but this has not been directly confirmed.

1.3.1.3. Primary and Secondary biogenesis processes make piRNA system adaptive

piRNA clusters function as transposon sequence archives. When a new transposon invades the germline, there are no matching copies in clusters and the element remains active. Transposition compromises genome integrity and fertility, and continues until a copy of the active element inserts into a piRNA cluster. The sequence is then incorporated into cluster transcripts, producing piRNAs that trans-silence full length elements that are dispersed throughout the genome. Subsequent invasion by the same element, or a close relative, would presumably lead to rapid silencing. The system also appears poised to respond to increased activity of an existing element. For example, if a resident element acquires a mutation that increases transcription, Ping-Pong cycle ensures that more piRNAs are produced, suppressing expression. The piRNA response to a new transposon is therefore conceptually similar to the adaptive immune response to a pathogenic bacteria or viruses. When a new pathogen first invades, rapid propagation compromises host fitness, but also triggers expansion of T-cells and B-cells that recognize the invader (similar to transposon jumping into piRNA cluster), suppressing the pathogen and providing immunity to subsequent infections. Thus, the *Drosophila* germline piRNA pathway acts an adaptive immune system directed against genomic pathogens.

1.3.2. piRNA biogenesis in mouse

piRNAs and PIWI clade Argonautes appear to be universal germline components, and have been extensively studies in mice (Fu and Wang, 2014; Pillai and Chuma, 2012).

In *Drosophila,* a subset of piRNA pathway mutants, including mutations in *piwi*, are male and female sterile. In mice, by contrast, the well characterized piRNA mutations are male sterile and female fertile (Aravin et al., 2006; Aravin et al., 2008). Mice also have three Piwi clade Argonaute proteins: MILI, MIWI2 and MIWI, the piRNAs are produced from piRNA clusters in Dicer independent fashion, and have 2'-O-Me modification at 3' ends (Aravin et al., 2006; Aravin et al., 2008; Carmell et al., 2007; De Fazio et al., 2011; Girard et al., 2006; Kirino and Mourelatos, 2007). The Mouse Papi homolog TDRKH mediates 3' end trimming of piRNA precursors (Han et al., 2015; Saxe et al., 2013). There are two classes of piRNAs in mouse testis based upon the developmental stage in mouse: pre-pachytene piRNAs and pachytene piRNAs. Pre-pachytene piRNAs are made in gonocytes and pachytene piRNAs at pachytene stage of meiosis (Fu and Wang, 2014).

Unlike flies, both pre-pachytene and pachytene piRNAs are produced from only one strand of piRNA clusters, which can be either unidirectional or bi-directional (transcription in opposite direction from start site) (Li et al., 2013). The piRNA clusters are transcribed by RNA Polymerase II to produce precursor transcripts having 5' cap and poly-A tail. These clusters resemble uni-strand clusters in *Drosophila*. A-Myb serves as a transcription factor for Pachytene clusters, however such factor is unknown for pre-Pachytene piRNA clusters.

The pre-pachytene piRNA biogenesis resembles the *Drosophila* germline piRNA biogenesis. Most pre-pachytene piRNAs target different repeats and are produced by Ping-Pong amplification cycle (Aravin et al., 2008). MILI bound primary piRNA slices a transcript and generates secondary piRNAs which can be either MILI or MIWI2 bound.

The MIWI2 bound secondary piRNAs lead to DNA methylation and silencing of repeats such as transposons. Mutations in MILI or MIWI2 leads to demethylation and activation of transposons. It is proposed that these piRNAs establish the methylation patterns after methylation is erased during the initial developmental stages. Adult testes consists of mostly pachytene piRNAs (95% of total piRNAs), bound by either MILI or MIWI (Aravin et al., 2008; Li et al., 2013). Contrary to pre-pachytene piRNAs, pachytene piRNAs contain mostly intergenic sequences and less of repeat associated piRNAs. MIWI bound piRNAs cleave the transposon transcripts and defects in its catalytic activity leads to LINE1 over-expression (Reuter et al., 2011). It is proposed that pre-pachytene piRNAs establish transcriptional silencing in embryos, whereas pachytene piRNAs maintain this silencing after birth at post-transcriptional level.

1.3.3. piRNA biogenesis in *C. elegans*

C. elegans piRNA pathway have many similarities and notable differences when compared to other organisms (Figure 1.2). *C. elegans* 21U-RNAs are piRNAs which bind to PIWI protein PRG-1, named as such due to their 21nt length and 5' U bias (Batista et al., 2008; Das et al., 2008; Ruby et al., 2006; Wang and Reinke, 2008). They are produced in Dicer independent fashion and have 2'-O-methyl modification produced by Hen1 homolog HENN-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). Unlike other organisms, each individual piRNA is produced from monocistronic piRNA gene with its own promoter identified as Ruby motif (Ruby et al., 2006). Most of these piRNA genes are clustered together on 4^{th} chromosome (Billi et al., 2013; Cecere et al., 2012; Gu et al., 2012; Ruby et al., 2006). These are transcribed by RNA Pol II, which

is regulated by Forkhead and TOFU (Twenty-One-u Fouled Ups) proteins (Cecere et al., 2012; Goh et al., 2014). Worms do not have Ping-Pong amplification cycle. However, piRNA amplification is achieved by PRG-1 bound 21U-RNAs targeting a transcript and production of secondary 22G-RNAs by RNA dependent RNA polymerase (RdRP), which can bind to worm specific AGOs (WAGOs) (Gu et al., 2009). These can further mediate transcriptional silencing via repressive histone modification H3K9me3 (Bagijn et al., 2012; Kasper et al., 2014; Lee et al., 2012).

1.3.4. piRNA biogenesis in silkworm

For silkworms, the piRNA function is studied in *Bombyx mori* BmN4 germline cell lines (Kawaoka et al., 2009). These cells also have both primary and secondary piRNA biogenesis mechanisms (Kawaoka et al., 2009; Sakakibara and Siomi, 2018). They contain piRNA clusters, but it is not known how these piRNA clusters are specified, as no Rhi and Deadlock homologs are identified in silkworms (Kawaoka et al., 2008; Kawaoka et al., 2012; Wang et al., 2017). Similar to flies, the Ping-Pong amplification is orchestrated by anti-sense piRNAs bound Siwi and sense piRNAs bound Ago3 and is regulated by Vasa and Qin (Honda et al., 2013; Izumi et al., 2016; Nishida et al., 2015; Xiol et al., 2014). The piRNAs show signatures of phasing, which is thought to be the results of Zuc, Trimmer and Nibbler (Izumi et al., 2016; Sakakibara and Siomi, 2018). The silencing is mainly achieved at post-transcriptional level during Ping-Pong cycle. It is not known whether transcriptional silencing occurs in silkworms, as these cells lines are germline derived, which lack *Drosophila* Piwi homolog (Kawaoka et al., 2009; Sakakibara and Siomi, 2018).

1.3.5. Diversity in piRNA biogenesis mechanisms

Despite the fact that piRNA pathway performs a conserved function of transposon repression, different animals use diverse mechanisms to achieve the same goal. Flies use Rhino and Deadlock to direct transcription from piRNA clusters (Mohn et al., 2014). No such Rhino and Deadlock homologs are found outside genus *Drosophila*. Mice use A-MYB as the transcription factor for pachytene piRNA clusters (Li et al., 2013). Worms use Forkhead and TOFU proteins for cluster transcription, which are not related to piRNA cluster factors from other organisms (Weick and Miska, 2014). Many piRNA clusters in flies and mice are more than 100kb long and subsequently lead to production of thousands of piRNAs (Brennecke et al., 2007; Li et al., 2013). However, each piRNA cluster in *C. elegans* encodes for only one piRNA and each piRNA has its own promoter (Ruby et al., 2006). This shows the diversity at the piRNA source. In flies, mice and silkworms, the piRNA repertoire is amplified by Ping-Pong amplification cycle (Sakakibara and Siomi, 2018; Weick and Miska, 2014). Whereas in *C. elegans*, piRNA amplification is achieved by RNA dependent RNA polymerases (Gu et al., 2009). Thus, different organisms use diverse mechanisms for primary and secondary piRNA generation. This variety in mechanism is also reflected in sequence divergence among the piRNA pathway proteins and rapid evolution of piRNA source loci, i.e. piRNA clusters.

1.4. Rapid evolution of piRNA pathway

1.4.1. Rapid evolution in piRNA pathway proteins

Multiples studies in *Drosophila* show evidence of adaptive evolution in the piRNA pathway (reviewed in (Blumenstiel et al., 2016)). Obbard et al calculated ratios of non-synonymous (K_A) to synonymous (K_S) substitutions for all the genes by comparing sibling species *D. melanogaster* and *D. simulans*. Most genes showed lower values indicative of neutral evolution or purifying selection. Many piRNA pathway genes, *krimper, maelstrom, aubergine, piwi, armitage, spnE,* showed higher K_A/K_S values indicative of adaptive evolution (Obbard et al., 2009). Before Rhino was found to be involved in piRNA pathway, Malik lab had shown the evidence of adaptive evolution in all the domains of this HP1 homolog (Vermaak et al., 2005). Lee and Langley also observed adaptive evolution *for rhino, krimper, maelstrom, aubergine, armitage, vasa* and *spindle-E* (Lee and Langley, 2012). Simkin et al used Phylogenetic Analysis by Maximum Likelihood method (PAML) to study the evolution of 10 piRNA pathway genes in 6 *Drosophila* species. They observed positive selection among *rhino, aubergine* and *krimper* genes across multiple *Drosophila* lineages (Simkin et al., 2013). Rapid evolution in piRNA pathway is also observed in Teleost fishes (Yi et al., 2014). Thus sequence comparison of various piRNA pathway proteins shows evidence of adaptive evolution.

Do these substitutions lead to functional differences in piRNA pathway proteins? One way to test this is to swap piRNA pathway genes between species and study their

function. Two reports have addressed this. Kelleher et al swapped *aubergine* between *D. melanogaster* and *D. simulans* (Kelleher et al., 2012). They expressed *D. simulans aub* (*sim-aub*) in a *D. melanogaster aub* mutant background. *aub* mutants are sterile. The control, *mel-aub* can rescue both fertility and transposon silencing in an *aub* mutant. When expressed in *D. melanogaster aub* mutant, *sim-aub* can only partially rescue fertility and transposon silencing. Thus, adaptive evolution in *aub* has produced fixed differences in *aub*, leading to functional changes in Aubergine between species. Our lab swapped rapidly evolving Rhino and its interacting partner Deadlock between *D. melanogaster* and *D. simulans* (Parhad et al., 2017). Both *sim*-Rhino and *sim*-Deadlock fail to function in *D. melanogaster*. In *D. melanogaster*, *sim*-Rhino does not bind to *mel*-Deadlock and does not localize to piRNA clusters. We identified the Deadlock interacting domain to be shadow domain, which shows strong signature of adaptive evolution (Vermaak et al., 2005). On the contrary, when *mel*-Rhi is expressed in *D. simulans*, it can bind to *sim*-Deadlock and find piRNA clusters also. This shows that Rhino shows directional incompatibility for Deadlock interaction. Deadlock swap shows that *sim*-Deadlock binds to *mel*-Rhino in *D. melanogaster*, but it fails to recruit downstream piRNA pathway components and does not work. Thus, fixed substitutions observed in both nuclear and cytoplasmic components of piRNA pathway have made these piRNA proteins non-functional in the nearest species, indicating that sequence divergence has functional consequences. Most genes with conserved functions between species show purifying selection and can functionally substitute their ortholog in different species. For instance, many human genes can functionally substitute fly and yeast

orthologs (Fernandez-Hernandez et al., 2016; Kachroo et al., 2015). In contrast, the piRNA pathway proteins are evolving are so fast that these orthologs fail to function in the closest species, suggesting piRNA pathway is involved in an evolutionary arms race and positive selection is necessary to stay ahead in the game.

1.4.2. Rapid evolution in piRNA clusters

Such adaptive evolution is also observed for piRNA generating loci. As piRNA clusters contain truncated copies of transposons, piRNAs produced from these clusters can mount sequence specific immune response against transposons. Bergman et al studied the distribution of transposons in the *Drosophila* genome and observed high transposon density in the β-heterochromatin. They proposed that this nesting of transposons can produce a "co-suppression network" which can regulate transposons (Bergman et al., 2006). Brennecke et al found piRNAs originating from these nested transposon rich regions and coined the term "piRNA clusters" (Brennecke et al., 2007). Using comparative analysis, Bergman et al found that these clusters evolve by transposition of transposons and by duplication of these clusters, but not by inversion of TEs from constitutive heterochromatin. This study provides clues about mechanisms for adaptation of host immune pathway against the threat pathogenic transposons.

Dynamic transposition leads to cluster evolution. Zanni et al reported structural diversity in flamenco piRNA clusters between different *D. melanogaster* strains (Zanni et al., 2013). The presence or absence of a transposon in given strain is correlated with the silencing of that particular transposon. They also show that many copies in flamenco

locus are from recent transposon insertions and many transposons are thought to move by horizontal transfer between *Drosophila* species. piRNA clusters can adapt as follows (Figure 1.3b): a newly introduced transposon can transpose and make multiple copies in naïve genome in the absence of piRNAs. Once it gets inserted into piRNA cluster, piRNAs can silence these transposons. However, as piRNA clusters are dynamic, loss of TE copy in the cluster will result in remobilization of transposons. These can be again silenced upon insertion in clusters. Such cluster dynamics has also been shown for telomere associated clusters in *Drosophila* (Asif-Laidin et al., 2017), highlighting role of transposition in making clusters dynamic.

Duplication of clusters can enhance the piRNA repertoire. Assis and Kondrashov studied piRNA cluster evolution in rodents (Assis and Kondrashov, 2009). They observed expansion of piRNA clusters in rodents. Almost half the clusters were new and originated after rodent-primate split. Similar to observation in *Drosophila* by Bergman et al (Bergman et al., 2006), new rodent piRNA clusters are proposed to be produced by duplication due to ectopic recombination. The presence of intervening repetitive elements can lead to both deletions and insertions. The observation that no piRNA clusters were lost in rodent species, but many are gained, led them to propose that piRNA clusters are advantageous to the host and are evolving under positive selection. A study by Chirn et al compared piRNA clusters in Eutherian mammals (Chirn et al., 2015). They observed two types of piRNA clusters in mammals: 1) clusters conserved between species may have a regulatory role in various reproductive processes and 2) diverged piRNA clusters may be due to gain of clusters to rapidly expand the piRNA repertoire

under positive selection pressure. Unlike mammals, absence of conserved piRNA clusters in distant *Drosophila* species can be due to the nature of piRNA targets in mammals and in *Drosophila*. Unlike mammals, most of piRNAs in flies target transposons. They propose that it can be due to more dynamic host-pathogen arms race between piRNA pathway and transposons in flies than in mammals. Thus, by increasing the number of piRNA clusters by duplications, the organism can raise the chance that a new TE can hop into a cluster. Similar to piRNA pathway proteins, the piRNA clusters themselves seem to be under positive selection, consistent with involvement in a genomic conflict.

1.5. Possible mechanisms leading to the arms race

Transposon behave like pathogens by the virtue of being deleterious to host and spreading rapidly between populations. Host piRNA pathway needs to continuously evolve to ever changing transposon threat. Both transposons and piRNA pathway are thought to be involved in "Red Queen arms race", just like viruses which are involved in continuous battle with the host immune system. We discuss below potential ways leading to this arms race.

1.5.1. Conflicts at piRNA clusters

1.5.1.1. Conflicts to modulate piRNA cluster insertions

piRNA clusters act as the source of primary piRNAs to silence transposons. If a transposon can inhibit the source itself, it would beneficial not only to that transposon, but also all the transposon in the genome. Thus, piRNA clusters can be one of the sites at which this arms race can take place. Hybrid dysgenesis offers one example where one transposon desilencing can be advantageous to multiple transposons (Khurana et al., 2011). F1 hybrids obtained from crosses between wild *D. melanogaster* males carrying P-elements and lab *D. melanogaster* females lacking the P-element show sterility syndrome termed as hybrid dysgenesis. P-elements become active in dysgenic hybrids as they lack the maternal P-element mapping piRNAs. It is observed that many resident transposons also become active in hybrid dysgenesis. Thus one transposon inhibiting the piRNA pathway can help other transposons as well. The dysgenic hybrids recover their fertility with time. The flies make piRNAs from the new transposons insertions in the

clusters. piRNAs are also generated from paternal piRNA clusters, which contain P-elements. Thus, flies can adapt their piRNA cluster repertoire in response to a new transposon invasion and transposon mobilization.

It is advantageous for the host to incorporate transposons into clusters, whereas transposons would favor to land outside the clusters. Many transposons shows target site preferences (Linheiro and Bergman, 2012). The observations that P-elements prefer to insert in germline expressed genes in *Drosophila* (Bownes, 1990) and the piRNA cluster transcriptional machinery is only expressed in the germline (Mohn et al., 2014) would suggest an interesting possibility that piRNA clusters may recruit components of normal gene transcriptional machinery such as RNA Polymerase II and TRF2 (Andersen et al., 2017) not only for transcription, but also to attract transposons for insertions into clusters. The germline cells are better suited to allow inheritance of immunity in the form of these insertions into clusters. The transposons may change their site preference to avoid cluster insertions. piRNA clusters can also enhance their advantage by increasing the number of clusters. Consistently there has been gain of piRNA clusters during the course of evolution (Assis and Kondrashov, 2009; Chirn et al., 2015). Thus the arms race would take place to increase or avoid cluster insertions.

1.5.1.2. Conflicts with proteins defining piRNA clusters

piRNA pathway can also be targeted by inhibiting the proteins necessary for piRNA cluster transcription. Rhino-Deadlock-Cutoff complex specifies piRNA clusters in *Drosophila*. The species specific interaction between rapidly evolving Rhino and

Deadlock suggests that cluster specification process may be a target of host-pathogen arms race (Parhad et al., 2017). Preventing interaction between these two cluster specifying proteins can inhibit the normal functioning of piRNA clusters and give an advantage to transposons. Host pathogen arms race leads to adaptive evolution in host and pathogen proteins (Daugherty and Malik, 2012). Adaptation of protein-protein interactions in the host led us to think of inhibition by molecular mimicry (Elde and Malik, 2009) as the possible mechanism by which Rhino and Deadlock are co-evolving in response to transposon threat. A transposon encoded protein can mimic Deadlock binding surface, so that it can interfere in the Rhino-Deadlock interaction. Rhino mutates so that it can avoid the mimic binding, by sacrificing affinity for specificity. The transposon can mimic Rhino surface to again inhibit the Rhino-Deadlock interaction. Deadlock can mutate further to restore the interaction. This reciprocal mutations in host proteins lead to adaptive evolution of the interacting proteins (Figure 1.4). Thus, transposons can target the proteins involved in piRNA cluster specification and function, to gain advantage in the arms race.

Figure 1.4

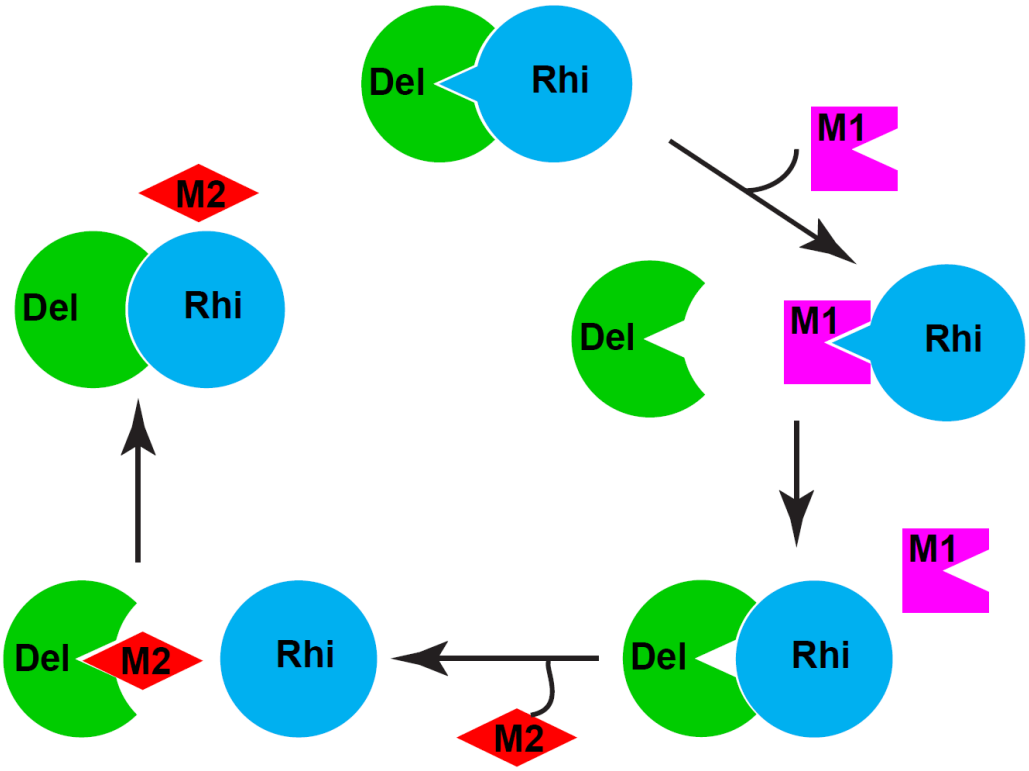


Figure 1.4. Possible conflicts between transposon and piRNA pathway in the form of molecular mimicry.

In *Drosophila*, Rhino (Rhi) and Deadlock (Del) bind to define piRNA clusters. Transposon protein can mutate to generate a protein (M1) that mimics Del binding surface. In response, Rhi mutates to avoid mimic binding. This restores Rhi-Del interaction. The transposon proteins can mutate further generating a new protein (M2) that mimics Rhi binding surface. Del further mutates to restore Rhi-Del binding. Thus, host-pathogen arms race in the form of molecular mimicry can lead to rapid evolution of the piRNA pathway proteins.

1.5.2. Conflicts at piRNA mediated silencing

The conflicts can also occur at the effector phase of piRNAs, i.e. post-transcriptional and transcriptional silencing. The nature of the silencing by piRNAs makes it more likely that the evolution would favor transposon protein sequence and structure change rather than changes in RNA sequence. As piRNAs mostly map to the whole length of transposons, a base change to prevent sequence complementarity would not save the transposon from inhibition. An amino acid substitution in a transposon protein can make it inhibitory for piRNA pathway protein(s) and would benefit transposons. As transposon proteins are associated with its RNA, the protein mutations can change the subcellular localization of RNA or inhibit cleavage by the PIWI proteins. As Aubergine localizes in nuage and cleaves transposon, the transposon can avoid the nuclear pores associated with nuage, thereby avoid getting chopped up. Consistently with this hypothesis, both nuclear pore proteins and Aubergine are rapidly evolving and shows evidence of species specific function (Kelleher et al., 2012; Presgraves et al., 2003; Presgraves and Stephan, 2007; Tang and Presgraves, 2009). A transposon encoded protein can also avoid silencing if it can inhibit Aubergine function. Aubergine mediated cleavage of transposon transcript leads to amplification of piRNAs against active transposon by Ping-Pong cycle. Other proteins involved in this amplification loop, such as Vasa and Argonaute3, also show signatures of adaptive evolution (Blumenstiel et al., 2016; Simkin et al., 2013). This suggests that post-transcriptional silencing by piRNAs, can be a target of inhibition by transposons.

The proteins involved in transcriptional silencing also show adaptive evolution. Piwi directs transcriptional silencing of TEs. Armitage is required for piRNA loading into Piwi (Saito et al., 2010; Sienski et al., 2012) and shows signatures of positive evolution. Thus the process of Piwi loading and thus subsequent transcriptional silencing can also be targets for transposon mediated inhibition. Thus, transposons can not only target the source of piRNAs, but also the transcriptional and post-transcriptional effector functions of piRNAs.

1.5.3. Can this host-pathogen arms race lead to reproductive isolation?

Hybrid dysgenesis is caused by activation of P-elements due to lack of maternal piRNAs. It is caused only when P-element is introduced through the father. Consider a hypothetical scenario where a population A has transposon α, but not β and population B has transposon β, but not α (Figure 1.5). The crosses in either direction between these populations would produce sterile F1 progeny, due to activation of either α or β transposon. Thus reproductive barriers can be formed due to transposon variation among populations. As transposons can jump between distant species, a population can adapt to this new transposon and evolve into a new species due to reproductive isolation. As many closely related species can have unique transposons (Bartolome et al., 2009), mutations in hybrid incompatibility genes and sterile hybrids between *D. melanogaster* and *D. simulans* show upregulation of many transposon families (Kelleher et al., 2012; Satyaki et al., 2014; Thomae et al., 2013), transposons can be the drivers of reproductive isolation and therefore speciation.

Figure 1.5

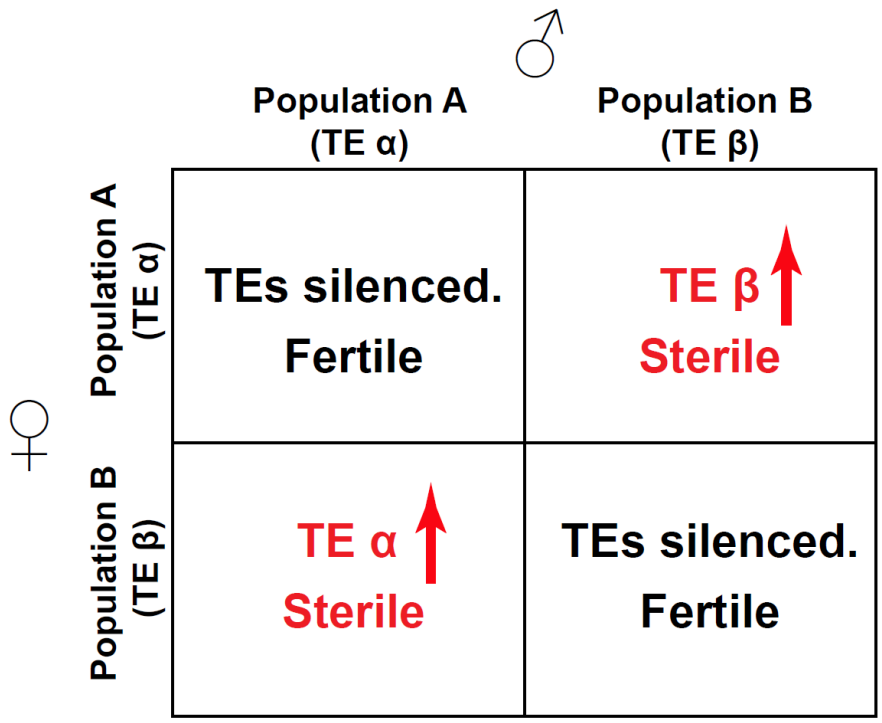

♂
Population A
(TE α)
Population B
(TE β)
♀
Population A
(TE α)
Population B
(TE β)
TEs silenced.
Fertile
TE β
Sterile
TE α
Sterile
TEs silenced.
Fertile

Figure 1.5. Transposon variation can drive reproductive isolation

Hypothetical scenario for crosses between populations A and B having unique transposons. Population A has transposon α, but not β and population B has transposon β, but not α. Crosses between populations lead to activation of transposons coming from father due to absence of maternal piRNAs against TE unique to father and sterility. This would establish reproductive barriers and lead to a speciation event.

Similar to transposons, piRNA pathway can also drive reproductive isolation. Rhino and Deadlock show directional incompatibility between *D. melanogaster* and *D. simulans* (Parhad et al., 2017). *sim*-Deadlock binds to *mel*-Rhino, but the complex is non-functional. In hybrids between the two species, this non-functional interaction may prevent *sim*-Deadlock to function with *sim*-Rhi and *mel*-Rhi with *mel*-Deadlock. This can provide one possible explanation for the defects in piRNA biogenesis and transposon over-expression in the interspecific hybrids. Mutations in the piRNA pathway can generate such biochemical incompatibilities in the crosses between different populations or strains. This can lead to bursts of transposition, which are linked to species divergence in animals and plants (Belyayev, 2014; Fontdevila, 2005). Thus, we think that arms race between piRNA pathway and transposons can drive speciation by generating biochemical incompatibilities that can set up reproductive barriers and cause bursts of transposon induced variations for natural selection to act on.

1.6. Concluding remarks:

Transposons are genomic pathogens which cause genomic instability. A class of small RNAs called piRNAs protects the genomes from these transposons. Just like pathogens such as viruses and host immune system, transposons and piRNA pathway are likely to be involved in host-pathogen arms race. Transposons are deleterious and can rapidly spread between the populations and different species, making them an effective pathogen. Thus, piRNA pathway has to continuously adapt to ever-changing threat of transposons. The organization of piRNA biogenesis includes: 1) piRNA clusters, which keep the database of all the transposons and directs sequence specific targeting of transposons, and 2) piRNA amplification loop, for example Ping-Pong amplification cycle, which leads to increase in piRNA repertoire against active TEs. Both these characteristics make immune response "adaptive". As the proteins and piRNA clusters directing both of these processes show hallmarks of adaptive evolution, pathogenic transposons and host piRNA pathway seem to be on opposite sides in the "Red Queen host pathogen arms race". This arms race can lead to evolution of genomes and may contribute in reproductive isolation and thus speciation.

Acknowledgements:

We would like to thank the members of Theurkauf lab and UMassMed RNA biology community for their insightful discussions. This work was supported by National Institute of Child Health and Human Development (R01HD049116 and P01HD078253).

Chapter II

Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery

Preface

Shikui Tu helped me with the bio-informatics analysis under the guidance of Zhiping Weng.

Summary

Reproductive isolation defines species divergence and is linked to adaptive evolution of hybrid incompatibility genes. Hybrids between *Drosophila melanogaster* and *Drosophila simulans* are sterile and phenocopy mutations in the piRNA pathway, which silences transposons and shows pervasive adaptive evolution, and *Drosophila rhino* and *deadlock* encode rapidly evolving components of a complex that binds to piRNA clusters. We show that Rhino and Deadlock interact and co-localize in *simulans* and *melanogaster*, but *simulans* Rhino does not bind *melanogaster* Deadlock, due to substitutions in the rapidly evolving Shadow domain. Significantly, a chimera expressing the *simulans* Shadow domain in a *melanogaster* Rhino backbone fails to support piRNA production, disrupts binding to piRNA clusters, and leads to ectopic localization to bulk heterochromatin. Fusing *melanogaster* Deadlock to *simulans* Rhino, by contrast, restores localization to clusters. Deadlock binding thus directs Rhino to piRNA clusters, and Rhino-Deadlock co-evolution has produced cross-species incompatibilities, which may contribute to reproductive isolation.

Introduction

Transposable elements (TEs) are ubiquitous genome constituents with the potential to mobilize and induce insertional mutations and double strand breaks (Belancio et al., 2008; Biemont and Vieira, 2006; Hedges and Deininger, 2007; Khurana and Theurkauf, 2010). Protecting the genome from these selfish elements is especially critical in the germline, which is dedicated to transmitting genetic information to the next generation. Germline transposon silencing is mediated by small PIWI interacting RNAs (piRNAs), which are bound by PIWI proteins and direct post-transcriptional and transcriptional silencing of target transposons (Brennecke et al., 2007; Gunawardane et al., 2007; Iwasaki et al., 2015; Kuramochi-Miyagawa et al., 2008; Sienski et al., 2012). Mutations that disrupt the piRNA pathway lead to sterility and transposon mobilization in worms, flies, fish and mice, and have been linked to human infertility (Batista et al., 2008; Carmell et al., 2007; Das et al., 2008; Gou et al., 2017; Gu et al., 2010; Heyn et al., 2012; Houwing et al., 2008; Lin and Spradling, 1997; Weick and Miska, 2014). In striking contrast, genes with essential functions in piRNA production and transposon silencing in established model systems are often very poorly conserved, and piRNA biogenesis and sequence composition show remarkable phylogenetic diversity (Chirn et al., 2015; Obbard et al., 2009; Simkin et al., 2013; Yi et al., 2014; Zanni et al., 2013). For example, the overwhelming majority of piRNAs in the *Drosophila* female germline map to transposons and are derived from heterochromatic domains that span hundreds of kilobases. These "clusters" produce long precursors that are processed into primary piRNAs, which are amplified by a ping-pong cycle driven by PIWI mediated RNA

cleavage (Brennecke et al., 2007; Gunawardane et al., 2007). In *C. elegans*, by contrast, most piRNAs appear to target protein coding genes, each piRNA is produced from a single gene, and amplification involves piRNA primed generation of secondary siRNAs by RNA-dependent RNA polymerases (RdRPs) (Batista et al., 2008; Das et al., 2008; Weick and Miska, 2014).

The divergence of piRNA biogenesis mechanisms is not understood, but many piRNA pathway genes are evolving rapidly under positive selection, which is a hallmark of a "Red Queen" host-pathogen arms race (Daugherty and Malik, 2012; Duggal and Emerman, 2012; Obbard et al., 2009). In a simple Red Queen arms race, mutations that allow a pathogen to evade the host defense system will propagate, compromise host fitness, and lead to selection of host alleles that restore pathogen control. This positive selection cycle continues, driving rapid evolution of host and pathogen genes. Rapid evolution of piRNA genes could therefore reflect a Red Queen arms race with the transposons the system controls. Transposons are a very significant source of genome variation between closely related species (Warren et al., 2015), suggesting that piRNA pathway adaptation to mobile elements could produce genes that are unable to function across recently diverged species. Supporting this hypothesis, hybrids between the sibling species *Drosophila melanogaster* (*melanogaster*) and *Drosophila simulans* (*simulans*) are viable but sterile (Sturtevant, 1920), and show defects in transposon silencing, piRNA production, and organization of the piRNA biogenesis machinery (Kelleher et al., 2012). In addition, *melanogaster* and *simulans* share over 100 transposon families, but also show significant differences in total transposon content, and several families are unique to each

species (Bartolome et al., 2009; Lerat et al., 2011). Together, these observations raise the possibility that adaptive evolution of piRNA pathway genes contributes to hybrid sterility (Kelleher et al., 2012).

The *Drosophila melanogaster rhino (rhi)* gene encodes an HP1 homolog that localizes to piRNA clusters and is required for piRNA biogenesis. The *rhi* gene shows elevated rates of non-synonymous substitution between *melanogaster* and *simulans*, consistent with adaptive evolution (Klattenhoff et al., 2009; Le Thomas et al., 2014; Mohn et al., 2014; Vermaak et al., 2005; Zhang et al., 2014). Rhino (Rhi) interacts with the linker protein Deadlock (Del) to promote piRNA precursor formation, and *del*, like *rhi*, is rapidly evolving. We show that *rhi* and *del* are co-evolving, and that this process has generated species-specific interactions that prevent function across the *melanogaster-simulans* species barrier. For *simulans* Rhi, this is reflected in a failure to bind *melanogaster* Del, or rescue *melanogaster rhi* mutations. This leads to significantly reduced binding of *sim*-Rhi to *melanogaster* piRNA clusters. Strikingly, fusing *melanogaster* Del to the C-terminal shadow domain of *simulans* Rhi restores cluster localization. Adaptive evolution thus targets a Rhi-Del interaction that directs assembly of cluster chromatin, and generates biochemical incompatibilities in the piRNA machinery. We speculate that these incompatibilities contribute to hybrid sterility, reproductive isolation, and speciation.

Results

simulans Rhino does not function in *melanogaster*

In *Drosophila,* piRNAs are derived from large heterochromatic domains composed of complex arrays of nested transposon insertions, supporting a model for transposon adaptation in which invading mobile elements remain active until a copy inserts into a cluster, leading to sequence incorporation into piRNA precursors and trans-silencing (Bergman et al., 2006; Brennecke et al., 2007). *Drosophila* clusters are marked by histone H3 tri-methylated at lysine 9 (H3K9me3) and the HP1 homolog Rhi, which anchors a complex that includes Del and the Rai/DXO homolog Cutoff (Cuff), and all three proteins are required to suppress piRNA precursor splicing and for piRNA biogenesis (Chen et al., 2016; Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014). Precursor transcripts from clusters are bound by the DEAD box protein UAP56, and transported across the nuclear pore for processing into piRNAs in the perinuclear nuage (Zhang et al., 2012a). Rhi thus anchors the core of the adaptive transposon silencing system. The *rhi* gene is also rapidly evolving under positive selection (Vermaak et al., 2005), raising the intriguing possibility that it is engaged in host-pathogen arms race with transposons. To determine the functional consequences of *rhi* divergence, we genetically replaced the *rhi* gene in *melanogaster* (*mel-rhi*) with *rhi* from the sibling species *simulans* (*sim-rhi*), assayed gene function, and characterized interacting proteins.

To functionally replace *melanogaster rhi*, we expressed GFP tagged *sim*-Rhi, under control of the *melanogaster rhi* promoter, in a *melanogaster rhi* null mutant background (Figure 2.1A). As a control, an analogous GFP:*mel*-Rhi fusion was expressed in the same *rhi* background. Mutations in *rhi*, and most other piRNA pathway genes, lead to female sterility and dorsal-ventral egg patterning defects (Klattenhoff et al., 2007; Klattenhoff et al., 2009). The patterning defects are secondary to DNA damage and Chk2 kinase activation, and serve as an easily quantifiable biological readout of germline genome instability (Klattenhoff et al., 2007). The *mel-rhi* control rescued both embryo viability as measured by hatch rate, and D-V patterning scored by dorsal appendage formation (Figures 2.1B and 2.S1A). By contrast, *sim-rhi* failed to rescue D-V patterning or fertility, and by these assays was equivalent to a null *rhi* allelic combination.

Mutations in *rhi* lead to transposon over-expression in the ovary, which appears to be the primary cause of sterility and D-V patterning defects (Klattenhoff et al., 2009). We therefore measured transposon and gene expression by RNA sequencing (RNA-seq). The scatterplot in Figure 2.1C shows expression of TE families in the ovaries of *rhi* mutants vs. wild type (WT). Increased transposon expression is reflected in points that fall above the diagonal. Consistent with previous observations, *rhi* mutations led to over-expression of numerous transposon families (Figure 2.1C), but did not alter gene expression (Figure 2.S1B). Transposon silencing is restored by expression of *mel-rhi* (Figure 2.1D), but not by expression of *sim-rhi* (Figure 2.1E). Strikingly, transposon expression in *rhi* mutants and *rhi* mutants carrying the *sim-rhi* transgene were essentially

identical (Figure 2.S1C). Consistent with these observations, small RNA sequencing demonstrates that the *sim-rhi* transgene does not restore piRNA production from germline clusters or target transposons. By contrast, the *mel-rhi* transgene restores essentially WT piRNA levels (see below). Therefore, by both biological and molecular measures, the *sim-rhi* gene, when placed within *melanogaster*, is equivalent to a null allele.

We next determined fusion protein localization by direct imaging of the GFP tags. Endogenous Rhino localizes to piRNA clusters and forms distinct foci in the germline nuclei, and this pattern is seen with GFP tagged *mel*-Rhi (Figure 2.1F). By contrast, GFP tagged *sim*-Rhi does not localize to distinct nuclear or cytoplasmic structures in germline cells. Western blotting shows that both the *mel*-Rhi and *sim*-Rhi proteins are expressed at similar levels (Figure 2.S1D), indicating that the *simulans* protein is stable, but unable to localize to clusters or promote piRNA production in a *melanogaster* genetic background.

Figure 2.1

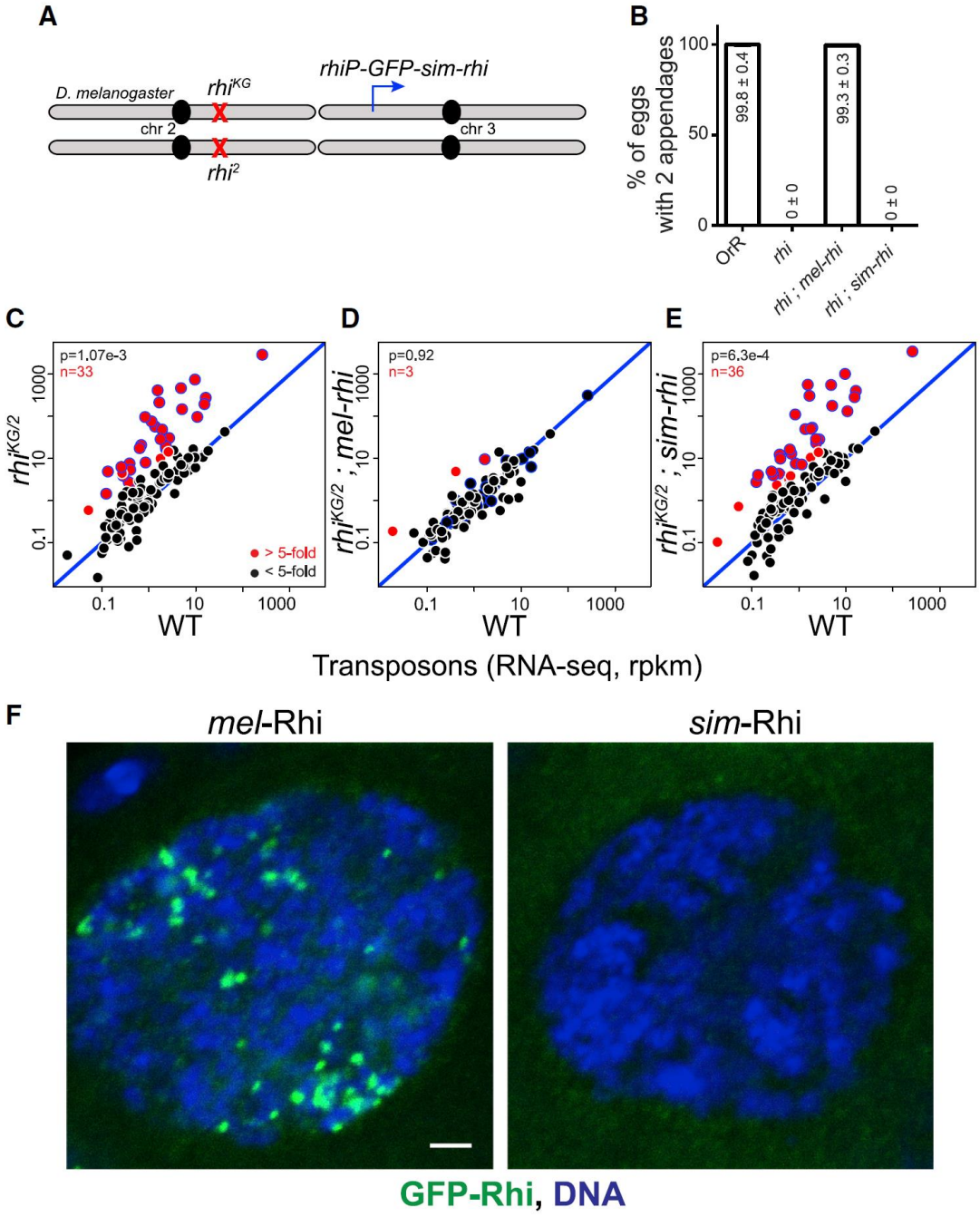

A
D. melanogaster
rhi^KG
rhiP-GFP-sim-rhi
chr 2
chr 3
rhi^2
B
% of eggs with 2 appendages
100
50
0
99.8 ± 0.4
0 ± 0
99.3 ± 0.3
0 ± 0
OrR
rhi
rhi ; mel-rhi
rhi ; sim-rhi
C
D
E
p=1.07e-3
n=33
p=0.92
n=3
p=6.3e-4
n=36
rhi^KG/2
rhi^KG/2 ; mel-rhi
rhi^KG/2 ; sim-rhi
0.1
10
1000
WT
> 5-fold
< 5-fold
Transposons (RNA-seq, rpkm)
F
mel-Rhi
sim-Rhi
GFP-Rhi, DNA

Figure 2.1. *simulans* Rhi does not function in *melanogaster*

(A) Genetic complementation strategy. The *sim-rhi* gene was expressed under endogenous *rhi* promoter in a *melanogaster* $rhi^{KG/2}$ trans-heterozygous null background.

(B) Bar graphs showing percentages of eggs with normal dorsoventral patterning produce by OrR (wild type (WT) control), *rhi* mutants, and *rhi* mutants rescued by either *mel-rhi* or *sim-rhi*. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *rhi* mutants and *rhi* mutants rescued by *sim-rhi* where average of at least 30 eggs were scored.

(C-E) Transposon expression in *rhi* mutants (C), *rhi* mutant rescued by *mel-rhi* (D), and *rhi* mutants rescued by *sim-rhi* (E). RNA-seq was performed on ovaries, and each point on the scatterplots shows rpkm values for a transposons family in ovaries of the indicated mutant/transgene combination relative to WT control. Diagonal represents x=y. Points in red show y/x>5 (n is number of these transposons). Blue bordered points are over-expressed by 5 fold or more in *rhi* mutants and *rhi* mutants expressing *sim-rhi*. p value for differences obtained by Wilcoxon test.

(F) Localization of *rhi* promoter driven GFP tagged *mel*-Rhi and *sim*-Rhi in *melanogaster* germline. GFP-Rhi is in green and DNA is in blue. Scale bar: 2 μm

Evolution of the Shadow domain restricts cross-species function

HP1 family proteins are composed of chromo, hinge and shadow domains (Vermaak and Malik, 2009). The Chromo domain of HP1a, the founding member of the family, binds to histone H3 tri-methylated at Lysine 9 (H3K9me3) (Bannister et al., 2001; Lachner et al., 2001). Structural and biochemical studies indicate that H3K9me3 binding is shared by the Chromo domain of Rhi (Le Thomas et al., 2014; Mohn et al., 2014; Yu et al., 2015a). The Hinge domain is a variable linker that may also bind RNA or DNA, and the Shadow domain mediates protein-protein interactions (Meehan et al., 2003; Muchardt et al., 2002; Smothers and Henikoff, 2000). To localize changes in *simulans* Rhi that prevent function in *melanogaster*, we generated transgenes expressing chimeric proteins composed of individual functional domains from *sim*-Rhi in a *mel*-Rhi backbone (Figure 2.2A). All chimeric proteins were fused to GFP, and the native *rhi* promoter was used to drive expression. Both chromo and hinge domain chimeras were able to fully rescue D-V patterning of eggs (Figure 2.2B), but only partially rescued hatching (Figure 2.S2A). However, the shadow domain chimera failed to rescue D-V patterning or hatching.

To determine if the fertility and patterning defects are linked to transposon over-expression, used RNA-seq to analyze the transcriptome in *rhi* mutant flies expressing the chimeric Rhi proteins (Figures 2.2C, 2.2D and 2.2E). Consistent with our phenotypic data, the hinge domain chimera completely restored transposon silencing, and the chromo domain chimera silenced all transposon families with the exception of the Tirant retrotransposon. This does not reflect a defect in piRNA production (see below), and the mechanism leading to Tirant over-expression in this background remains to be explored.

We speculated that over-expression of Tirant could contribute to the low hatch rate with the chromo-domain chimera, but forced expression of a full length element in WT, using the *nanos*-Gal4 driver and *UASp* promoter, did not reduce hatch rate or lead to D-V patterning defects (unpublished observation). In contrast to the chromo and hinge chimeras, the shadow chimera failed to restore silencing and was comparable to the *rhi* null allelic combination (Figure 2.S2B). The chromo and hinge domain substitutions also localize to nuclear foci in germline cells (Figure 2.2F), while the shadow chimera is expressed but fails to localize. Changes in the Rhi shadow domain have therefore disrupted the ability of the protein to function across the *melanogaster-simulans* species barrier.

To determine if defects in transposon silencing are due to loss of piRNAs, we sequenced small RNAs from ovaries of *rhi* mutants expressing chimeric Rhino proteins. The *mel-rhi,* chromo and hinge domain chimeras were able to restore production of piRNAs mapping to transposon repeats (Figures 2.3A to 2.3F). By contrast, *sim-rhi* and the shadow chimera failed to rescue transposon mapping piRNA expression, and were comparable to the null allelic combination. The primary piRNAs that initiate piRNA biogenesis are derived from clusters, and 42AB is the major dual strand piRNA cluster in *melanogaster* germline cells. As shown in Figure 2.3G, 42AB piRNAs are lost in *rhi* mutants. *mel-rhi* and the chromo and hinge domain chimeras are able to rescue piRNA production, but *sim-rhi* and shadow domain chimera do not. Rhi functions with Del and Cuff to suppress splicing of piRNA cluster transcripts, and *mel*-Rhi and the chromo and

hinge chimeras are able to suppress the splicing at clusters (Figure 2.S3C). By contrast, *sim*-Rhi and the shadow domain chimera fail to suppress splicing of piRNA precursors.

WT Rhi localizes specifically to piRNA clusters. We therefore performed ChIP-seq to determine if the chimeric proteins support cluster localization. To assay localization independent of the ability to promote cluster assembly, these studies were done in a WT genetic background. As shown in Figure 2.3H, full length *mel*-Rhi, the chromo and hinge chimeras bind to the 42AB piRNA cluster. By contrast, *sim*-Rhi and shadow chimera show signal comparable to GFP-nls control (Figure 2.3H). Intriguingly, the shadow domain shows the strongest signature of positive selection (Vermaak et al., 2005). Adaptive evolution of this domain thus prevents Rhi function across the *simulans-melanogaster* species barrier, and appears to alter sequences that direct localization to piRNA clusters.

Figure 2.2

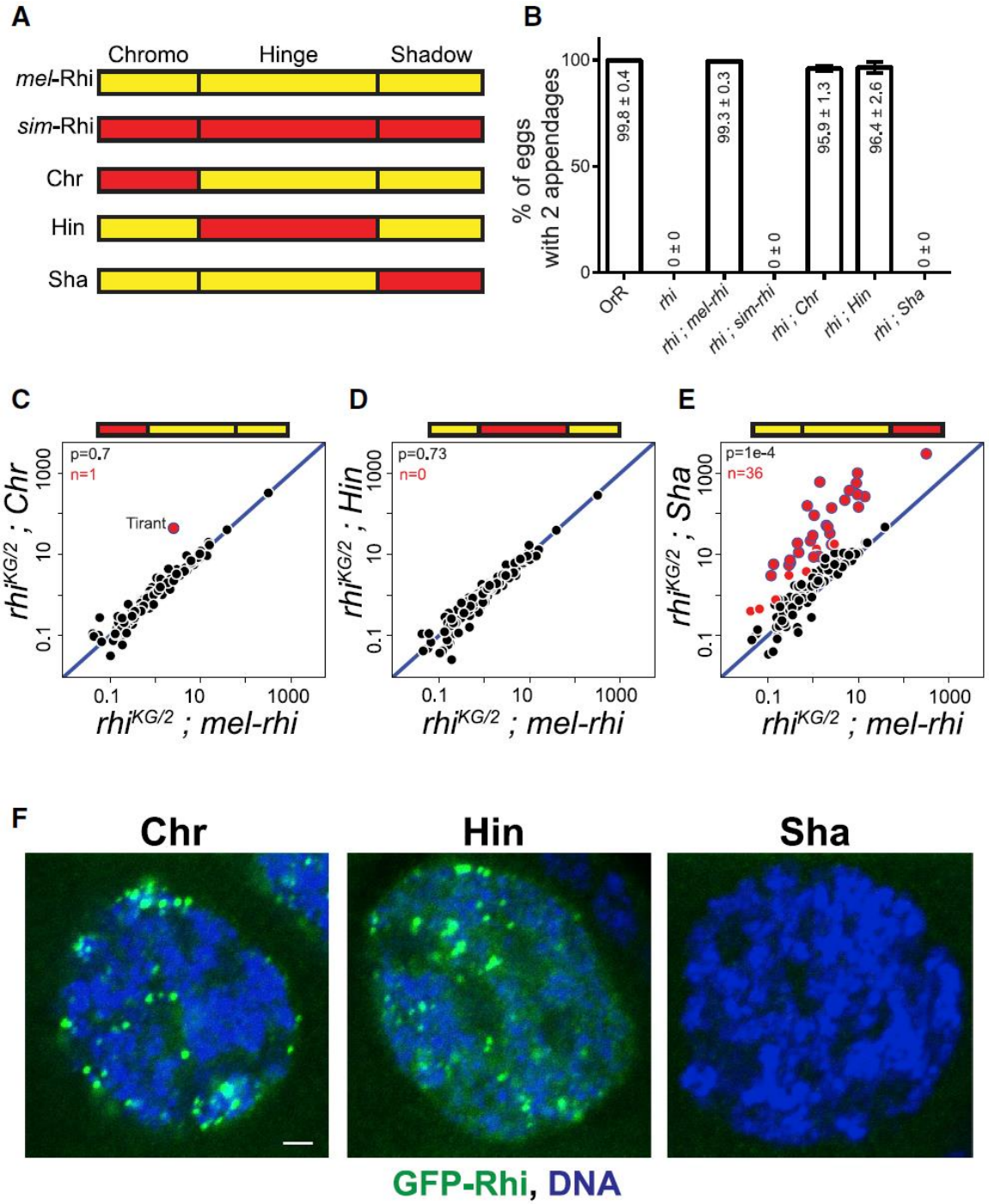

A
Chromo
Hinge
Shadow
mel-Rhi
sim-Rhi
Chr
Hin
Sha
B
% of eggs with 2 appendages
100
50
0
99.8 ± 0.4
0 ± 0
99.3 ± 0.3
0 ± 0
95.9 ± 1.3
96.4 ± 2.6
0 ± 0
OrR
rhi
rhi ; mel-rhi
rhi ; sim-rhi
rhi ; Chr
rhi ; Hin
rhi ; Sha
C
D
E
p=0.7
n=1
Tirant
p=0.73
n=0
p=1e-4
n=36
rhiKG/2 ; Chr
rhiKG/2 ; Hin
rhiKG/2 ; Sha
rhiKG/2 ; mel-rhi
0.1
10
1000
F
Chr
Hin
Sha
GFP-Rhi, DNA

Figure 2.2. The Shadow domain of *sim*-Rhi does not function in sibling species *melanogaster*

(A) Design of Rhino chimeras: Chromo (Chr), Hinge (Hin) and Shadow (Sha) domains are shown for *mel*-Rhi (yellow) and *sim*-Rhi (red). Each domain from *sim*-Rhi is placed in the *mel*-Rhi backbone and expressed as a GFP tagged transgene driven by the *rhi* promoter.

(B) Bar graphs showing percentages of eggs with normal dorsoventral patterning produced by OrR (WT control), *rhi* mutant, and *rhi* mutants expressing *mel-rhi*, *sim-rhi* or the chimeric Rhi variants. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *rhi* mutants and *rhi* mutants rescued by *sim-rhi* or Shadow chimera where average of at least 30 eggs were scored.

(C-E) Scatterplots showing transposon expression, measured by RNA-seq, in ovaries of *rhi* mutant expressing the chimeras vs. *mel-rhi*. Each point represents rpkm values for a different transposon family. Diagonal represents x=y. Points in red show y/x>5 (n is number of these transposons). Blue bordered points are over-expressed in *rhi* mutants, *rhi* mutants expressing *sim-rhi*, and *rhi* mutants expressing the *Sha* chimera. p value for differences obtained by Wilcoxon test.

(F) Localization of *rhi* promoter driven GFP tagged Rhino variants in *melanogaster* germline. GFP-Rhi is in green and DNA is in blue. Scale bar: 2 μm.

Figure 2.3

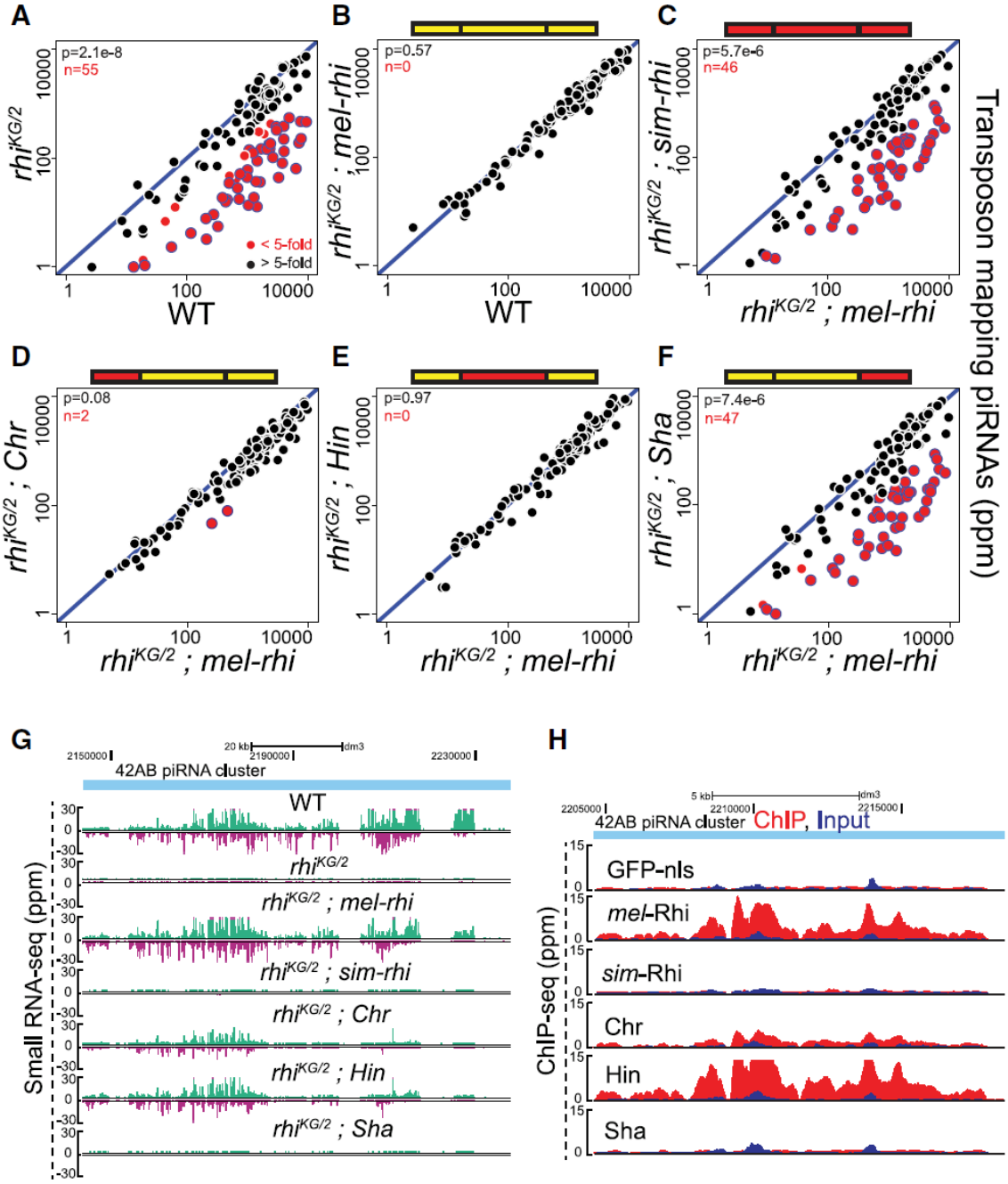

A
B
C
D
E
F
G
H
p=2.1e-8
n=55
p=0.57
n=0
p=5.7e-6
n=46
p=0.08
n=2
p=0.97
n=0
p=7.4e-6
n=47
< 5-fold
> 5-fold
WT
rhiKG/2 ; mel-rhi
rhiKG/2 ; sim-rhi
rhiKG/2 ; Chr
rhiKG/2 ; Hin
rhiKG/2 ; Sha
Transposon mapping piRNAs (ppm)
42AB piRNA cluster
Small RNA-seq (ppm)
ChIP-seq (ppm)
GFP-nls
mel-Rhi
sim-Rhi
Chr
Hin
Sha
ChIP, Input

Figure 2.3. *sim*-Rhi and Shadow chimera do not bind to piRNA clusters and fail to support piRNA production

(A-F) Scatterplots showing abundance of transposon mapping piRNAs in ovaries of *rhi* mutant (A), *rhi* mutant expressing either *mel-rhi* (B) vs. WT control, *sim-rhi* (C) or the chimeras (D-F) vs. *mel-rhi*. Points in red show x/y>5 (n is number of these transposons). Blue bordered points have reduced expression in *rhi* mutants, *rhi* mutants expressing *sim*-Rhi, and *rhi* mutants expressing the Sha chimera. p value for differences is obtained by Wilcoxon test.

(G) Genome browser view showing abundance of piRNAs uniquely mapping to 42AB piRNA cluster in WT, *rhi* mutant and *rhi* mutants expressing *mel-rhi*, *sim-rhi* or chimeric proteins. The Watson strand is in green, and Crick strand in magenta.

(H) Genome browser view of ChIP-seq profiles at 42AB cluster for *mel*-Rhi, *sim*-Rhi, chimeras, and GFP-nls control. All ChIP done under identical conditions, using the same anti-GFP antibody. ChIP signal in red, input signal in blue.

Evolution of the Rhi-Del interaction

To determine if species-specific substitutions in Rhi alter interactions with partner proteins, we expressed GFP tagged full length *simulans* and *melanogaster* Rhi, and each of the chimeras, in WT *melanogaster* ovaries, immuno-precipitated the tagged proteins, and identified associated polypeptides by mass spectrometry. We expressed these transgenes using both the *rhi* promoter, and the inducible *UASp* promoter with the germline specific *nanos*-Gal4 driver. Rhi is expressed at low levels and peptide counts with the *rhi* promoter were low, making statistical analysis difficult. As the interacting proteins were very similar with two systems, we focused our analysis on the *nanos*-Gal4 driven transgenes (Figures 2.4A, 2.4B and 2.S4A). Spectrum counts for GFP tag and the Rhi protein variants were similar, indicating that the fusion proteins were expressed at comparable levels, and precipitated with similar efficiencies. For each of the fusions, we quantified co-precipitating proteins (normalized iBAQ), and divided normalized protein levels in the *mel*-Rhi control by normalized protein levels with each experimental fusion (see Methods). Pseudocounts were used to avoid dividing by zero, and proteins with reduced binding relative to control produce a ratio greater than 1. Figures 2.4A and 2.4B show rank order of this ratio for full length *sim*-Rhi and the shadow chimera. The vast majority of proteins were present at similar levels in all samples, including the GFP control, and ratios clustered around 1. A single protein, Del, was not detected with *sim*-Rhi or the shadow chimera, and but showed essentially identical binding to full length *mel*-Rhi and the chromo and hinge chimeras (Figure 2.S4A). We have been unable to obtain antibodies to Rhi or Del that work reliably on Western blots, and therefore

confirmed these observations by expressing Rhi:GFP with Del:FLAG:mRFP fusions in *melanogaster* ovaries, and performing reciprocal co-immunoprecipitation and Western blotting. As shown in Figure 2.S4B, *melanogaster* Del co-precipitated with *mel*-Rhi, and *mel*-Rhi precipitated with *melanogaster* Del. However, *melanogaster* Del did not co-IP with *sim*-Rhi, and *sim*-Rhi did not co-IP with *melanogaster* Del. Substitutions in *sim*-Rhi shadow thus prevent binding to Del from *melanogaster*. Previous studies indicate that Del is required for primary piRNA biogenesis and interacts with Rhi through the shadow domain (Mohn et al., 2014). In yeast two hybrid assays, Del also interacts with the Rai1/DXO homolog Cuff, which functions with Rhi and Del in piRNA precursor processing. Adaptive evolution has therefore targeted a Rhi-Del interaction that is essential to the assembly of the nuclear piRNA precursor processing machinery, and generated a barrier to function across closely related species.

Figure 2.4

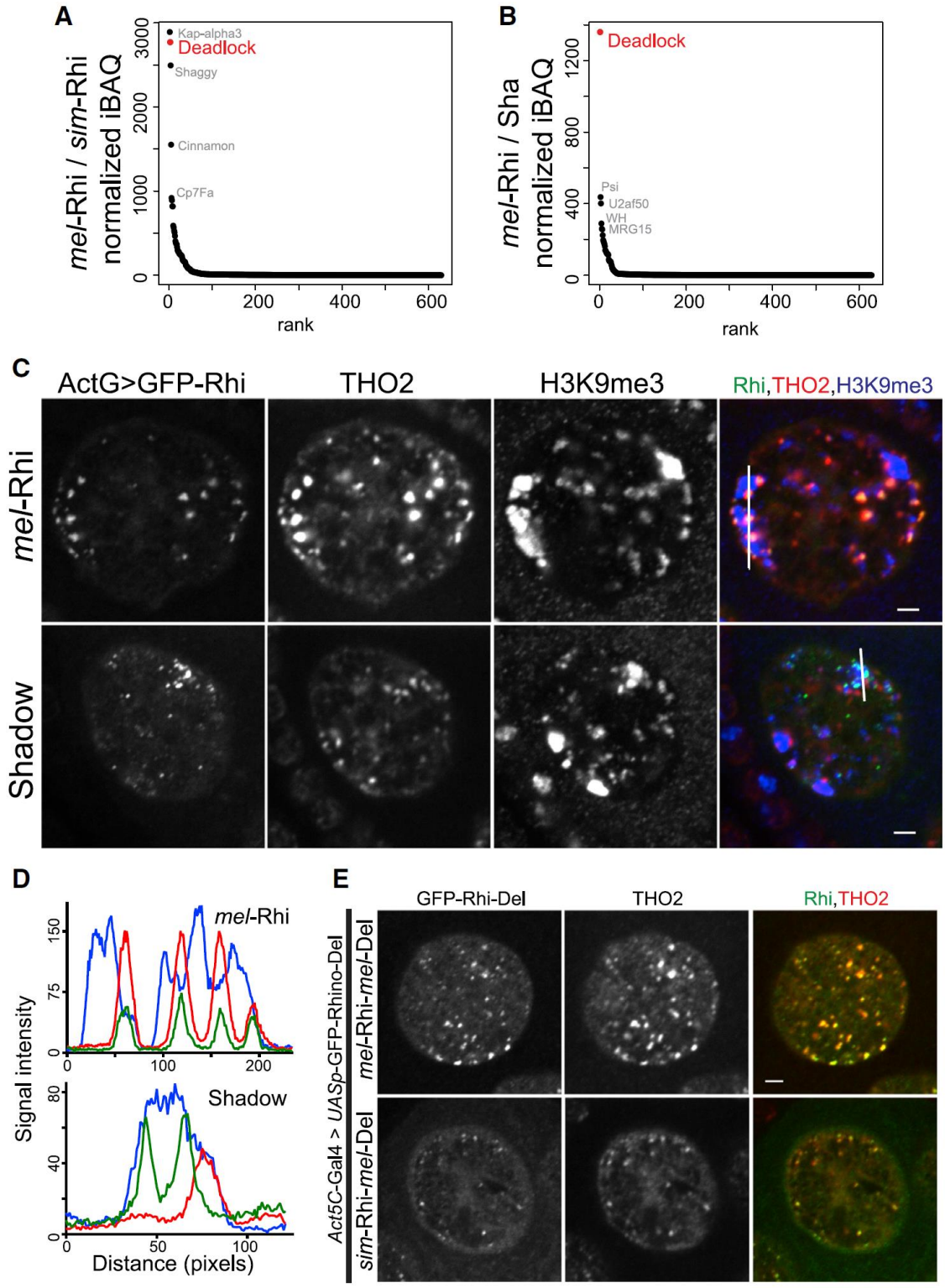

A
mel-Rhi / sim-Rhi normalized iBAQ
Kap-alpha3
Deadlock
Shaggy
Cinnamon
Cp7Fa
rank
B
mel-Rhi / Sha normalized iBAQ
Deadlock
Psi
U2af50
WH
MRG15
rank
C
ActG>GFP-Rhi
THO2
H3K9me3
Rhi,THO2,H3K9me3
mel-Rhi
Shadow
D
Signal intensity
mel-Rhi
Shadow
Distance (pixels)
E
GFP-Rhi-Del
THO2
Rhi,THO2
Act5C-Gal4 > UASp-GFP-Rhino-Del
mel-Rhi-mel-Del
sim-Rhi-mel-Del

Figure 2.4. Cross species incompatibility in Rhi-Del interaction

(A, B) Mass spectrometric analysis of Rhi binding proteins. Graphs showing ratios of GFP normalized iBAQ values for *mel*-Rhi vs. *sim*-Rhi (A), and *mel*-Rhi vs. Sha chimera (B), ranked by ratio values. Transgenes were expressed in *melanogaster* using the germline specific *nanos*-Gal4 driver.

(C) Localization of THO2 (piRNA cluster marker), H3K9me3 marked chromatin in the germline nuclei expressing *Act5C*-Gal4 driven Rhi:GFP. Color assignments for merged image shown on top. Scale bar: 2μm. Fluorescence intensities are measured along the line shown in merged image for Rhi:GFP (green), THO2 (red) and H3K9me3 (blue) as depicted in (D).

(E) Localization of *Act5C*-Gal4 driven Rhi-Del fusion proteins with respect to THO2 (piRNA cluster marker) in the female germline nuclei. Scale bar: 2μm.

Rhino localization to piRNA clusters

simulans Rhi fails to function in *melanogaster* and does not interact with *melanogaster* Del, which is required for Rhi localization to the nucleus (Mohn et al., 2014). To determine if the requirement for Del could be bypassed by driving Rhi into the nucleus by an independent mechanism, we generated a transgene expressing *simulans* Rhi with 3 copies of the nuclear localization signal (nls) from SV40 large T antigen (Kalderon et al., 1984). The *sim*-Rhi-nls protein localized to germline nuclei and formed foci (Figure 2.S4C). However, it did not rescue *rhi* mutations (Figure 2.S4D) and the TREX complex component THO2, which co-localizes with Rhino at clusters (Hur et al., 2016), did not co-localize with the foci formed by *sim*-Rhi-nls (Figure 2.S4C). The Shadow domain chimera does not bind Del, but is not detected in nuclear foci when expressed by the *rhi* promoter and imaged using our standard confocal procedures. However, the GFP tagged shadow chimera, over-expressed using either the constitutive *Act5C*-Gal4 driver or the germline specific *nanos*-Gal4 driver, localizes to distinct nuclear foci. Significantly, these foci do not co-localize with THO2 (Figures 2.4C and 2.4D). By contrast, similarly expressed full-length *melanogaster* Rhi consistently co-localizes with downstream components of the piRNA machinery (Figure 2.4C). In addition, ChIP-seq confirms that over-expressed shadow chimera fails to localize to clusters, while over-expressed *mel*-Rhi shows normal cluster binding (Figure 2.S4F). These data strongly suggest that binding to Del is required to direct Rhino to clusters.

The chromo domain of Rhi binds to H3K9me3, and this mark and Rhi show similar distributions over germline clusters. However, H3K9me3 also marks constitutive

centromeric heterochromatin and a number of euchromatic loci, which do not bind Rhi or produce piRNAs. This restricted distribution is reflected in cytological localization of Rhi foci to the periphery of large H3K9me3 domains, but not within these domains (Figures 2.4C and 2.4D). By contrast, the foci formed by the shadow domain chimera are embedded within the prominent H3K9me3 domains, and do not accumulate to the periphery of these domains (Figures 2.4C and 2.4D). These findings strongly suggest that binding to Del directs Rhi to H3K9me3 marks on clusters, and away from bulk heterochromatin. To test this hypothesis, we generated a transgene expressing full length *sim*-Rhi fused through the C-terminal shadow domain to full length *melanogaster* Del. We included a flexible linker, to reduce conformational constraints. As a control, we generated an analogous transgene expressing *mel*-Rhi fused to *mel*-Del. In WT ovaries, both fusions localize to germline nuclei and formed distinct foci. Significantly, the foci formed by *sim*-Rhi fused to *mel*-Del also co-localized with THO2 (Figure 2.4E). Forced binding to Del is therefore sufficient to direct *sim*-Rhi to clusters.

To assay for fusion protein function, we crossed both transgenes into *rhi, del, and rhi,del* double mutants and assayed fertility. Surprisingly, neither fusion transgenes rescued *rhi, del* or the double mutant combination (Figure 2.S4E), and the fusion proteins show very weak localization to nuclear foci. Directly linking Rhi to Del is therefore sufficient for cluster localization in WT ovaries, where these chromatin domains appear to be established by endogenous gene products. However, the fusions are not able to promote assembly of these domains. The fusion could disrupt the function of critical domains near the junction, but both proteins localize to clusters and a transgene

expressing GFP fused to N-terminus of Del rescues *del* mutants. We therefore speculate that a dynamic interaction between Rhi and Del is critical for the function of both proteins, which may shuttle between biochemically distinct complexes during biogenesis. For example, Del could bind to Rhi before moving to a distinct complex with downstream piRNA components, including Cuff, UAP56 and THO. While the reason the fusion proteins fail to rescue full biological function remains to be determined, these data indicate that the interaction with Del is required for Rhi localization specifically to piRNA cluster chromatin.

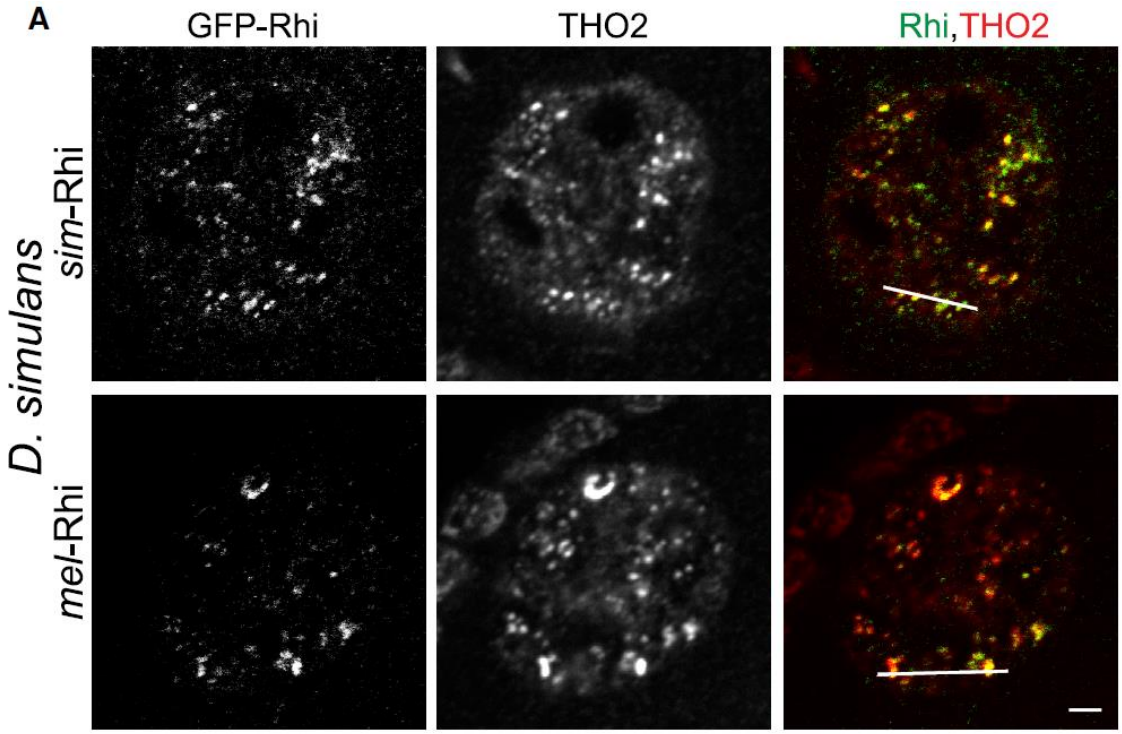
Directional incompatibility in Rhi-Del interaction

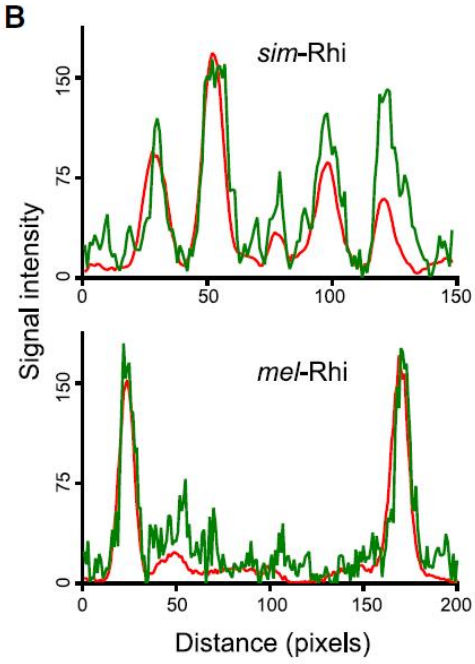
The inability of *sim*-Rhi to rescue *rhi* mutants or interact with *melanogaster* Del raised the possibility that Rhi-Del co-evolution had generated species-specific interaction interfaces that prevent cross-species heterodimer formation. Alternatively, the stable interaction of Rhi and Del could have evolved in the *melanogaster* lineage. To differentiate between these alternatives, we generated transgenic *simulans* lines expressing GFP tagged *sim*-Rhi and *mel*-Rhi, under control of the *rhi* promoter, and directly assayed subcellular localization and interactions with Del and other proteins. IP-mass spectrometry demonstrated that *simulans* Del co-precipitates with *sim*-Rhi, indicating that the interaction is conserved. Surprisingly, *simulans* Del also co-precipitated with *mel*-Rhi (Figure 2.5C). Consistent with these studies, *sim*-Rhi and *mel*-Rhi co-localized with THO2 in germline nuclear foci (Figures 2.5A and 2.5B). The interaction between Rhi and Del thus shows cross-species directionality: *mel*-Rhi is able to bind to Del from both *simulans* and *melanogaster*, and localizes in both species, while *sim*-Rhi binds to *sim*-Del and localizes to nuclear foci in *simulans*, but cannot interact with *mel*-Del or localize in the *melanogaster* germline.

These observations suggested that the *sim-del* gene may function in *melanogaster*. We therefore expressed mRFP or GFP tagged *sim*-Del and *mel*-Del in *melanogaster*, and assayed interacting proteins, subcellular localization, and the ability to rescue *del* mutations. IP and mass spectrometry on the tagged proteins expressed in WT ovaries showed that *mel*-Del and *sim*-Del interact with *mel*-Rhi (Figure 2.6A). The *mel*-Del and *sim*-Del proteins also formed nuclear foci that co-localized with THO2 (Figure 2.S5B).

To directly assay biological function, we expressed both tagged proteins in a *melanogaster del* mutant background. Mutations in *del* disrupt oocyte/embryo D-V patterning, embryo viability, and transposon silencing (Mohn et al., 2014; Wehr et al., 2006). The *mel-del* transgene rescued all three of these defects (Figures 2.6B, 2.6C and 2.6F). By contrast, the *sim-del* was similar to a null allele by all three measures (Figure 2.6B, 6C and 6G). These findings were surprising, given the robust cluster localization of *sim*-Del in WT ovaries (Figure 2.S5B). In the *del* mutant background, however, the *sim*-Del showed very weak localization to nuclear foci, which were present in only a subset of germline nuclei. These foci also showed very weak localization of Rhi, and the downstream proteins THO2 and UAP56 (Figures 2.6D, 2.S5A and 2.S5C). In WT ovaries, *sim*-Del thus appears to localize to clusters through endogenous Rhi, which functions with endogenous Del to promote cluster assembly. In the *del* mutants, by contrast, the *mel*-Rhi interacts with *sim*-Del, but the complex is unable to promote cluster assembly. This is likely due to defects in recruitment of downstream components of the pathway. We therefore speculate that Del has a minimum of two rapidly evolving functional domains, which mediate binding to Rhi and interactions with downstream proteins. Together, evolution of these domains creates a barrier to cross-species function.

Figure 2.5





C

Bait protein	Spectrum counts		
	GFP	Rhi	Deadlock
sim-Rhi	17	28	8
mel-Rhi	10	19	9
No GFP	0	0	0

Figure 2.5: *mel*-Rhi binds to Del in *simulans* and localizes to piRNA clusters

(A) Localization of *sim*-Rhi:GFP and *mel*-Rhi:GFP in the *simulans* female germline. Egg chambers were double labeled for the piRNA cluster marker THO2. Both forms of Rhi co-localize to nuclear foci with THO2. Scale bar: 2μm. Fluorescence intensities are measured along the line shown in merged image for Rhi:GFP (green), THO2 (red) as depicted in (B).

(C) Total spectrum counts for GFP, Rhi and Del co-precipitating with *sim*-Rhi and *mel*-Rhi expressed in *simulans*. Expression was driven with the *rhi* promoter. Both *sim*-Rhi and *mel*-Rhi coprecipitate with the endogenous *simulans* Del ortholog.

Figure 2.6

A

Bait protein	Spectrum counts		
	RFP	Del	Rhi
mel-Del	85	189	16
sim-Del	61	121	27
mRFP ctrl	115	2	0

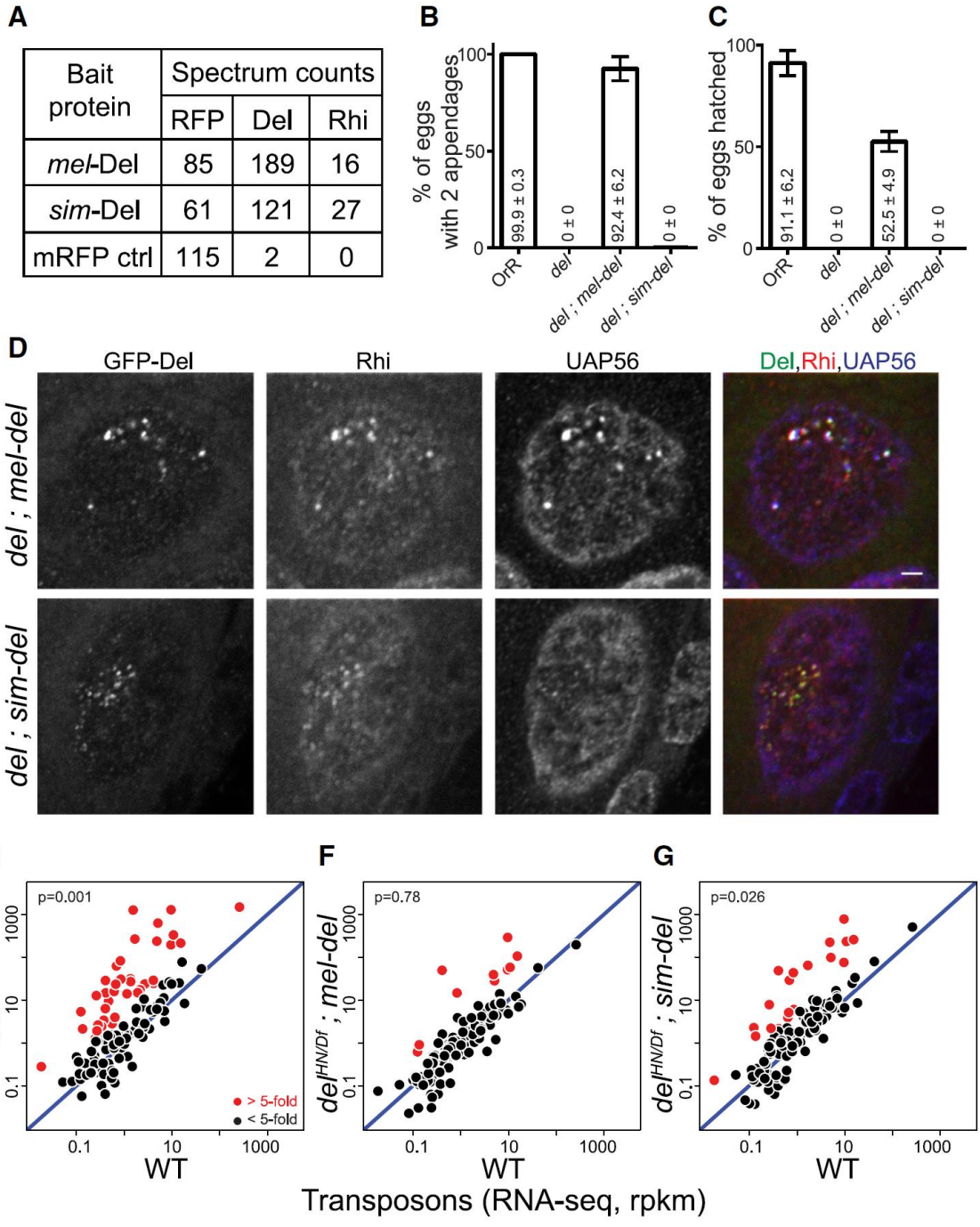


Figure 2.6. *sim*-Del binds to *mel*-Rhi, but fails to function in *melanogaster*

(A) Mass spectrometric analysis of Del binding proteins. Table shows total spectrum counts for the mRFP tag, Rhi and Del in immunoprecipitates of *mel*-Del, *sim*-Del and mRFP control, expressed in the *melanogaster* germline under *nanos*-Gal4 driver. *sim*-Del co-precipitates with *mel*-Rhi.

(B, C) Bar graphs showing percentages of eggs with normal dorsoventral patterning (B) and percentages of hatched eggs (C) produced by females of the following genotypes: OrR (WT control); *del* mutant; *del* mutants expressing either *mel-del* or *sim-del*. *sim-del* fails to rescue embryo patterning and hatching. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 100 embryos scored per replicate, except for *del* mutants where average of 7.33 eggs were scored.

(D) Localization of Rhi and UAP56 in *del* mutants expressing either *mel*-Del or *sim*-Del. Scale bar: 2μm (all images at same scale).

(E-G) Scatterplots showing transposon expression levels measured by RNA-seq in ovaries of *del* mutant (E), *del* mutant rescued by either *mel-del* (F) or *sim-del* (G) vs. WT control. Each point represents rpkm values for a different transposon. Diagonal represents x=y. Points in red show y/x>5. p value for differences is obtained by Wilcoxon test. *sim-del* fails to rescue transposon silencing.

Discussion

piRNA clusters determine sequence specificity for transposon silencing by the piRNA pathway, and thus function at the heart of the adaptive genome immune system (Brennecke et al., 2007). Rhi localizes specifically to piRNA clusters, where it promotes cluster transcription, suppresses cluster transcript splicing, and promotes piRNA production (Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014). Our data indicate that co-evolution of this key component of the piRNA machinery, with its partner Del, prevents assembly of functional complexes across the sibling species barrier. The *simulans rhi* and *del* genes do not rescue *melanogaster* mutants, *sim*-Rhi does not interact with *mel*-Del, and *sim*-Del forms a non-functional complex with *mel*-Rhi. These studies also provide intriguing insights into the mechanisms that localize Rhi to clusters. Substituting the *mel*-Rhi rapidly evolving Shadow domain with the *simulans* Rhi shadow domain is sufficient to block piRNA production, transposon silencing, and binding to Del. Significantly this Shadow domain chimera does not localize to germline piRNA clusters and forms ectopic nuclear foci that do not co-localize with the downstream processing machinery. By contrast, a fusion between *sim*-Rhi and *mel*-Del localizes to clusters. Binding to Del, and likely subsequent recruitment of downstream piRNA biogenesis proteins, thus allows Rhi to discriminate between H3K9me3 marks at clusters and other repeats (Figure 2.7A).

How Del directs Rhi to clusters remains to be determined, but piRNA clusters are actively transcribed, while most heterochromatin marked by H3K9me3 is transcriptionally silent. In addition, many of the piRNA biogenesis factors that co-

localize with Rhino at clusters, including Cuff, UAP56 and the THO complex, are RNA binding proteins that appear to be loaded on cluster transcripts co-transcriptionally. We therefore propose that Rhino, though it's Chromo domain, samples available H3K9me3 marks, moving between centromeric heterochromatin and clusters. At clusters, however, interactions between Del and Cuff, a putative RNA end binding protein, lead to recruitment of UAP56 and THO, which bind single stranded RNA. We propose that Rhino in the resulting chromatin bound RNA complex does not exchange with the soluble pool, driving specific accumulation at clusters (Figure 2.7A). However, covalent binding of Rhi to Del prevents function. We therefore propose that as piRNA precursors are released from these complexes, Rhi and Del dissociate and reinitiate assembly of the chromatin bound RNPs, and that fusion of Del to Rhi prevents this recycling step. While speculative, this model provides a useful framework for future studies.

Figure 2.7

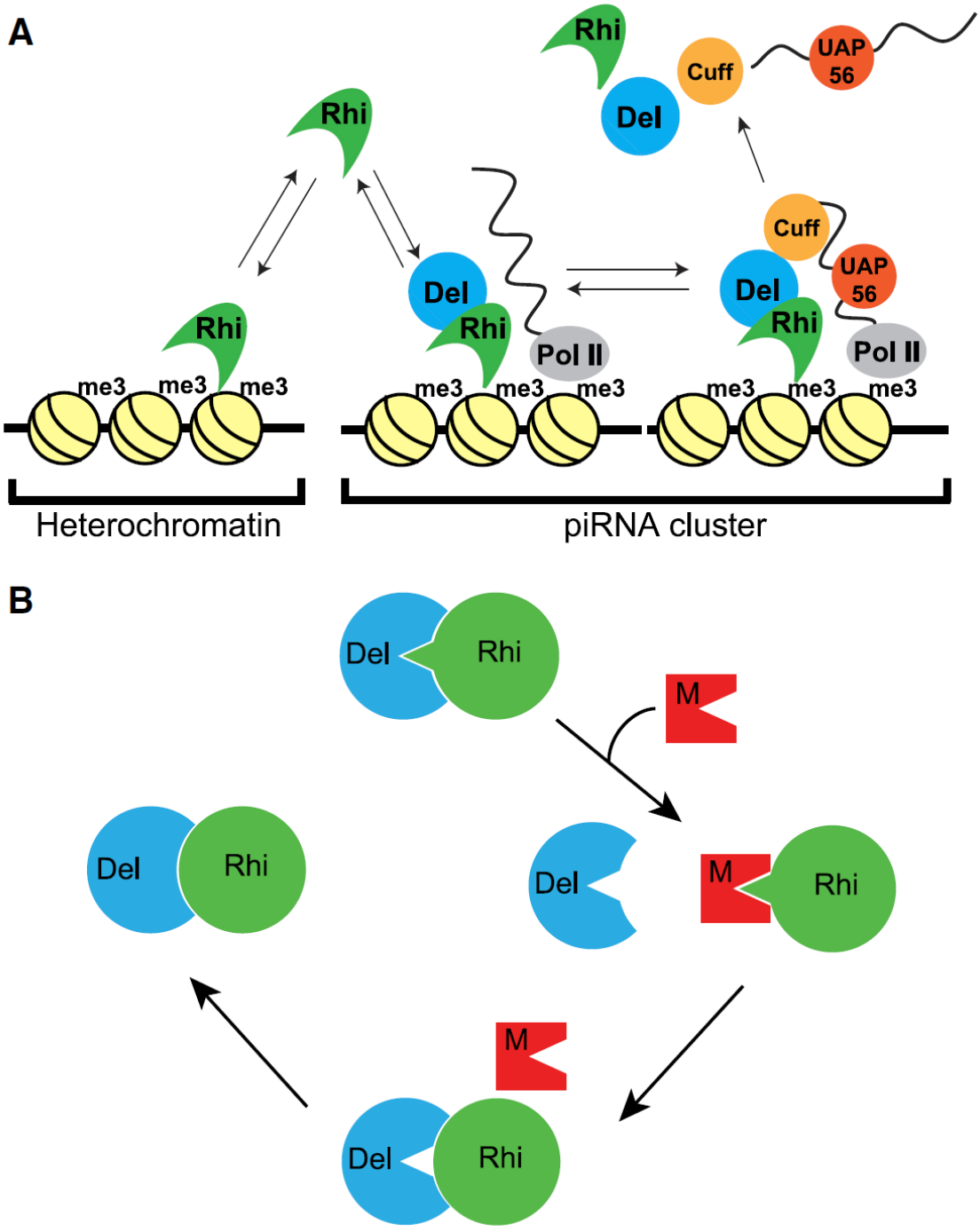

A
Rhi
Del
Cuff
UAP
56
Rhi
Del
Rhi
Pol II
Cuff
UAP
56
Del
Rhi
Pol II
me3
me3
me3
me3
me3
me3
me3
me3
me3
Heterochromatin
piRNA cluster
B
Del
Rhi
M
Del
Rhi
Del
M
Rhi
M
Del
Rhi

Figure 2.7. Model for co-evolution of the Rhino-Deadlock interface

(A) Model for Del function in Rhi localization to piRNA clusters. The Rhi Chromo domain interact with H3K9me3 marks throughout the genome, but most of these marks are in transcriptionally silent regions. At clusters, which are transcribed, Del interactions with Cuff, a putative RNA end binding protein, leads to formation of a chromatin bound complex, which recruits additional RNA binding components (i.e. UAP56). Assembly of these complexes leads to Rhi accumulation at clusters. We further propose that release of piRNA precursor complexes is accompanied by release and recycling of Rhi, Del and Cuff, which then re-initiate the cycle.

(B) A transposon mutation generates a protein that mimics the Del surface that binds to Rhi. Competition for productive Rhi-Del complex formation disrupts piRNA biogenesis and causes increased transposition. Reduced fertility leads to selection of Rhi mutations that reduce mimic binding, at the expense of Rhi-Del affinity. "Leaky" transposition leads to selection of Del mutations that completely restore Rhi binding. Pathogen mimicry thus leads to rapid evolution of Rhi-Del interface.

piRNA pathway evolution

The piRNA pathway has a conserved function in germline development and transposon silencing, but piRNA sequence composition and biogenesis mechanisms show remarkable phylogenetic diversity, and many genes in the piRNA pathway are evolving rapidly under positive selection, and are poorly conserved (Chirn et al., 2015; Khurana and Theurkauf, 2010; Obbard et al., 2009; Simkin et al., 2013; Yi et al., 2014; Zanni et al., 2013). We directly determined the functional consequences of piRNA gene divergence over a short evolutionary time scale, by expressing the *simulans rhi* and *del* genes in *melanogaster* and assaying protein interactions, subcellular localization, and the ability to rescue chromosomal mutations. Our data indicate that rapid co-evolution of *rhi* and *del* has generated orthologs that do not form functional complexes across the sibling species barrier.

How did this incompatibility arise? *rhi* and *del* are evolving rapidly under positive selection, characteristic of genes engaged in a "Red Queen" host-pathogen arms race, presumably with transposons, which function as mobile genome pathogens. A simple Red Queen system, however, leads to rapid co-evolution of host and pathogen genes that encode interacting proteins (Daugherty and Malik, 2012; Duggal and Emerman, 2012; Elde and Malik, 2009), and *rhi* and *del* encode interacting components of the host defense system. This could arise through pathogen mimicry. In this variant of an arms race, mutations in the pathogen generate a protein with a surface that structurally mirrors a host partner engaged in an interaction required for defense. Competition with the functional interaction leads to pathogen propagation, reduced host fitness, and selection of host

mutations that reduce host binding to the mimic (Daugherty and Malik, 2012; Elde and Malik, 2009). Figure 2.7B outlines a speculative model for an evolutionary cycle driven by a mimic targeting the Rhi-Del interface. In this model, mutations in a transposon gene (retrotransposon gag, pol or env, for example) generate a mimic of the Del surface that interacts with Rhi. Mimic competition with Del for Rhi reduces productive dimer formation and piRNA production, leading to increased transposition. Reduced fertility then leads to selection of mutations in Rhi that reduce mimic binding and increase the Del interaction, but at a cost of biding affinity. "Leaky" transposon silencing could then lead to selection of Del mutations that restore high affinity binding to Rhi (Figure 2.7B). This "mimicry cycle" thus remodels the Rhi-Del interface. As diagrammed in Figure 2.S6, this process also has the potential to produce "directional incompatibility" at the dimerization interface.

Our data, with the finding that *melanogaster- simulans* inter-species hybrids phenocopy piRNA mutations (Kelleher et al., 2012), raise the intriguing possibility that adaptive evolution of piRNA pathway genes directly contributes to the reproductive barrier between these sibling species. For example, we show that *sim*-Del binds to *mel*-Rhi, but is unable to rescue *melanogaster del* mutations or direct Rhino to cluster chromatin. In hybrids, non-functional complexes between *sim*-Del and *mel*-Rhi may therefore complete with productive complexes between proteins from the same species. It is unclear if the resulting reduction in functional Rhi-Del complexes would be sufficient to trigger the observed transposon silencing defects. However, our preliminary data indicate that adaptive evolution of additional piRNA genes prevents cross-species

function (Parhad and Theurkauf, unpublished), and mimics could target any non-redundant interaction in the piRNA pathway. We therefore speculate that adaptive evolution of several piRNA pathway genes leads to multiple biochemical incompatibilities, which together disrupt piRNA function in hybrids.

It is difficult to directly test the hypothesis that transposon-encoded mimics drive piRNA pathway evolution, as mimics would arise over an evolutionary time scale (millions of years), and once bypassed by a compensating host mutation, provide no selective benefit and would be lost. However, mimics targeting piRNA biogenesis are predicted to mobilize all transposon families with functional copies in the genome, not just the transposon family that produced the mimic, leading to bursts of global transposon activity. Consistent with this prediction, *melanogaster* has retained active copies of almost all transposon families, while *simulans* has retained very few functional transposons (Lerat et al., 2011). Furthermore, global bursts of transposition are associated with species divergence in plants and animals (Belyayev, 2014; Fontdevila, 2005), which could reflect periodic relaxation of germline transposon silencing. We therefore speculate that an ongoing arms race between transposons and the piRNA pathway facilitates speciation by producing biochemical incompatibilities that help build reproductive barriers, and by triggering bursts of insertional mutations that serve as substrates for natural selection.

Experimental procedures

Table 2.1. Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GFP Booster_ATTO488 (Immuno-staining, 1:200)	ChromoTek	Cat# gba488-100
Rat anti-THO2 (Immuno-staining, 1:2000)	(Rehwinkel et al., 2004)	N/A
Rabbit anti-UAP56 (Immuno-staining, 1:1000)	(Eberl et al., 1997)	N/A
Guinea pig anti-Rhi (Immuno-staining, 1:1000)	(Klattenhoff et al., 2009)	N/A
Mouse anti-GFP (Western, 1:1000)	Santacruz	Cat# sc-9996
Mouse anti-FLAG (Western, 1:4000)	Sigma	Cat# F3165
Rabbit anti-GFP (ChIP)	ThermoFisher Scientific	Cat# A11122
Chemicals, Peptides, and Recombinant Proteins		
Superscript III	ThermoFisher Scientific	Cat# 18080-085
dNTP mix	NEB	Cat# N0447L
RNase OUT	ThermoFisher Scientific	Cat# 10777-019
TURBO DNase	ThermoFisher Scientific	Cat# AM2238
dUTP mix	Bioline	Cat# BIO-39041
RNaseH	ThermoFisher Scientific	Cat# 18021-071
DNA polymerase I	NEB	Cat# M0209S
T4 DNA polymerase	NEB	Cat# M0203L
Klenow DNA polymerase	NEB	Cat# M0210S
T4 PNK	NEB	Cat# M0201L
Klenow 3’ to 5’ exo	NEB	Cat# M0212L
T4 DNA ligase	Enzymatics Inc.	Cat# L6030-HC-L
UDG	NEB	Cat# M0280S
Phusion Polymerase	NEB	Cat# M0530S
16% formaldehyde	Ted Pella Inc	Cat# 18505
Gateway® LR Clonase® Enzyme mix	ThermoFisher Scientific	Cat# 11791019
In-Fusion® HD Cloning Kit	Clontech	Cat# 639648
Critical Commercial Assays		

mirVANA™ miRNA isolation kit	ThermoFisher Scientific	Cat# AM1560
Dynabeads® Protein G	ThermoFisher Scientific	Cat# 10004D
GFP-Trap®_A beads	Chromotek	Cat# gta-100
RFP-Trap®_A beads	Chromotek	Cat# rta-100
RNeasy Mini Kit	Qiagen	Cat# 74104
RNA Clean & Concentrator-5	Zymo Research	Cat# R1015
Ribo-Zero™ Gold rRNA removal kit	Illumina	Cat# MRZG12324
Deposited Data		
High throughput Sequencing	This study	NCBI Trace archive SRP111075
Raw data	This study	http://dx.doi.org/10.17632/w2ym383bp8.1
Experimental Models: Organisms/Strains		
D. melanogaster: *rhiP* > GFP-*mel*-Rhi	This study	N/A
D. melanogaster: *rhiP* > GFP-*sim*-Rhi	This study	N/A
D. melanogaster: *rhiP* > GFP-Chr	This study	N/A
D. melanogaster: *rhiP* > GFP-Hin	This study	N/A
D. melanogaster: *rhiP* > GFP-Sha	This study	N/A
D. melanogaster: *rhiP* > GFP-*sim*-Rhi-nls	This study	N/A
D. melanogaster: *UASp* > GFP-*mel*-Rhi	This study	N/A
D. melanogaster: *UASp* > GFP-*sim*-Rhi	This study	N/A
D. melanogaster: *UASp* > GFP-Chr	This study	N/A
D. melanogaster: *UASp* > GFP-Hin	This study	N/A
D. melanogaster: *UASp* > GFP-Sha	This study	N/A
D. melanogaster: *UASp* > GFP-*sim*-Rhi-nls	This study	N/A
D. melanogaster: *rhiP* > GFP-*mel*-Del	This study	N/A
D. melanogaster: *rhiP* > GFP-*sim*-Del	This study	N/A
D. melanogaster: *UASp* > mRFP-FLAG-*mel*-Del	This study	N/A
D. melanogaster: *UASp* > mRFP-FLAG-*sim*-Del	This study	N/A
D. melanogaster: *rhiP* > GFP-*mel*-Rhi-*mel*-Del	This study	N/A
D. melanogaster: *rhiP* > GFP-*sim*-Rhi-*mel*-Del	This study	N/A
D. melanogaster: *UASp* > GFP-*mel*-Rhi-*mel*-Del	This study	N/A

D. melanogaster: *UASp* > GFP-*sim*-Rhi-*mel*-Del	This study	N/A
D. simulans: *rhiP* > GFP-*mel*-Rhi	This study	N/A
D. simulans: *rhiP* > GFP-*sim*-Rhi	This study	N/A
D. melanogaster: rhi$^{KG/2}$	(Klattenhoff et al., 2009)	N/A
D. melanogaster: del^{HNDf}	(Wehr et al., 2006)	N/A
D. melanogaster: Oregon-R	William Theurkauf lab	N/A
D. melanogaster: w^{1}	William Theurkauf lab	N/A
D. melanogaster: Act5C > Gal4	William Theurkauf lab	N/A
D. melanogaster: nanos > Gal4	William Theurkauf lab	N/A
D. simulans: w^{501}	*Drosophila* species stock center UCSD	Stock# 14021-0251.011
D. melanogaster: *vasP*-GFP-nls	(Zhang et al., 2014)	N/A
D. melanogaster: *UASp-Sep5.mRFP*	Bloomington stock	Stock # 56492
Oligonucleotides		
Sequences given in Method details	Integrated DNA Technologies (IDT)	N/A
Random primers	ThermoFisher Scientific	Cat# 48190011
Recombinant DNA		
pENTR™/D-TOPO®	ThermoFisher Scientific	Cat# K2400-20
pUAST-attB	(Bischof et al., 2007)	N/A
Drosophila gateway vector: pPGW	Drosophila Genomic Resource Center (DGRC)	Stock# 1077
Drosophila gateway vector: pPW	DGRC	Stock# 1130
Drosophila gateway vector: pPRW	DGRC	Stock# 1137
Drosophila gateway vector: pPFW	DGRC	Stock# 1117
Software and Algorithms		

GraphPad Prism	https://www.graphpad.com/scientific-software/prism/
RStudio	https://www.rstudio.com/
ImageJ	https://imagej.nih.gov/ij/
Adobe Creative Suite 6	Adobe Systems Inc.
Scaffold	http://www.proteomesoftware.com/products/scaffold/
UCSC Genome Browser	https://genome.ucsc.edu/cgi-bin/hgGateway
Microsoft Office	Microsoft
Bowtie	(Langmead et al., 2009)
BEDTools	(Quinlan and Hall, 2010)
TopHat	(Trapnell et al., 2009)
BWA	(Li and Durbin, 2009)

Fly stocks

All experiments were performed on females of two *Drosophila* species: *Drosophila melanogaster* and *Drosophila simulans*. All flies were kept at 25^{O}C on cornmeal medium. All *D. melanogaster* transgenic lines were generated by ϕC31 integration at 3L-68A4. All *D. simulans* transgenic lines in w^{501} strain were generated by random P-element mediated transformation. rhi^{KG} and rhi^{2} alleles were described in (Klattenhoff et al., 2009; Volpe et al., 2001). del^{HN} and del^{Df} alleles were obtained from Trudi Schüpbach (Princeton University). F1 females from OregonR crossed to w^{1} were used as wild type (WT) control, unless mentioned otherwise.

Transgenic flies

Gateway Technology from Invitrogen was used to generate the plasmids. ϕC31 attB was added to *Drosophila* gateway transformation vector pPGW to get attB-pPGW. ϕC31 attB was PCR amplified from pUASTattB plasmid (Bischof et al., 2007) (by using primers 5'-CGA TTA TGC ATG TCG ACG ATG TAG GTC ACG GTC TC -3' and 5'- AAT CGA TGC ATG TCG ACA TGC CCG CCG TGA CCG TC -3'). Both the PCR product and pPGW vector were digested by NsiI, gel purified and ligated to get the final vector attB-pPGW. This serves as entry vector for expressing N' GFP tagged proteins under *UASp* promoter. The procedure for generation of *rhiP*-attB-pPGW (for expressing N' GFP tagged proteins under *rhi* promoter) is as follows: *rhi* promoter was PCR amplified by using primers (PCR1: 5'- CTC TCT TTC TCG AGG TCA TCA AGC TTA GGC ATG TAC CAA GTT GTT AAC TCT ATC GAA TTA-3', 5'- GAA GAT TTC TCC TTG

ACG TTT CGG ACA CCC AAG GTT AGC CCA AAT CGA TGG ATT TCT GGG ACA TGA TC -3'; PCR2: 5'- GAT CAT GTC CCA GAA ATC CAT CGA TTT GGG CTA ACC TTG GGT GTC CGA AAC GTC AAG GAG AAA TCT TC -3', 5'- CTC ACC ATG GTG GCG GGC TTC TCT AGA CAG GAA CTT ATC CGC TCA CAG GAC GCC GAG CAA AAG -3') to introduce STOP codons in the upstream *Oxp* gene, from w^{1} genomic DNA. PCR1, PCR2 and StuI digested attB-pPGW were ligated by Clontech In-Fusion® HD Cloning Kit. Same cloning strategy was used to express Rhino in *D. simulans*, except for *rhi* promoter cloned from *D. simulans* w^{501}(by using primers: PCR1:5'- CTC TCT TTC TCG AGG TCA TCA AGC TTA GGC ATG TAC CGA GTT GTT AAC TCT ATC GAA TTA -3', 5'- GAG GAT TTT TCC TTG ACG TTG CGG ACA CCA AGG GTT ATC CCA GAT CAA CTG ATT TCT TGG GCA TGA TC -3'; PCR2:5'- GAT CAT GCC CAA GAA ATC AGT TGA TCT GGG ATA ACC CTT GGT GTC CGC AAC GTC AAG GAA AAA TCC TC -3', 5'- CTC ACC ATG GTG GCG GGC TTC TCT AGA CAG GAA CTT AAA CGC TGA AAG GAC GCC GAG CAA ATG -3'). For expression of N' mRFP-FLAG tagged proteins, the vector attB-pPFVR was generated as follows: The FLAG tag was PCR amplified from pPFW (primers: CGG ACG AAT TTT TTT TTG AAA ACC GGT GAT AGA GCC TGA ACC AGA AAA G and GGA CTG GAA GTA CAG GTT CTC CTT GTC ATC GTC ATC CTT GTA ATC) and mRFP tag from pPRW (primers: GAG AAC CTG TAC TTC CAG TCC ATG GCC TCC TCC GAG GAC GTC ATC AAG and CAG CTT TTT TGT ACA AAC TTG TAT ACC GGT GGG CG). The FLAG and mRFP tag PCRs and AgeI digested attB-pPFW were ligated by Clontech In-Fusion® HD Cloning Kit. The resulting plasmid is gateway

destination vector attB-pPFVR, having ϕC31 attB site and expressing N' FLAG and mRFP tagged protein under UASp promoter. TEV protease site is cloned between FLAG and mRFP tag. *mel-rhi* (primers: CAC CAT GTC TCG CAA CCA TCA GCG ACC AAA TC and TTA CTT GGG CAC AAT GAT CCT CAA GCT C) from cDNA clone obtained from DGRC clone RE36324 (from Riken *y; cn, bw, sp* strain), *sim-rhi* (primers: CAC CAT GTC TCG CAA AAA TCA ACG ACC AAA TCT TG and TTA CTT GAG CAC AGT GGT CCT CAA GCT C) from cDNA from *simulans* C167.4 strain ovaries, *mel-del* (primers: CAC CAT GGA AAA GTT GGA CAA AAT AAG GAT G and TTA ATC AAA ATT ATG TAT ATT GAT CGC ATA TTC ATT GG) from OregonR genomic DNA, *sim-del* (primers: CAC CAT GGA AAA CTT GGC TAA AAT AAG GAT G and TTA ATC AAA ATG ATG TAT ATT GGT CGT A) from *simulans* C167.4 genomic DNA were cloned in the gateway entry vector with pENTR directional TOPO cloning kit (Invitrogen). (The sequences are provided in Mendeley Data). *sim*-Rhi-nls was made by PCR of *sim*-Rhi with reverse primer encoding for nls sequence (TTA CAC CTT GCG CTT CTT CTT TGG ATC CAC CTT GCG CTT CTT CTT TGG ATC CAC CTT GCG CTT CTT CTT TGG ATC AGC TCG GGA TCT GAG TCC GGA CTT GAG CAC AGT GGT CCT CAA GCT C) and forward primer mentioned above. Rhi and Del fusions were made by gene synthesis and cloned into pENTR vector. Del was at C' end of Rhi, with a flexible linker in the middle. The plasmids obtained after LR gateway cloning reaction were sequenced and injected into respective fly strains.

Fertility assays

2-4 day old flies were kept on grape juice agar plates for 1 or 2 days. After removal of flies, the eggs were scored for fused appendages and the hatching was measured after 2 days. The bar graphs show mean and standard deviation for 3 biological replicates, with the indicated number of scored embryos.

Immuno-staining

Immuno-staining and image analysis was done as described in (McKim et al., 2009; Zhang et al., 2012a). In brief, 2-4 day old female ovaries were fixed with 4% formaldehyde, washed, incubated overnight with primary antibody, washed, incubated with secondary antibody with fluorophore overnight and mounted on slide. ChromoTek anti-GFP Booster (Atto-488) antibody added with secondary antibody to enhance GFP signal.

Immuno-precipitation

2-4 day old female ovaries were dissected in Robb's buffer. The ovaries were washed once with lysis buffer with composition: HEPES (pH 7.5) 50mM, NaCl 150mM, $MgCl_2$ 3.2mM, NP40 0.5%, PMSF 1mM, Proteinase Inhibitor (Roche) 1X. The ovaries suspended in lysis buffer were homogenized, sonicated in bioruptor (5 min, 30sec on and 30 sec off), centrifuged at 13200 rpm for 30min at 4^0C. The supernatant was used as input and added to chromotek GFP-Trap®_A or RFP-Trap®_A beads suspended in lysis buffer. The lysate and beads were kept rotating at 4^0C for 3 hours and then washed 4 times with lysis buffer. The beads were resuspended SDS-PAGE lysis buffer. The procedure for mass spectrometry of IPed samples is descried in (Vanderweyde et al.,

2016). In brief, the IPed samples were resolved on a 10% SDS-PAGE gel. The gel pieces were processed for trypsin digestion to get the peptides, which were further analyzed by LC-MS/MS. For Rhi-Del co-IP westerns, the samples were separated on a SDS-PAGE gel, transferred onto nitrocellulose membrane, incubated with anti-GFP and anti-FLAG antibodies and imaged with LI-COR Odyssey system.

Small RNA-seq

Small RNA libraries were prepared as described in (Zhang et al., 2014). In brief, total RNA prepared from 2-4 day old female ovaries by mirVANA kit (Ambion) were size selected for 18-30nt small RNAs by gel purification. They were further 3' and 5' ligated by adapters, reverse transcribed, PCR amplified and sequenced by Illumina platform.

RNA-seq

RNA-seq libraries were prepared as described in (Zhang et al., 2012b). In brief, ribosomal rRNA depleted (Ribo-Zero kit (Illumina)) RNA samples were fragmented, reverse transcribed, ligated by adapters and PCR amplified to make libraries. dUTP incorporation done for strand specificity. Sequenced by Illumina platform.

ChIP-seq

ChIP-seq libraries were prepared as described in (Zhang et al., 2014). In brief, ovaries were fixed with 2% formaldehyde, sonicated for 2 hours in Bioruptor. The lysate after centrifugation was added to Dynabeads Protein G (Invitrogen) bound by anti-GFP antibody (Invitrogen #A11122). After overnight incubation, the beads were washed,

reverse crosslinked and DNA was purified for library preparation. The libraries for input and ChIP samples were prepared by adapter ligation and PCR amplification for sequencing by Illumina platform.

Image analysis for immuno-staining

Image processing was done by Adobe Photoshop and ImageJ. For fluorescence intensity quantification in ImageJ, the GFP-Del foci were defined after background subtraction, thresholding. Using these foci as reference, the fluorescence intensity was quantified in other channels.

Bioinformatics analysis

Reads from small RNA-seq libraries were aligned to the genome (dm3) by bowtie (Langmead et al., 2009), after removing the 3´end linkers. The transcriptome annotations were collected from Flybase r5.50. The piRNA cluster coordinates were taken from (Brennecke et al., 2007). Reads that were mapped to known non-coding RNAs (ncRNAs, such as rRNAs, tRNAs, etc.) and miRNAs were excluded for the quantification of piRNA abundance of clusters. The counts of reads were obtained using BEDTools (Quinlan and Hall, 2010) and normalized by the total number of reads aligned to the genome excluding known ncRNAs. A read is counted proportionally if it has multiple mapping locations. RNA-seq reads were aligned to the genome by TopHat (Trapnell et al., 2009), and rRNA reads were removed before the quantification of expression levels of genes, piRNA clusters, and transposons. ChIP-seq reads were

aligned by BWA (Li and Durbin, 2009), and duplicate reads were marked and removed by Picard tools.

Analysis of Proteome obtained by Mass spectrometry

The raw data was processed through Proteome Discoverer and Mascot Server before display on Scaffold Viewer (Proteome Software, Inc.). iBAQ values (Schwanhausser et al., 2011) of each IPed protein were normalized to corresponding GFP iBAQ values after adding pseudocount. The ratios of these normalized values were ranked and plotted with R.

Statistical analysis

The error bars in the bar graphs represent standard deviation for 3 biological replicates.

Data and software availability

High throughput sequencing files are available from NCBI short read archive (SRA) SRP111075. The scaffold files for mass spectrometry analysis and cloned gene sequences are available at Mendeley Data (http://dx.doi.org/10.17632/w2ym383bp8.1).

Acknowledgements

I would like to thank the members of Theurkauf and Weng labs and UMassmed RNA Biology community for their insightful discussions and critical comments throughout the project; Shikui Tu for help in the Bioinformatics analysis under guidance of Zhiping Weng, Trudi Schüpbach for *del* stocks; Bloomington and UCSD *Drosophila* stock centers for various fly stocks, John Leszyk of UMassmed Proteomics facility for mass spectrometry. This work was supported by National Institute of Child Health and Human Development (R01HD049116 and P01HD078253).

Figure 2.S1

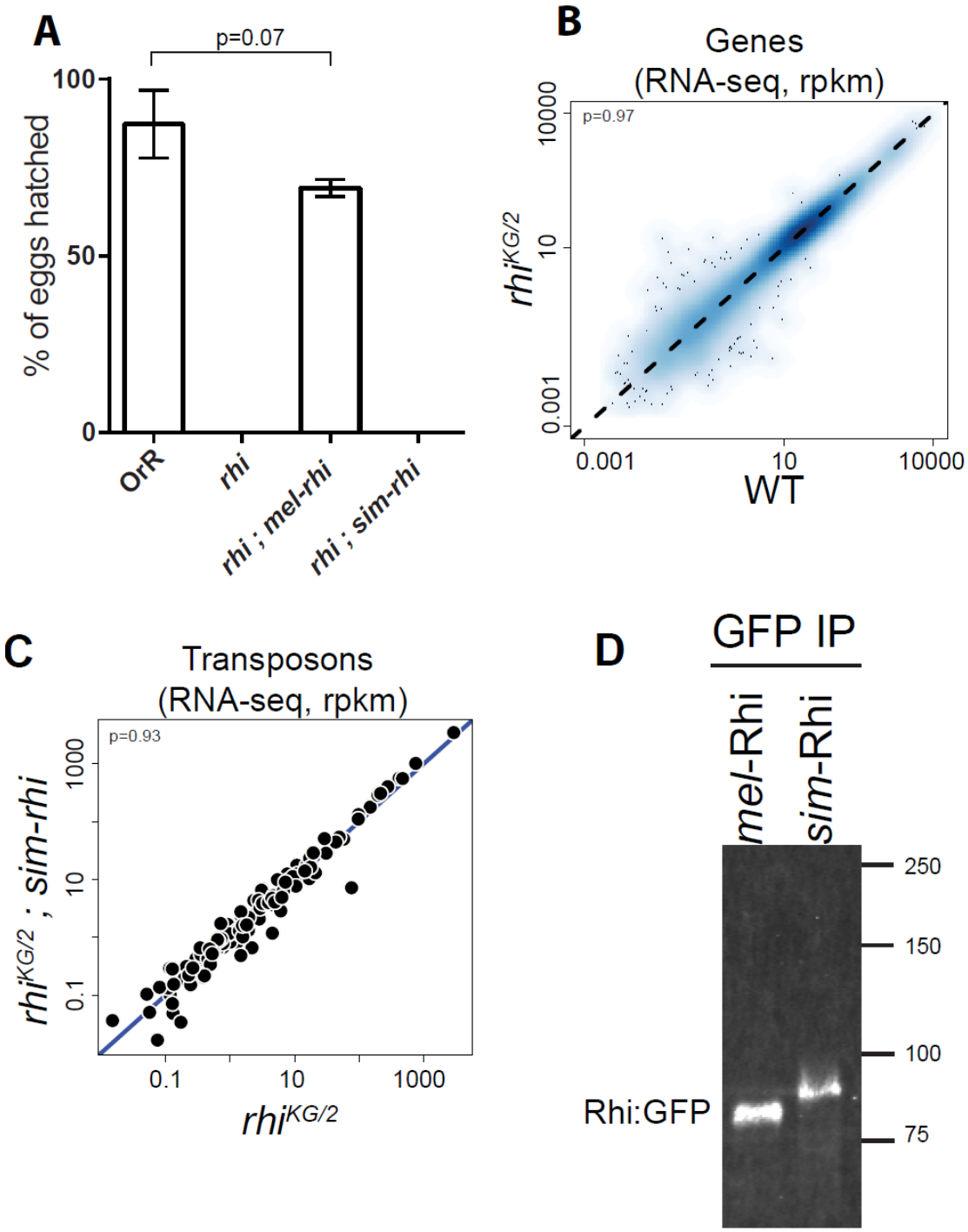

A
p=0.07
100
50
0
% of eggs hatched
OrR
rhi
rhi ; mel-rhi
rhi ; sim-rhi
B
Genes
(RNA-seq, rpkm)
p=0.97
10000
10
0.001
rhiKG/2
0.001
10
10000
WT
C
Transposons
(RNA-seq, rpkm)
p=0.93
1000
10
0.1
rhiKG/2 ; sim-rhi
0.1
10
1000
rhiKG/2
D
GFP IP
mel-Rhi
sim-Rhi
250
150
100
75
Rhi:GFP

Figure 2.S1. *sim-rhi* behaves like a null *rhi* allele in *melanogaster*

(A) Bar graphs showing percentages of hatched eggs from females of genotypes: OrR (WT control), *rhi* mutant, and *rhi* mutants expressing either *mel-rhi* or *sim-rhi*. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *rhi* mutants and *rhi* mutants rescued by *sim-rhi* where average of at least 30 eggs were scored.

(B) Scatterplots showing gene expression levels measured by RNA-seq in ovaries of *rhi* mutant vs. WT control. Each point represents rpkm value for a different gene. Diagonal represents x=y. p value for differences is obtained by Wilcoxon test.

(C) Scatterplots showing transposon expression levels measured by RNA-seq in ovaries of *rhi* mutant vs. *rhi* mutant expressing *sim-rhi*. Diagonal represents x=y. p value for differences is obtained by Wilcoxon test.

(D) Western blot showing *mel*-Rhi and *sim*-Rhi IPed from *melanogaster* ovaries by the GFP tag.

Figure 2.S2

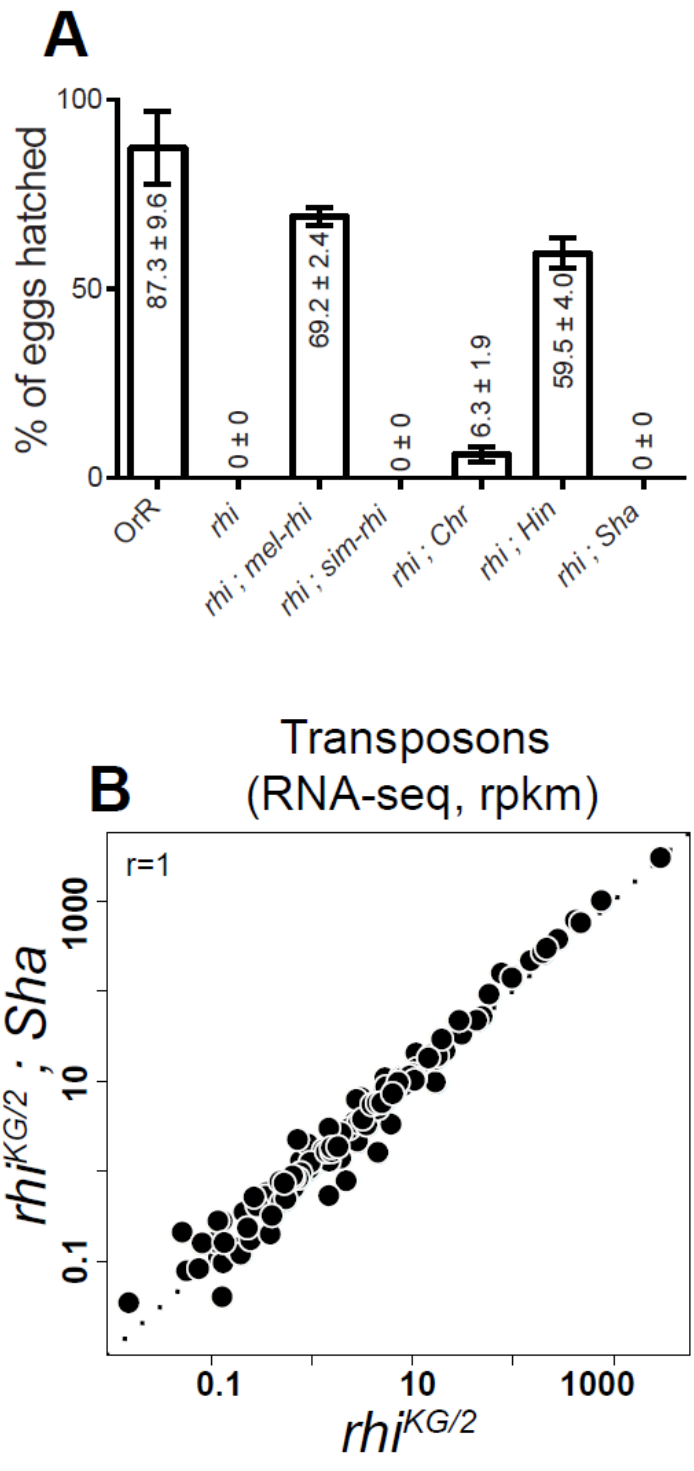

A
% of eggs hatched
100
50
0
87.3 ± 9.6
0 ± 0
69.2 ± 2.4
0 ± 0
6.3 ± 1.9
59.5 ± 4.0
0 ± 0
OrR
rhi
rhi ; mel-rhi
rhi ; sim-rhi
rhi ; Chr
rhi ; Hin
rhi ; Sha
B
Transposons
(RNA-seq, rpkm)
r=1
rhiKG/2 ; Sha
1000
10
0.1
0.1
10
1000
rhiKG/2

Figure 2.S2. Shadow chimera transgene behaves like a null *rhi* allele in *melanogaster*

(A) Bar graphs showing percentages of hatched eggs from females of genotypes: OrR (WT control), *rhi* mutant, and *rhi* mutants expressing either *mel-rhi* or *sim-rhi* or different chimeras. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *rhi* mutants and *rhi* mutants rescued by *sim-rhi* or Shadow chimera where average of at least 30 eggs were scored.

(B) Scatterplots showing transposon expression levels measured by RNA-seq in ovaries of *rhi* mutant vs. *rhi* mutant expressing shadow chimera. Diagonal represents x=y. p value for differences is obtained by Wilcoxon test.

Figure 2.S3

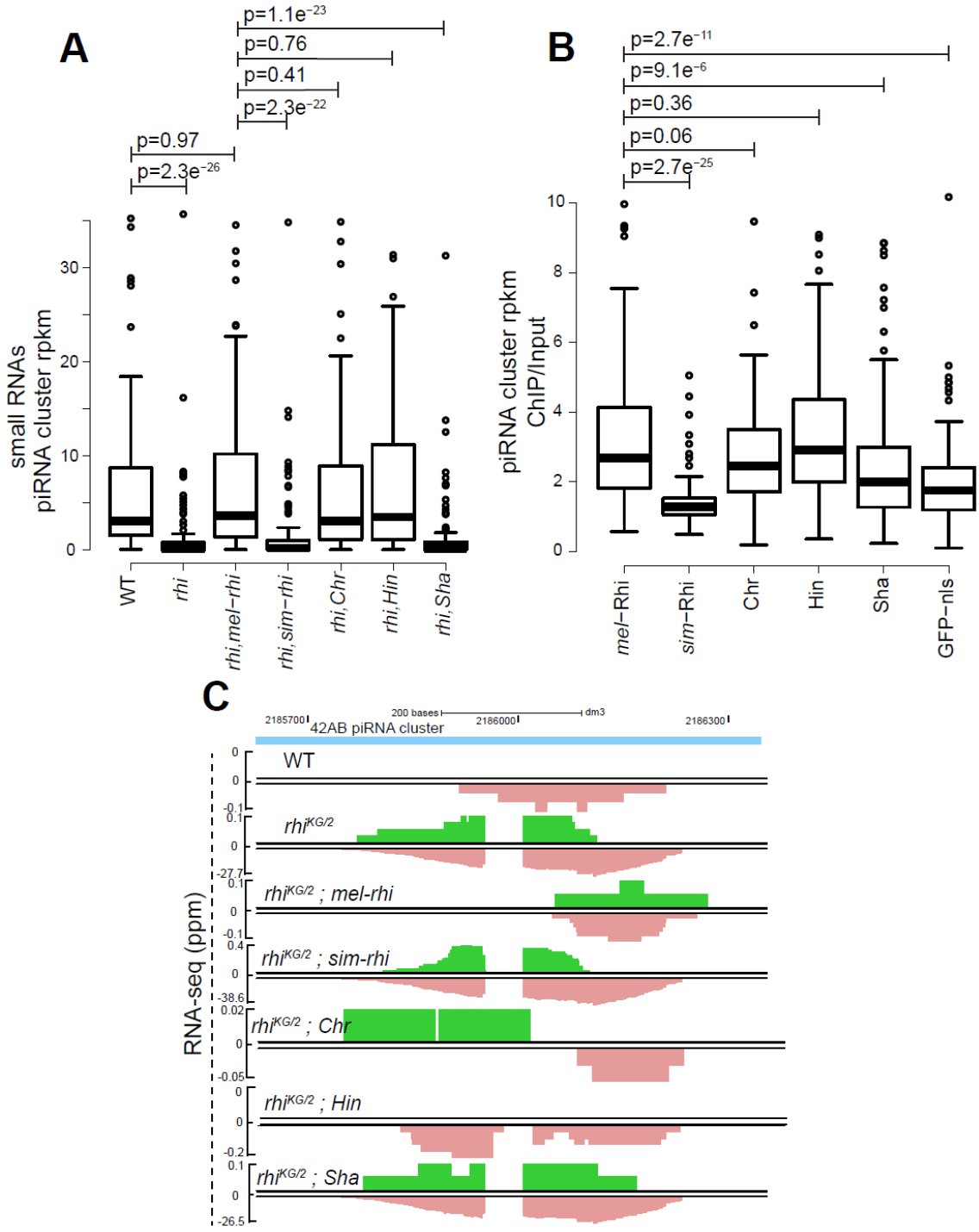

A
p=1.1e−23
p=0.76
p=0.41
p=2.3e−22
p=0.97
p=2.3e−26
small RNAs
piRNA cluster rpkm
WT
rhi
rhi,mel−rhi
rhi,sim−rhi
rhi,Chr
rhi,Hin
rhi,Sha
B
p=2.7e−11
p=9.1e−6
p=0.36
p=0.06
p=2.7e−25
piRNA cluster rpkm
ChIP/Input
mel−Rhi
sim−Rhi
Chr
Hin
Sha
GFP−nls
C
200 bases
dm3
2185700
2186000
2186300
42AB piRNA cluster
RNA-seq (ppm)
WT
rhiKG/2
rhiKG/2 ; mel-rhi
rhiKG/2 ; sim-rhi
rhiKG/2 ; Chr
rhiKG/2 ; Hin
rhiKG/2 ; Sha

Figure 2.S3. piRNA, ChIP-seq and RNA-seq splicing profiles for different Rhi variants

(A) Boxplot showing piRNA cluster rpkm values for small RNAs, in genotypes: WT, *rhi* mutant, and *rhi* mutants expressing either *mel-rhi* or *sim-rhi* or different chimeras. p value for differences is obtained by Wilcoxon test.

(B) Boxplot showing piRNA cluster rpkm ratios for ChIP signal to input signal, for *mel*-Rhi, *sim*-Rhi, chimeras or GFP-nls control. p value for differences is obtained by Wilcoxon test.

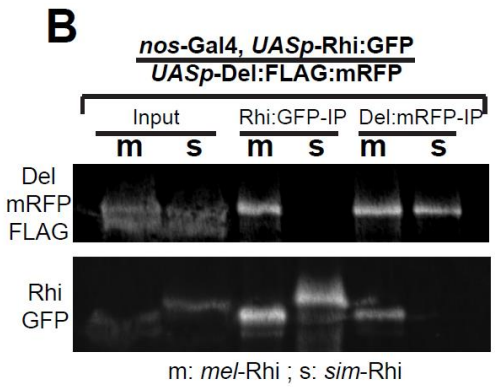
(C) Genome browser view of RNA-seq profiles at 42AB cluster in WT, *rhi* mutant and *rhi* mutants rescued by *mel*-Rhi, *sim*-Rhi or different chimeras. Green: Watson strand, pink: Crick strand. The scales are adjusted to prevent peak clipping.

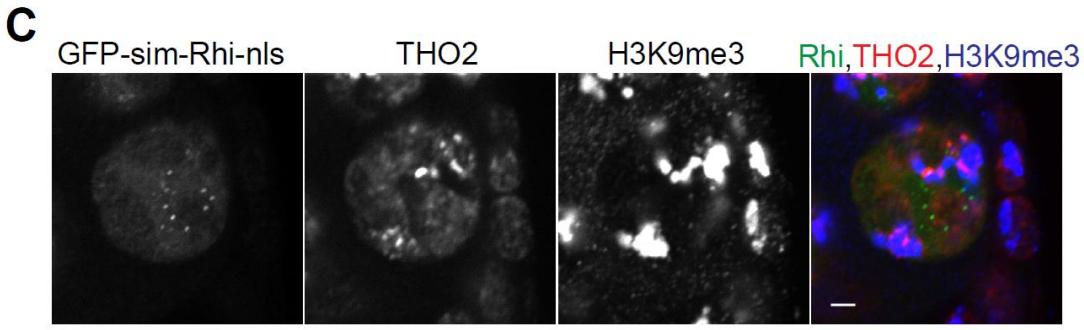
Figure 2.S4

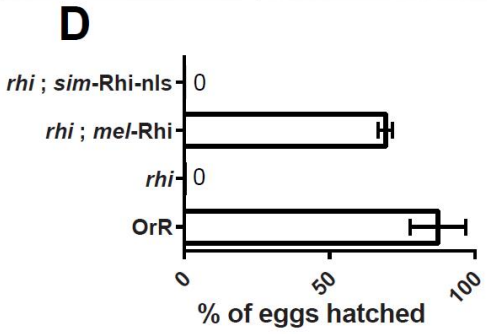
A

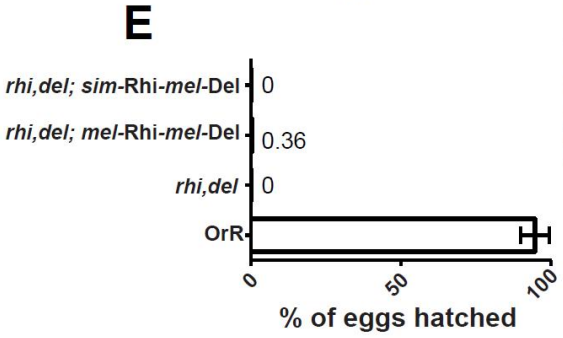
nosG > GFP-Rhi

Bait protein	Spectrum counts		
	GFP	Rhi	Deadlock
mel-Rhi	60	79	24
sim-Rhi	75	73	0
GFP ctrl	187	0	0
Chromo	74	87	23
Hinge	83	89	26
Shadow	48	67	0









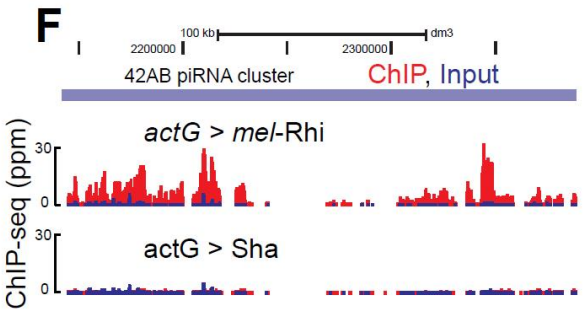


Figure 2.S4. *sim*-Rhi does not bind to *mel*-Del

(A) Total spectrum counts corresponding to GFP, Rhino and Deadlock that co-precipitated with the indicated tagged proteins and GFP control. The *sim*-Rhi and shadow chimera fail to bind to *melanogaster* Del.

(B) Western blot showing that *sim*-Rhi does not bind to *mel*-Del, observed by reciprocal IP of Rhi:GFP and Del:FLAG:mRFP. m: *mel*-Rhi:GFP, s: *sim*-Rhi:GFP. Probing was done by anti-GFP (for Rhi:GFP) and anti-FLAG (for Del:FLAG:mRFP) antibodies.

(C) Localization of THO2 (piRNA cluster marker), H3K9me3 marked chromatin in the germline nuclei expressing *Act5C*-Gal4 driven GFP tagged *sim*-Rhi-nls. Color assignments for merged image shown on top. Scale bar: 2μm.

(D) Bar graphs showing percentages of hatched eggs from females of genotypes: OrR (WT control), *rhi* mutant, and *rhi* mutants expressing either *mel*-Rhi or *sim*-Rhi-nls. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *rhi* mutants and *rhi* mutants rescued by *sim*-Rhi-nls where average of at least 30 eggs were scored.

(E) Bar graphs showing percentages of hatched eggs from females of genotypes: OrR (WT control), *rhi,del* double mutant, and *rhi,del* double mutants expressing Rhi-Del fusions. The numbers in/above the bars show mean ± standard deviation of three or more biological replicates, with a minimum of 40 embryos scored per replicate.

(F) Genome browser view of ChIP-seq profiles at 42AB cluster for *act5C*-Gal4 driven GFP tagged *mel*-Rhi and shadow chimera. Both ChIP done under identical conditions, using the same anti-GFP antibody. ChIP signal in red, input signal in blue.

Figure 2.S5

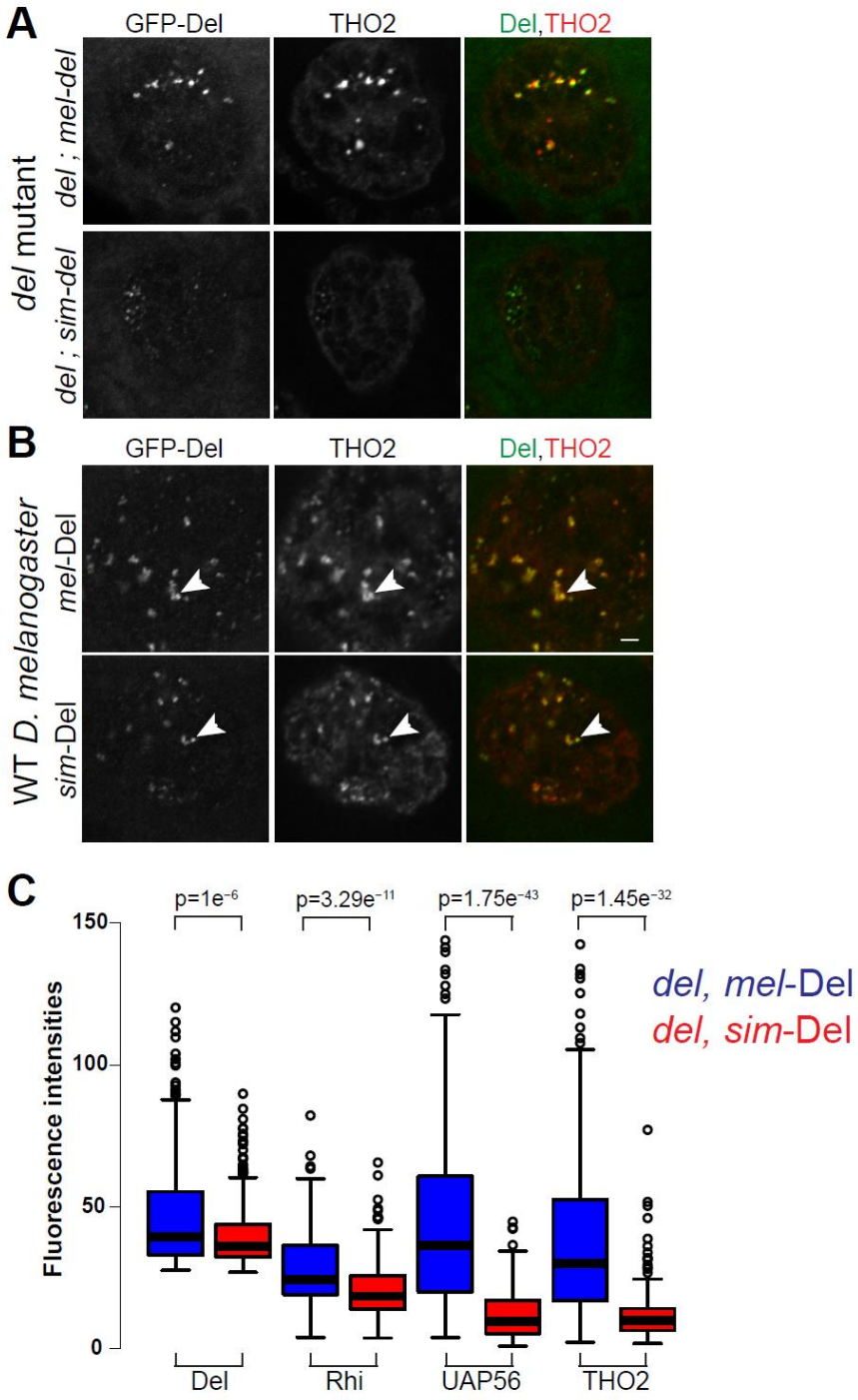

A
GFP-Del
THO2
Del,THO2
del mutant
del ; mel-del
del ; sim-del
B
GFP-Del
THO2
Del,THO2
WT D. melanogaster
mel-Del
sim-Del
C
p=1e⁻⁶
p=3.29e⁻¹¹
p=1.75e⁻⁴³
p=1.45e⁻³²
Fluorescence intensities
150
100
50
0
del, mel-Del
del, sim-Del
Del
Rhi
UAP56
THO2

Figure 2.S5. *sim*-Del fails to recruit components of piRNA pathway

(A) Localization of THO2 in *del* mutants expressing either *mel*-Del or *sim*-Del.

(B) Localization of GFP tagged *mel*-Del and *sim*-Del in WT *melanogaster* female germline nuclei. Egg chambers were double labeled for THO2 (piRNA cluster marker). Both forms of Del co-localize to nuclear foci with THO2. Scale bar: 2μm.

(C) Boxplot showing fluorescence intensities of Del:GFP, Rhi, UAP56 and THO2 foci in del mutants expressing *mel*-Del (blue) or *sim*-Del (red).

Figure 2.S.6

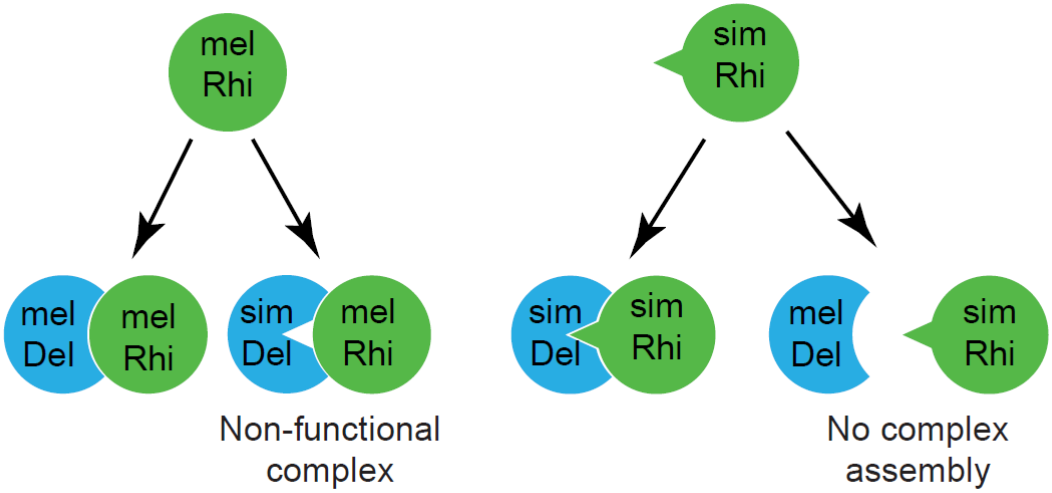

mel
Rhi
mel
Del
mel
Rhi
sim
Del
mel
Rhi
Non-functional
complex
sim
Rhi
sim
Del
sim
Rhi
mel
Del
sim
Rhi
No complex
assembly

Figure 2.S6. Model for directional binding incompatibility in Rhi-Del

mel-Rhi can bind to Del from both the species, but *sim*-Rhi can bind to *sim*-Del, but is incompatible with *mel*-Del. Significantly, our data indicate that the *mel*-Rhi/*sim*-Del complex is not functional.

Chapter III

Species swap of Cutoff reveals dynamic complex assembly in the piRNA pathway

Preface

Tianxiong Yu (Bear) and Shikui Tu helped me with the bio-informatics analysis under the guidance of Zhiping Weng.

Summary

PIWI interacting RNAs (piRNAs) protect the genome from transposons. Many piRNA pathway genes are rapidly evolving, suggesting that the piRNA pathway may be involved in a host-pathogen arms race with transposons. To test whether adaptive evolution leads to any functional divergence, we swapped a piRNA pathway gene, *cutoff*, between sibling species *Drosophila simulans* and *Drosophila melanogaster*. Cutoff, along with Rhino and Deadlock forms RDC complex, which specifies the piRNA producing loci in *Drosophila* and also the piRNA precursor transcripts. We found that Cutoff from *D. simulans* (*sim*-Cuff) does not function in *D. melanogaster*. Unlike *mel*-Cuff, it stably binds to Deadlock and TRF2. Consistent with this stable interaction, overexpression of *sim*-Cuff leads to a dominant negative effect on fertility. We propose that Cutoff normally forms a transient interaction with Deadlock and TRF2 to promote transcription from piRNA clusters. Dominant negative *sim*-Cuff traps the RDC complex, which may have a role in hybrid sterility.

Introduction

Transposons or transposable elements (TEs) are major components of eukaryotic genomes (Bao et al., 2015; Canapa et al., 2015). For humans, almost half of the genome encodes for transposons. These transposons can move and insert into genes leading to mutations. Recombination between different transposon copies can cause insertions and deletions. Thus transposons can be a major source of genomic instability (Ayarpadikannan and Kim, 2014; Hedges and Deininger, 2007). TEs can rapidly spread in populations once introduced into a new species (Gilbert and Feschotte, 2018). Over a span of few decades, P-element transposon has spread through wild populations of *D. melanogaster* and is currently in the process of sweeping through wild *D. simulans* populations (Daniels et al., 1990; Engels, 1992; Kofler et al., 2015). The capacity to damage the host and spread between individuals makes transposons successful pathogens. Animals produce a class of small RNAs called PIWI interacting RNAs (piRNAs) to regulate these transposons, especially in the germline which is responsible for inheritance of genetic information (Ghildiyal and Zamore, 2009; Weick and Miska, 2014). These piRNAs can silence transposons both transcriptionally and post-transcriptionally. Thus host piRNA pathway forms an immune defense against pathogenic transposons.

The organization of piRNA machinery makes it possible to adapt to ever-changing threat of transposons (Huang et al., 2017). In *Drosophila*, piRNAs are produced from piRNA clusters, which contain truncated transposon copies. Thus piRNAs produced from clusters can silence transposons in sequence specific manner. Any new transposon

jumped into species can be effectively silenced once it jumps into a cluster (Khurana et al., 2011). piRNA production is amplified by Ping-Pong cycle which consists of reciprocal cleavage of transposons and corresponding piRNA precursor transcripts (Brennecke et al., 2007; Gunawardane et al., 2007). It leads to more piRNA production against active transposons. As piRNA machinery adaptively inhibits transposons, they are likely to be involved in a host-pathogen arms race. Other pathogens such as viruses and the immune system adapt continuously to stay ahead in the Red Queen host pathogen arms race (Daugherty and Malik, 2012). As many piRNA pathway proteins show hallmarks of adaptive evolution, it is proposed that the piRNA pathway is rapidly evolving in response to arms race with transposons (Blumenstiel et al., 2016; Lee and Langley, 2012; Obbard et al., 2009; Simkin et al., 2013; Vermaak et al., 2005). Transposons can be a major source of genome variation in recently diverged species. Many closely related *Drosophila* species have different transposon profiles and transposons unique to those species (Bartolome et al., 2009; Lerat et al., 2011). This transposon environment variation between the species would necessitate functional variation in the piRNA pathway, leading to rapid evolution and species divergence in the piRNA pathway. Crosses between recently diverged species *Drosophila* melanogaster and *Drosophila* simulans are sterile (Sturtevant, 1920) and lead to defects in piRNA biogenesis and transposon silencing (Kelleher et al., 2012). Thus, sequence divergence in piRNA pathway and differences in transposons can have a role in hybrid incompatibility. To test the effect of this sequence divergence and possible implications in hybrid

incompatibility, we swapped a piRNA pathway gene between species and checked its function in a different transposon environment.

In *Drosophila*, piRNA clusters are specified by RDC complex composed of HP1 homolog Rhino, Deadlock and Rai homolog Cutoff (Chen et al., 2016; Le Thomas et al., 2014; Mohn et al., 2014; Parhad et al., 2017; Yu et al., 2015a; Zhang et al., 2014). Rhino binds to histone modification H3K9me3 at the clusters. It directly binds to Deadlock and which further recruits Moonshiner and TATA box related protein 2 (TRF2) and promotes transcription from piRNA clusters (Andersen et al., 2017). Cutoff is proposed to suppress termination at the truncated transposon termination sites within the clusters (Chen et al., 2016). This complex also promotes piRNA cluster transcript processing by suppressing splicing of the cluster transcripts (Zhang et al., 2014). The whole RDC complex is rapidly evolving and Rhino and Deadlock fail to function in sibling species, indicating that they are actively involved in host pathogen arms race with TEs (Blumenstiel et al., 2016; Parhad et al., 2017). To understand whether the sequence variation in RDC complex leads to functional diversity, we swapped *cutoff* gene between the two *Drosophila* species. We found that Cuff from *D. simulans* (*sim*-Cuff) does not function in *D. melanogaster*. It stably binds to Deadlock and TRF2, allowing us to identify a transient complex assembly at piRNA clusters. As stable interaction of *sim*-Cuff with *mel*-Deadlock leads to a dominant effect on fertility, we propose that this stable interaction may have a role in hybrid incompatibility manifested in the form of sterility.

Results

D. simulans cutoff does not function in *D. melanogaster*

In *Drosophila*, RDC complex is at the heart of piRNA biogenesis machinery as it specifies clusters and controls piRNA cluster transcription and processing (Andersen et al., 2017; Mohn et al., 2014; Zhang et al., 2014). All Rhi, Del and Cutoff show hallmarks of adaptive evolution. We previously observed that rapid evolution has led to functional divergence in Rhi and Del (Parhad et al., 2017). We wondered whether the whole complex has functionally evolved. To test that, we swapped Cutoff between sibling species *D. melanogaster* and *D. simulans* and studied its function.

In a *D. melanogaster cuff* mutant background, we expressed GFP tagged *sim*-Cuff. Similar expression of *mel*-Cuff served as a control. Both these Cuff variants were expressed in the germline by using *rhi* promoter, in transgenes obtained by PhiC31 mediated transformation at the same chromosomal location (Figure 3.1A). Mutations in piRNA pathway genes, including *cuff* leads to D-V patterning defects and female sterility (Chen et al., 2007b; Klattenhoff et al., 2007). *mel-cuff* can rescue both D-V patterning and hatching, but *sim-cuff* fails to rescue both (Figures 3.1B and 3.S1). Thus *sim-cuff* behaves like a null *cuff* allele in *D. melanogaster*. The sterility in *cuff* mutants is due to transposon over-expression (Chen et al., 2007b). To test transposon expression, we sequenced strand specific RNA-seq libraries from ovaries of these flies. Figure 3.1C shows transposon expression in *cuff* mutant ovaries vs. WT ovaries. Many transposons are over-expressed as shown by points above the diagonal. *mel-cuff* can rescue

transposon silencing, but *sim-cuff* does not (Figures 3.1D and 3.1E). The transposon expression in *cuff* mutants expressing *sim-cuff* is highly correlated to *cuff* mutants (Figure 3.1F). Thus, *sim-cuff* fails to silence transposons in *D. melanogaster* and behaves like a null allele.

Figure 3.1

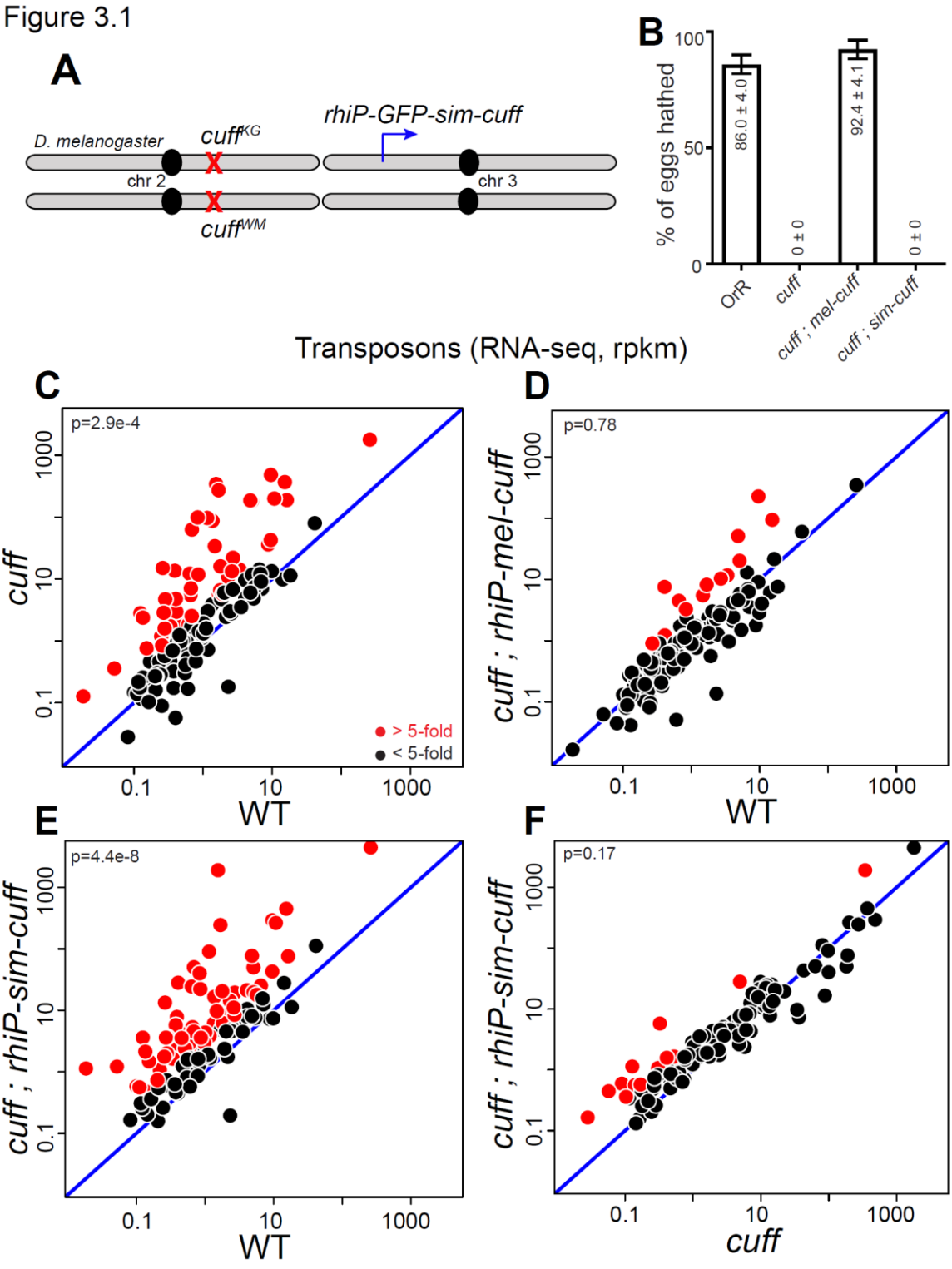

A
D. melanogaster
cuff^KG
rhiP-GFP-sim-cuff
chr 2
chr 3
cuff^WM
B
% of eggs hatched
100
50
0
86.0 ± 4.0
0 ± 0
92.4 ± 4.1
0 ± 0
OrR
cuff
cuff ; mel-cuff
cuff ; sim-cuff
Transposons (RNA-seq, rpkm)
C
p=2.9e-4
cuff
WT
> 5-fold
< 5-fold
D
p=0.78
cuff ; rhiP-mel-cuff
WT
E
p=4.4e-8
cuff ; rhiP-sim-cuff
WT
F
p=0.17
cuff ; rhiP-sim-cuff
cuff
0.1
10
1000

Figure 3.1: *sim*-cuff does not function in *D. melanogaster*

(A) Genetic complementation strategy. The *sim-cuff* gene was expressed under germline specific *rhi* promoter in a *D. melanogaster* $cuff^{KG/WM}$ mutant background.

(B) Bar graphs showing percentages of hatched eggs produced by OrR (wild type (WT) control), *cuff* mutants, and *cuff* mutants rescued by either *mel-cuff* or *sim-cuff*. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *cuff* mutants and *cuff* mutants rescued by *sim-cuff* where average of 230 and 23 eggs were scored respectively.

(C-F) Transposon expression by RNA-seq in *cuff* mutants (C), *cuff* mutant expressing either *mel-cuff* (D) or *sim-cuff* (E, F) in ovaries. Each point on the scatterplots shows rpkm values for a transposons family in ovaries of the indicated mutant/transgene combination relative to WT control or *cuff* mutant. Diagonal represents x=y. Points in red show y/x>5. p value for differences obtained by Wilcoxon test.

Functional RDC complex requires dynamic interaction

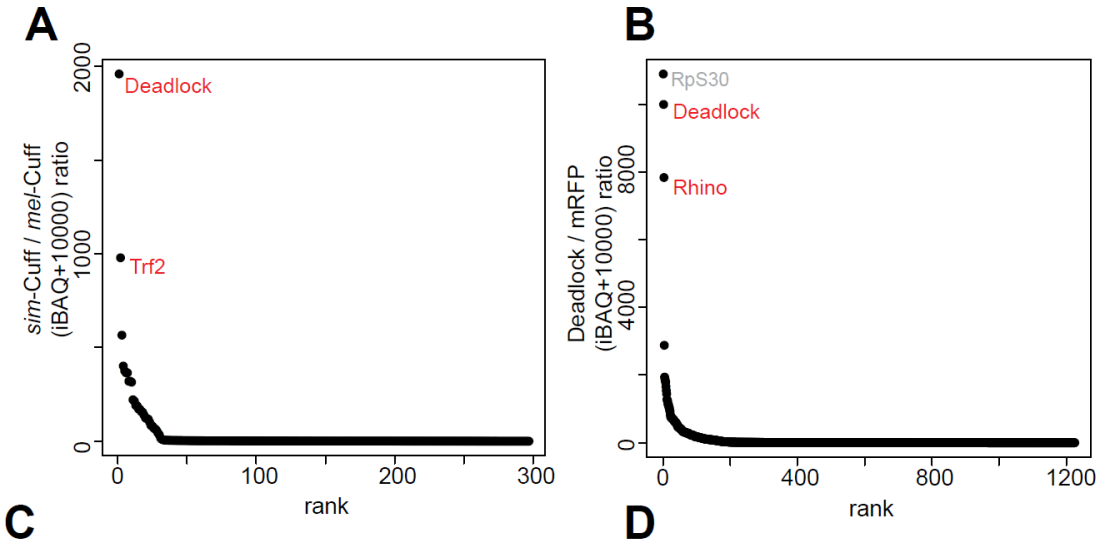
In order to understand why *sim*-Cuff fails to function in *D. melanogaster*, we checked the Cuff interacting proteins. A few publications have reported Cuff interactors. Pane et al showed that GFP tagged Cuff binds to Rhi by immuno-precipitation (IP) and Western blotting (Pane et al., 2011). Mohn et al studied the interaction between all the components of RDC complex by yeast two hybrid assay (Mohn et al., 2014). They observed that Cuff does not interact with Rhi, instead it interacts with Del. Hur et al showed that TREX complex component Thoc5 is required for piRNA biogenesis in *Drosophila* and it co-IPs with Cuff, but not Rhi when expressed in S2 cells (Hur et al., 2016). To dissect Cuff function in light of these contradictory results, we immuno-precipitated both GFP tagged *sim*-Cuff and *mel*-Cuff from fly ovaries, where it is normally expressed and identified interacting proteins by mass spectrometry.

We observed that *mel*-Cuff does not bind to Del and Rhi (Figure 3.2C). IP and mass spec of mRFP tagged Del also does not show any binding to Cuff (Figure 3.2D) and Rhi IP also shows no binding between Rhi and Cuff (Figure 3.S2A). This led us to think that either Cuff does not bind to Del or Rhi, or the interaction of Cuff with either Del or Rhi is transient which cannot be detected by immuno-precipitation. Surprisingly, we observed that *sim*-Cuff binds to Del. Del was the most differentially bound protein in *sim*-Cuff IP vs. *mel*-Cuff IP (Figure 3.2A). However, Del binds to Rhi (Figures 3.2B and 3.2D). Even though *sim*-Cuff binds Del, we did not observe any Rhi peptides in *sim*-Cuff IP. This led us to hypothesize that Cuff only transiently binds to Del. *sim*-Cuff stabilizes this Cuff-Del interaction. We think that Del is either bound to Rhi or Cuff. It does not

bind to Rhi and Cuff at the same time. This leads to two predictions: 1) as *sim*-Cuff forms a stable complex with Del, it can have a dominant negative effect in the presence of WT copy of *mel-cuff* and 2) the RDC complex will be non-functional if we stabilize the interaction between Rhi-Del or Del-Cuff. To test the first prediction, we over-expressed *sim*-Cuff in the germline with *UASp* promoter and *nanos*-Gal4 driver. Note that these flies contain WT copy of *mel-cuff* gene. Consistent with our prediction, over-expressed *sim*-Cuff leads to a dominant effect on fertility of flies (Figure 3.2E and S2B). Similar over-expression of *mel*-Cuff has no effect on fertility. Thus, *sim*-Cuff seems to be trapping Del in a non-functional complex. We have previously tested the first part of second prediction. We fused Rhi and Del as a single polypeptide chain and expressed the fusion protein in a *rhi, del* double mutant background (Parhad et al., 2017). The fusion protein fails to rescue the fertility defect of the double mutant, indicating that Rhi-Del fusion does not work. The fusion could affect the function, however the fusion protein localizes to clusters and tagging Del by N terminal GFP does not affect its function, which is the site where Rhi is fused. To test the second part, we promoted stable interaction between Cuff and Del by co-expressing GFP tagged *mel*-Cuff and Del fused to GFP nanobody that binds to GFP (Figure 3.2F). *cuff* mutants are sterile and GFP tagged *mel*-Cuff can rescue the fertility defect of *cuff* mutants. However, when both GFP tagged *mel*-Cuff and GFP nanobody tagged Del are co-expressed in a *cuff* mutant background, *mel*-Cuff fails to rescue the fertility defect of *cuff* mutants. The only Cuff protein in these flies is expressed as GFP tagged transgene. However, when Del stably binds to the only source of Cuff protein, it does not work. Thus we propose that dynamic

interactions within RDC complex are necessary for its function. When *sim*-Cuff stabilizes this Cuff-Del interaction, the complex does not work.

Figure 3.2



C

Identified protein	Spectrum counts		
	mel-Cuff	*sim*-Cuff	GFP ctrl
GFP	13	19	56
Cuff	14	23	0
Del	0	12	0
Rhi	0	0	0
Trf2	0	8	0

D

Identified protein	Spectrum counts	
	Del	mRFP
mRFP	29	36
Del	83	1
Rhi	16	0
Cuff	0	0

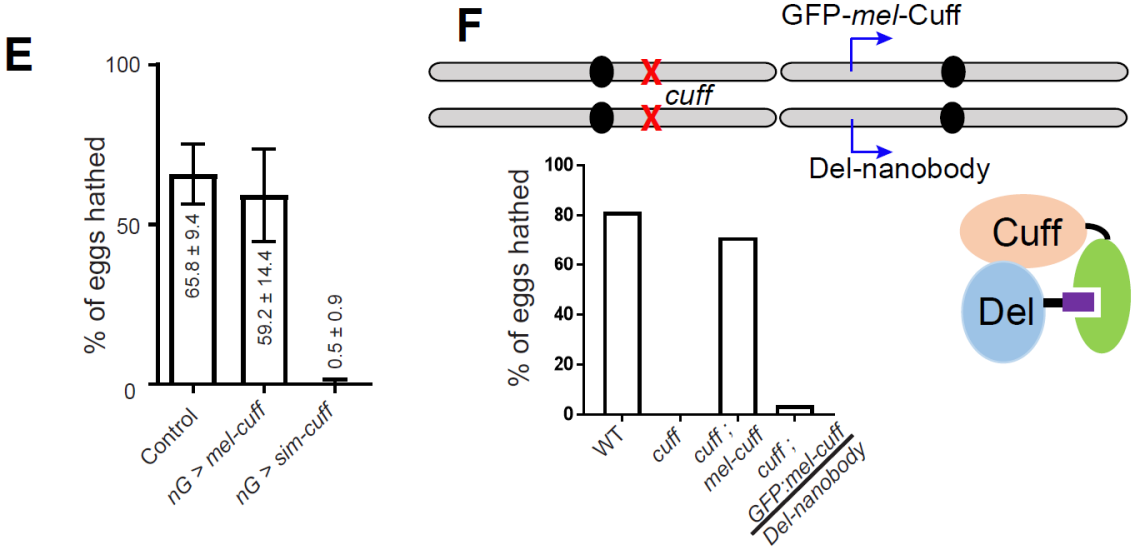


Figure 3.2: *sim*-Cuff stably interacts with Del, TRF2 and leads to a dominant effect on fertility

(A-D) Mass spectrometric analysis of Cuff and Del binding proteins. Graphs showing ratios of GFP normalized iBAQ values for *mel*-Cuff vs. *sim*-Cuff (A), and normalized iBAQ values for *mel*-Del vs. mRFP (B), ranked by ratio values. Spectrum counts in Cuff IP-mass spec (C) and Del IP-mass spec (D) with respective controls.

(E) *sim*-Cuff has a dominant effect on fertility. Bar graphs showing percentages of hatched eggs produced by control, flies over-expressing either *mel-cuff* or *sim-cuff* by germline specific *nanos*-Gal4 driven. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *nanos*-Gal4 driven *sim-cuff* where average of 50 eggs were scored.

(F) Strong binding *mel*-Cuff and *mel*-Del leads to a fertility defect. Schematic for *cuff* mutant co-expressing GFP tagged *mel*-Cuff and GFP binding nanobody tagged *mel*-Del. Bar graphs shows percentages of hatched eggs in WT control, *cuff* mutants, *cuff* mutants expressing *mel*-Cuff, *cuff* mutants expressing both GFP:*mel*-Cuff and GFP-nanobody: *mel*-Del.

sim-Cuff re-organizes the RDC complex in the nucleus

RDC complex is shown to localize to piRNA clusters in the germline (Mohn et al., 2014). The piRNA clusters are marked by histone modification H3K9me3, just like constitutive heterochromatin (Le Thomas et al., 2014; Mohn et al., 2014; Zhang et al., 2014). Rhi localizes to the periphery of the H3K9me3 marked domains (Parhad et al., 2017). As all the components of RDC complex co-localize, Cuff also shows such peripheral localization as seen for GFP tagged *mel*-Cuff (Figure 3.3A). However, *sim*-Cuff when expressed in a *D. melanogaster cuff* mutant background, shows diffused signal within the H3K9me3 domains. Such loss of peripheral localization and spread to non-peripheral histone marks are also observed when *sim*-Cuff is over-expressed (Figure 3.S3A). The localization of all the RDC complex components is inter-dependent. Mutation in one gene leads to disruption of the whole complex (Mohn et al., 2014). To test whether the defect in Cuff localization leads to a defect for the entire RDC complex, we stained the ovaries for Rhi and Del. *mel*-Cuff co-localizes with Rhi and Del and forms distinct foci in the germline (Figure 3.3B). Diffusely localized *sim*-Cuff retains co-localization with Rhi and Del. Thus the entire RDC complex shows diffused signal over H3K9me3 domains. The same mislocalization is also observed for over-expressed *sim*-Cuff (Figure 3.S3B).

Figure 3.3

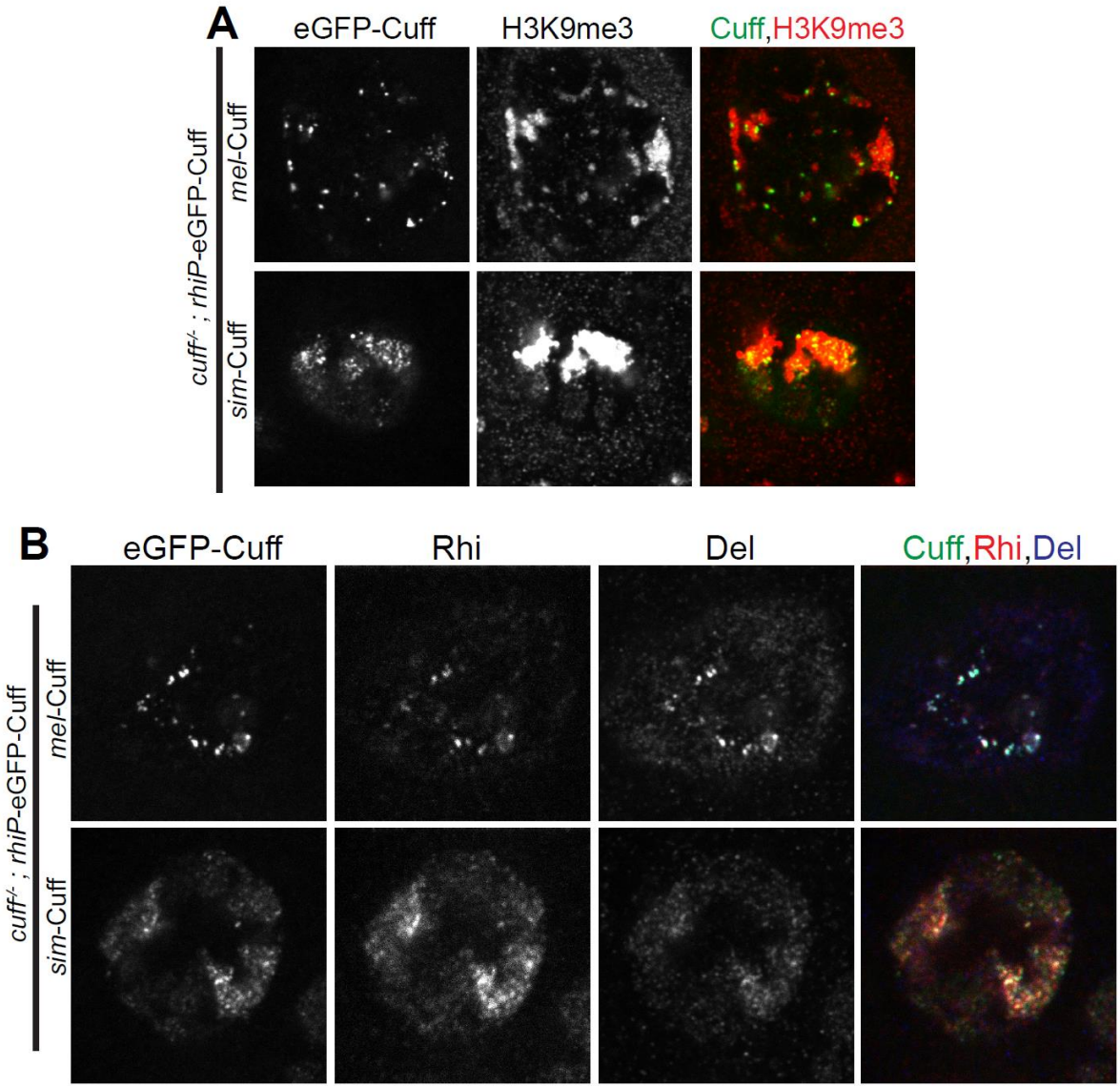

A
eGFP-Cuff
H3K9me3
Cuff,H3K9me3
cuff- ; rhiP-eGFP-Cuff
mel-Cuff
sim-Cuff
B
eGFP-Cuff
Rhi
Del
Cuff,Rhi,Del
cuff- ; rhiP-eGFP-Cuff
mel-Cuff
sim-Cuff

Figure 3.3: *sim*-Cuff containing RDC complex loses specificity for piRNA clusters

(A) Localization of GFP tagged Cuff with respect to H3K9me3 marked chromatin in the germline nuclei of *cuff* mutants expressing *rhi* promoter driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged image shown on top.

(B) Localization of GFP tagged Cuff with respect to Rhi and Del in the germline nuclei of *cuff* mutants expressing *rhi* promoter driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged image shown on top.

sim-Cuff changes transcriptional profile in _D. melanogaster_ by also trapping TRF2

Andersen et al showed that Deadlock forms a complex with TRF2 through Moonshiner and licenses transcription from piRNA clusters (Andersen et al., 2017). TRF2 is a transcription factor which binds to TCT core promoters and regulates transcription of ribosomal genes (Kedmi et al., 2014; Wang et al., 2014). We observed that *sim*-Cuff stably binds to Del. This trapped complex also contains TRF2 (Figure 3.2C). In the comparison of protein iBAQ values for *sim*-Cuff IP vs. *mel*-Cuff IP, TRF2 is second highly enriched protein after Del (Figure 3.2A). As *sim*-Cuff traps Del and leads to its mislocalization, we wondered whether the same is true for trapped TRF2. To test this, we checked the localization of TRF2 with respect to GFP-Cuff in the germline. In WT and *cuff* mutants rescued by *mel*-Cuff, TRF2 forms a few distinct foci in the germline (Figure 3.4A). Even though TRF2 is required for transcription from piRNA clusters, it does not enrich at these clusters and the observed TRF2 foci do not co-localize with *mel*-Cuff. This further supports the idea of transient assembly of transcriptional complex at piRNA clusters. *cuff* mutants expressing *sim*-Cuff show reduced intensity of these TRF2 foci. However, this delocalized TRF2 does spread across *rhi* promoter driven *sim*-Cuff. When *sim*-Cuff is over-expressed, TRF2 also shows diffused signal which co-localizes with *sim*-Cuff (Figure 3.4B). Thus *sim*-Cuff leads to mislocalization of TRF2 along with the RDC complex.

Figure 3.4

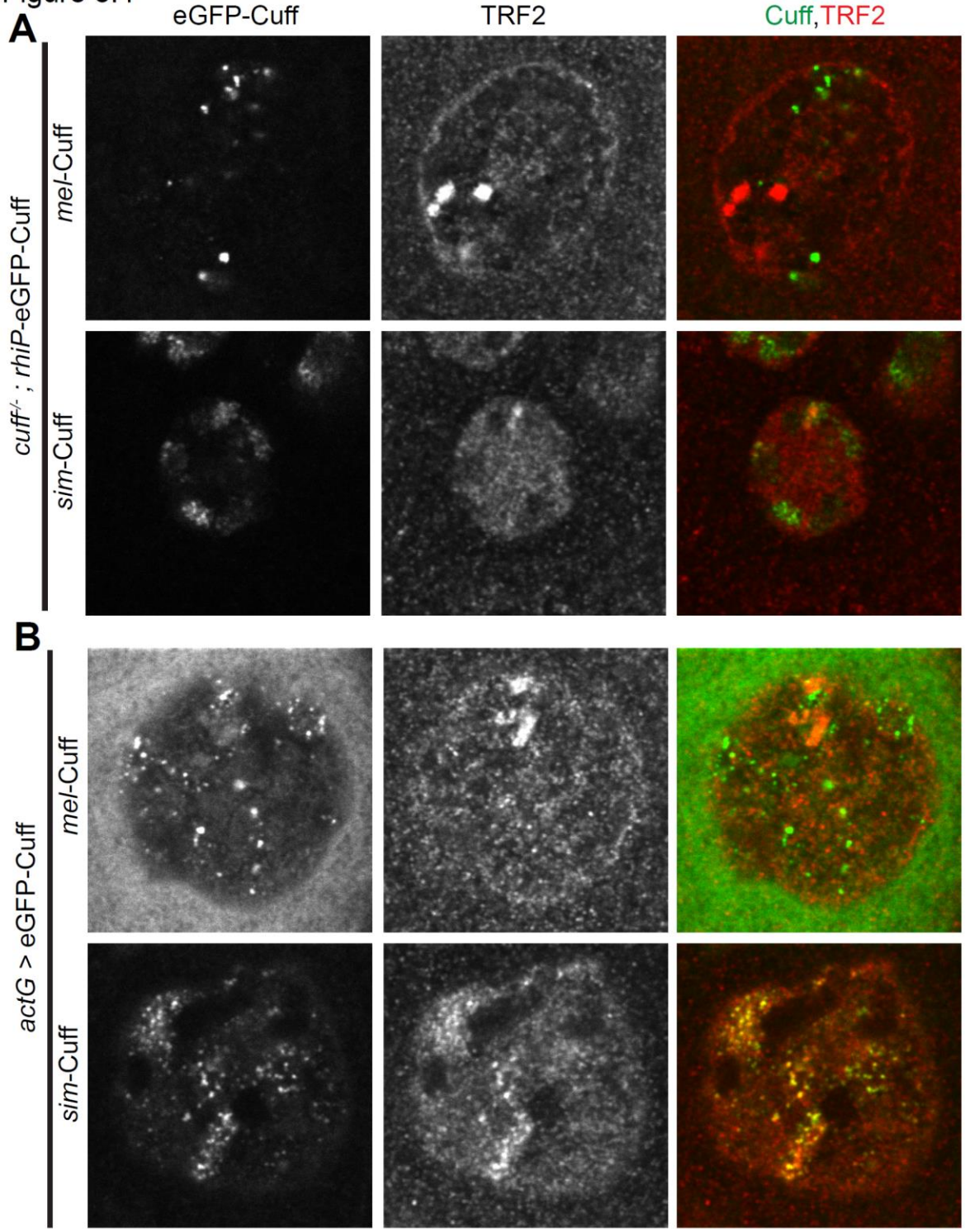

A
eGFP-Cuff
TRF2
Cuff,TRF2
cuff-/- ; rhiP-eGFP-Cuff
mel-Cuff
sim-Cuff
B
actG > eGFP-Cuff
mel-Cuff
sim-Cuff

Figure 3.4: *sim*-Cuff leads to TRF2 mislocalization

(A) Localization of TRF2 with respect to GFP tagged Cuff in the germline nuclei of *cuff* mutants expressing *rhi* promoter driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged image shown on top. TRF2 signal is reduced in *cuff* mutants expressing *sim*-Cuff.

(B) Localization of TRF2 with respect to *act5C*-Gal4 driven GFP tagged Cuff in the germline nuclei. Color assignments for merged image shown on top. TRF2 co-localizes with over-expressed *sim*-Cuff.

As TRF2 acts as a transcription factor at piRNA clusters and also for many genes, we wondered whether *sim*-Cuff driven mislocalization of TRF2 leads to changes in the transcriptional profile in the ovaries. To test this, we checked both piRNA cluster and gene transcript levels in ovaries. Cuff has been shown to be required for transcription from piRNA clusters and piRNA cluster transcript processing by inhibiting splicing (Chen et al., 2016; Zhang et al., 2014). *cuff* mutants show reduced transcription and leads to production of spliced transcripts from clusters. Quantification of the cluster transcript levels is complicated by the peaks of these spliced transcripts in *cuff* mutants (Figure 3.5A). The signal at the spliced transcripts in *cuff* mutants is more than 50 fold compared to WT. *mel*-Cuff can rescue the piRNA transcript production and suppress splicing at clusters. *sim*-Cuff fails to suppress splicing at clusters, as observed by the increase in spliced transcripts. In order to quantify piRNA cluster transcripts in light of this increase in spliced reads from specific locations, we chopped the piRNA clusters into 1kb bins and calculated the rpkm value for each bin. We divided the bins into two groups based on rpkm values in WT control: first group has rpkm ≥ 5 and the second group with rpkm < 5. For bins with rpkm ≥ 5 in WT, we observed a decrease in piRNA cluster transcripts in *cuff* mutants (Figures 3.5B and 3.5D). Both *mel*-Cuff and *sim*-Cuff can partially rescue the transcription from piRNA cluster regions with rpkm ≥ 5 in WT. For the bins which have rpkm < 5 in WT (Figure 3.5C), *cuff* mutants and *cuff* mutants expressing *sim*-Cuff show a significant increase in cluster transcripts, due to production of spliced transcripts. *mel*-Cuff does not show an increase in transcripts from these low expressing loci. RDC complex has been proposed to suppress splicing and promote non-canonical transcription

from clusters (Mohn et al., 2014; Zhang et al., 2014). Our observations suggest that *sim*-Cuff does not suppress splicing and leads to canonical transcription from piRNA clusters.

Figure 3.5

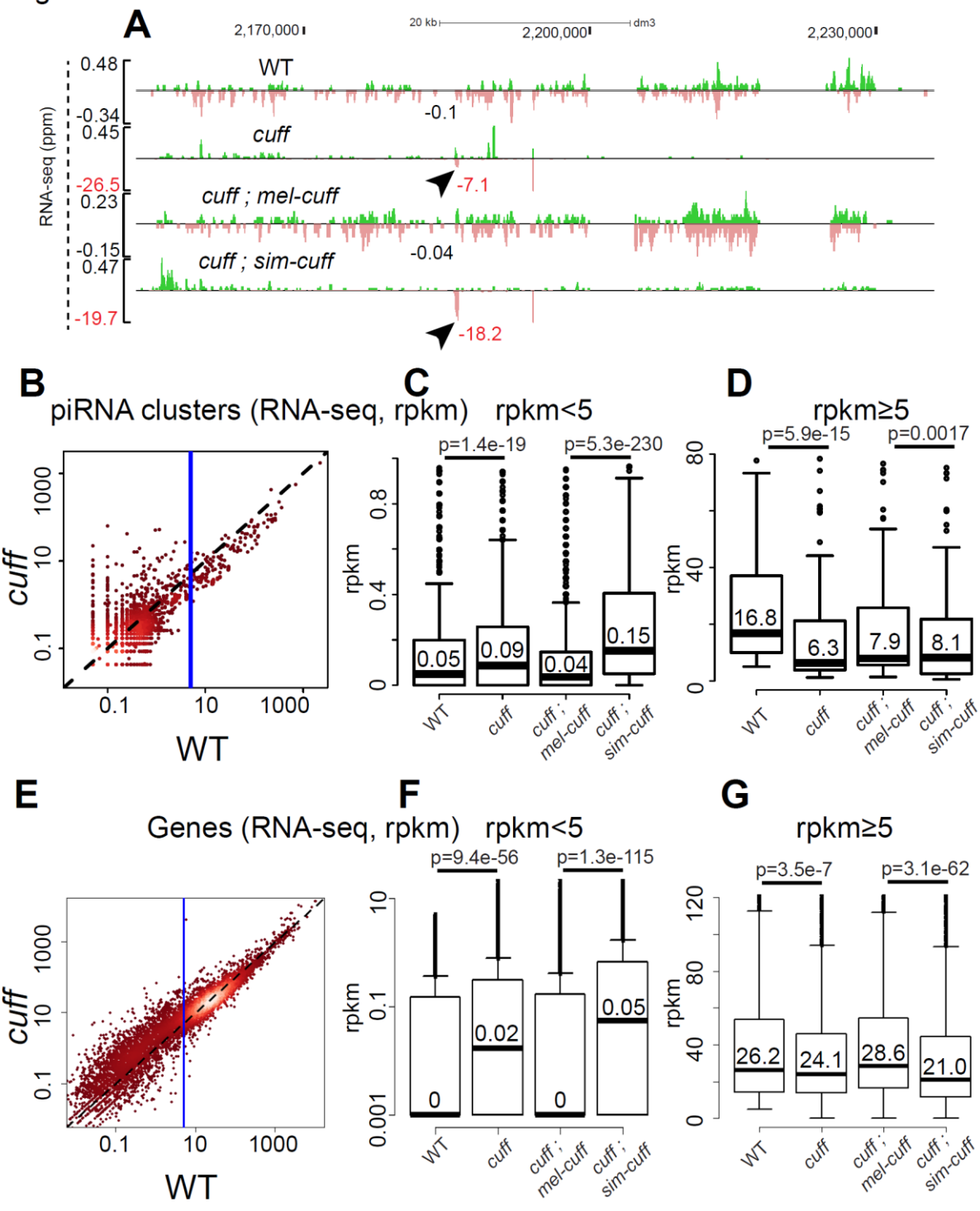

A
2,170,000
20 kb
2,200,000
dm3
2,230,000
RNA-seq (ppm)
WT
cuff
cuff ; mel-cuff
cuff ; sim-cuff
0.48
-0.34
0.45
-26.5
0.23
-0.15
0.47
-19.7
-0.1
-7.1
-0.04
-18.2
B
piRNA clusters (RNA-seq, rpkm)
cuff
WT
C
rpkm<5
p=1.4e-19
p=5.3e-230
rpkm
0.05
0.09
0.04
0.15
D
rpkm≥5
p=5.9e-15
p=0.0017
16.8
6.3
7.9
8.1
E
Genes (RNA-seq, rpkm)
F
rpkm<5
p=9.4e-56
p=1.3e-115
0
0.02
0
0.05
G
rpkm≥5
p=3.5e-7
p=3.1e-62
26.2
24.1
28.6
21.0
WT
cuff
cuff ; mel-cuff
cuff ; sim-cuff

Figure 3.5: *sim*-Cuff changes transcriptional profile in ovaries

(A) Genome browser view showing abundance of piRNA precursor transcripts uniquely mapping to 42AB piRNA cluster in WT, *cuff* mutant and *cuff* mutants expressing *mel-cuff*, *sim-cuff*. The Watson strand is in green, and Crick strand in magenta. The arrowheads show the location of spliced transcripts in *cuff* mutants and *cuff* mutant expressing *sim-cuff*. The numbers next to arrowheads show the signal at the spliced transcript.

(B) Scatterplot showing rpkm values for 1kb piRNA cluster bins in *cuff* mutant vs. WT control. Each point on the scatterplots shows rpkm values for a piRNA cluster bin in ovaries. Diagonal black dotted line represents x=y. The vertical blue line is for x=5.

(C, D) Boxplots showing piRNA cluster expression by RNA-seq in WT, *cuff* mutants, *cuff* mutant expressing either by *mel-cuff* or *sim-cuff* in ovaries. (C) shows the bins for which rpkm < 5 and (D) shows bins with rpkm ≥ 5 in WT. The numbers in boxplots show the median values. p value for differences obtained by Wilcoxon test.

(E) Scatterplot showing rpkm values for genes in *cuff* mutant vs. WT control. Each point on the scatterplots shows rpkm values for a gene in ovaries. Diagonal black dotted line represents x=y. The vertical blue line is for x=5. The points are merged based on density.

(F, G) Boxplots showing gene expression by RNA-seq in ovaries of WT, *cuff* mutants, *cuff* mutant expressing either *mel-cuff* or *sim-cuff* flies. (F) shows the genes for which

rpkm < 5 and (G) shows genes with rpkm ≥ 5 in WT. The numbers in boxplots show the median values. p value for differences obtained by Wilcoxon test.

Cuff has been proposed to be involved in the regulation of subset of genes (Pritykin et al., 2017). As *sim*-Cuff leads to mislocalization of TRF2, which is a general transcription factor, we wondered whether *sim*-Cuff leads to transcriptional changes in genes. In order to differentiate between highly expressed and low expressed genes, we divided the genes into 2 groups based on their rpkm values in WT. One group had rpkm ≥ 5 and other had rpkm < 5. For high expressed genes (rpkm ≥ 5), we observed a significant decrease in transcript levels in *cuff* mutants (Figures 3.5E and 3.5G). *cuff* mutants expressing *sim*-Cuff showed an even higher decrease, consistent with TRF2 displacement (Figure 3.5G). To check whether the TRF2 mislocalization leads to an increase in low expressed genes, we checked the expression of genes with rpkm < 5 in WT. We observed an increase in gene transcripts in *cuff* mutants (Figures 3.5E and 3.5F) and even greater increase in cuff mutants expressing *sim*-Cuff. Thus, *sim*-Cuff re-organizes the transcriptional machinery, especially TRF2 and changes the transcriptional profile of ovaries.

sim-Cuff also changes piRNA profile

As *sim*-Cuff traps Del, TRF2 and leads to changes in the transcriptional profile, we wondered whether this change in transcriptional profile affects piRNAs. To test this, we sequenced strand specific small RNA libraries from the ovaries. *cuff* mutation leads to loss of piRNAs (Figure 3.6A). Both *mel*-Cuff and *sim*-Cuff can rescue the piRNA production from clusters, however, *sim*-Cuff still makes less piRNAs compared to *mel*-Cuff rescue (Figures 3.6A and 3.6B). *sim*-Cuff expression in *cuff* mutants leads to a decrease in piRNAs uniquely mapping to major dual strand cluster 42AB and shows no

change in flamenco mapping piRNAs which is a uni-strand cluster. As *sim*-Cuff increases transcript levels from low expressed genes, we checked genic piRNAs. *sim*-Cuff significantly increases piRNAs mapping to genes (Figures 3.6C and 3.6D). Thus, *sim*-Cuff seems to change the organization of RDC complex and leads to production of piRNAs from non-cluster sites.

Figure 3.6

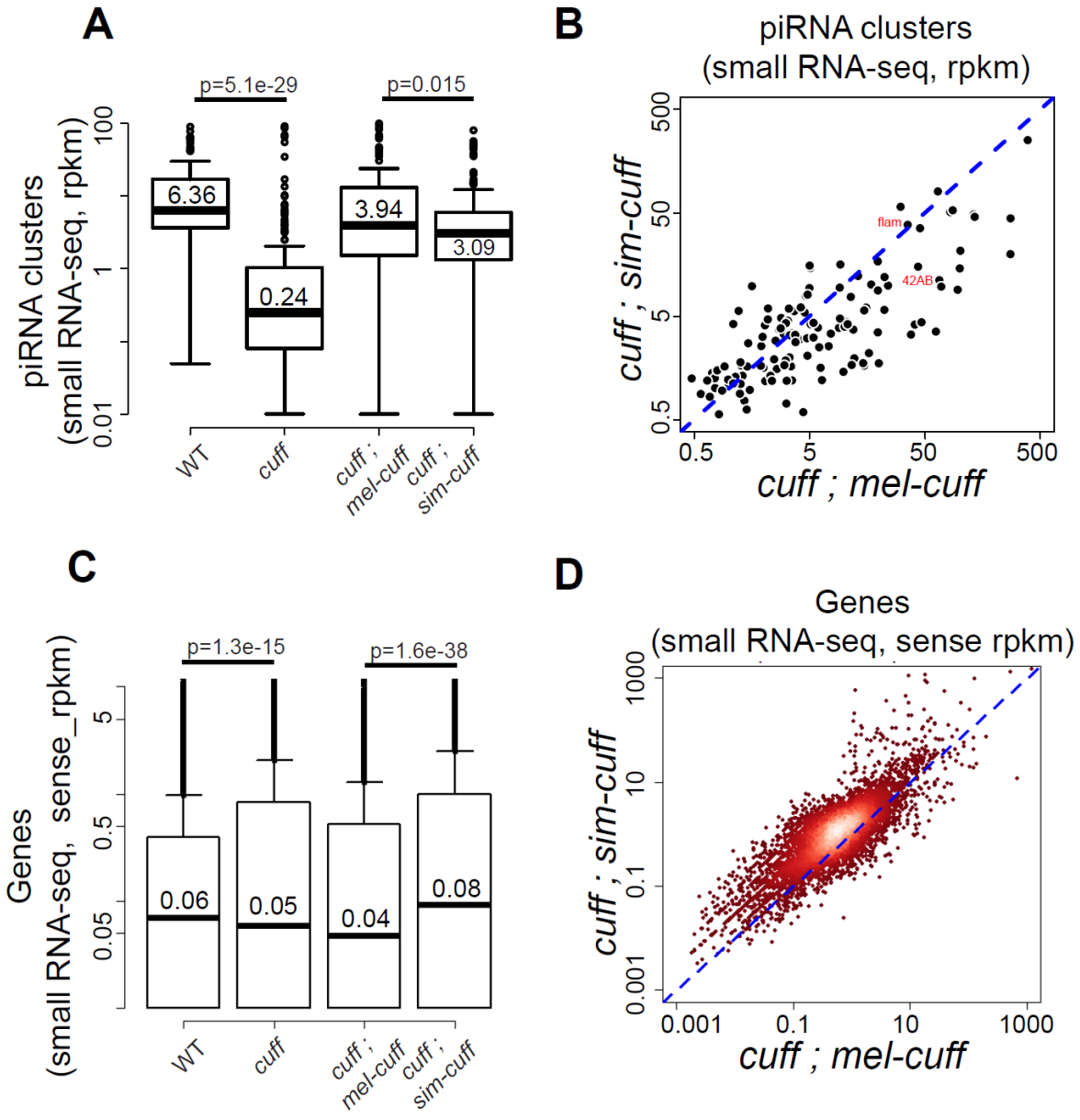

A
p=5.1e-29
p=0.015
piRNA clusters
(small RNA-seq, rpkm)
0.01
1
100
6.36
0.24
3.94
3.09
WT
cuff
cuff ; mel-cuff
cuff ; sim-cuff
B
piRNA clusters
(small RNA-seq, rpkm)
cuff ; sim-cuff
cuff ; mel-cuff
0.5
5
50
500
flam
42AB
C
p=1.3e-15
p=1.6e-38
Genes
(small RNA-seq, sense_rpkm)
0.05
0.5
5
0.06
0.05
0.04
0.08
WT
cuff
cuff ; mel-cuff
cuff ; sim-cuff
D
Genes
(small RNA-seq, sense rpkm)
cuff ; sim-cuff
cuff ; mel-cuff
0.001
0.1
10
1000

Figure 3.6: *sim*-Cuff causes aberrant piRNA production

(A) Boxplots showing rpkm for cluster mapping piRNAs in ovaries of WT, *cuff* mutants, *cuff* mutant expressing either *mel-cuff* or *sim-cuff* flies. The numbers in boxplots show the median values. p value for differences is obtained by Wilcoxon test.

(B) Scatterplot showing rpkm values for cluster mapping piRNAs in ovaries of *cuff* mutant expressing *sim-cuff* vs. *cuff* mutant expressing *mel-cuff*. Each point on the scatterplot shows rpkm values for a piRNA cluster in ovaries. Diagonal black dotted line represents x=y. The points for 42AB and flamenco piRNA clusters are labelled.

(C) Boxplots showing rpkm for genic piRNAs in ovaries of WT, *cuff* mutants, *cuff* mutant expressing either by *mel-cuff* or *sim-cuff* flies. The numbers in boxplots show the median values. p value for differences is obtained by Wilcoxon test.

(D) Scatterplot showing rpkm values for genic piRNAs in ovaries of *cuff* mutant expressing *sim-cuff* vs. *cuff* mutant expressing *mel-cuff*. Each point on the scatterplot shows rpkm values for a gene in ovaries. Diagonal black dotted line represents x=y. The points are merged based on density.

Discussion

Dynamic assembly of piRNA pathway proteins

piRNA clusters are the source of piRNAs. Many protein factors are known to facilitate production of piRNAs from the dual strand clusters (Huang et al., 2017). Rhi binds to histone modification H3K9me3 at these clusters (Yu et al., 2015a). It also binds to Del, which further recruits Moonshiner and TRF2 (Andersen et al., 2017). This whole complex is required for transcription from clusters. Cutoff is proposed to prevent transcriptional termination during the transcription of long piRNA precursor transcripts (Chen et al., 2016). The RDC complex suppresses splicing of the piRNA precursor transcripts (Zhang et al., 2014). Thus, the unique or "non-canonical" transcription from piRNA clusters is different from the canonical transcription from many genes. It is still unclear how the components of RDC complex work together to achieve this non-canonical transcription.

All these three proteins are interdependent for their localization (Mohn et al., 2014). Rhi-Del interaction is validated by multiple assays (Mohn et al., 2014; Parhad et al., 2017). When Rhi and Del are expressed as a single polypeptide chain, the fusion protein localizes to piRNA clusters, but fails to work in the either *rhi* or *del* or double mutant background. This indicates that Rhi-Del interaction is dynamic and both the proteins do not function if they are strongly bound. Del has been shown to bind to Cuff by yeast two hybrid assay (Mohn et al., 2014). However, we failed to observe any interaction between Del and Cuff in *D. melanogaster*. When we swap Cuff between

species, *sim*-Cuff does not function in *D. melanogaster*. Unlike *mel*-Cuff, *sim*-Cuff stably binds to Del. Over-expressed *sim*-Cuff has a dominant effect on fertility of flies, indicating that *sim*-Cuff traps Del and having wild type copies of *mel-cuff* does not matter. We propose that Del-Cuff interaction is transient, so that we cannot detect it in either Del or Cuff IP. It is observed only when *sim*-Cuff traps Del. As we see no Cuff in Rhi IP and no Rhi in the *sim*-Cuff-Del complex, we think Del shuttles back and forth between Rhi-Del and Del-Cuff complexes. Del can bind to either Rhi or Cuff, but not at the same time. This would explain why the interactions within RDC complex need to be transient for proper function.

Why these interactions need to be dynamic? Rhi binds to H3K9me3 which is a histone mark for both constitutive heterochromatin and piRNA clusters. If Rhi is not interdependent with Del and Cuff, but is upstream of both Del and Cuff, any non-specific binding of Rhi at heterochromatin would begin the assembly of piRNA transcription machinery at wrong sites. piRNA precursor transcription also uses the components used for normal gene transcription, such as RNA polymerase II and TRF2 (Andersen et al., 2017). If Cuff is upstream of Del and Rhi, it can lead to assembly of RDC complex at the genes, due to its interaction with TRF2 and production of genic piRNAs. Thus we propose that the interactions within RDC complex need to be transient to direct piRNA production only from piRNA clusters and not from any ectopic sites. Consistent with this model, stable binding of Del and *sim*-Cuff leads to the spread of RDC complex to other genomic locations, unlike dynamic RDC complex which only localizes to piRNA clusters (Figure 3.7). The observations that TRF2 localization is altered in *cuff* mutants expressing

sim-Cuff and over-expression of *sim*-Cuff leads to mislocalization of TRF2, suggest that *sim*-Cuff traps TRF2 at ectopic sites. As TRF2 is a transcription factor, such ectopic localization would lead to changes in transcription profile, which is what we observe. Expression of *sim*-Cuff in *cuff* mutants leads to decrease in transcript levels for highly expressed genes and significant increase in transcripts for low expressed genes. Thus, we think that non-canonical transcription at piRNA clusters requires the transcription machinery used for canonical transcription at genes. To prevent the establishment of non-canonical transcription and thus piRNA production from other canonical transcription sites, especially genes, RDC complex segregates the functions of cluster binding and transcriptional activation among different proteins, such that their dynamic interaction would prevent generation of wrong set of piRNAs.

Figure 3.7

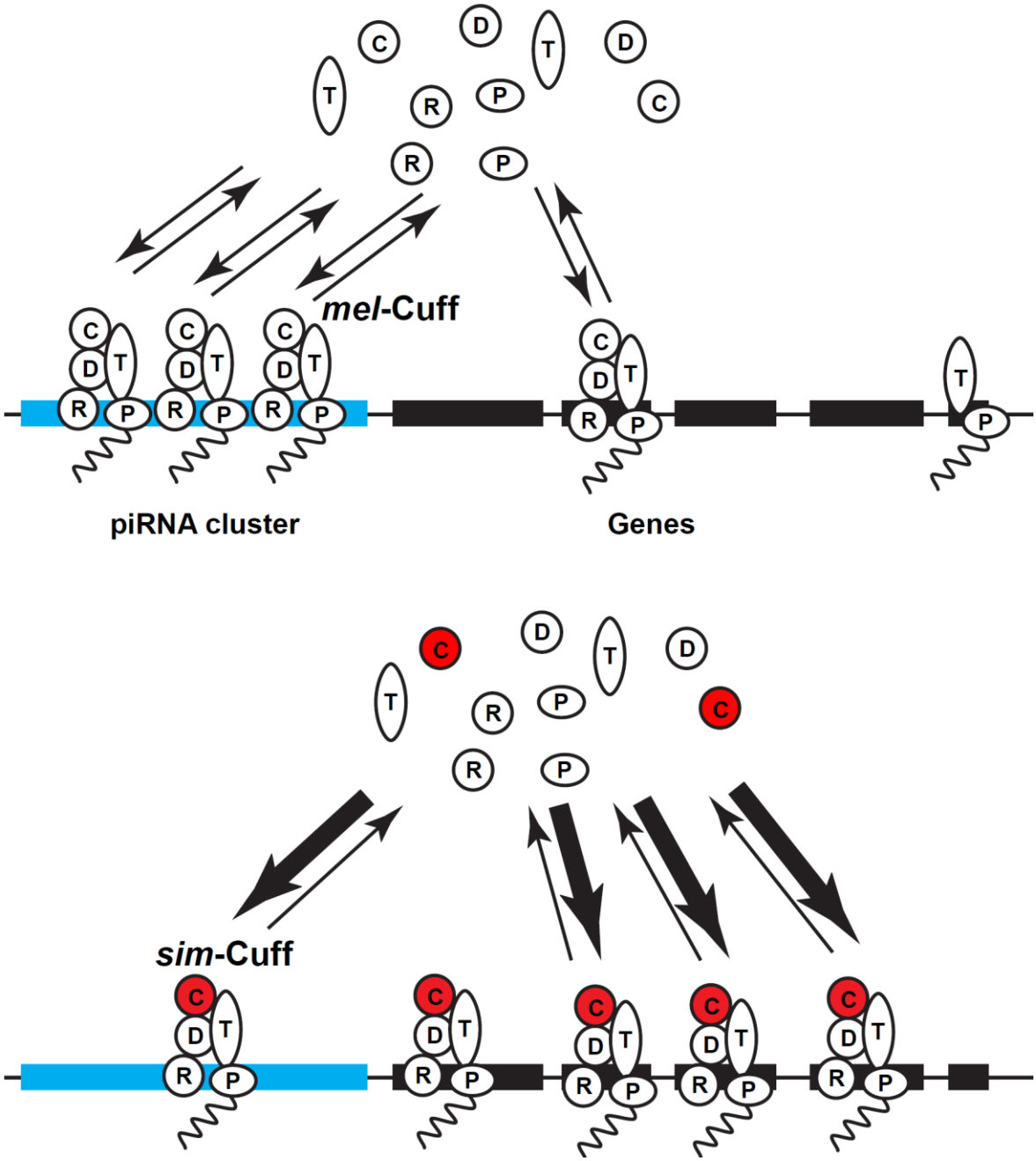


Figure 3.7: Model for species specific function of *sim*-Cuff

Symbols and proteins are as follows: R: Rhi, D: Del, C: *mel*-Cuff (white background), *sim*-Cuff (red background), T: TRF2, P: RNA Polymerase II. Top: RDC complex can dissociate fast from the chromatin binding sites. With *mel*-Cuff, it mostly organizes at piRNA clusters (blue) and promotes non-canonical transcription and piRNA precursor processing from clusters. It rarely affects transcription from other genes. TRF2 acts as transcription factor for some genes. Bottom: *sim*-Cuff traps the whole complex at various locations in the genome, due to stable Del binding. This leads to the mislocalization of RDC complex from piRNA cluster to other genes in the genome. *sim*-Cuff also traps TRF2 at these ectopic sites leading to increased transcripts from these gene and loss of TRF2 foci present in WT. Dynamic interaction between RDC complex is therefore necessary for its normal function.

Significance of adaptive evolution of Cuff

Many piRNA pathway genes are rapidly evolving indicative of involvement in a host pathogen arms race (Blumenstiel et al., 2016). *cuff* gene encodes for Rai1 homolog Cutoff in flies (Chen et al., 2007b). Many Cutoff residues show sequence divergence between closely related species *D. melanogaster* and *D. simulans*. When we swap the *cuff* gene between these two species, it fails to function, indicating that the sequence divergence has functional consequences. Such species swap renders Rhi and Del non-functional (Parhad et al., 2017). As we see species specific function for Rhi, Del and Cuff, we think that the whole RDC complex is the target of transposons in a host pathogen arms race. As *sim*-Cuff strongly binds to Del, adaptive evolution has led to incompatibility in the Del-Cuff interface.

What led to this incompatibility? As piRNA pathway forms an adaptive immune defense against pathogenic transposons, we hypothesize that Del-Cuff and other components in piRNA pathway are rapidly evolving in response to host-pathogen arms race. Host pathogen arms race leads to adaptive evolution of host and pathogen genes, however we see incompatibility in the Del-Cuff interface. We propose that this incompatibility could arise due to molecular mimicry (Figure 3.S4). In *D. melanogaster*, Del and Cuff transiently bind. Mutations in transposon encoded proteins generate an inhibitor protein that mimics Cuff binding surface. This leads to reduction in Del-Cuff binding and disruption in piRNA pathway. With time, Del mutates so that the mimic can no longer bind to Del. This restores affinity by sacrificing specificity. Cuff mutates further to increase Cuff-Del interaction. These continuous changes in the Del and Cuff

proteins may lead to sequence divergence between species. Consistent with this model, we observe a stable interaction between *sim*-Cuff and *mel*-Del.

The hybrids from crosses between *D. melanogaster* and *D. simulans* are sterile and phenocopy piRNA pathway mutants (Kelleher et al., 2012), even though these hybrids contain all the copies of piRNA genes from both the species. As *sim*-Cuff causes a dominant effect on fertility in *D. melanogaster*, it can trap *mel*-Del and TRF2 in the hybrids and lead to defects in piRNA production. Thus adaptive evolution in piRNA pathway may contribute to establish and maintain reproductive isolation between species.

Experimental Procedures

Fly stocks

All experiments were performed in *Drosophila melanogaster* females, except mentioned otherwise. All flies were maintained at 25^{O}C on cornmeal medium. All transgenic lines were generated by ϕC31 integration at chr3L-68A4. *cuff*^{WM25} (*cuff*^{WM}) and *cuff*^{QQ37} (*cuff*^{QQ}) alleles were obtained from Trudi Schüpbach (Princeton University). *cuff*^{KG05951} (*cuff*^{KG}) was obtained from Bloomington (Stock # 14462). Del-Nanobody stock was obtained from Julius Brennecke (IMBA, Vienna). F1 females from OregonR crossed to *w*^{1} were used as wild type (WT) control, unless mentioned otherwise.

Generation of transgenic flies

mel-cuff was cloned from OregonR ovary cDNA and *sim-cuff* from *D. simulans* C167.4 ovary cDNA. The reverse primer for the PCR reaction was used for making cDNA with Superscript III reverse transcriptase (Thermo Fisher Scientific). *mel-cuff* was PCR amplified from cDNA by using forward primer: CAC CAT GAA TTC TAA TTA CAC AAT ATT AAA C and reverse primer: TTA AAC TAT AGA AGA CAT GGT TTG C and cloned into pENTR-D-TOPO vector by directional TOPO cloning kit (Thermo Fisher Scientific). Similarly, *sim-cuff* was PCR amplified from cDNA using forward primer: CAC CAT GAA TTC TAA TTA CAA AAT ATT GAA C and reverse primer: TTA TTG GTA AAC TGT GGA AGA CAT GG and cloned into pENTR-D-TOPO vector. These served as entry vectors for Gateway cloning. The destination vectors *rhiP*-attB-pPGW (for expressing N' GFP tagged proteins under *rhi* promoter) and attB-pPGW (for

expressing N' GFP tagged proteins under *UASp* promoter) were used as described in (Parhad et al., 2017). The plasmids obtained after LR gateway cloning reaction were sequenced and injected by ϕC31 integration at chromosomal location 3L-68A4 (Bischof et al., 2007).

Fertility assays

2-4 day old flies were maintained on grape juice agar plates for 1 or 2 days. After removing flies, the eggs were counted for fused appendages and the hatching was measured after 2 days. The bar graphs indicate mean and standard deviation for 3 biological replicates, with the shown number of scored embryos.

Immuno-staining

Immuno-staining and image analysis was performed as described in (McKim et al., 2009; Zhang et al., 2012a). In brief, 2-4 day old female ovaries were dissected in Robb's buffer, fixed with 4% formaldehyde, washed, incubated overnight with primary antibody, washed, incubated with secondary antibody with the fluorophore overnight and mounted on slide. ChromoTek anti-GFP Booster (Atto-488) antibody added with secondary antibody to enhance the GFP signal. Antibodies used: anti-GFP Booster (ChromoTek) at 1:200, guinea pig anti-Rhi (our lab) at 1:1000, rabbit anti-Del (from Julius Brennecke) at 1:1000, rabbit anti-TRF2 (from James Kadonaga) at 1:500, rabbit anti-H3K9me3 (abcam) at 1:1000.

Immuno-precipitation

IP was performed as described in (Parhad et al., 2017). Briefly, 2-4 day old female ovaries were dissected in Robb's medium, lysed by homogenization and sonication and centrifuged to get input for IP. chromotek GFP-Trap®_A beads were used for GFP IP. The lysate was incubated with beads for 3 hours at 4^0C and subsequently washed 4 times with lysis buffer. Finally the beads were suspended in SDS-PAGE lysis buffer. The procedure for mass spectrometry is descried in (Vanderweyde et al., 2016). Briefly, the IPed samples were resolved on a 10% SDS-PAGE gel. The gel pieces were processed for trypsin digestion to get the peptides, which were further analyzed by LC-MS/MS. Rhi and Del IP data was used from (Parhad et al., 2017).

Small RNA-seq

Small RNA libraries were prepared as described in (Zhang et al., 2014). Briefly, total RNA prepared from 2-4 day old female ovaries by mirVANA kit (Ambion). Size selection was done for 18-30nt by gel purification. Further 3' and 5' ligation by adapters, reverse transcription, PCR amplification was performed to obtain libraries. Sequencing was done by Illumina platform.

RNA-seq

RNA-seq libraries were prepared as described in (Zhang et al., 2012b). Briefly, RNA samples were depleted for ribosomal rRNA by Ribo-Zero kit (Illumina), fragmented, reverse transcribed, ligated by adapters and PCR amplified to make libraries. dUTP incorporation done for strand specificity. Sequencing was done by Illumina platform.

Bioinformatics analysis

Reads from small RNA-seq libraries were aligned to the *D. melanogaster* genome (dm3) by bowtie (Langmead et al., 2009), after removing the 3´end linkers. The transcriptome annotations were obtained from Flybase r5.50. The piRNA cluster coordinates were from (Brennecke et al., 2007). Reads mapping to known non-coding RNAs (ncRNAs, such as rRNAs, tRNAs, etc.) and miRNAs were excluded for the quantification of piRNA abundance of clusters. The read counts were obtained using BEDTools (Quinlan and Hall, 2010) and normalized by the total number of reads aligned to the genome excluding known ncRNAs. A read is counted proportionally if it has multiple mapping locations. TopHat (Trapnell et al., 2009) was used to align RNA-seq reads to the genome. rRNA reads were removed before the quantification of expression levels of genes, piRNA clusters, and transposons.

Analysis of Proteome obtained by Mass spectrometry

The raw data was processed through software programs Proteome Discoverer and Mascot Server before display on Scaffold Viewer (Proteome Software, Inc.). iBAQ values (Schwanhausser et al., 2011) of each IPed protein were used to make the graphs with R after adding pseudocount.

Acknowledgements

We would like to thank the members of Wend and Theurkauf labs and UMassmed RNA Bio. community for their insightful discussions and critical comments; Trudi Schüpbach for *cuff* stocks; Julius Brennecke for Del antibody and Del-nanobody stock; James Kadonaga for TRF2 antibody; Bloomington and UCSD *Drosophila* stock centers for various fly stocks; John Leszyk of UMassmed Proteomics facility for mass spectrometry. This work was supported by National Institute of Child Health and Human Development (R01HD049116 and P01HD078253).

Figure 3.S1

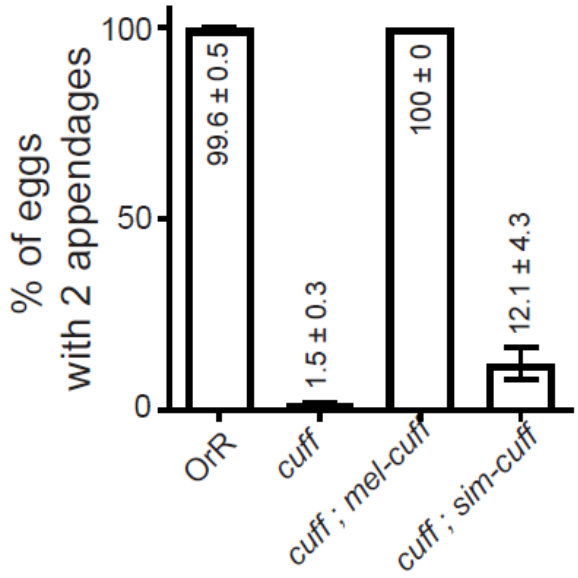

% of eggs
with 2 appendages
100
50
0
99.6 ± 0.5
1.5 ± 0.3
100 ± 0
12.1 ± 4.3
OrR
cuff
cuff ; mel-cuff
cuff ; sim-cuff

Figure 3.S1: *sim-cuff* behaves like a null allele in *D. melanogaster*

Bar graphs showing D-V patterning of eggs, in terms of percentages of eggs with two appendages, produced by OrR (wild type (WT) control), *cuff* mutants, and *cuff* mutants rescued by either *mel-cuff* or *sim-cuff*. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *cuff* mutants and *cuff* mutants rescued by *sim-cuff* where average of 230 and 23 eggs were scored respectively.

Figure 3.S2

A

Identified protein	Spectrum counts	
	Rhi	GFP ctrl
GFP	60	187
Rhi	79	0
Del	24	0
Cuff	0	0

B

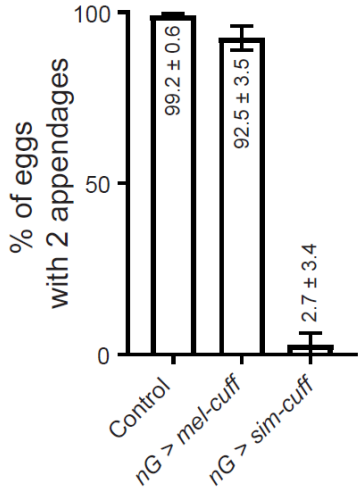


Figure 3.S2: Dynamic interaction within RDC complex

(A) Mass spectrometric analysis of Rhi binding proteins. Spectrum counts in different proteins in Rhi IP and GFP control IP mass spec. The IP-mass spec data was used from (Parhad et al., 2017).

(B) *sim*-Cuff has a dominant effect on D-V patterning. Bar graphs showing percentages of eggs with 2 appendages, produced by control, flies over-expressing either *mel-cuff* or *sim-cuff* by germline specific *nanos*-Gal4 driven. The numbers in/above the bars show mean ± standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *nanos*-Gal4 driven *sim-cuff* where average of 50 eggs were scored.

Figure 3.S3

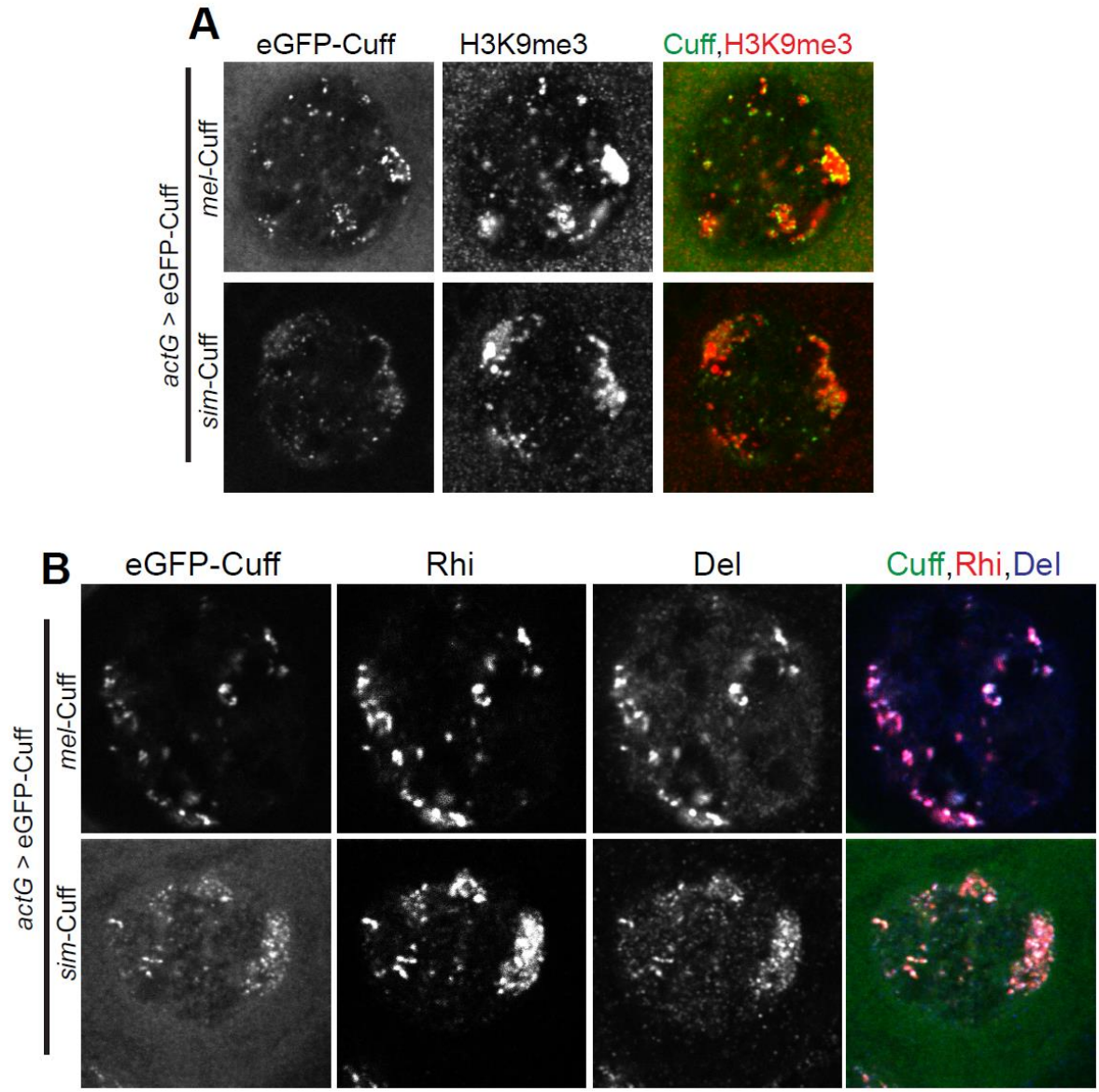

A
eGFP-Cuff
H3K9me3
Cuff,H3K9me3
actG > eGFP-Cuff
mel-Cuff
sim-Cuff
B
eGFP-Cuff
Rhi
Del
Cuff,Rhi,Del
actG > eGFP-Cuff
mel-Cuff
sim-Cuff

Figure 3.S3: *sim*-Cuff containing RDC complex loses specificity for piRNA clusters

(A) Localization of GFP tagged Cuff with respect to H3K9me3 marked chromatin in the germline nuclei of *act5C*-Gal4 driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged image shown on top.

(B) Localization of GFP tagged Cuff with respect to Rhi and Del in the germline nuclei of *act5C*-Gal4 driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged image shown on top.

Figure 3.S4

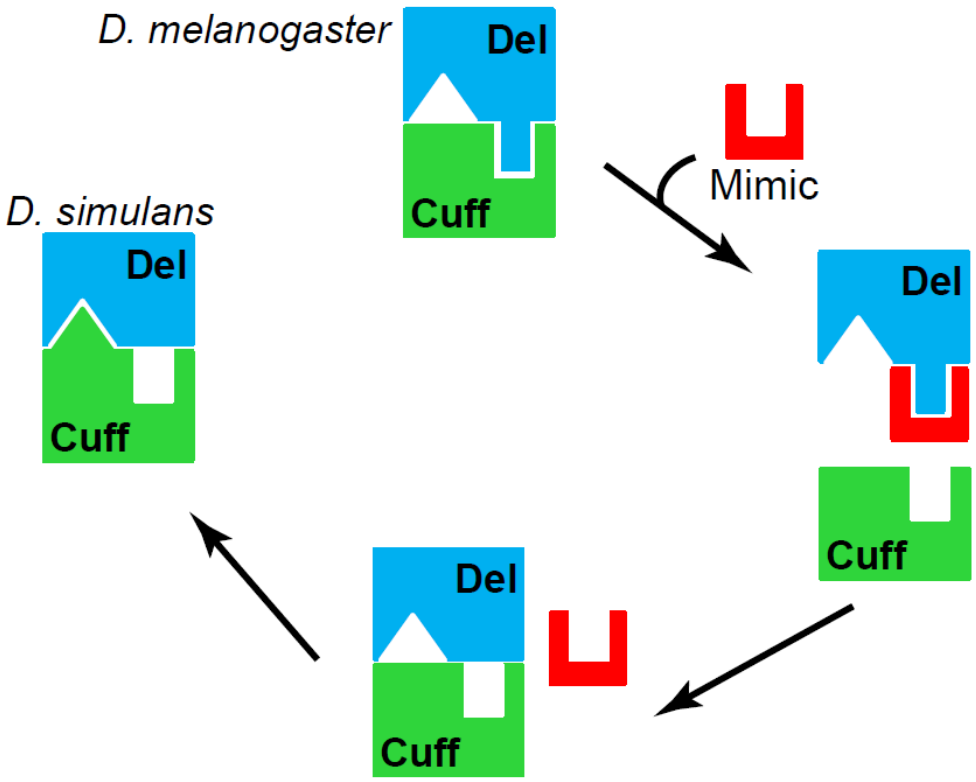

D. melanogaster
Del
Cuff
Mimic
D. simulans
Del
Cuff
Del
Cuff
Del
Cuff

Figure 3.S4: Model for adaptive evolution of Del-Cuff interface

Initially Del and Cuff bind. A transposon encoded protein mutates so that it can mimic Cuff binding surface. This prevents Del-Cuff interaction and disrupts piRNA biogenesis. The host protein Del mutates so that the transposon encoded mimic cannot bind. It reduces the affinity at the cost of specificity. Cuff mutates further at a different site to increase affinity. We think that *D. melanogaster* and *D. simulans* are at different stages of this cycle. Host-pathogen arms race in the form of molecular mimicry can lead to adaptive evolution of Del-Cuff interface.

Chapter IV: Conclusions, discussion and future directions

Our data sheds light on the nature of the host-pathogen arms race between piRNA pathway and transposons. We show few instances of how this conflict defines the host immune system and speculate about the nature of this arms race.

Dynamic interaction within the RDC complex

piRNA clusters are the chief source of piRNAs (Brennecke et al., 2007). In *Drosophila*, piRNA clusters are specified by RDC complex composed of Rhino (Rhi), Deadlock (Del) and Cutoff (Cuff) (Mohn et al., 2014). These proteins promote transcription from clusters and also direct the piRNA precursor transcripts for piRNA processing (Chen et al., 2016; Mohn et al., 2014; Zhang et al., 2014). It is not clear how these proteins interact within the complex to carry out this function. Our data provides clues about the nature of interactions within the complex, which is at the heart of piRNA biogenesis machinery.

There were conflicting reports about the nature of interactions within this RDC complex. Pane et al showed that Cuff binds to Rhi by immuno-precipitation (IP) and Western blotting (Pane et al., 2011). By using yeast two hybrid assay, Mohn et al showed that Rhi binds to Del and not to Cuff and Cuff binds to Del (Mohn et al., 2014). To resolve this issue, we expressed all these components of RDC complex as tagged transgenes in *D. melanogaster* ovaries, which is the normal site where these proteins are expressed, immuno-precipitated the tagged proteins and studied their protein interactors by mass spectrometry. We confirmed the Rhi-Del binding. We found Del in Rhi IP and Rhi in Del IP. We observed that the interaction between these proteins needs to be dynamic for normal function. When we expressed these proteins as a single polypeptide chain, i.e. fusion with a covalent linkage, both these proteins fail to function. The fusion protein can localize to piRNA clusters, but it fails to rescue *rhi*, *del* or *rhi*, *del* double

mutants. Strongly bound Rhi-Del localize to clusters, but fail to rescue respective mutants, suggesting roles other than cluster identification for Rhi and Del.

The role of Cuff was rather peculiar. We did not see Cuff in either Rhi or Del IP. There was no Del or Rhi in Cuff IP either. It is unlikely due to aberrant Cuff protein produced from the transgene, because it can rescue cuff mutants. We could observe Cuff-Del interaction when we expressed Cuff from sibling species *D. simulans* in *D. melanogaster*. *sim*-Cuff shows strong binding to *mel*-Del. This led us to conclude that Cuff-Del interaction is transient, because of which we could not detect interaction between Cuff-Del from *D. melanogaster*. When we stabilized the Cuff-Del interaction by expressing GFP tagged *mel*-Cuff in a *cuff* mutant background alongwith GFP nanobody tagged Del, we observed that *mel*-Cuff fails to rescue *cuff* mutation. Thus, we observed evidence for transient Cuff-Del interaction by two different assays. The species swap of Cuff allowed us to speculate about potential role of Cuff. When *sim*-Cuff traps Del, we also observe TRF2 (TATA box-binding protein (TBP)-related factor 2) to be the part of the same complex. TRF2 is a general transcription factor, especially for ribosomal genes (Kedmi et al., 2014; Wang et al., 2014). Recently it was shown to act as transcription factor at piRNA clusters in complex with Del and Moonshiner (Andersen et al., 2017). Our data confirms the interaction between Del and TRF2. We think that the natural function of Cuff is to dissociate this complex. Taken together, we propose that Del acts as a linker protein between Rhi and Cuff. It binds to either Rhi or Cuff and not both the proteins at the same time. Rhi provides the binding ability to piRNA cluster histones. Del

and Cuff couple the transcriptional machinery to these loci. It would be good to test this models *in vitro* by using purified proteins.

Why the interactions within the RDC complex need to be dynamic? It is known that Rhi, Del and Cuff localization to piRNA clusters is interdependent (Mohn et al., 2014). Rhi binds to H3K9me3 marks. The same histone modification marks both piRNA clusters and constitutive heterochromatin. If Rhi is upstream of Del and Cuff, then any accidental biding of Rhi to non-cluster H3K9me3 mark would lead to assembly of piRNA proteins at wrong locations. If Cuff is upstream of Rhi and Del, then it can assemble RDC complex at genes due to its interaction with TRF2. Thus, we think that to prevent the assembly of piRNA cluster transcription machinery and thus piRNA production from wrong locations, RDC complex components interact dynamically.

Adaptive evolution in piRNA pathway

piRNA pathway protects the genome from pathogenic transposons. Just like viruses and our immune system, it is thought that piRNA pathway and transposons are involved in a host-pathogen arms race, as many piRNA pathway genes show hallmarks of adaptive evolution (Blumenstiel et al., 2016). It was not known whether the diversity in sequence has any functional consequences. Our data shows that the rapid evolution has led to functional divergence in the piRNA pathway.

In *Drosophila*, *rhino* was shown to be under positive selection (Vermaak et al., 2005). *rhino* from *D. simulans* does not function in *D. melanogaster*. *sim*-Rhi does not interact with Del, resulting in loss of cluster binding, loss of piRNAs and transposons

upregulation. This highlights importance of Del binding for directing Rhi to clusters. Domain swaps between *mel*-Rhi and *sim*-Rhi showed that shadow domain is responsible for Del binding. Interestingly, chromo domain swap can rescue D-V patterning, piRNA production and transposon silencing, except for tirant transposon. This chromo domain swap fails to rescue the fertility, but transgenic tirant over-expression has no effect on the fertility on the flies. This leads to the possibility that Rhi has functions other than piRNA cluster specification. Exploring why chromo domain fails to rescue fertility would help to find this unknown function.

To check whether Rhi-Del interaction is species specific in both directions, we expressed *mel*-Rhi in *D. simulans*. The control was similar expression of *sim*-Rhi. We found that *sim*-Rhi binds to *sim*-Del. Thus Rhi-Del interaction is conserved between species. This highlights the possibility that even though the piRNA pathway is rapidly evolving, the mechanisms of piRNA production may be conserved, at least in species where these piRNA protein homologs are present. It would be interesting to check whether this interaction is conserved in distant *Drosophila* species. In *D. simulans*, we found that *mel*-Rhi binds to *sim*-Del and it also localizes to piRNA clusters. Thus, incompatibility in Rhi-Del interaction seems to be directional. *mel*-Rhi can bind to Del from both species, but *sim*-Rhi can only bind to *sim*-Del and not *mel*-Del. This directional incompatibility indicates that these two species may be at different stages in host-pathogen arms race. The GFP tagged transgenes would be useful to study piRNA clusters' organization in *D. simulans*.

As *mel*-Rhi binds to *sim*-Del in *D. simulans,* we wanted to ask whether *sim*-Del can bind to *mel*-Rhi in *D. melanogaster*. To test this, we expressed *sim*-Del in *D. melanogaster* as GFP or mRFP tagged protein. Consistent with observation in *D. simulans,* we found that *sim*-Del can bind to *mel*-Rhi in *D. melanogaster*. However, *sim*-Del fails to complement *del* mutation in *D. melanogaster*. It shows reduced number of cluster foci in a *del* mutant and subsequent reduction in Rhi localization. Many downstream components such as UAP56 and THO2 show even reduced cluster localization, indicating that *sim*-Del fails to assemble the piRNA cluster transcription and processing complex. It would be interesting to check which downstream component(s) of the piRNA pathway *sim*-Del is incompatible with. It may direct to other potential targets of the arms race.

Cuff species swap also shows species specific function. Thus adaptive evolution has led to functional divergence in the entire RDC complex. *sim*-Cuff incompatibility is due to aberrant interaction with Del. It is interesting that *sim*-Rhi does not bind to Del, but *sim*-Cuff strongly binds to Del. These are opposite effects for proteins swap between same two species. We have tried co-expressing two components of *D. simulans* RDC complex in *D. melanogaster*, to check whether co-expression can rescue any mutant phenotype and found that it does not work. In future, we should express all the *D. simulans* RDC complex components in *D. melanogaster* simultaneously. Rather than doing it in the triple mutant and three transgenes, it would be convenient to replace these genes in *D. melanogaster* by that of *D. simulans* genes by using CRISPR.

We found evidence of functional divergence within RDC complex. Many other piRNA pathway components are also rapidly evolving (Blumenstiel et al., 2016). We should systematically swap these components between species to study how extensive this functional divergence is. RDC complex is nuclear. Is the functional divergence also true for the cytoplasmic components? I have tried to answer it in Appendix I. Here I have studied species specific function for other piRNA pathway proteins, such as Piwi, Aubergine (Aub), Vasa and Zucchini (Zuc). Piwi and Zuc species swap don't seem have much effect on fertility. Aub swap shows mild effect on fertility, consistent with (Kelleher et al., 2012). Vasa however seems to have a greater effect on fertility and it also leads to expression of Stellate crystals in testes, indicative of loss of *Suppressor of Stellate* piRNAs. Aub and Vasa both are involved in Ping-Pong amplification cycle at the nuage assembled around nuclear periphery. My observations indicate that this Ping-Pong amplification cycle may also be the target of the host-pathogen arms race. However, this needs to be confirmed rigorously.

D. simulans RDC complex does not function in *D. melanogaster*. But converse does not seem to be true, at least in case of *mel*-Rhi. We should check whether such is the case for all the components of RDC complex. Even though *mel*-Rhi localizes to clusters in *D. simulans*, it may not complement *D. simulans rhi* mutants. piRNA pathway mutants are sterile in *D. melanogaster*. We should check whether the piRNA pathway mutations have the same effect in *D. simulans* and other species.

Possible nature of arms race

What could be driving this host-pathogen arms race between piRNA pathway and transposons? In our mass spectrometry data, we looked for possible transposon encoded inhibitor. However, we did not see any transposon encoded proteins binding to RDC components from both species. We propose a model for this arms race based on molecular mimicry. In this, a transposon encoded protein mutates such that it can mimic a host piRNA pathway protein. It would inhibit the interactions within the piRNA pathway. The piRNA protein can correspondingly mutate so as to evade this inhibition by mimic. This would prompt selection of transposon mutations which can again inhibit the host piRNA pathway. This reciprocal mutations in the host piRNA pathway and transposons would result in the observed adaptive evolution in the piRNA pathway. We don't know the possible nature of a transposon encoded inhibitor or mimic. It would be wonderful to identify this transposon inhibitor.

As transposons are direct targets of piRNAs, we propose that rapid evolution of piRNA pathway is due to the host-pathogen arms race with transposons. However, it is possible that some other pathogens, such as viruses, can be the cause this arms race and rapid evolution.

Possible role of piRNA pathway and transposons in speciation

Host pathogen arms race between piRNA pathway and transposons has led to sequence divergence in piRNA pathway which has resulted in functional divergence. As the populations diverge, rapid evolution in the piRNA pathway can build biochemical

incompatibilities which may establish reproductive barriers between populations and lead to formation of new species. These incompatibilities in piRNA pathway may help not only to establish, but also to maintain the reproductive isolation. RDC complex alone show two instances of biochemical incompatibilities which can lead to sterility. *sim*-Cuff forms a stable complex with *mel*-Del and has a dominant effect on the fertility. In hybrids between *D. melanogaster* and *D. simulans*, Cuff-Del complexes between two species can trap the piRNA proteins in a non-functional complexes leading to sterility phenotype observed in case of hybrids. *sim*-Del forms a non-functional complex with *mel*-Rhi. Thus in hybrids *sim*-Del can sequester *mel*-Rhi into a non-functional complex leading to defects in piRNAs and subsequent sterility. It would be interesting to study the role of piRNA pathway in hybrid incompatibility by knocking down piRNA pathway gene from one species in the hybrid germline. Any improvement in the fertility would lead to identification of a piRNA factor in hybrid sterility. There are multiple transgenic lines in *D. melanogaster* to knock down each gene. We can do a mini-screen by crossing a few of these lines to *D. simulans nanos*-Gal4 (this line already exists) males.

Similar to piRNA pathway, transposons can also contribute to establishment of reproductive barriers. Hybrid dysgenesis is caused by absence of maternal piRNAs to target a paternal transposon (Khurana et al., 2011). If two populations have unique transposons, then the crosses between them in either orientation would lead to sterile hybrids. Not much has been done to study the contributions of piRNA pathway and transposons in speciation. Our data points to the possible links between them.

Bigger picture

Change is an essential component of life. Species need to continuously evolve themselves to survive in the changing environments. Host-pathogen arms race is the result of the conflicts between survival instincts of the two sides with opposing interests. Transposons are selfish DNA elements which have infected almost all eukaryotes. piRNA pathway in animals tries to tame these pathogens. Despite having this conserved function of transposon silencing, many aspects of piRNA biogenesis are very different in various organisms. Flies have RDC complex to specify piRNA clusters which is not present in other organisms. Worms have RdRPs to amplify piRNA repertoire, however other animals use Ping-Pong cycle. As we know more about the variety in piRNA systems, we would appreciate why the nature adopted these different ways to face the same enemy. Our experiments takes us one step closer to understand how such conflicts shape life.

Appendix I

Studying species specific function of various piRNA pathway proteins by using introgression lines

Preface

Nadine Schultz helped me with the egg counting. I performed the rest of the experiments.

Summary

Transposons encode for major fraction of eukaryotic genomes. These transposons have deleterious effects on host and can rapidly spread through populations, characteristic of a successful pathogen. A class of small RNAs called PIWI interacting RNAs (piRNAs) protects the animal genomes from these pathogenic transposons, especially in the germline. As piRNA pathway acts as a defense against pathogenic transposons, the rapid evolution in the piRNA pathway genes is thought to be the result of host pathogen arms race with transposons. To test whether this rapid evolution results in functional divergence within piRNA pathway, we swapped different piRNA pathway genes between sibling *Drosophila* species, *D. melanogaster* and *D. simulans*. This species swap was achieved by means of introgression lines, where a large portion of *D. simulans* chromosome is incorporated into *D. melanogaster* chromosome. 4 piRNA pathway genes were within the introgression lines: *vasa*, *aubergine*, *piwi* and *zucchini*. For species swap, we crossed the introgression lines with the mutations in each of these genes. We observed no effect on fertility with *piwi* and *zucchini* swap, mild fertility defect with *aubergine* and prominent defect in fertility with *vasa*. Stellate expression was observed in *vasa* species swap. We conclude that piRNA pathway gene *vasa* has species specific function.

Introduction

Transposons encode for a large part of eukaryotic genomes. For humans, almost half of the genome consists of transposons (Bao et al., 2015; Canapa et al., 2015). These transposable elements (TEs) can have deleterious effects on host by either mobilization or recombination between repeats (Ayarpadikannan and Kim, 2014). Transposon insertions can disrupt genes by inserting into either the coding or regulatory sequences. Being repetitive in nature, recombination between different transposon copies can lead to genomic rearrangements in the form of deletion, duplication, translocation, inversion. Thus, transposons can be a major cause of genomic instability (Hedges and Deininger, 2007). Regulation of these TEs is especially important in the germline, as it is the tissue that is responsible for transfer of genetic material to the next generation. In animals, a class of small RNAs called PIWI interacting RNAs (piRNAs) control these transposons (Ghildiyal and Zamore, 2009). The piRNAs are produced from specific regions inside the genome called piRNA clusters (Brennecke et al., 2007). These piRNAs are loaded into PIWI class of Argonaute proteins, which can either cleave the transposon transcripts or identify the transposon transcript in the nucleus and direct heterochromatin formation at the transposon site. Thus piRNA pathway can regulate transposons both post-transcriptionally and transcriptionally (Huang et al., 2017).

As host piRNA pathway acts a defense against pathogenic transposons, we wondered whether there is any host-pathogen arms race between the two. Many piRNA pathway genes are rapidly evolving indicative of involvement in an arms race (Blumenstiel et al., 2016). To understand whether the sequence divergence has any

functional consequences, we decided to swap piRNA pathway genes between species having different transposon profiles. One prediction of the arms race is that a piRNA pathway gene adapted to one transposon environment would not function in a different transposon environment. *Drosophila* offers a great system to address this. Two sibling *Drosophila* species, *D. melanogaster* and *D. simulans* have very different transposon profiles (Bartolome et al., 2009; Lerat et al., 2011). Many TEs in *D. melanogaster* are full length, whereas majority of *D. simulans* transposons are truncated. Thus, these two species seem to be at a different stages of the arms race. Many piRNA pathway genes show sequence divergence between the two species (Lee and Langley, 2012; Obbard et al., 2009; Simkin et al., 2013). Thus to test this arms race, we swapped different piRNA pathway genes between species.

Introgression lines offer a simple tool to achieve this. As *D. melanogaster* and *D. simulans* are different species, the hybrids between the two are either inviable or sterile depending on the direction of the cross (Sturtevant, 1920). The fertility in hybrids can be partially rescued with the use of specific strains (Davis et al., 1996). This gives unique opportunity to generate organism containing parts of chromosomes from two species, as a result of meiotic recombination between chromosomes from two species in hybrids. By using this strategy, a fly line containing parts of *D. simulans* chromosome into *D. melanogaster* genome was obtained and is called introgression line (Sawamura et al., 2000). We observed that *D. simulans* region within these introgression line contains many piRNA pathway genes. When mutations in these genes are kept in trans with the introgression line, we essentially would swap a piRNA pathway gene between species.

By this approach, we swapped 4 piRNA pathway genes between the two species to study the functional consequences of sequence divergence.

piRNAs are produced specific regions inside the genome called piRNA clusters. These clusters contain truncated copies of transposons (Brennecke et al., 2007). The piRNAs produced as a result can direct sequence specific silencing of transposons. The piRNA precursor transcripts are transported across nuclear envelope and in the cytoplasm they are processed into piRNA length and loaded into corresponding Argonaute proteins. This is primary piRNA biogenesis. It generates piRNAs bound to PIWI proteins Piwi and Aubergine (Aub). Piwi enters nucleus, identifies transposon transcripts and directs heterochromatin formation at these transposon generating loci (Le Thomas et al., 2013; Sienski et al., 2012). Aub bound piRNAs cleave the transposon transcript. These cleaved transcript is loaded into a PIWI protein Argonaute3 (Ago3) and processed into piRNA length. This Ago3 bound piRNA further cleaves the piRNA precursor transcript, which can be loaded into Aub. This reciprocal cleavage of transposon and piRNA precursor is called Ping-Pong amplification cycle or secondary piRNA biogenesis and increases the piRNA abundance especially for active transposons (Brennecke et al., 2007; Gunawardane et al., 2007). DEAD box helicase protein Vasa regulates this Ping-Pong cycle by clamping onto RNA and orchestrating its transfer between PIWI partners (Xiol et al., 2014). After cleavage by the PIWI proteins, Zucchini (Zuc) mediates phased piRNA production from the cleaved RNAs (Han et al., 2015; Mohn et al., 2015). 4 of the proteins involved in the piRNA biogenesis, Piwi, Aub, Vasa and Zuc are expressed by genes within the introgression lines. Mutations in these genes placed in trans with

introgression lines allowed us to do species swap. We observed that Vasa has a prominent defect on fertility and expression of piRNA target, indicating that Vasa function is diverged between species.

Results

Using introgression lines to swap piRNA pathway genes between species

Introgression line replaces part of *D. melanogaster* 2L chromosome with *D. simulans* 2L chromosome. There are two introgressed segments and were mapped from 21A1 to 22D1–23A2 (Int(2L)D) and 30F1-31E7 to 35D7-36A14 (Int(2L)S) by using deficiencies for screening (Sawamura et al., 2000). To precisely map the introgressed regions, we prepared genomic libraries from ovaries of homozygous introgression lines, sequenced them by Illumina and mapped the reads onto *D. melanogaster* genome. The reads obtained from regions from *D. simulans* genome would show increased SNP frequency and reduced number of reads mapping onto *D. melanogaster* genome. By this comparison, we observed the break site for Int(2L)D to be from 21A1 to 22F1 and for Int(2L)S from 30A2 to 36A5 (Figure AI-1A). This introgression line is called as Int in further text.

Within Int(2L)S, we found 4 piRNA pathway genes: *piwi* (32C1), *aub* (32C1), *zuc* (33B5) and *vasa* (35C1) (Figure AI-1B). These genes express *D. simulans* proteins in *D. melanogaster*. For species swap, we used mutants in *D. melanogaster* to cross with Int lines, so that only *D. simulans* proteins are expressed (Figure AI-1C).

Figure AI-1

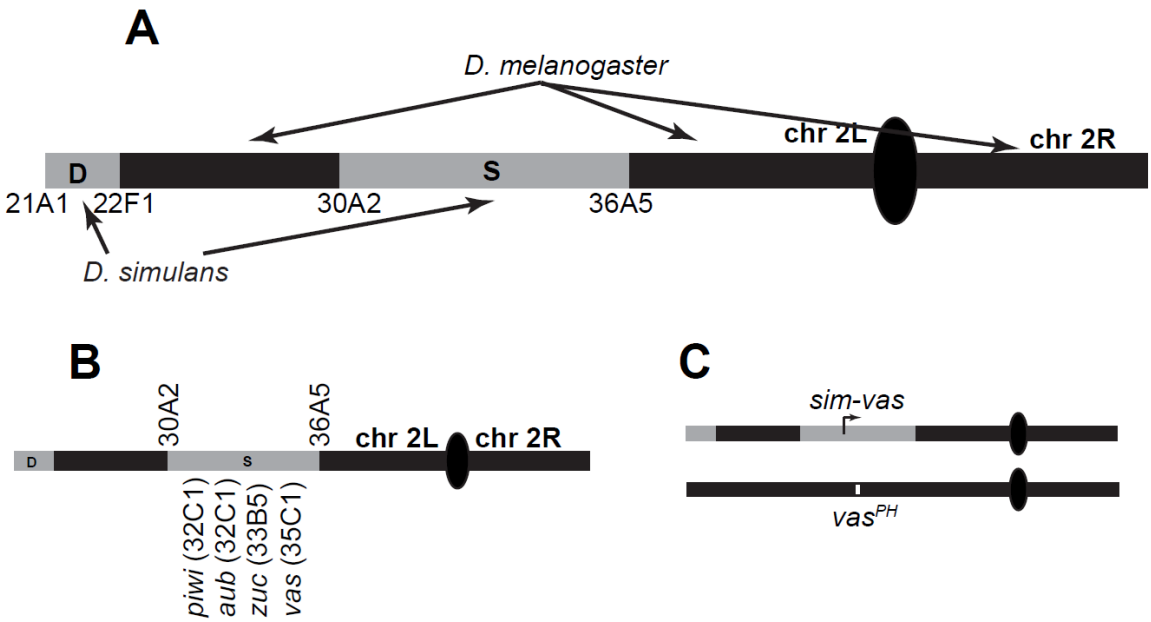

A
D. melanogaster
chr 2L
chr 2R
D
S
21A1
22F1
30A2
36A5
D. simulans
B
30A2
36A5
chr 2L
chr 2R
D
S
piwi (32C1)
aub (32C1)
zuc (33B5)
vas (35C1)
C
sim-vas
vasPH

Figure AI-1: Characterization of introgression lines

(A) 2nd chromosome is shown for introgression lines. *D. melanogaster* chromosome is shown in black and *D. simulans* in grey. Regions D and S are two different introgressions. The cytological map of the breakpoints observed from genomic sequencing is shown below the chromosome.

(B) The cytological location of piRNA pathway genes *piwi*, *aub*, *zuc* and *vas* is shown with respect to introgression line.

(C) Species swap of *vasa*: vas^{PH} mutation is kept in trans to Int. The only Vasa expressed in these flies is *D. simulans* Vasa.

Female fertility after species swap of piRNA proteins

For female fertility testing, balanced Int female virgins were collected and crossed with males of a balanced stock of a piRNA pathway mutant. From the F1 progeny, non-CyO flies were collected for fertility testing as they contain Int in trans with a piRNA pathway mutant. *piwi*^{1}/CyO and Int/CyO served as controls. 2-4 day old females were kept with *w*^{1} males on grape juice agar plates to collect eggs and assay fertility (Figure AI-2). Mutations in the piRNA pathway genes lead to D-V patterning defects in eggs, observed in the form of fused appendages (Klattenhoff et al., 2007). Int/*piwi* and Int/*zuc* did not affect the D-V patterning of eggs. Int/*aub* showed average of 8% eggs with fused appendages (Figure AI-2A), whereas Int/*vas* showed average of 27% eggs with fused appendages. Int/*piwi* has no effect on hatching of eggs (Figure AI-2B). Average of 89% and 92% eggs hatch for Int/*aub* and Int/*zuc* respectively. Thus, *aub* has a mild defect on both D-V patterning and fertility of files. For Int/*vas*, only average of 30% eggs hatch. This mild defect in fertility by *aub* swap is consistent with previous report by (Kelleher et al., 2012). Vasa species swap shows a distinct effect on female fertility. Thus, adaptive evolution of Vasa has led to functional divergence between *D. melanogaster* and *D. simulans*.

Figure AI-2

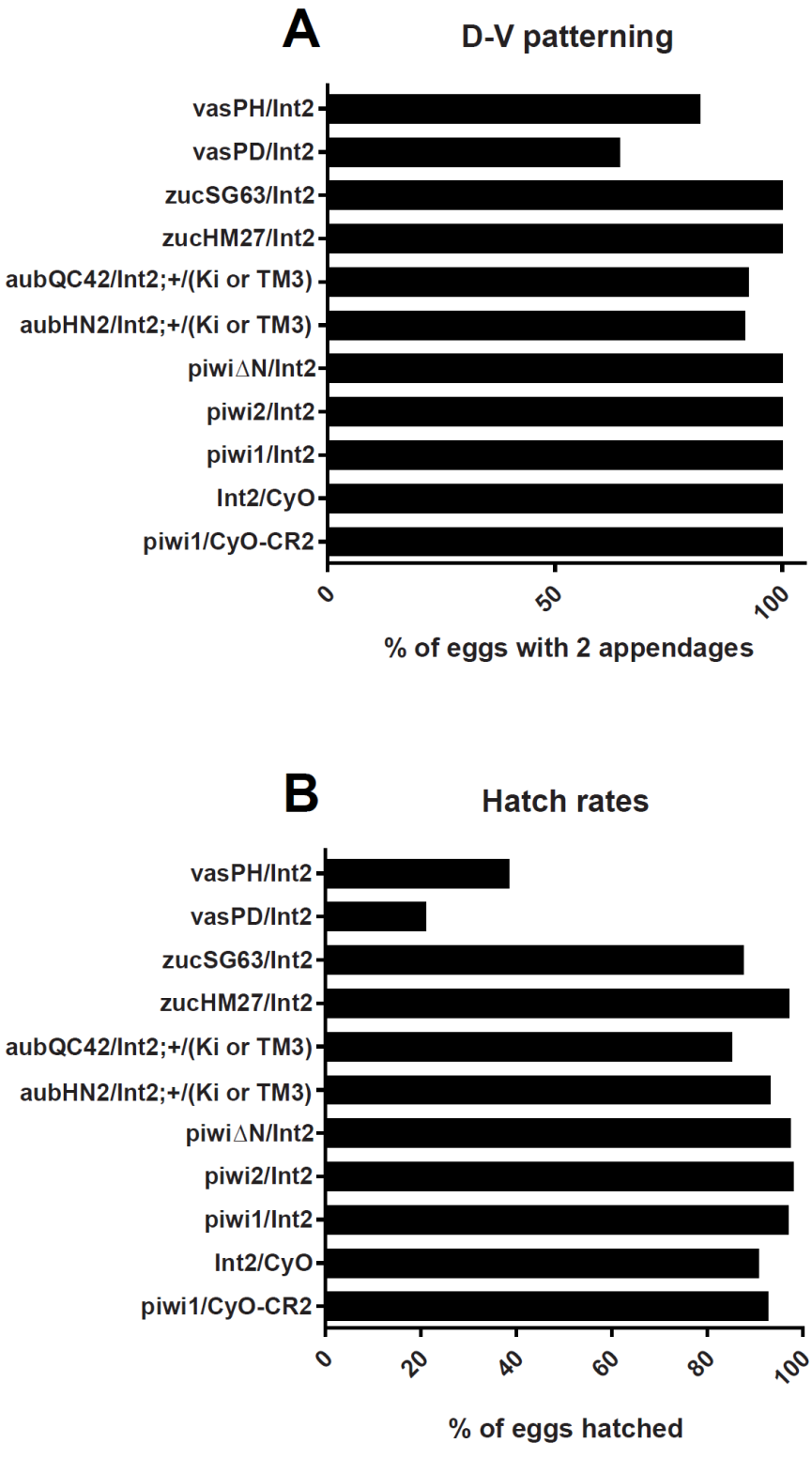

A
D-V patterning
vasPH/Int2
vasPD/Int2
zucSG63/Int2
zucHM27/Int2
aubQC42/Int2;+/(Ki or TM3)
aubHN2/Int2;+/(Ki or TM3)
piwiΔN/Int2
piwi2/Int2
piwi1/Int2
Int2/CyO
piwi1/CyO-CR2
0
50
100
% of eggs with 2 appendages
B
Hatch rates
vasPH/Int2
vasPD/Int2
zucSG63/Int2
zucHM27/Int2
aubQC42/Int2;+/(Ki or TM3)
aubHN2/Int2;+/(Ki or TM3)
piwiΔN/Int2
piwi2/Int2
piwi1/Int2
Int2/CyO
piwi1/CyO-CR2
0
20
40
60
80
100
% of eggs hatched

Figure AI-2: Female fertility

(A) D-V patterning of eggs laid by the females in shown in term of percentages of eggs with 2 appendages.

(B) Percentages of hatched eggs in different genotypes.

Male fertility after species swap of piRNA proteins

Males from piRNA pathway mutants show reduced fertility (Li et al., 2009). To check whether similar male fertility defects are observed after species swap, we checked the fertility of Int flies in trans with different piRNA pathway mutants. To assay male fertility, we crossed the males with OregonR virgin females and counted the progeny obtained from the cross. The controls *piwi*^{1}/CyO and Int/CyO produced average of 98 and 65 progeny respectively (Figure AI-3). Int/*piwi* and Int/*zuc* did not show a huge decrease in male fertility. For *aub*, two different alleles were used *aub*^{HN2} and *aub*^{QC42}. Int/*aub*^{HN2} and Int/*aub*^{QC42} males produced average of 83 and 25 individuals respectively. For *vasa*, Int/*vasa*^{PH} and Int/*vasa*^{D5} produced 46 and 18 individuals respectively. Int/*vasa*^{D5} shows significant decrease in the progeny when Int/CyO was used as control. Thus, Vasa may have a role in male fertility. However, this data is preliminary and it would be necessary to repeat the experiments with more biological replicates to conclusively test the role of piRNA proteins species swap as these experiments were done with n=3 each.

Figure AI-3

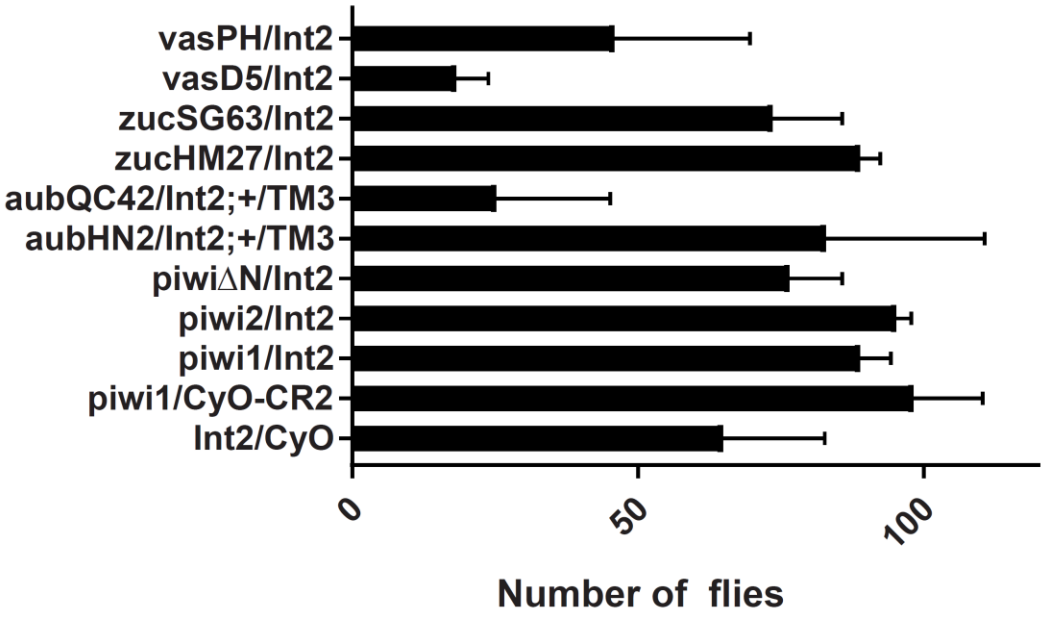

vasPH/Int2
vasD5/Int2
zucSG63/Int2
zucHM27/Int2
aubQC42/Int2;+/TM3
aubHN2/Int2;+/TM3
piwiΔN/Int2
piwi2/Int2
piwi1/Int2
piwi1/CyO-CR2
Int2/CyO
0
50
100
Number of flies

Figure AI-3: Male fertility

The number of flies obtained by crossing the denoted males with OregonR virgin females. Error bars show standard deviation.

Vasa species swap leads to expression of Stellate crystals

First piRNAs were discovered for *Stellate* locus (Aravin et al., 2001). *Suppressor of Stellate* piRNAs control expression of *Stellate* locus. In the absence of these piRNAs, *Stellate* locus is expressed which leads to formation of Stellate crystals in testes. Many piRNA pathway mutants lead to loss of *Su(Ste)* piRNAs and production of Stellate crystals in testes (Pane et al., 2007; Tomari et al., 2004). To test whether the species swap of piRNA pathway proteins leads to similar Stellate expression, we immuno-stained testes for Stellate (Figure AI-4A). Stellate crystals are not expressed in control, Int/+ due to functional piRNA pathway. No Stellate crystals were observed with *piwi*, *aub* and *zuc* swap. However, *vasa* swap leads to production of Stellate crystals. To check whether the effect is due to vasa, we expressed Vasa-GFP in the same background (Figure AI-4B). Expression of GFP tagged Vasa partially suppresses the Stellate crystal production. Vasa seems to have species specific function in Stellate silencing and thus, piRNA mediated silencing.

Figure AI-4

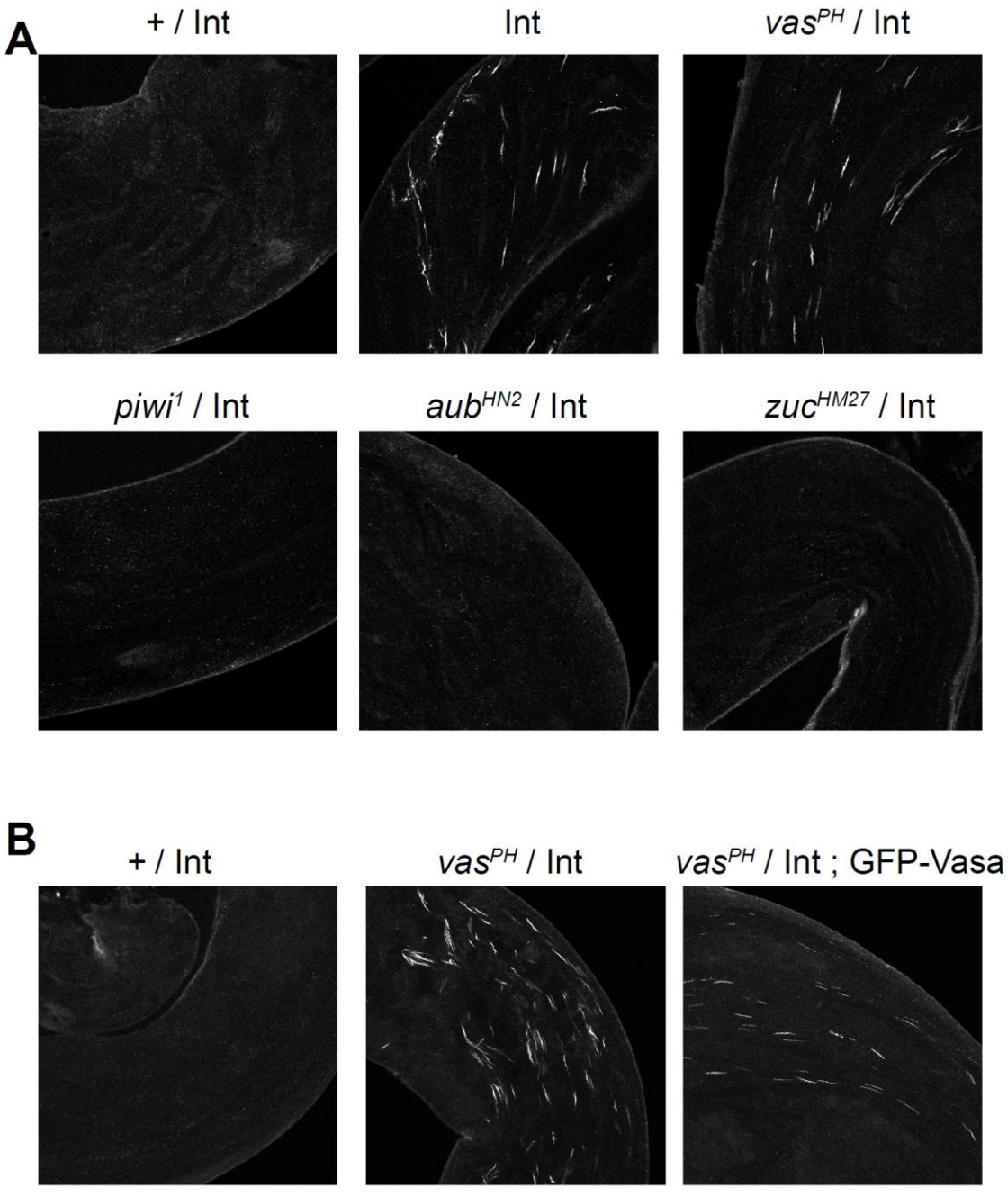

A
+ / Int
Int
vasPH / Int
piwi1 / Int
aubHN2 / Int
zucHM27 / Int
B
+ / Int
vasPH / Int
vasPH / Int ; GFP-Vasa

Figure AI-4: Stellate expression in testes

(A, B) Stellate immuno-staining in testes of indicated genotype males.

Discussion

Transposons behave as pathogens and occupy major fraction of eukaryotic genomes. piRNAs form an adaptive immune defense against these pathogenic transposons. The rapid evolution of the piRNA pathway genes is thought to be the result of host-pathogen arms race with transposons (Blumenstiel et al., 2016). To test whether the sequence divergence in the piRNA pathway proteins has any functional consequences, we swapped *vasa*, *aub*, *piwi* and *zuc* between *D. melanogaster* and *D. simulans* by means of introgression lines. Many studies have shown adaptive evolution in these genes when compared between different *Drosophila* species (Blumenstiel et al., 2016; Lee and Langley, 2012; Obbard et al., 2009; Simkin et al., 2013). Species swap of *piwi* and *zuc* show no effect on fertility and piRNA mediated silencing. Zuc mediated cleavage of RNAs leads to production of phased piRNA, which are loaded into Piwi and Piwi bound piRNAs mediate transcriptional silencing at transposon loci (Han et al., 2015; Mohn et al., 2015). Thus, even though both these proteins show evidence of adaptive evolution, they are not functionally diverged.

Vasa and Aub are involved in the Ping-Pong amplification cycle and thus post-transcriptional silencing of transposons (Brennecke et al., 2007; Gunawardane et al., 2007; Xiol et al., 2014). *aub* from *D. simulans* (*sim-aub*) has been shown to partially complement *aub* function in *D. melanogaster* (Kelleher et al., 2012). We also observe a mild defect in the fertility with *aub* swap. However, *aub* swap does not lead to Stellate crystal formation. Thus *aub* may have led to divergence in function at least in female germline. Vasa swap shows reduced fertility in both males and females. It also shows

Stellate crystal formation in testes, indicative of the loss of *Su(Ste)* piRNAs. Thus, Vasa seems to have species specific function in both male and female germlines. Ping-Pong amplification cycle leads to amplification of piRNAs especially for active transposons and makes the piRNA immune system adaptive. As inhibition of this cycle would be advantageous for all the active transposons, it can be a likely target in the host-pathogen arms race. The transposon encoded proteins are manufactured in the cytoplasm and need to be transported in the nucleus to carry out transposition. As both Vasa, Aub and other machinery involved in Ping-Pong cycle is localized around nuclear envelope in the form of nuage (Li et al., 2009; Malone et al., 2009; Xiol et al., 2014), nuclear envelope can be the site of arms race between the Ping-Pong machinery and transposons.

Introgression line gives us a simple tool to study the effects of sequence divergence in the piRNA pathway. However, one caveat of the experiment is that all these *D. simulans* piRNA pathway proteins and other proteins within introgression interval are expressed while carrying out the species swap with introgression lines. In Int/*vas*, the only Vasa protein expressed in *D. melanogaster* is from *D. simulans*. However, for the proteins in the introgression interval, the proteins from both the species are expressed. Thus, it should be noted that the sterility phenotype observed with Vasa swap may be compounded by expression of other proteins from *D. simulans*.

When we swapped Rhino, Deadlock and Cutoff between these two species, all of them fail to completely rescue fertility. By comparison, the effects of Aub and Vasa swap are modest. It is possible that the RDC complex which is responsible for the beginnings in the primary piRNA biogenesis, may be the target of many transposon encoded

inhibitors and under more selective pressure, whereas the Aub and Vasa which are involved in the secondary piRNA biogenesis may be the targets of fewer transposon targets. Nonetheless our experiments show a species specific function for Vasa, suggesting that the Ping-Pong amplification cycle may be a target of host-pathogen arms race with transposons.

Experimental Procedures:

Experimental model and subject details

All flies were kept at 25^{O}C on cornmeal medium during crosses and fertility testing. Following stocks were used from the lab stocks: *vas*^{PH}/CyO, *vas*^{D5}/CyO, *vas*^{PH}/CyO; *vasP*-GFP-Vasa, *w*^{1}; *aub*^{HN2}/CyO; *Ki*/TM3, *w*^{1}; *aubQC*^{32}/CyO; *Ki*/TM3, *piwi*^{1}/CyO, *piwi*^{2}/CyO, *piwi*$^{\Delta N}$/CyO, *zuc*^{HM27}/CyO, *zuc*^{SG63}/CyO. Int lines were kind gift from Dr. Kyoichi Sawamura (University of Tsukuba, Japan).

Female Fertility testing

2-4 day old flies were kept on grape juice agar plates for 1 day. After removal of flies, the eggs were checked for fused appendages and the hatching was measured after 2 days.

Male fertility testing

5 OregonR virgin females were mated with 2 males of given genotype in a vial. Parents were removed after 5 days. The number of individuals in vials were counted 13 days after setting up the cross. The crosses were kept at 25^{o}C. Three biological replicates were done for each genotype.

Immuno-staining

Immuno-staining and image analysis was done as described in (McKim et al., 2009; Zhang et al., 2012a). In brief, 2-4 day old testes were fixed with 4% formaldehyde, washed, incubated overnight with primary antibody, washed, incubated with secondary

antibody with fluorophore overnight and mounted on slide. Rabbit anti-Stellate antibody was used for Stellate staining (Klattenhoff et al., 2007) at 1:1000 dilution.

Acknowledgements

I would like to thank Nadine Schultz for helped me with the egg counting and Dr. Kyoichi Sawamura (University of Tsukuba, Japan) for generously sharing Int line with us.

Appendix II

Non-Mendelian transposon inheritance in *Drosophila melanogaster*

Preface

Nadine Schultz for helped me with the fly crosses. Tianxiong Yu (Bear) and Jiali Zhuang helped me with the bio-informatics analysis under the guidance of Zhiping Weng.

Summary

Transposons are major genome constituents that can produce deleterious mutations, generate beneficial genetic diversity, and are the source of significant population variation. To directly assay inheritance of fixed and de novo transposon insertions, we used *Drosophila* P-M hybrid dysgenesis to transiently activate germline transposition, and whole genome sequencing and qPCR to follow inheritance of both de novo insertions and fixed insertions present in the parental strains. For the vast majority of insertions, backcrossing to parental strains led to expected reductions in population frequency. However, a subset of both fixed parental and de novo insertions defied Mendelian predictions and increased in frequency through multiple generations. This could result from recurrent transposition into an insertion hot spot, or transposon linked gene conversion or meiotic drive. The SNPs that were directly linked to the original insertion by paired end sequencing reads remained with the insertions through 5 generations. Thus, we think that the observed non-Mendelian inheritance of transposons maybe the result of gene conversion or meiotic drive.

Introduction

Since the initial discovery of transposons by Barbara McClintock (Fedoroff, 2012; McClintock, 1950), many transposable elements (TEs) have been identified in essentially all organisms (Bao et al., 2015; Canapa et al., 2015). For humans, almost half of the genome encodes for these transposons (Biemont and Vieira, 2006). The transposons can jump to various locations in the genome and cause recombination between different repeats leading to deleterious effects in organisms (Hedges and Deininger, 2007). Insertions within genes can disrupt them in various ways. Insertions into exons can produce truncated proteins, promoter or enhancer insertions can lead to change in the expression pattern, intron insertions can change the splicing patterns, and insertions in UTRs can alter the regulatory sequences (Ayarpadikannan and Kim, 2014). Recombination between transposon copies at different locations can cause genomic rearrangements, leading to inversions, deletions, duplications and translocations. Thus transposons can be a major source of genomic instability and are found to be the cause of many human diseases (Ayarpadikannan and Kim, 2014; Belancio et al., 2008). Despite having such negative consequences, most transposon insertions are neutral and some have been found to have beneficial role. Thus, transposons play an important role in genome evolution and population variation.

Transposons can spread by both vertical and horizontal transmission (Gilbert and Feschotte, 2018). Just like genes in the genome, transposons can be vertically transmitted from one generation to the next. Many transposons are proposed to be active in the germline, so that the newly generated copies can be transmitted to the next generation

(Haig, 2016; Tiwari et al., 2017). To prevent this transposon selfish behavior, many animals produce small RNAs called PIWI interacting RNAs (piRNAs) which control these transposons especially in the germline (Ghildiyal and Zamore, 2009). Comparison of transposons between species shows very less divergence, indicative of horizontal transfer of TEs (Bartolome et al., 2009; Lerat et al., 2011). Various cases have found horizontal transfer of transposons. In fishes and frogs, horizontal transfer is observed for Tc1 like transposon (Leaver, 2001). P-element has been shown to have jumped from *Drosophila willistoni* into *Drosophila melanogaster* (*D. melanogaster*) (Engels, 1992). The sequence comparison of P-element in these two species shows only one nucleotide difference (Daniels et al., 1990). It has swept through wild populations of *D. melanogaster* in a span of few decades (Engels, 1992) and currently is sweeping through *Drosophila simulans* populations worldwide (Kofler et al., 2015). Thus, transposons can jump between species and a new transposon introduced into a species can rapidly spread through populations. Thus transposons behave like a genomic pathogens as they have deleterious effects on host and can rapidly spread between individuals.

As transposons shape the genomic landscape, we wanted to study how the genomes adapt after the transposon mobilization. To test this, we wanted a system where we can activate transposons and follow their inheritance for multiple generations. Mutations in piRNA pathway lead to transposon mobilization, however they are not useful for the study of inheritance, because they are sterile (Klattenhoff et al., 2007; Zhang et al., 2012a). Hybrid dysgenesis also leads to transposon activation (Khurana et al., 2011) and offers a unique system to address this question. Crosses between *D.*

melanogaster lab strain females and *D. melanogaster* wild strain males produce progeny which is sterile. This is due to activation of transposon P-element in the progeny because of absence of maternally deposited P-element piRNAs. This sterility syndrome is called hybrid dysgenesis (Kelleher, 2016). Along with P-element, many other TEs also become active in the dysgenic hybrids (Khurana et al., 2011). In the dysgenic cross of lab strain w^1 and wild strain Harwich, the fertility of hybrids improves with age. This gives us perfect tool to study the inheritance of activated transposons. By using this hybrid dysgenesis as our model system, we back crossed the dysgenic hybrids with either of the parents and assayed transposon activity by sequencing genomes from the parents, dysgenic hybrids and their progeny. We observed large scale mobilization of transposons. Many of these transposons are lost after multiple generations of outcrossing. However, frequency of some of these active transposons increases after successive outcrossing, defying Mendelian inheritance. As the dysgenic hybrids contain a mixture of DNA from two parents, we can use the single nucleotide polymorphism (SNP) information to follow the inheritance of DNA surrounding the transposon. The SNPs identified with the original insertions remained linked to these TEs even after multiple generations of back-crossing. This shows how transposons can selfishly increase their copy numbers.

Results

Most transposons show Mendelian inheritance

Wild strains of *D. melanogaster* contain P-elements whereas the lab strains are devoid of it. As maternal deposition of piRNAs is important for suppressing transposons in the progeny, the crosses between lab *D. melanogaster* females and wild strains from the same species lead to P-element activation and sterility, as lab strain females cannot deposit P-element piRNAs in the progeny. However, the reciprocal crosses do not show increased P-element activity and are fertile (Khurana et al., 2011). In the dysgenic cross of lab strain w^1 and wild strain Harwich, the F1 dysgenic females are sterile and show increased transposon activity not only for P-element, but also for many other transposons. By 2-3 weeks, the fertility improves and the transposons are silenced. We used these hybrid dysgenic females to follow the inheritance of activated TEs. To study the inheritance of activated transposons, we collected virgin dysgenic hybrid females (wH1) and crossed them to males from either of the parents, w^1 or Har (Figure AII-1). The successive generations obtained after crossing to Har or w^1 are termed as wHH or wHw. After the dysgenic hybrids regain their fertility, eggs hatch to produce the second generation progeny wHH2 and wHw2. Again virgin females are collected from these flies and crossed with either Har or w^1 males to obtain wHH3 and wHw3 respectively. Such crosses were continued for multiple generations. In order to assay transposon insertions, we isolated genomic DNA from ovaries, prepared paired end genomic libraries and sequenced by Illumina platform. Ovaries were chosen as the tissue for assay because: 1) hybrid dysgenesis sterility syndrome is a germline defect, 2) germline tissue

is responsible for inheritance of genetic material to the next generation and 3) hybrid dysgenesis syndrome is much severe in females compared to males (Engels and Preston, 1979).

Figure AII-1

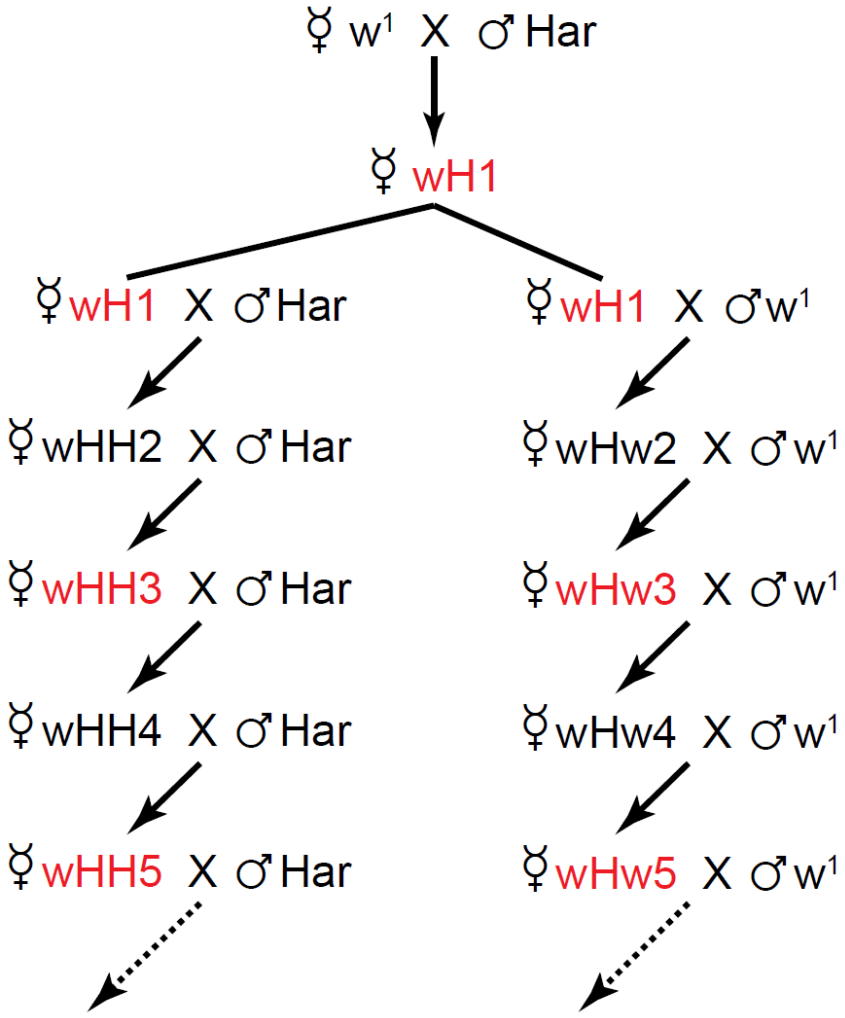

♀ w¹ X ♂ Har
♀ wH1
♀ wH1 X ♂ Har
♀ wH1 X ♂ w¹
♀ wHH2 X ♂ Har
♀ wHw2 X ♂ w¹
♀ wHH3 X ♂ Har
♀ wHw3 X ♂ w¹
♀ wHH4 X ♂ Har
♀ wHw4 X ♂ w¹
♀ wHH5 X ♂ Har
♀ wHw5 X ♂ w¹

Figure AII-1: Experimental setup for fly crosses.

Lab strain w^1 virgin females were mated with wild strain Har males. The resulting dysgenic progeny is called wH1. Female wH1 virgins were back-crossed with either Har or w^1 to get progeny wHH2 and wHw2 respectively. Such crossing was continued with either Har or w^1 males. The genomic DNA was sequenced from ovaries of Har and w^1 parents and the progeny shown in red.

The transposon insertions were identified by TEMP method which was designed to identify transposon insertions based on paired end read information (Zhuang et al., 2014). First we identified the transposon insertions in all the sequenced genotypes. The abundance of a particular insertion is characterized by "penetrance" which is defined as sum of the reads confirming the insertion divided by the sum of these reads and the reads spanning the insertion site. A transposon insertion which is fixed a genotype would have penetrance of 1. The transposons which were present in w^1 parent, but not in Har parent or wH1 were called as w^1 specific TEs. Similarly Har parent specific TEs were identified. The TEs which were not present in either of the parents, but were newly generated in wH1 were called as novel insertions. Figure AII-2 shows inheritance of w^1 specific insertions in wH1 and in subsequent outcrossed generations. The scatterplots depict penetrance values of w^1 specific TEs on X-axis vs. the penetrance of same TEs in respective genotypes on Y-axis. In wH1, the penetrance of most of these w^1 insertions is halved, consistent with equal amounts of DNA from *w1* and Har. The fixed insertions in w^1 have penetrance about 1 and in wH1, they cluster just above 0.5. Thus many fixed insertions shows biased inheritance. After outcrossing to w^1 for multiple generations, the penetrance of these TEs increase and by 5^{th} generation, the penetrance values resemble w^1 parent. Reverse trend is observed when outcrossed to Har. The penetrance of w^1 insertions decrease with each successive generation of outcrossing and by 5^{th} generation, the penetrance of most of the TEs is about 0. This is consistent with Mendelian inheritance of transposons. Similar trend is also observed for Har specific parental insertions (Figure AII-3). Many fixed insertions, with penetrance 1, show penetrance

values slightly greater than 0.5 in wH1. But, most of the Har insertions are lost upon out-crossing to w^1 and increase in frequency after outcrossed with Har. Many new transposon insertions generated in wH1 have very low penetrance (Figure AII-4). Most of these TEs are lost with successive generations of outcrossing. This increase or decrease in transposon penetrance after successive generations of outcrossing shows that they follow the Mendel's laws like genes in the genome.

Figure AII-2

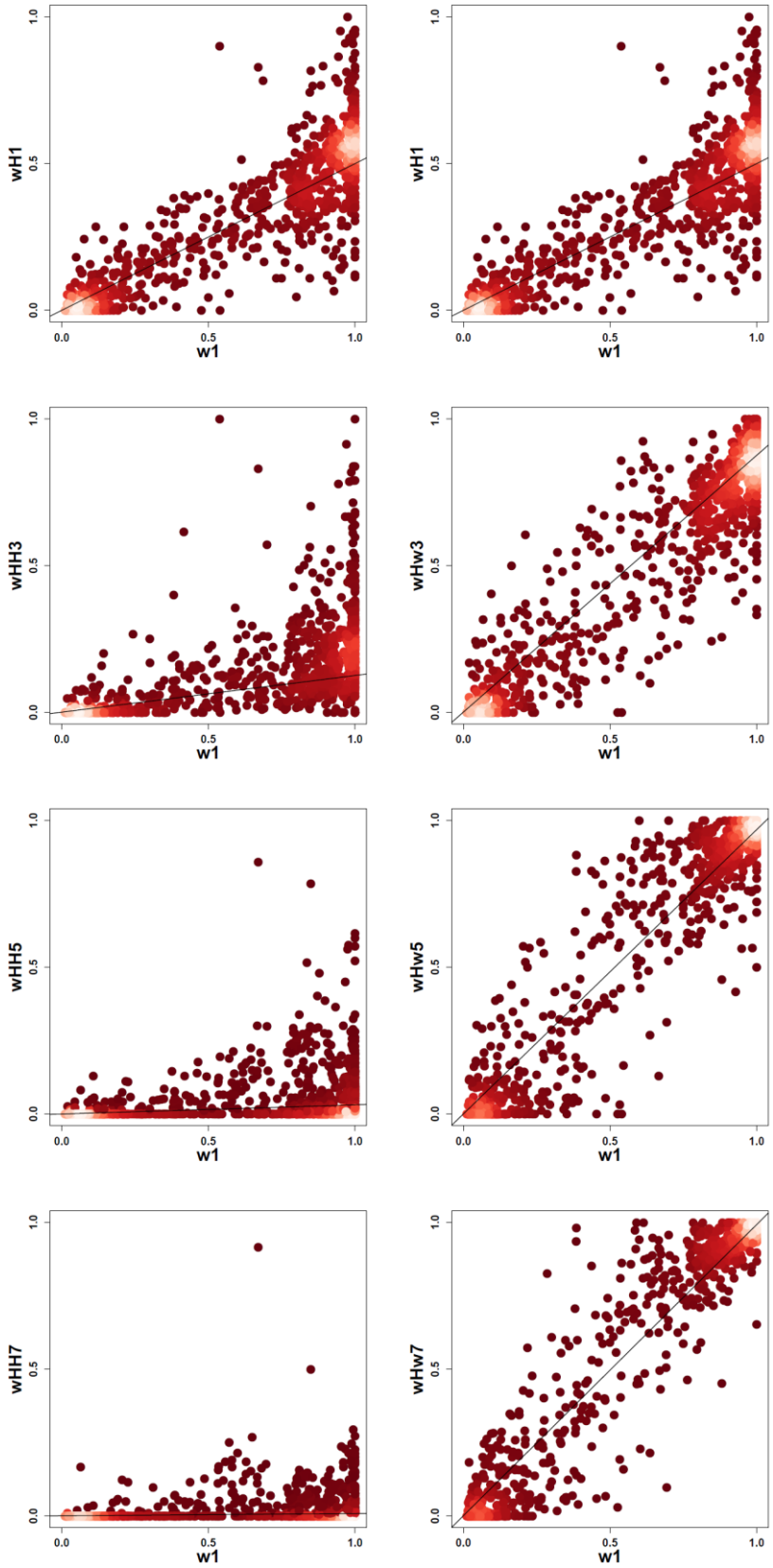

wH1
w1
wH1
w1
wHH3
w1
wHw3
w1
wHH5
w1
wHw5
w1
wHH7
w1
wHw7
w1
0.0
0.5
1.0

Figure AII-2: Most *w*^{1} parental transposons show Mendelian inheritance.

Scatterplots show penetrance of *w*^{1} specific parental transposon insertions on X-axes. Y-axes show the penetrance of these *w*^{1} parental insertions in the respective genotypes. Outcrossing to Har leads to progressive loss of *w*^{1} insertions, whereas outcrossing to *w*^{1} increases the penetrance of *w*^{1} insertions. In wH1, many fixed *w*^{1} insertions show biased inheritance. Line represents values expected by Mendelian inheritance. The points are merged based on density.

Figure AII-3

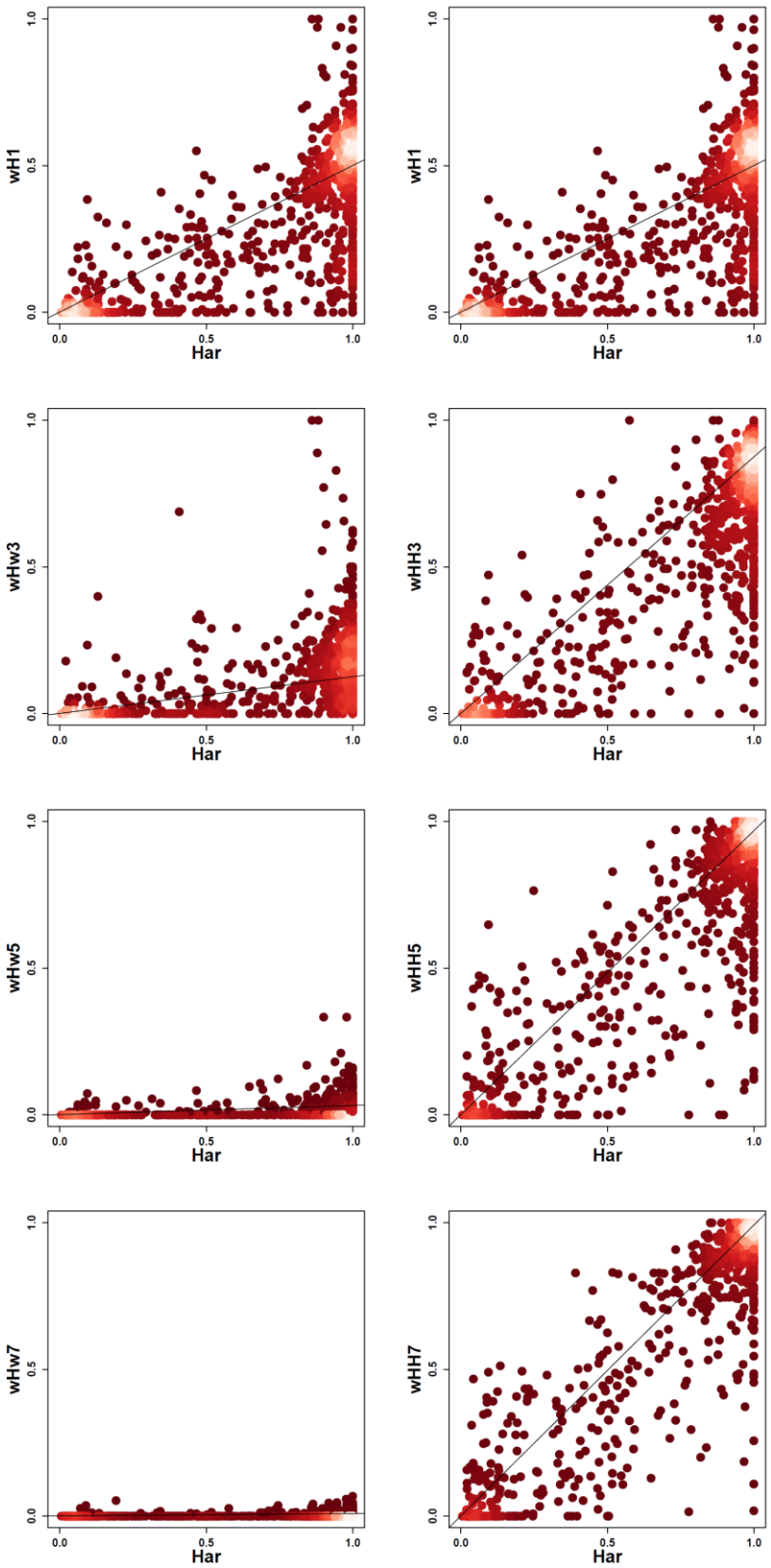

wH1
Har
wH1
Har
wHw3
Har
wHH3
Har
wHw5
Har
wHH5
Har
wHw7
Har
wHH7
Har

Figure AII-3: Most Har parental transposons show Mendelian inheritance.

Scatterplots show penetrance of Har specific parental transposon insertions on X-axes. Y-axes show the penetrance of these Har parental insertions in the respective genotypes. Outcrossing to w^1 leads to progressive loss of Har insertions, whereas outcrossing to Har increases the penetrance of Har insertions. In wH1, many fixed Har insertions show biased inheritance. Line represents values expected by Mendelian inheritance. The points are merged based on density.

Figure AII-4

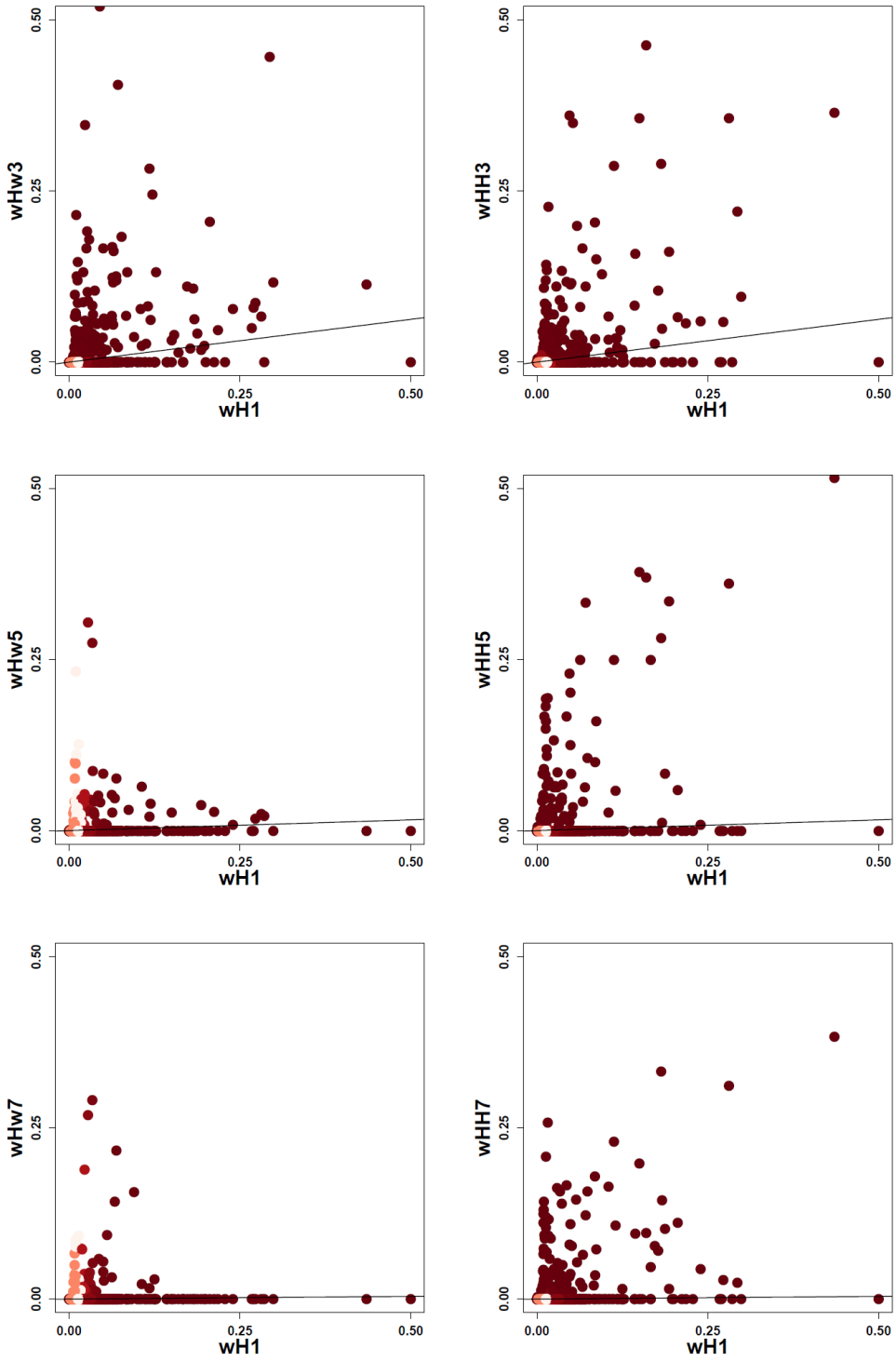


Figure AII-4: Most novel transposons show Mendelian inheritance.

Scatterplots show penetrance of novel insertions generated in wH1 on X-axes. Most insertions have very low penetrance values. Y-axes show the penetrance of these new insertions in the respective genotypes. Outcrossing to either parents leads to progressive loss of these TEs. Line shows x=y. Line represents values expected by Mendelian inheritance. The points are merged based on density.

Few transposons show non-Mendelian inheritance

Although most transposons follow Mendel's law, a few transposons show non-Mendelian inheritance and increase in penetrance after successive outcrossing (Figure AII-5). Such instances are observed for both the parental and newly generated insertions. Shown in Figure AII-5C is a new P-element insertion generated in wH1. We could only detect one read confirming this transposon. When outcrossed to Har, we observe an increase in the number of P-element confirming reads. Interestingly, this increase is only observed while outcrossing into Har and not w^1. We further confirmed this penetrance by qPCR by using primers for both 5' and 3' ends of P-element (Figure AII-5E). Out of 2512 novel TEs identified in F1 dysgenic hybrids, 31 TEs were retained in wHH5 and wHH7, with reads mapping to both the ends of transposons. This increase in transposon frequency compared to expected frequency defies Mendel's laws.

Similar behavior was also observed for a few fixed parental insertions. Figure AII-5D shows Jockey insertion in the w^1 genome. Outcrossing to w^1 increases frequency of this insertion. We could detect the reads supporting this insertion for several generations even after outcrossing to Har. For parental TEs, we focused on TEs which were confirmed by reads mapping to both the ends of transposons. We identified 931 such TE insertions in w^1. Out of these, 189 TEs were retained in Har after 7 generations of outcrossing. On the contrary, out of 921 Har insertions, only 23 were identified in w^1 after 7 generations of outcrossing. Thus both fixed and novel insertions can show non-Mendelian inheritance.

Figure AII-5

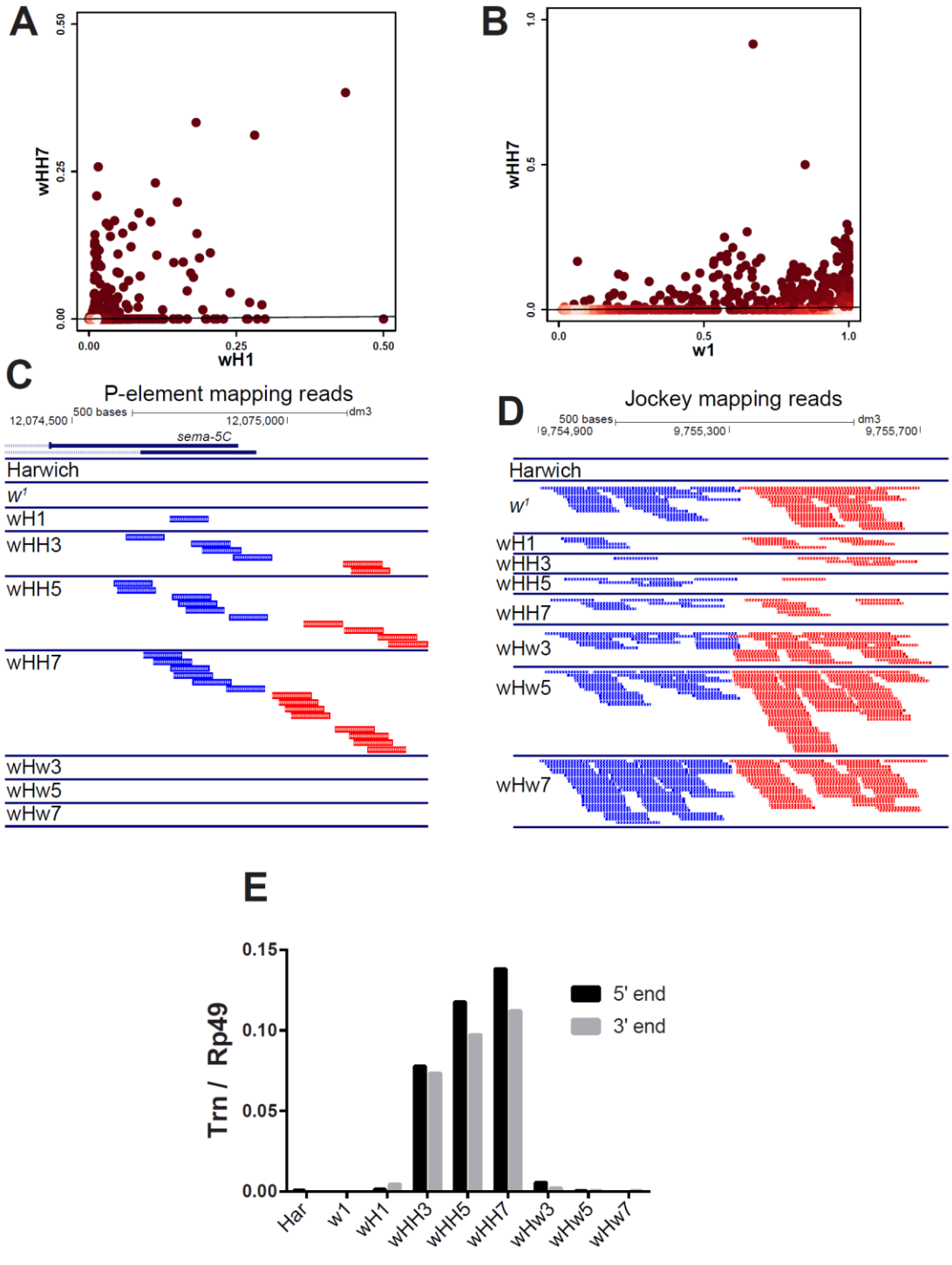

A
wHH7
wH1
0.00
0.25
0.50
B
wHH7
w1
0.0
0.5
1.0
C
P-element mapping reads
12,074,500
500 bases
12,075,000
dm3
sema-5C
Harwich
w1
wH1
wHH3
wHH5
wHH7
wHw3
wHw5
wHw7
D
Jockey mapping reads
9,754,900
500 bases
9,755,300
dm3
9,755,700
Harwich
w1
wH1
wHH3
wHH5
wHH7
wHw3
wHw5
wHw7
E
Trn / Rp49
0.15
0.10
0.05
0.00
5' end
3' end
Har
w1
wH1
wHH3
wHH5
wHH7
wHw3
wHw5
wHw7

Figure AII-5: Some TEs show non-Mendelian inheritance.

(A) Scatterplot showing penetrance of new transposon insertions in wH1 on X-axis vs. the penetrance of these transposon insertions in wHH7 on Y-axis. (B) Scatterplot showing penetrance of w^1 specific parental insertions on X-axis vs. the penetrance of these transposon insertions in wHH7 on Y-axis. (C) Discordant reads mapping to a new P-element insertion in the promoter of *sema-5C* gene in the mentioned genotypes. Blue and red represent discordant reads mapping to opposite ends of P-element. (D) Discordant reads for w^1 specific Jockey transposon in all the generations. Blue and red represent discordant reads mapping to opposite ends of Jockey. (E) qPCR analysis comparing the copy numbers of this *sema-5C* P-element to Rp49 in respective genotypes. qPCR was done for both 5' end (black bars) and 3' ends (grey bars) of the P-element.

Transposition into a hotspot is not the mechanism behind non-Mendelian inheritance

What could be the mechanism behind this non-Mendelian inheritance? One possibility is recurrent insertions into the same site. Some transposons have insertion site preferences (Linheiro and Bergman, 2012). P-elements tend to insert in promoters of genes (Spradling et al., 1995). The sites which show recurrent transposition are termed as transposition hotspots. The observed penetrance increase can be due to recurrent insertions in the transposition hotspots. This scenario is shown for a newly generated transposon (Figure AII-6A) and a parental transposon (Figure AII-6C). The Figure AII-6 shows the pools of DNA strands for Har and w^1 genotype. wH1 contains equal proportion from both the parents. After outcrossing to Har, most of the w^1 DNA would be lost. Even though the initial insertion(s) was in w^1 DNA (Figures AII-6A and 4.6C), transposition hotspot leads to an increase in the frequency of this transposon in the Har DNA. The identity of the surrounding DNA can be assayed by the adjoining SNPs particular to either w^1 or Har DNA. Another way this non-Mendelian inheritance could occur is by gene conversion (Figures AII-6B and 4.6D). It occurs when one strand of DNA copies itself into another strand. In case of transposons, this can be observed by checking SNPs linked to transposons. In case of a new transposon which is inserted into w^1 DNA, gene conversion would lead to retention of this TE in spite of being outcrossed with Har for several generations (Figure AII-6B). The linked SNPs show that the adjoining DNA from w^1 is also copied during outcrossing with Har, suggesting gene conversion of the site harboring a transposon. Similar observation can be made for a parental insertion (Figure

AII-6D) where a SNP linked w^1 specific transposon would confirm gene conversion when it shows non-Mendelian inheritance along with the transposon insertion. One more mechanism could be meiotic drive, which leads to preferential inheritance of one chromosome over the other chromosomes, during the process of meiosis. In this case, a big part of the chromosome(s) would show non-Mendelian inheritance (Figure AII-6E). In F1, both Har and w^1 chromosomes are in equal proportion. After 7 generations of out-crossing to Har, the w^1 chromosome should have been lost. If it retained after 7 generations, non-Mendelian inheritance of SNPs would be observed for a big portion of chromosome(s). Thus, the SNPs linked to transposons can reveal the identity of the surrounding DNA and indicate the mechanisms behind the non-Mendelian inheritance. Another way to rule out first possibility would be to precisely map the transposon insertion sites. Recurrent transposition into a hotspot would differ in the exact base position and gene converting insertion would show the same base position in different generations. However, even though the newly generated transposons and the parental insertions retained after outcrossing to a different genome show increased penetrance than expected, very few transposon mapping reads are obtained crossing the precise junctions at the insertion sites. Thus, to find the mechanism, we analyzed the SNPs linked to transposons and in the entire genome.

Figure AII-6

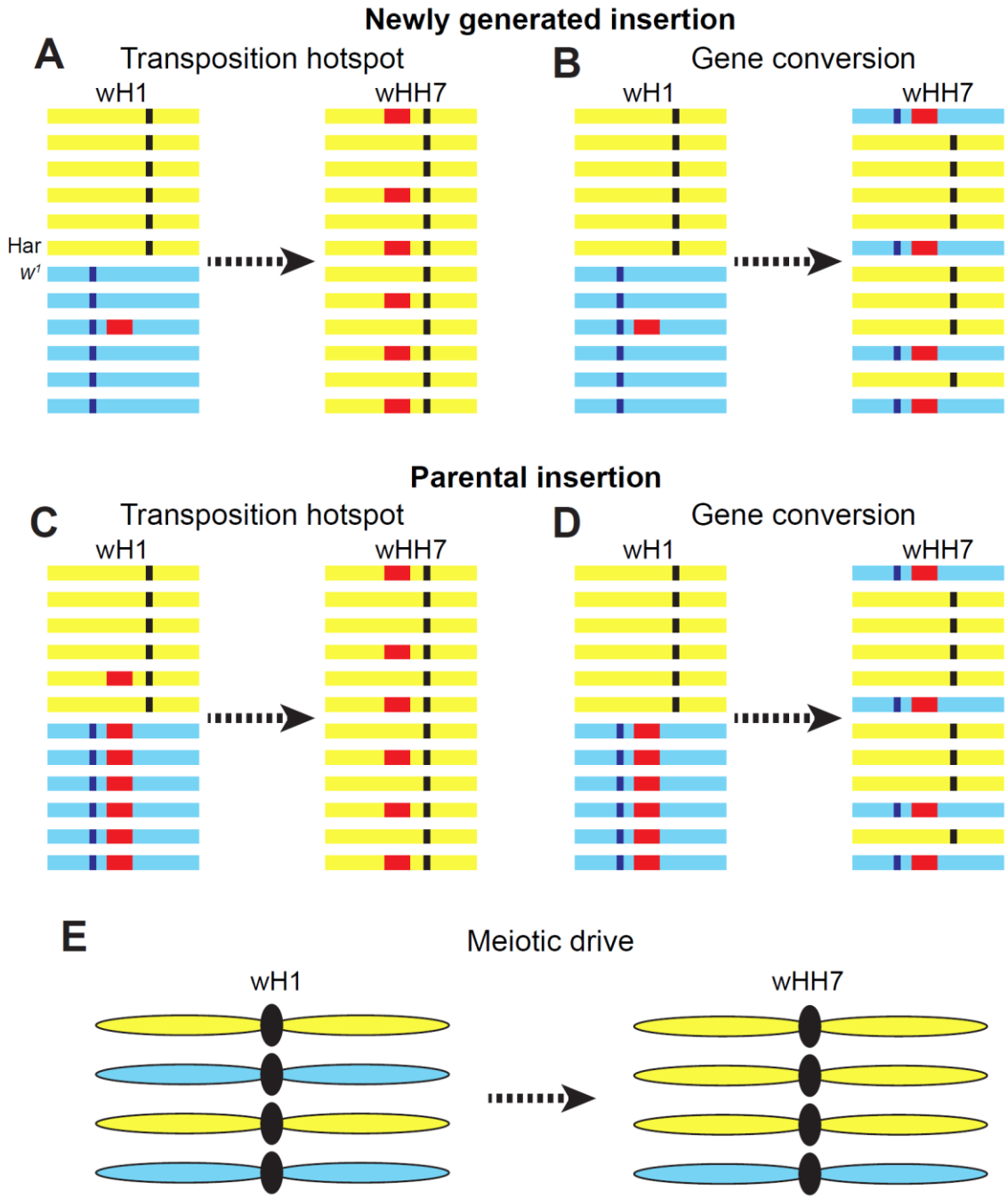

Newly generated insertion
A
Transposition hotspot
wH1
wHH7
Har
w1
B
Gene conversion
wH1
wHH7
Parental insertion
C
Transposition hotspot
wH1
wHH7
D
Gene conversion
wH1
wHH7
E
Meiotic drive
wH1
wHH7

Figure AII-6: Possible mechanisms for non-Mendelian inheritance

Non-Mendelian inheritance can be observed either by recurrent transposition in a transposition hotspot or by gene conversion or by meiotic drive. First two scenarios are shown for new transposon in F1 (A, B) and for a transposon fixed in w^1 parent (C, D). Har DNA is shown in yellow and w^1 in sky blue. wH1 contains equal proportion of Har and w^1 DNA. The SNPs distinguishing the DNA based on origin are shown by black color for Har and blue color for w^1. If non-Mendelian inheritance is due to recurrent transposition in the transposon hotspot, then all the SNPs linked to TE will be from Har in wHH7 (A, C). Whereas if non-Mendelian inheritance is due to gene conversion, then w^1 SNPs will be linked to the transposon in wHH7 (B, D). Meiotic drive can lead to retaining of w^1 chromosome in spite of outcrossing for 7 generations (E).

Figure AII-7 shows a new Springer insertion in wH1 and its non-Mendelian inheritance after outcrossing to Har. This insertion is lost after outcrossing to w^1. There are multiple Har mapping SNPs which can allow us to distinguish between Har and w^1. The only one Springer mapping read observed in wH1 shows absence of Har mapping SNPs, indicating that insertion occurred in w^1 DNA. Analysis of SNPs in springer mapping reads still shows absence of Har SNPs after multiple generations of outcrossing. Thus this springer insertion seems to carry the surrounding w^1 DNA after outcrossing to Har. This shows that non-Mendelian inheritance is not due to recurrent transposition.

Such case is also observed for parental TEs. Figure AII-8 shows inheritance of w^1 specific 412 insertion in multiple generations. The SNPs flanking the TE are present only in the Har genome and the surrounding w^1 DNA matches with the reference genome. In F1, all the discordant reads mapping to this transposon lack the Har SNPs. The same insertion retained in future generations is also devoid of the Har SNPs, indicating transposition into a hotspot cannot explain non-Mendelian inheritance of the transposon and surrounding DNA.

When we looked the SNP frequency at multiple generations, we observed non-Mendelian inheritance surrounding centromere of chromosome 3L. Figure AII-9 shows w^1 specific SNP frequency in different genotypes. w^1 SNP frequency is slightly higher in F1. Upon outcrossing to w^1, the w^1 SNP frequency increases according to Mendelian

inheritance. However, when outcrossed to Har, a big portion of chr3L next to centromere shows non-Mendelian inheritance. Such an increase in indicative of meiotic drive.

Figure AII-7

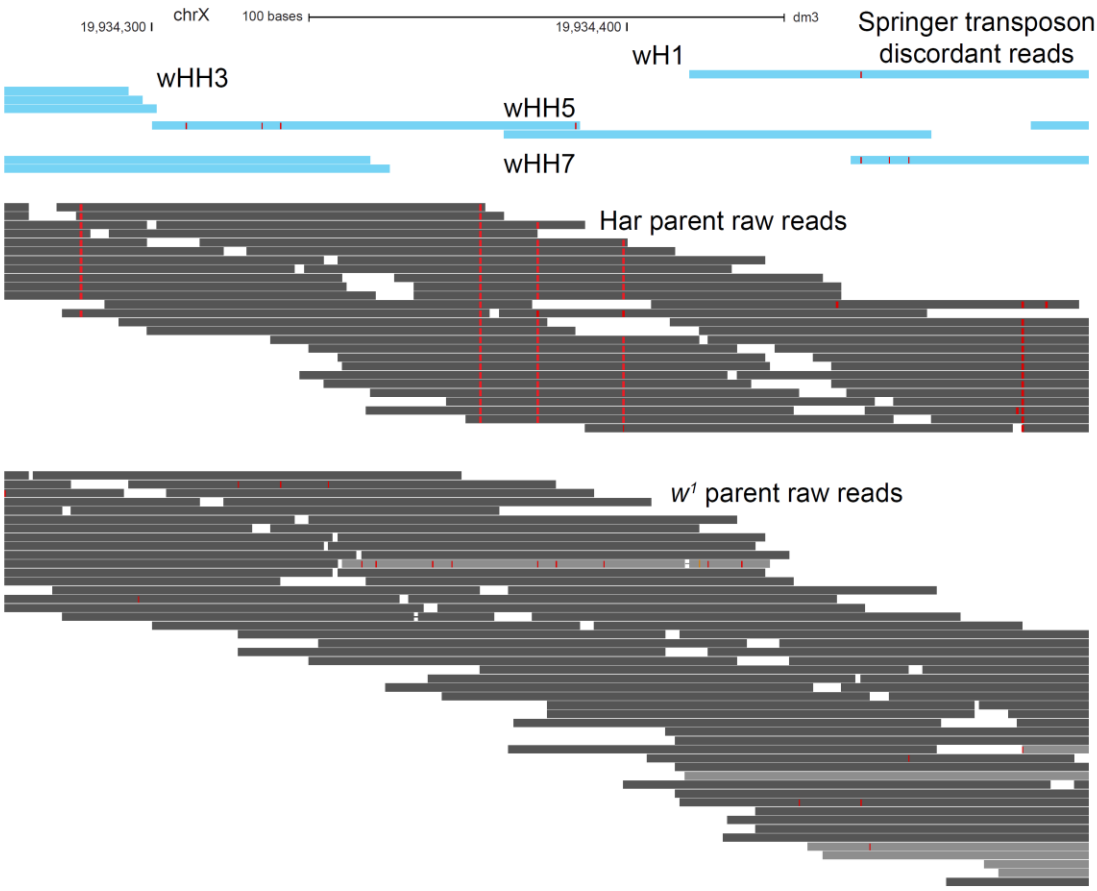

chrX
19,934,300
100 bases
19,934,400
dm3
Springer transposon discordant reads
wH1
wHH3
wHH5
wHH7
Har parent raw reads
w1 parent raw reads

Figure AII-7: SNPs surrounding new TE insertion showing non-Mendelian inheritance.

The springer insertion generated in wH1 shows non-Mendelian inheritance. The TE mapping discordant reads are shown in sky blue for different genotypes. Raw reads for Har and w^1 DNA are shown in grey. The SNPs are shown as red lines within the raw reads. The Har specific SNPs are absent in the TE discordant reads, indicating that the transposon inserted into w^1 DNA and the subsequent springer mapping discordant reads lack Har specific SNPs, suggesting that this TE retains w^1 DNA despite being crossed to Har for 7 generations.

Figure AII-8

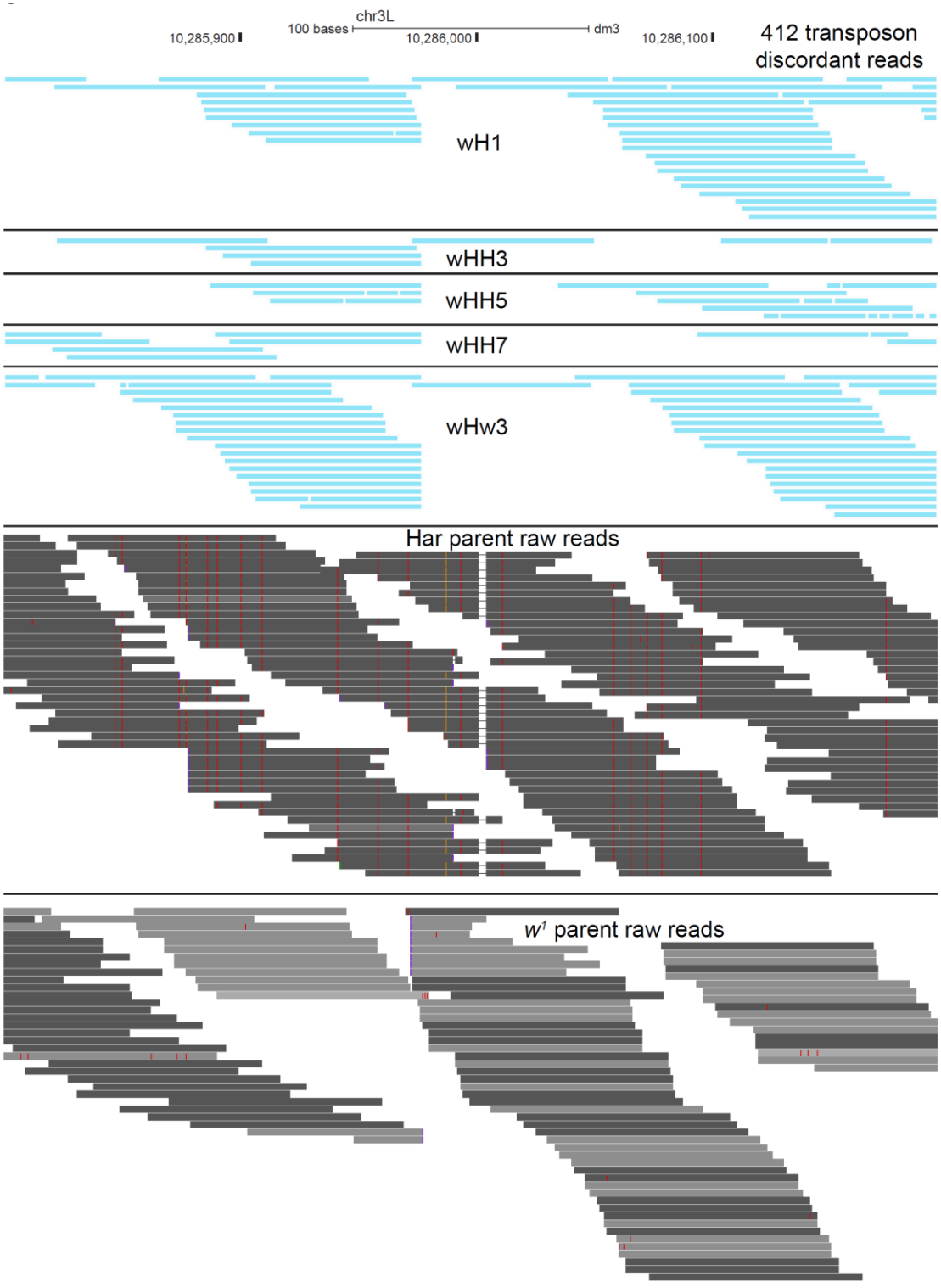

chr3L
10,285,900
100 bases
10,286,000
dm3
10,286,100
412 transposon discordant reads
wH1
wHH3
wHH5
wHH7
wHw3
Har parent raw reads
w1 parent raw reads

Figure AII-8: SNPs flanking a parental TE showing non-Mendelian inheritance

The 412 insertion from w^1 parent shows non-Mendelian inheritance. The TE mapping discordant reads are shown in sky blue for different genotypes. Raw reads for Har and w^1 DNA are shown in grey. The SNPs are shown as red lines within the raw reads. The Har specific SNPs are absent in the TE discordant reads. The linkage of w^1 SNPs to the 412 transposon retained after outcrossed into Har background.

Figure AII-9

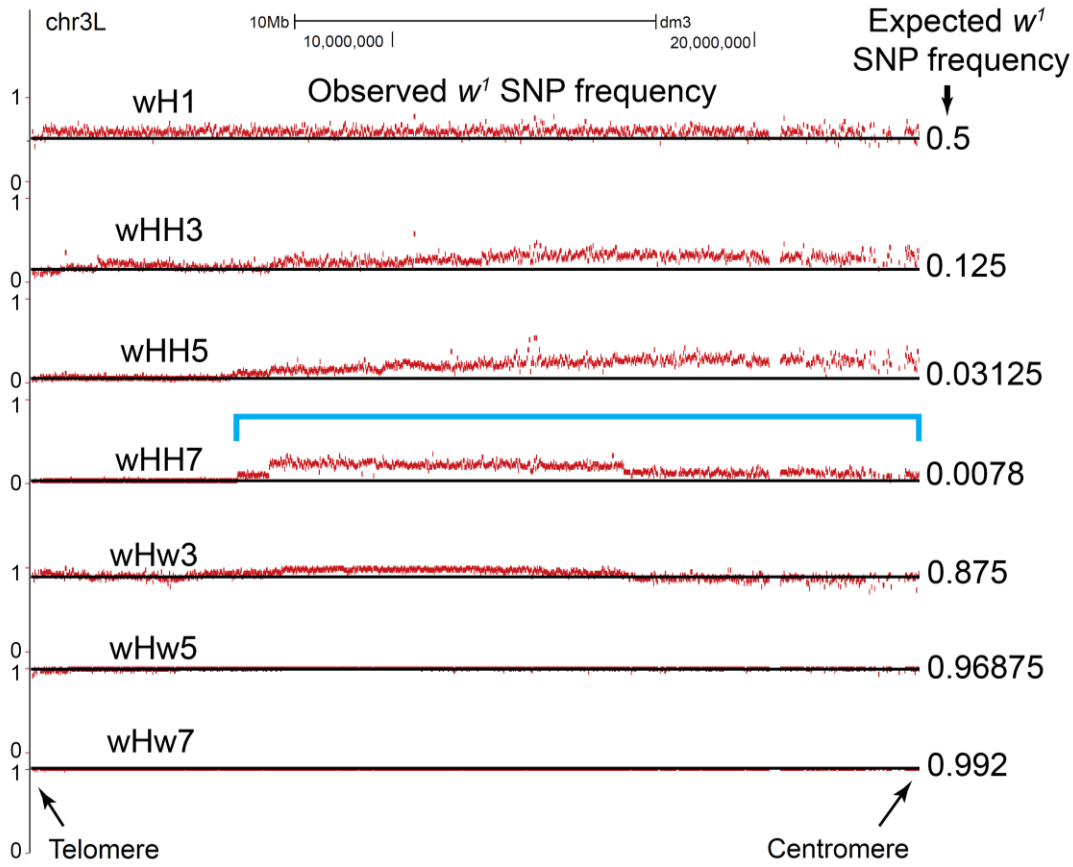

chr3L
10Mb
10,000,000
20,000,000
dm3
Expected w1 SNP frequency
Observed w1 SNP frequency
wH1
0.5
wHH3
0.125
wHH5
0.03125
wHH7
0.0078
wHw3
0.875
wHw5
0.96875
wHw7
0.992
Telomere
Centromere

Figure AII-9: Non-Mendelian inheritance on chr3L adjacent to centromere

Genome browser view showing w^1 specific SNP frequency in different genotypes. The red dots show observed w^1 SNP frequency. Black line shows expected w^1 SNP frequency according to Mendelian inheritance. As the browser shot is for chr3L, the telomere is to the left and centromere is to the right as shown. w^1 SNPs follow Mendelian inheritance when crossed to w^1. However, they show non-Mendelian inheritance in the right arm of chr3L. The region showing non-Mendelian inheritance in wHH7 is denoted by blue bracket.

Discussion

Transposon mobilization can cause mutations and genomic instability. Despite these deleterious effects on host, transposons have invaded almost all the eukaryotic genomes (Bao et al., 2015). Transposons can move between individuals by both horizontal and vertical transmission (Gilbert and Feschotte, 2018). Horizontal transmission leads to rapid spread of transposons between individuals and populations. One example is P-element which jumped from *D. willistoni* to *D. melanogaster* and has rapidly spread through wild *D. melanogaster* strains within a few decades (Engels, 1992). It recently jumped into D. *simulans* and is currently sweeping through this species (Kofler et al., 2015). Numerous instances of such horizontal transfer is observed in different organisms. This shows success of transposons as genomic pathogens.

Vertical transmission leads to inheritance of TEs from parents to the progeny. Most of the genome shows Mendelian inheritance. However, many selfish genetic elements have been observed to show biased inheritance, leading to increased frequency of selfish elements in successive populations (Werren, 2011). For examples, centromere drive can lead to asymmetry during meiosis and can give advantage to selfish centromere over others during gamete formation (Kursel and Malik, 2018). Being pathogens, transposons can also show selfish behavior. Transposons are proposed to be active in the germline, which is the tissue responsible for transmission of genetic information to the next generation (Haig, 2016; Tiwari et al., 2017). Increased copy numbers in germline would lead to increase in frequency of TEs in progeny. The germline activity would also help dump transposon transcripts in the oocyte, which can insert in the naïve genome

during early stages of embryogenesis. Thus, selfish behavior can give undue advantage to transposons during vertical transmission.

We wanted to study the vertical transmission of TEs. We could do this by crossing different strains of *D. melanogaster*, which produce fertile progeny, and follow their inheritance for multiple generations. However, our observations would only be limited to the existing parental TEs. We needed a system to activate transposons, so that we can follow the existing and many new insertions. Mutations in piRNA pathway lead to transposon over-expression (Zhang et al., 2012a). However, we cannot use them for inheritance studies as they lead to sterility. Hybrid dysgenesis leads to transposon activation and sterility. However, as sterility improves with age (Khurana et al., 2011), we used this as a model to study inheritance of existing and novel TEs. Most resident and newly generated transposons show Mendelian inheritance. A subset of transposons defy this Mendelian inheritance and increase in frequency during successive generations of outcrossing. As w^1 and Har strains used in the dysgenic cross differ by multiple fixed SNPs, we could follow the inheritance of DNA surrounding transposons during outcrossing. SNPs linked to the TEs showing non-Mendelian inheritance were also retained after multiple generations of outcrossing. This shows that non-Mendelian inheritance is not due to recurrent transposition into transposition hotspots.

The other possible mechanisms are gene conversion and centromeric drive. Gene conversion involves copying of part of one chromosome into another chromosome (Chen et al., 2007a). This copying increases the frequency of the donor DNA. It can occur during the process of DNA double strand break repair due recombination between

homologous chromosomes. New transposon insertions can produce double strand breaks, which can initiate recombination and subsequent gene conversion. As the SNPs linked to transposons show non-Mendelian inheritance, gene conversion could be one mechanism for non-Mendelian inheritance. Another possible mechanism is meiotic drive. Major portion of 3rd chromosome shows non-Mendelian inheritance when outcrossed to Har. This may be because of w^1 chromosome centromere leads to centromere drive and increase the inheritance of w^1 chromosome. Transposons may just hitch-hike these regions causing centromeric drive. We think that gene conversion and/or centromeric drive leads to non-Mendelian inheritance of transposons. These mechanisms can allow transposons to increase their copy numbers and provide more ways to behave like a selfish genetic element apart from transposition.

Methods

Fly stocks:

All flies were kept at 25^{o}C during crosses. w^{1} and Har strains were obtained from our lab stocks published previously (Khurana et al., 2011). For dysgenic cross, virgin w^{1} females were crossed with Har males. The F1 dysgenic females became fertile in about 2-3 weeks. After the eggs laid by these females hatched, their ovaries were dissected in 1X Robb's medium (McKim et al., 2009). For all the subsequent crosses, virgins were collected and mated with respective males. They were dissected after 2-3 weeks. The eggs laid before dissections led to adult of next generation.

Genomic library preparation:

Genomic DNA was purified from Qiagen DNeasy Blood and tissue DNA isolation kit using Manufacturer's protocol. ProteinaseK digestion was done for 3 hours in the tissue lysis step. Genomic libraries were prepared by Beijing Genomics Institute (BGI) or by me. The DNA was fragmented, gel purified, adapters were ligated and PCR amplification was done to get the genomic libraries. The libraries were sequenced by Illumina HiSeq2000 or NextSeq500.

TEMP analysis:

Transposon insertions were identified by TEMP method (Zhuang et al., 2014). The program gives the list of TE insertions in the given genotype by using the paired end read information, where one read mapped to transposon and other to the reference genome.

For analyzing parental insertions, only the transposons with reads mapping to both the ends were used. For newly generated TEs in wH1, all the transposons used except chrU_extra. For each transposon, penetrance value is assigned which is the number of transposon confirming reads divided by total number of reads in that genomic region.

SNP analysis:

All the reads were first mapped to *D. melanogaster* dm3 genome. The SNPs were identified which were different than the reference genome for both w^1 and Har parents. This provided with SNPs which are different in the two genotypes. We used the SNPs which were fixed in either of the parents. Their inheritance was followed for multiple generations. The bigwig files for SNP frequency were loaded onto UCSC genome browser for visual representation.

Acknowledgements

I would like to thank Nadine Schultz for helping me with the crosses and Tianxiong Yu (Bear) and Jiali Zhuang for help in Bioinformatics analysis.

Bibliography:

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol *11*, 1017-1027.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Asif-Laidin, A., Delmarre, V., Laurentie, J., Miller, W.J., Ronsseray, S., and Teysset, L. (2017). Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in Drosophila. DNA Res *24*, 459-472.

Assis, R., and Kondrashov, A.S. (2009). Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. Proc Natl Acad Sci U S A *106*, 7079-7082.

Ayarpadikannan, S., and Kim, H.S. (2014). The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genomics Inform *12*, 98-104.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA *6*, 11.

Bartolome, C., Bello, X., and Maside, X. (2009). Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. Genome Biol *10*, R22.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., *et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

Belancio, V.P., Hedges, D.J., and Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res *18*, 343-358.

Belyayev, A. (2014). Bursts of transposable elements as an evolutionary driving force. J Evol Biol *27*, 2573-2584.

Bergman, C.M., Quesneville, H., Anxolabehere, D., and Ashburner, M. (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol *7*, R112.

Biemont, C., and Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature *443*, 521-524.

Billi, A.C., Alessi, A.F., Khivansara, V., Han, T., Freeberg, M., Mitani, S., and Kim, J.K. (2012). The Caenorhabditis elegans HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. PLoS Genet *8*, e1002617.

Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K. (2013). A conserved upstream motif orchestrates autonomous, germline-enriched expression of Caenorhabditis elegans piRNAs. PLoS Genet *9*, e1003392.

Bingham, P.M., Kidwell, M.G., and Rubin, G.M. (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell *29*, 995-1004.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A *104*, 3312-3317.

Blumenstiel, J.P., Erwin, A.A., and Hemmer, L.W. (2016). What Drives Positive Selection in the Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis. Yale J Biol Med *89*, 499-512.

Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491-500.

Bownes, M. (1990). Preferential insertion of P elements into genes expressed in the germ-line of Drosophila melanogaster. Mol Gen Genet *222*, 457-460.

Bozzetti, M.P., Massari, S., Finelli, P., Meggio, F., Pinna, L.A., Boldyreff, B., Issinger, O.G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., *et al.* (1995). The Ste locus, a component of the parasitic cry-Ste system of Drosophila melanogaster, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. Proc Natl Acad Sci U S A *92*, 6067-6071.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Canapa, A., Barucca, M., Biscotti, M.A., Forconi, M., and Olmo, E. (2015). Transposons, Genome Size, and Evolutionary Insights in Animals. Cytogenet Genome Res *147*, 217-239.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Cecere, G., Zheng, G.X., Mansisidor, A.R., Klymko, K.E., and Grishok, A. (2012). Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Mol Cell *47*, 734-745.

Chen, J.M., Cooper, D.N., Chuzhanova, N., Ferec, C., and Patrinos, G.P. (2007a). Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet *8*, 762-775.

Chen, Y., Pane, A., and Schupbach, T. (2007b). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Curr Biol *17*, 637-642.

Chen, Y.C., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J., *et al.* (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell *63*, 97-109.

Chirn, G.W., Rahman, R., Sytnikova, Y.A., Matts, J.A., Zeng, M., Gerlach, D., Yu, M., Berger, B., Naramura, M., Kile, B.T., *et al.* (2015). Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. PLoS Genet *11*, e1005652.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from mu to kangaroo. Nat Rev Mol Cell Biol *4*, 865-877.

Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell *50*, 749-761.

Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A. (1990). Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics *124*, 339-355.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., *et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell *31*, 79-90.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet *46*, 677-700.

Davis, A.W., Roote, J., Morley, T., Sawamura, K., Herrmann, S., and Ashburner, M. (1996). Rescue of hybrid sterility in crosses between D. melanogaster and D. simulans. Nature *380*, 157-159.

De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. Nature *480*, 259-263.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol *12*, 687-695.

Eberl, D.F., Lorenz, L.J., Melnick, M.B., Sood, V., Lasko, P., and Perrimon, N. (1997). A new enhancer of position-effect variegation in Drosophila melanogaster encodes a putative RNA helicase that binds chromosomes and is regulated by the cell cycle. Genetics *146*, 951-963.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol *7*, 787-797.

Engels, W.R. (1992). The origin of P elements in Drosophila melanogaster. Bioessays *14*, 681-686.

Engels, W.R., and Preston, C.R. (1979). Hybrid dysgenesis in Drosophila melanogaster: the biology of female and male sterility. Genetics *92*, 161-174.

Fanning, T.G., and Singer, M.F. (1987). LINE-1: a mammalian transposable element. Biochim Biophys Acta *910*, 203-212.

Fedoroff, N.V. (2012). McClintock's challenge in the 21st century. Proc Natl Acad Sci U S A *109*, 20200-20203.

Fernandez-Hernandez, I., Scheenaard, E., Pollarolo, G., and Gonzalez, C. (2016). The translational relevance of Drosophila in drug discovery. EMBO Rep *17*, 471-472.

Feschotte, C., and Wessler, S.R. (2001). Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc Natl Acad Sci U S A *98*, 8923-8924.

Fontdevila, A. (2005). Hybrid genome evolution by transposition. Cytogenet Genome Res *110*, 49-55.

Friedlander, M.R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T.A., Hirst, M., Marra, M., Nusbaum, C., Lee, W.L., Jenkin, J.C., *et al.* (2009). High-resolution profiling and discovery of planarian small RNAs. Proc Natl Acad Sci U S A *106*, 11546-11551.

Fu, Q., and Wang, P.J. (2014). Mammalian piRNAs: Biogenesis, function, and mysteries. Spermatogenesis *4*, e27889.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

Gilbert, C., and Feschotte, C. (2018). Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev *49*, 15-24.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Goh, W.S., Seah, J.W., Harrison, E.J., Chen, C., Hammell, C.M., and Hannon, G.J. (2014). A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis. Genes Dev *28*, 797-807.

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep *15*, 411-418.

Gou, L.T., Kang, J.Y., Dai, P., Wang, X., Li, F., Zhao, S., Zhang, M., Hua, M.M., Lu, Y., Zhu, Y., *et al.* (2017). Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. Cell *169*, 1090-1104 e1013.

Gu, A., Ji, G., Shi, X., Long, Y., Xia, Y., Song, L., Wang, S., and Wang, X. (2010). Genetic variants in Piwi-interacting RNA pathway genes confer susceptibility to spermatogenic failure in a Chinese population. Hum Reprod *25*, 2955-2961.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell *151*, 1488-1500.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., *et al.* (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell *36*, 231-244.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Ha, H., Song, J., Wang, S., Kapusta, A., Feschotte, C., Chen, K.C., and Xing, J. (2014). A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. BMC Genomics *15*, 545.

Haig, D. (2016). Transposable elements: Self-seekers of the germline, team-players of the soma. Bioessays *38*, 1158-1166.

Han, B.W., Wang, W., Li, C., Weng, Z., and Zamore, P.D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817-821.

Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and Brennecke, J. (2013). The genetic makeup of the Drosophila piRNA pathway. Mol Cell *50*, 762-777.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J *30*, 3977-3993.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S.L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. Nature *539*, 588-592.

Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res *616*, 46-59.

Heyn, H., Ferreira, H.J., Bassas, L., Bonache, S., Sayols, S., Sandoval, J., Esteller, M., and Larriba, S. (2012). Epigenetic disruption of the PIWI pathway in human spermatogenic disorders. PLoS One *7*, e47892.

Honda, S., Kirino, Y., Maragkakis, M., Alexiou, P., Ohtaki, A., Murali, R., Mourelatos, Z., and Kirino, Y. (2013). Mitochondrial protein BmPAPI modulates the length of mature piRNAs. RNA *19*, 1405-1418.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Houwing, S., Berezikov, E., and Ketting, R.F. (2008). Zili is required for germ cell differentiation and meiosis in zebrafish. EMBO J *27*, 2702-2711.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Huang, C.R., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annu Rev Genet *46*, 651-675.

Huang, H., Li, Y., Szulwach, K.E., Zhang, G., Jin, P., and Chen, D. (2014). AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. J Cell Biol *206*, 217-230.

Huang, X., Fejes Toth, K., and Aravin, A.A. (2017). piRNA Biogenesis in Drosophila melanogaster. Trends Genet *33*, 882-894.

Hur, J.K., Luo, Y., Moon, S., Ninova, M., Marinov, G.K., Chung, Y.D., and Aravin, A.A. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev *30*, 840-855.

Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem *84*, 405-433.

Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., Katsuma, S., and Tomari, Y. (2016). Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. Cell *164*, 962-973.

Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte, E.M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science *348*, 921-925.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499-509.

Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E., and Ketting, R.F. (2012). Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. PLoS Genet *8*, e1002702.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A *98*, 8714-8719.

Kasper, D.M., Gardner, K.E., and Reinke, V. (2014). Homeland security in the C. elegans germ line: insights into the biogenesis and function of piRNAs. Epigenetics *9*, 62-74.

Kawaoka, S., Hayashi, N., Katsuma, S., Kishino, H., Kohara, Y., Mita, K., and Shimada, T. (2008). Bombyx small RNAs: genomic defense system against transposons in the silkworm, Bombyx mori. Insect Biochem Mol Biol *38*, 1058-1065.

Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. RNA *15*, 1258-1264.

Kawaoka, S., Mitsutake, H., Kiuchi, T., Kobayashi, M., Yoshikawa, M., Suzuki, Y., Sugano, S., Shimada, T., Kobayashi, J., Tomari, Y., *et al.* (2012). A role for transcription from a piRNA cluster in de novo piRNA production. RNA *18*, 265-273.

Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Waldman Ben-Asher, H., Muster, N., Thompson, J., *et al.* (2014). Drosophila TRF2 is a preferential core promoter regulator. Genes Dev *28*, 2163-2174.

Kelleher, E.S. (2016). Reexamining the P-Element Invasion of Drosophila melanogaster Through the Lens of piRNA Silencing. Genetics *203*, 1513-1531.

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

Khurana, J.S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. J Cell Biol *191*, 905-913.

Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell *147*, 1551-1563.

Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nat Struct Mol Biol *14*, 347-348.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic axis

specification through activation of an ATR/Chk2 DNA damage response. Dev Cell *12*, 45-55.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., *et al.* (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Kofler, R., Hill, T., Nolte, V., Betancourt, A.J., and Schlotterer, C. (2015). The recent invasion of natural Drosophila simulans populations by the P-element. Proc Natl Acad Sci U S A *112*, 6659-6663.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., *et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

Kursel, L.E., and Malik, H.S. (2018). The cellular mechanisms and consequences of centromere drive. Curr Opin Cell Biol *52*, 58-65.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

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

Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Toth, K.F. (2013). Piwi induces piRNA-guided

transcriptional silencing and establishment of a repressive chromatin state. Genes Dev *27*, 390-399.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C., Luo, Y., Sachidanandam, R., Toth, K.F., *et al.* (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev *28*, 1667-1680.

Leaver, M.J. (2001). A family of Tc1-like transposons from the genomes of fishes and frogs: evidence for horizontal transmission. Gene *271*, 203-214.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell *150*, 78-87.

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

Lerat, E., Burlet, N., Biemont, C., and Vieira, C. (2011). Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene *473*, 100-109.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., *et al.* (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., *et al.* (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell *50*, 67-81.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463-2476.

Linheiro, R.S., and Bergman, C.M. (2012). Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. PLoS One *7*, e30008.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McKim, K.S., Joyce, E.F., and Jang, J.K. (2009). Cytological analysis of meiosis in fixed Drosophila ovaries. Methods Mol Biol *558*, 197-216.

Meehan, R.R., Kao, C.F., and Pennings, S. (2003). HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J *22*, 3164-3174.

Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science *348*, 812-817.

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

Montgomery, T.A., Rim, Y.S., Zhang, C., Dowen, R.H., Phillips, C.M., Fischer, S.E., and Ruvkun, G. (2012). PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. PLoS Genet *8*, e1002616.

Muchardt, C., Guilleme, M., Seeler, J.S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep *3*, 975-981.

Muerdter, F., Guzzardo, P.M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G.J. (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol Cell *50*, 736-748.

Murota, Y., Ishizu, H., Nakagawa, S., Iwasaki, Y.W., Shibata, S., Kamatani, M.K., Saito, K., Okano, H., Siomi, H., and Siomi, M.C. (2014). Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly. Cell Rep *8*, 103-113.

Nishida, K.M., Iwasaki, Y.W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., Siomi, H., and Siomi, M.C. (2015). Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in Bombyx germ cells. Cell Rep *10*, 193-203.

Obbard, D.J., Gordon, K.H., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci *364*, 99-115.

Pace, J.K., 2nd, Gilbert, C., Clark, M.S., and Feschotte, C. (2008). Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A *105*, 17023-17028.

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of Drosophila melanogaster. Genetics *138*, 1181-1197.

Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J *30*, 4601-4615.

Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell *12*, 851-862.

Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev Cell *43*, 60-70 e65.

Pillai, R.S., and Chuma, S. (2012). piRNAs and their involvement in male germline development in mice. Dev Growth Differ *54*, 78-92.

Ponomarenko, N.A., Bannikov, V.M., Anashchenko, V.A., and Tchurikov, N.A. (1997). Burdock, a novel retrotransposon in Drosophila melanogaster, integrates into the coding region of the cut locus. FEBS Lett *413*, 7-10.

Presgraves, D.C., Balagopalan, L., Abmayr, S.M., and Orr, H.A. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature *423*, 715-719.

Presgraves, D.C., and Stephan, W. (2007). Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. Mol Biol Evol *24*, 306-314.

Priimagi, A.F., Mizrokhi, L.J., and Ilyin, Y.V. (1988). The Drosophila mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene *70*, 253-262.

Pritykin, Y., Brito, T., Schupbach, T., Singh, M., and Pane, A. (2017). Integrative analysis unveils new functions for the Drosophila Cutoff protein in noncoding RNA biogenesis and gene regulation. RNA *23*, 1097-1109.

Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. J Biol Chem *286*, 3789-3797.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat Struct Mol Biol *11*, 558-566.

Rengaraj, D., Lee, S.I., Park, T.S., Lee, H.J., Kim, Y.M., Sohn, Y.A., Jung, M., Noh, S.J., Jung, H., and Han, J.Y. (2014). Small non-coding RNA profiling and the role of piRNA pathway genes in the protection of chicken primordial germ cells. BMC Genomics *15*, 757.

Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature *480*, 264-267.

Rosenzweig, B., Liao, L.W., and Hirsh, D. (1983). Sequence of the C. elegans transposable element Tc1. Nucleic Acids Res *11*, 4201-4209.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.

Sakakibara, K., and Siomi, M.C. (2018). The PIWI-Interacting RNA Molecular Pathway: Insights From Cultured Silkworm Germline Cells. Bioessays *40*.

Sanchez-Gracia, A., Maside, X., and Charlesworth, B. (2005). High rate of horizontal transfer of transposable elements in Drosophila. Trends Genet *21*, 200-203.

Satyaki, P.R., Cuykendall, T.N., Wei, K.H., Brideau, N.J., Kwak, H., Aruna, S., Ferree, P.M., Ji, S., and Barbash, D.A. (2014). The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet *10*, e1004240.

Sawamura, K., Davis, A.W., and Wu, C.I. (2000). Genetic analysis of speciation by means of introgression into Drosophila melanogaster. Proc Natl Acad Sci U S A *97*, 2652-2655.

Saxe, J.P., Chen, M., Zhao, H., and Lin, H. (2013). Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. EMBO J *32*, 1869-1885.

Schaack, S., Gilbert, C., and Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol *25*, 537-546.

Schupbach, T., and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. Genetics *121*, 101-117.

Schupbach, T., and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics *129*, 1119-1136.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964-980.

Simkin, A., Wong, A., Poh, Y.P., Theurkauf, W.E., and Jensen, J.D. (2013). Recurrent and recent selective sweeps in the piRNA pathway. Evolution *67*, 1081-1090.

Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr Biol *10*, 27-30.

Song, S.U., Kurkulos, M., Boeke, J.D., and Corces, V.G. (1997). Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila. Development *124*, 2789-2798.

Spradling, A.C., Stern, D.M., Kiss, I., Roote, J., Laverty, T., and Rubin, G.M. (1995). Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci U S A *92*, 10824-10830.

Sturtevant, A.H. (1920). Genetic Studies on DROSOPHILA SIMULANS. I. Introduction. Hybrids with DROSOPHILA MELANOGASTER. Genetics *5*, 488-500.

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

Thomae, A.W., Schade, G.O., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A pair of centromeric proteins mediates reproductive isolation in Drosophila species. Dev Cell *27*, 412-424.

Thomas, J., and Pritham, E.J. (2015). Helitrons, the Eukaryotic Rolling-circle Transposable Elements. Microbiol Spectr *3*.

Tiwari, B., Kurtz, P., Jones, A.E., Wylie, A., Amatruda, J.F., Boggupalli, D.P., Gonsalvez, G.B., and Abrams, J.M. (2017). Retrotransposons Mimic Germ Plasm Determinants to Promote Transgenerational Inheritance. Curr Biol *27*, 3010-3016 e3013.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

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

Van Valen, L.M. (1973). A new evolutionary law. Evolutionary Theory *1*, 1-30.

Vanderweyde, T., Apicco, D.J., Youmans-Kidder, K., Ash, P.E., Cook, C., Lummertz da Rocha, E., Jansen-West, K., Frame, A.A., Citro, A., Leszyk, J.D., *et al.* (2016). Interaction of tau with the RNA-Binding Protein TIA1 Regulates tau Pathophysiology and Toxicity. Cell Rep *15*, 1455-1466.

Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet *1*, 96-108.

Vermaak, D., and Malik, H.S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. Annu Rev Genet *43*, 467-492.

Volpe, A.M., Horowitz, H., Grafer, C.M., Jackson, S.M., and Berg, C.A. (2001). Drosophila rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. Genetics *159*, 1117-1134.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr Biol *18*, 861-867.

Wang, W., Ashby, R., Ying, H., Maleszka, R., and Foret, S. (2017). Contrasting Sex-and Caste-Dependent piRNA Profiles in the Transposon Depleted Haplodiploid Honeybee Apis mellifera. Genome Biol Evol *9*, 1341-1356.

Wang, Y.L., Duttke, S.H., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev *28*, 1550-1555.

Warren, I.A., Naville, M., Chalopin, D., Levin, P., Berger, C.S., Galiana, D., and Volff, J.N. (2015). Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome Res *23*, 505-531.

Wehr, K., Swan, A., and Schupbach, T. (2006). Deadlock, a novel protein of Drosophila, is required for germline maintenance, fusome morphogenesis and axial patterning in oogenesis and associates with centrosomes in the early embryo. Dev Biol *294*, 406-417.

Weick, E.M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. Development *141*, 3458-3471.

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

Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Coute, Y., Conn, S., Kadlec, J., Sachidanandam, R., *et al.* (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell *157*, 1698-1711.

Yi, M., Chen, F., Luo, M., Cheng, Y., Zhao, H., Cheng, H., and Zhou, R. (2014). Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. Genome Biol Evol *6*, 1393-1407.

Yu, B., Cassani, M., Wang, M., Liu, M., Ma, J., Li, G., Zhang, Z., and Huang, Y. (2015a). Structural insights into Rhino-mediated germline piRNA cluster formation. Cell Res *25*, 525-528.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015b). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339-342.

Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development *138*, 4039-4050.

Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C., and Jensen, S. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci U S A *110*, 19842-19847.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., *et al.* (2012a). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhang, Z., Theurkauf, W.E., Weng, Z., and Zamore, P.D. (2012b). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence *3*, 9.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell *157*, 1353-1363.

Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell *44*, 572-584.

Zhuang, J., Wang, J., Theurkauf, W., and Weng, Z. (2014). TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res *42*, 6826-6838.