

Protein–Protein Interactions Shape Genomic Autoimmunity in the Adaptively Evolving Rhino-Deadlock-Cutoff Complex

Erin S. Kelleher*

Department of Biology and Biochemistry, University of Houston, Texas, USA

*Corresponding author: E-mail: eskelleher@uh.edu.

Accepted: 4 June 2021

Abstract

The Piwi-interacting RNA (piRNA) pathway is a genomic defense system that controls the movement of transposable elements (TEs) through transcriptional and post-transcriptional silencing. Although TE defense is critical to ensuring germline genome integrity, it is equally critical that the piRNA pathway avoids autoimmunity in the form of silencing host genes. Ongoing cycles of selection for expanded control of invading TEs, followed by selection for increased specificity to reduce impacts on host genes, are proposed to explain the frequent signatures of adaptive evolution among piRNA pathway proteins. However, empirical tests of this model remain limited, particularly with regards to selection against genomic autoimmunity.

I examined three adaptively evolving piRNA proteins, Rhino, Deadlock, and Cutoff, for evidence of interspecific divergence in autoimmunity between *Drosophila melanogaster* and *Drosophila simulans*. I tested a key prediction of the autoimmunity hypothesis that foreign heterospecific piRNA proteins will exhibit enhanced autoimmunity, due to the absence of historical selection against off-target effects. Consistent with this prediction, full-length *D. simulans* Cutoff, as well as the *D. simulans* hinge and chromo domains of Rhino, exhibit expanded regulation of *D. melanogaster* genes. I further demonstrate that this autoimmunity is dependent on known incompatibilities between *D. simulans* proteins or domains and their interacting partners in *D. melanogaster*. My observations reveal that the same protein–protein interaction domains that are interfaces of adaptive evolution in Rhino and Cutoff also determine their potential for autoimmunity.

Significance

Pathways involved in genome defense present a liability to the organism in the form of autoimmunity. In this paper, I demonstrate that two adaptively evolving proteins that defend the genome from selfish transposable elements exhibit interspecific differences in host gene regulation, consistent with selection against genomic autoimmunity.

Introduction

The Piwi-interacting RNA (piRNA) pathway is an RNA-mediated silencing pathway that controls the mobilization of transposable elements (TEs) in metazoan germlines (reviewed in Czech et al. 2018; Ozata et al. 2019). piRNA pathway evolution is exceptionally dynamic, including both gene duplication and rapid adaptive protein evolution (Obbard et al. 2009; Kolaczkowski et al. 2011; Simkin et al. 2013; Yi et al. 2014; Lewis et al. 2016; Palmer et al. 2018; Crysanto and Obbard 2019). This adaptive evolution is likely required to maintain control of genomic TEs, which change rapidly in presence and abundance over short evolutionary

time periods (Kidwell 1983; Naito et al. 2006; Yang and Barbash 2008; El Baidouri and Panaud 2013; Gilbert et al. 2010; Reiss et al. 2019). However, it is equally crucial that the piRNA pathway avoids collateral damage in the form of off-target silencing of host genes. piRNA pathway evolution, therefore, is proposed to reflect a tradeoff between maximizing TE regulation whereas minimizing genomic autoimmunity (Blumenstiel et al. 2016).

The piRNA pathway has been most extensively characterized in *Drosophila melanogaster* and consists of >30 proteins with diverse functional roles in piRNA transcription, maturation, and enforcement of transcriptional and post-

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

transcriptional silencing (reviewed in Senti and Brennecke 2010; Ozata et al. 2019). Proteins that establish piRNA transcription play potentially critical roles in both adaptation to genomic TEs and avoidance of autoimmunity by determining the repertoire of cellular piRNAs (Blumenstiel et al. 2016; Palmer et al. 2018). Indeed, three key regulators of piRNA precursor transcription: Rhino (Rhi), Deadlock (Del), and Cutoff (Cuff) are among the most adaptively evolving piRNA proteins in the genus *Drosophila* (fig. 1A; Vermaak et al. 2005; Simkin et al. 2013; Blumenstiel et al. 2016; Palmer et al. 2018). Rhi recognizes piRNA-producing loci known as piRNA clusters, and together with Del and other cofactors, recruits RNA-polymerase II to initiate precursor transcription (Andersen et al. 2017). The Rhino-Deadlock-Cutoff (RDC) complex further suppresses mRNA transcription at piRNA clusters and ensures the transport of precursor transcripts to cytoplasmic sites of piRNA maturation (Mohn et al. 2014; Zhang et al. 2014; Chen et al. 2016). The RDC complex could therefore exhibit autoimmunity through suppressing mRNA transcription or misappropriating genic mRNAs to piRNA processing bodies.

Interspecific complementation, in which a loss-of-function mutant is rescued by a wild-type allele from another species, provides a powerful approach for uncovering functional differences in proteins resulting from adaptive evolution (Aruna et al. 2009; Flores et al. 2015; Parhad et al. 2017; Brand et al. 2018). Using this approach, it was revealed that *Drosophila simulans* alleles of Rhi, Del, and Cuff are unable to complement *D. melanogaster* mutant backgrounds and exhibit drastic defects in piRNA biogenesis and TE regulation (Parhad et al. 2017; Yu et al. 2018; Parhad et al. 2020). Furthermore, these functional deficits of *D. simulans* alleles result from incompatibilities with interacting cofactors in *D. melanogaster* (Parhad et al. 2017; Yu et al. 2018; Parhad et al. 2020). In the context of genomic autoimmunity, it is predicted that *D. simulans* alleles will also exhibit enhanced off-target effects in *D. melanogaster*, because there is no evolutionary history of purifying selection against targeting host mRNAs in a foreign genome. We recently uncovered this signature of expanded autoimmunity in *D. simulans* alleles of the cytoplasmic piRNA proteins Aubergine and Armitage (Wang et al. 2020). However, despite their stronger signatures of adaptive evolution and greater potential to initiate off-target effects, the autoimmunity of *D. simulans* alleles of Rhi, Del, and Cuff has never been investigated.

Here, I examined the autoimmunity of *D. simulans* Rhi, Del, and Cuff in a *D. melanogaster* background using published RNA-seq, small RNA-seq, and ChIP-seq data from interspecific complementation experiments (Parhad et al. 2017, 2020). For Rhi and Cutoff, I discovered disparate patterns of increased autoimmunity, which are determined by their incompatibilities with *D. melanogaster* cofactors. In the case of Rhi, increased autoimmunity is exhibited by the *D. simulans* hinge and chromo domains, but is masked by an incompatibility

between the *D. simulans* chromo shadow domain and *D. melanogaster* Del. In contrast, *D. simulans* Cuff increases the expression of hundreds of genes, potentially through the nonfunctional sequestration of *D. melanogaster* transcriptional regulators.

Results

Drosophila simulans Alleles of the RDC Complex Do Not Exhibit Expanded Silencing of *D. melanogaster* Host Genes

The genomic autoimmunity model predicts that in a *D. melanogaster* background foreign *D. simulans* alleles will exhibit expanded negative regulation of host genes when compared with their native *D. melanogaster* counterparts (Blumenstiel et al. 2016). To test this prediction, I identified genes that were upregulated and downregulated in ovaries by *D. melanogaster* and *D. simulans* transgenic rescues of *rhi*, *del*, and *cuff*, when compared with an unrescued mutant background, based on stranded, ribo-depleted total RNA-seq data (fig. 1B, supplementary table S1–S3, Supplementary Material online). Ribo-depleted stranded libraries include nonpolyadenylated RNAs (such as histones and ribosomal RNAs) and also allow for the differentiation of sense and antisense transcripts. I focused on sense transcripts that give rise to mRNAs and proteins, whose expression might be reduced by RDC function. Estimated abundance of the *D. melanogaster* and *D. simulans* transgenically expressed Rhi and Del proteins are similar (Parhad et al. 2017; Parhad et al. 2020), and I determined from the RNA-seq data that sense RNA expression levels are similar for *D. melanogaster* and *D. simulans* transgenes of all three proteins (supplementary fig. S1, Supplementary Material online). Therefore, genome-wide regulatory differences between transgenic rescues are best explained by differences in the encoded proteins.

In my initial analysis, I did not find any evidence of enhanced autoimmunity among *D. simulans* alleles. For both *del* and *rhi*, significantly fewer genes were negatively regulated by *D. simulans* transgenes than by *D. melanogaster* transgenes (Del: $\chi^2 = 27.27$, $df = 1$, P value = 1.77×10^{-7} , Rhi: $\chi^2 = 304.31$, $df = 1$, P value < 10^{-15}). For *cuff*, the *D. melanogaster* and *D. simulans* transgenes negatively regulate a similarly small number of genes; however, *D. simulans* *cuff* upregulates significantly more genes than *D. melanogaster* *cuff* (Fisher's exact test $P = 0.0005$). Although there are differences in the degree of replication in the *rhi* (two biological replicates) as compared with the *cuff* and *del* data sets (one biological replicate), these should not confound my inference of which transgenic rescue (*D. melanogaster* or *D. simulans*) negatively regulates more genes. Furthermore, read depths for different libraries are quite consistent for *D. melanogaster* and *D. simulans* transgenic rescues of the same mutation.

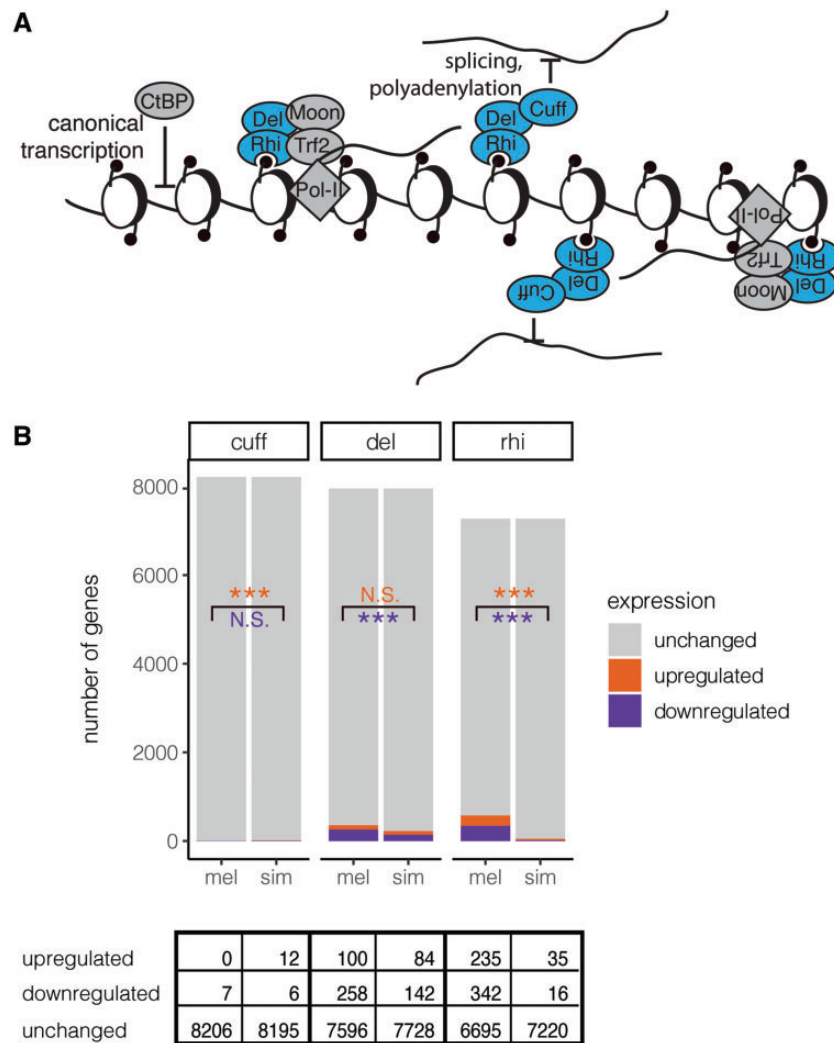


Fig. 1.—RDC regulation of piRNA precursors and host genes. (A) Schematic of known Cuff, Rhi, and Del functions in piRNA precursor transcription. Rhi and Del, together with Moon and Trf2 act to recruit bidirectional transcription of piRNA clusters (Andersen et al. 2017). Rhino, Del, and Cuff further specify piRNA precursors through suppressing splicing, polyadenylation, and termination (Mohn et al. 2014; Zhang et al. 2014; Chen et al. 2016). CtBP represses canonical, promoter-dependent transcription at piRNA clusters (Parhad et al. 2020). (B) Genes that are upregulated, downregulated, and unchanged by *cuff*, *del*, and *rhi* transgenic rescues as compared with the corresponding mutant background are indicated. Results of Fisher's exact test (*cuff*, upregulated) or χ^2 test-of-independence (all others, $df = 1$) indicating differences in the proportion of genes positively (orange, top) or negatively (purple, bottom) regulated between transgenic rescues are also indicated. N.S. denotes P value > 0.05 , * denotes P value < 0.05 , and *** denotes P value < 0.001 .

In the case of *D. simulans del* and *rhi*, reduced impacts on host gene expression are potentially explained by incompatibilities with *D. melanogaster* interactors, which prevent the production of piRNAs and therefore the manifestation of autoimmunity. *Drosophila simulans* Rhi is unable to interact with *D. melanogaster* Del, which abrogates piRNA transcription (Parhad et al. 2017; Yu et al. 2018). *Drosophila simulans* Del is similarly unable to promote piRNA biogenesis in *D. melanogaster*, most likely due to incompatibilities with other cofactors (Parhad et al. 2017). Robust examination of autoimmunity of *D. simulans* proteins may therefore require restoring their capacity to interact with cofactors, so that resulting differences in gene regulation are revealed.

Divergence in Gene Regulation by *D. simulans* Rhi Is Masked by Its Incompatibility with *D. melanogaster* Del

For *D. simulans* Rhi, the incompatibility with Del is caused by the chromo shadow domain (fig. 2A; Parhad et al. 2017). Chimeric transgenes combining the *D. melanogaster* chromo shadow domain with *D. simulans* domains elsewhere in the protein are functional for female fertility, as well as piRNA biogenesis and TE regulation (fig. 2B; Parhad et al. 2017). Rhi is a HP1 homolog that, in addition to the chromo shadow domain, contains a chromo and a hinge domain (Vermaak et al. 2005; reviewed in Vermaak and Malik 2009, fig. 2A). The chromo domain is responsible for binding to the histone

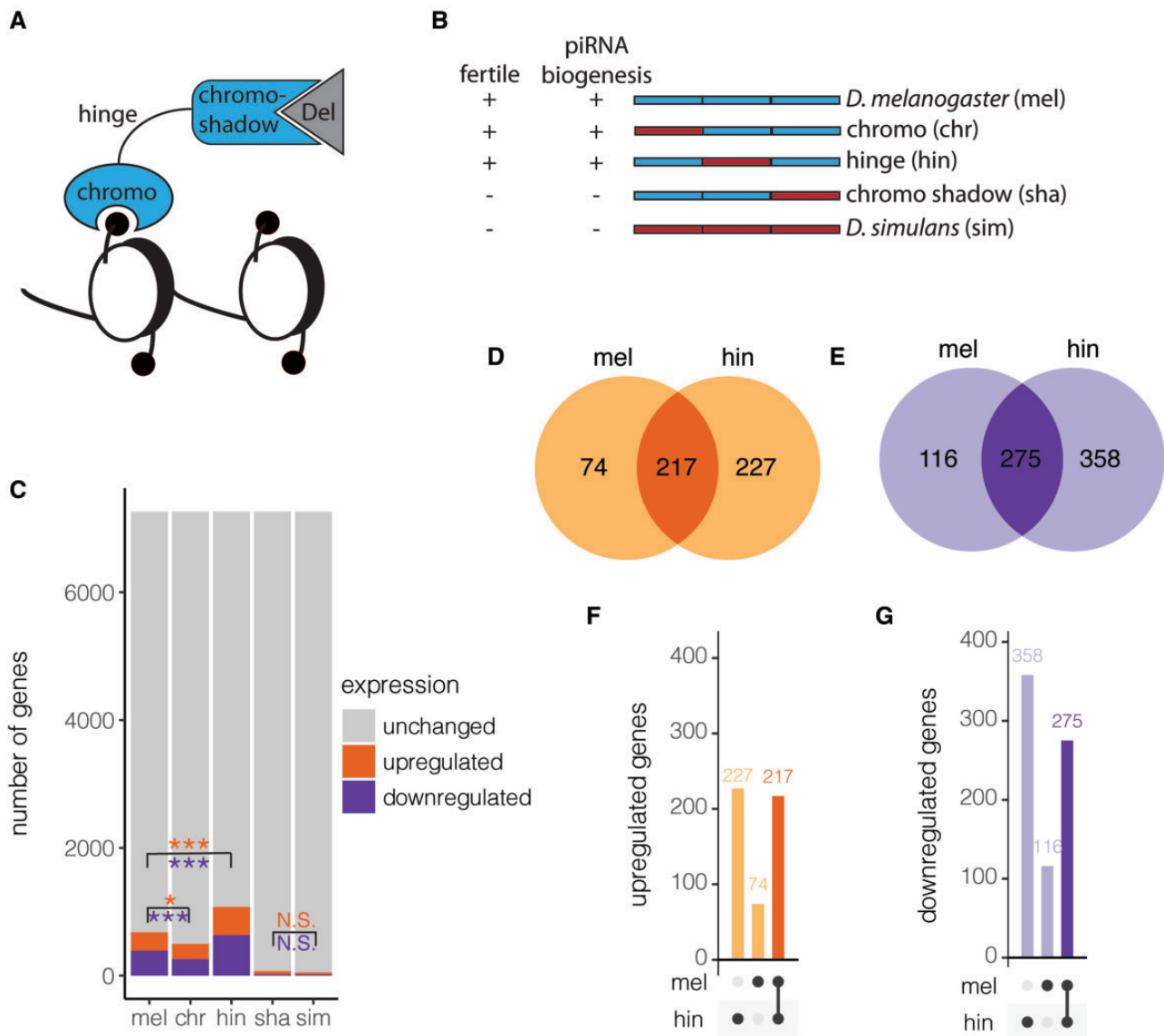


FIG. 2.—Divergence in host gene regulation individual Rhi domains. (A) Cartoon of three Rhino domains. (B) Schematic of *D. melanogaster*, *D. simulans*, and three fusion rescue proteins generated by Parhad et al. (2017). Rescue \pm indicates whether the fusion construct was previously reported to rescue female fertility or piRNA biogenesis and TE regulation (Parhad et al. 2020). (C) Genes that are upregulated, downregulated, and unchanged as compared with the corresponding mutant background are compared between *D. melanogaster* (mel) and *D. simulans* (sim) transgenic rescues as well as the chromo (chr), hinge (hin), and chromo shadow (sha) fusion transgenes. Results of χ^2 tests-of-independence indicating differences in the proportion of genes positively (orange, top) or negatively (purple, bottom) regulated between a fusion transgene and the *D. melanogaster* or *D. simulans* transgenes are indicated. (D and E) Venn diagrams and (F and G) upset plots comparing genes positively (D and F) and negatively (E and G) regulated by the *D. melanogaster* (mel) and hinge (hin) transgenic rescues of *rhi* as compared with unrescued mutants. Genes that are uniquely positively regulated by one of the two transgenic constructs are shaded lighter, whereas those regulated by both transgenic constructs in the same direction are shaded darker. N.S. denotes P value > 0.05 , * denotes P value < 0.05 , and *** denotes P value < 0.001 .

modification H3K9me3 (Le Thomas et al. 2014; Mohn et al. 2014; Yu et al. 2015), and the hinge domain plays an important role in determining the euchromatic versus heterochromatic localization of HP1 homologs (Smothers and Henikoff 2001). Both the chromo and hinge domains therefore have the potential to establish off-target effects by localizing Rhino to genic regions.

To isolate interspecific divergence in the hinge and chromo domains of *D. simulans* Rhi, I examined the gene regulatory effects of chimeric *rhi* transgenes using stranded, ribo-depleted total RNA-seq data (two biological replicates, fig. 2B and C, supplementary table S2, Supplementary Material online). Rhi protein abundance from each of these transgenes is similar to each other and to the *D. melanogaster* and *D.*

simulans transgenes, as are the estimated *rhi* transcript abundances based on the RNA-seq data (Parhad et al. 2017, supplementary fig. S1, [Supplementary Material](#) online). Consistent with abrogated function resulting from the Del incompatibility, the fusion transgene containing the *D. simulans* chromo shadow regulates only a small handful of host genes, similar to the pure *D. simulans* transgene (fig. 2C). In contrast, the hinge and chromo fusion transgenes, which are compatible with *D. melanogaster* Del, exhibit increased regulation of host genes when compared with the *D. simulans* transgene (fig. 2C). In particular, the hinge fusion transgene also positively and negatively regulates more host genes than the *D. melanogaster* transgene (positive regulation: $\chi^2 = 34.47$, $df = 1$, P value = 4.32×10^{-9} ; negative regulation: $\chi^2 = 61.46$, $df = 1$, P value = 4.53×10^{-15} ; fig. 2C–G). The pattern is stronger for negatively regulated genes, with the hinge fusion transgene reducing the expression of 633 genes, 358 of which are not negatively regulated by the *D. melanogaster* transgene (fig. 2E and G). Overall, this pattern is consistent with increased autoimmunity of the *D. simulans* hinge domain in a *D. melanogaster* background.

Although the expression of many genes could be indirectly affected by Rhi function (e.g., through reduced DNA damage resulting from TE activity), genic sites of Rhi occupancy are more likely to represent true examples of genomic autoimmunity. I therefore used ChIP-seq data from GFP-tagged *D. melanogaster*, hinge, and chromo fusion Rhi proteins to compare their occupancy proximal to genes. I detected 2,689, 1,757, and 611 occupancy peaks for *D. melanogaster* Rhi, the hinge fusion protein, and the chromo fusion protein, respectively (Parhad et al. 2017; [supplementary table S4, Supplementary Material](#) online), indicating that Rhi fusion proteins containing *D. simulans* domains do not necessarily occupy more overall genomic sites in the *D. melanogaster* genome than the native protein. It should be noted that comparatively low read depth for the chromo fusion protein input library may limit the power to detect peaks for this sample. Despite this limitation, occupancy peaks for the hinge and chromo fusion proteins are enriched within the gene bodies or up to 1 kb upstream of genes when compared with *D. melanogaster* Rhi (hinge: $\chi^2 = 14.79$, $df = 1$, P value = 1.26×10^{-4} , chromo: $\chi^2 = 5.8$, $df = 1$, P value = 0.016, fig. 3A). Thirty-three percent of hinge and chromo fusion protein occupancy sites occur within gene bodies or up to 1 kb upstream of genes, as compared with 28% (746) *D. melanogaster* Rhi occupancy sites. Hinge and chromo fusion proteins may therefore have greater potential to regulate genic sites, consistent with the autoimmunity model.

To directly evaluate the potential for Rhi occupancy to alter the regulation of adjacent genes, I considered the fraction of genes upregulated or downregulated by the *D. melanogaster* and fusion constructs that also exhibited a corresponding peak of Rhi occupancy. Genes negatively regulated by all three transgenes were enriched for Rhi occupancy, consistent

with a model in which Rhi reduces mRNA transcription (hinge: $\chi^2 = 130.94$, $df = 1$, P value < 10^{-15} , chromo: $\chi^2 = 1,231.8$, $df = 1$, P value < 10^{-15} , mel: $\chi^2 = 12.579$, $df = 1$, P value = 0.00039). However, genes negatively regulated by the hinge and chromo fusion constructs are significantly more enriched for Rhi occupancy, with >18% (118/633) and >35% (95/260) of negatively regulated genes also being occupied by the hinge and chromo fusion Rhi proteins, respectively, whereas only ~13% (51/391) of genes negatively regulated by *D. melanogaster* Rhi are also occupied by the *D. melanogaster* Rhi protein (hinge: $\chi^2 = 5.5$, $df = 1$, P value = 0.019, chromo: $\chi^2 = 45.74$, $df = 1$, P value = 1.35×10^{-11} , fig. 3C vs D). In contrast, upregulated genes are not enriched for Rhi occupancy for any transgenic rescue (fig. 3E–G). In fact, genes occupied by the chromo fusion Rhi protein are underrepresented among positively regulated genes ($\chi^2 = 3.86$, $df = 1$, P value = 0.05, fig. 3G). Taken together, these observations reveal that Rhi occupancy leads exclusively to negative regulation of adjacent genes and that this activity is more pronounced in proteins containing a hinge or chromo domain from *D. simulans*.

Expanded Autoimmunity at the Histone Gene Cluster

To better understand the expanded autoimmunity conferred by the *D. simulans* hinge and chromo domains, I examined the genes that are unique autoimmunity targets (occupied and negatively regulated) of each fusion protein when compared with *D. melanogaster* Rhi (fig. 4A). The majority of novel autoimmunity targets of the hinge fusion protein (68 of 85) and the chromo fusion protein (46 of 48) correspond to copies of replication-dependent histones (fig. 4A and 4B). *Drosophila melanogaster* Rhi also occupies and regulates some histone gene copies, suggesting that histone regulation is a property of Rhi that is shared between species, but has expanded in *D. simulans*. Excluding histones, the *D. melanogaster* and hinge fusion proteins exhibit a similar number of unique autoimmunity targets (17 and 18), whereas the chromo fusion protein exhibits only 2 nonhistone autoimmunity targets (fig. 4A). The expanded autoimmunity of the *D. simulans* hinge and chromo domains is therefore fully attributed to their expanded regulation of histone gene copies.

Replication-dependent histone genes reside in a coregulated tandem array, which includes 20–23 copies of each of the five histone genes. All five histones are negatively regulated (1.4- to 2.8-fold) by the hinge and chromo transgenes, while the *D. melanogaster* transgene exhibits only modest negative regulation of *his1* and *his2A* copies (1.5- and 1.3-fold reduction, respectively; fig. 4B). Rhi-dependent differences in histone regulation could reflect differential occupancy of the histone gene cluster, or differential downstream effects of Rhi occupancy. To discriminate between these alternatives, I examined both Rhi occupancy of the histone gene cluster, and its downstream effects on histone expression and piRNA

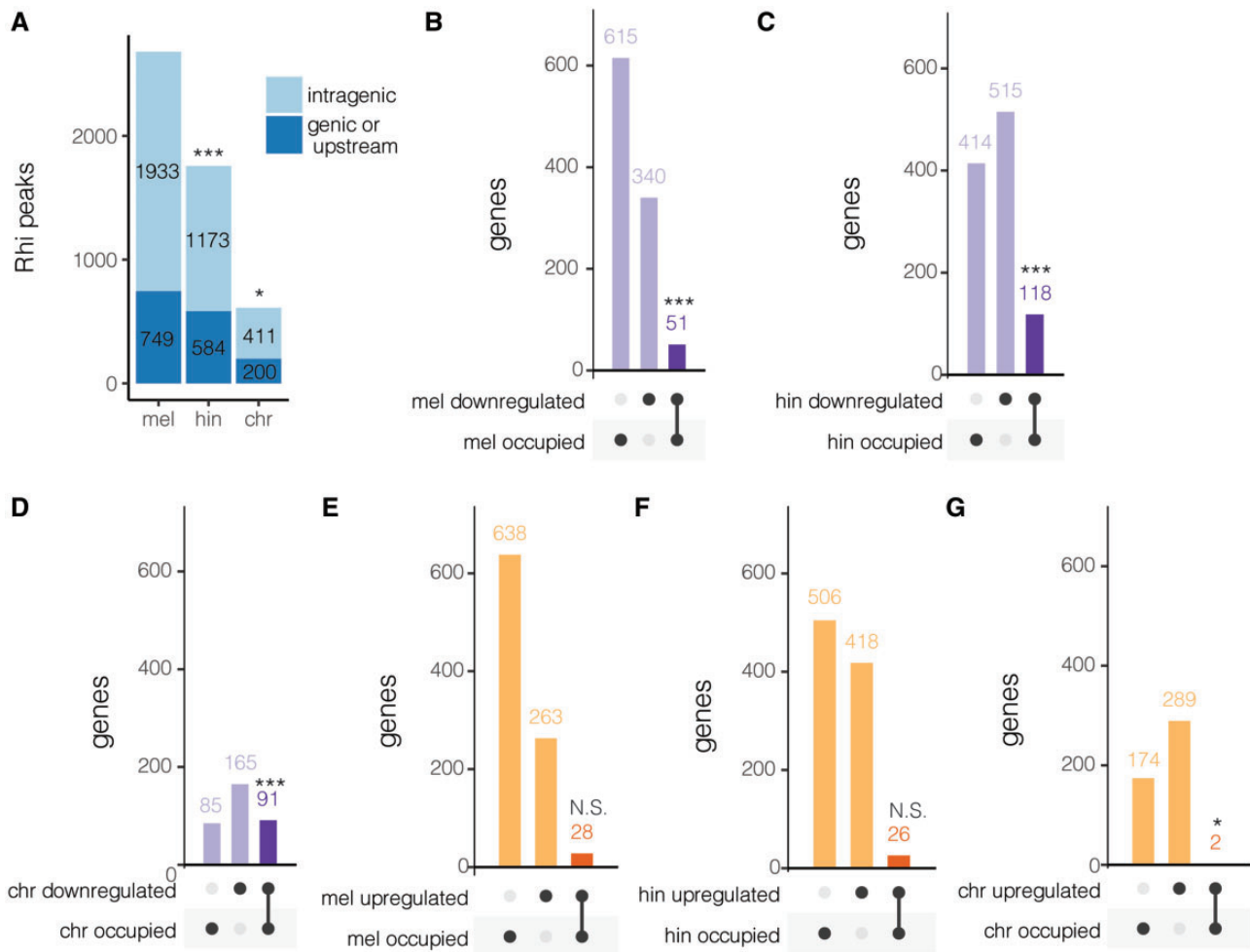


Fig. 3.—The hinge and chromo domain of *D. simulans* Rhi exhibit expanded autoimmunity of *D. melanogaster* genes. (A) Bar graph representing the number of intragenic and genic (including 1 kb upstream of any transcription start site) occupancy peaks of *D. melanogaster*, hinge, and chromo fusion Rhi proteins. (B–D) Upset plots comparing genes occupied and downregulated by *D. melanogaster* (mel, B), hinge fusion (hin, C), and chromo fusion Rhi (chr, D). (E–G) Genes occupied and upregulated by *D. melanogaster* (mel, E), hinge fusion (hin, F), and chromo fusion Rhi (chr, G). Darker colors (B–G) denote an overlapping set of genes in both groups: regulated and occupied, which are the strongest candidates for autoimmunity. *** denotes P value < 0.001 , * denotes P value < 0.05 , and N.S. denotes P value > 0.05 for χ^2 test-of-independence between transgene and occupancy close to genes (A) or regulation and occupancy (B–G).

production (fig. 4B–E). To avoid complications of sequence homology among histone gene copies, these analyses were performed using a genome containing a single representative histone gene cluster, as in McKay et al. (2015). Although two occupancy peaks within the histone cluster are observable for all 5 Rhi proteins, the hinge fusion protein is the most enriched (fig. 4C), and similarly shows the greatest abundance of histone genic piRNAs (fig. 4D and E). By contrast, the *D. melanogaster* protein is significantly enriched only upstream of *his1* (although a nonsignificant peak 3' to *his2A* and *his4* is observable) and exhibits only modest impacts on histone genic piRNAs. Expanded autoimmunity against the *D. melanogaster* histone gene cluster established by the *D. simulans* hinge domain is therefore associated with enhanced Rhi occupancy and downstream piRNA biogenesis.

Differences in histone gene regulation among Rhi proteins are not universally explained by differential occupancy, however. The chromo fusion construct exhibits negative regulation of all five histone genes that exceeds that of the hinge fusion construct (fig. 4A), yet the protein itself is less enriched at the histone cluster (fig. 4C). Therefore, increased negative regulation of histones established by the *D. simulans* chromo domain must occur downstream of occupancy, potentially by altering interactions between Rhi and other proteins.

Drosophila Simulans Cuff Regulates Host Genes through Sequestration of CtBP

Lastly, I considered the unusual regulatory effects of *D. simulans* cuff, which exhibits expanded upregulation of host genes

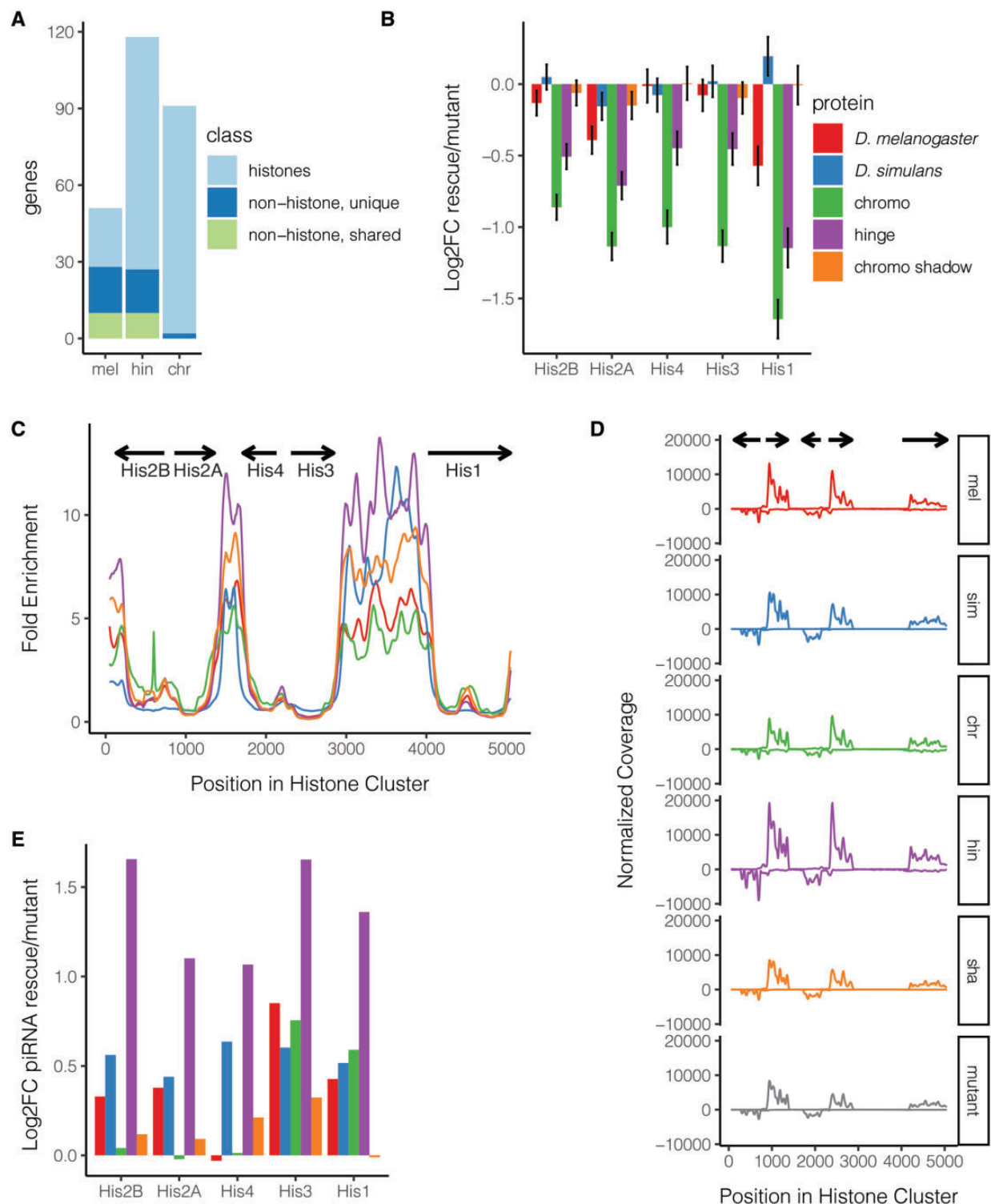


FIG. 4.—Autoimmunity at the histone gene cluster. (A) Autoimmunity targets, which are both occupied and negatively regulated, are compared between the *D. melanogaster* (mel), hinge (hin), and chromo (chr) fusion proteins. (B) Log₂FC expression of histone copies in transgenic rescues as compared with unrescued *rhi* mutants. Due to sequence similarity between histone gene copies as well as their coordinated regulation, estimated read counts were summed across all copies to calculate differential expression. (C) Sliding window of Rhi protein enrichment, as compared with input, across a single representative copy of the histone array (dm6, 2L: 21,482,367–21,487,518). (D) Sliding window analysis of small RNA coverage across a single representative copy of the histone array for *D. melanogaster* (mel), *D. simulans* (sim), chromo (chr), hinge (hin), and chromoshadow (sha) fusion rescues, as well as the unrescued mutant. (E) Log₂FC piRNA abundance for each of the histone genes in transgenic rescues as compared with unrescued mutants.

when compared with the *D. melanogaster* allele (fig. 1B). This expanded upregulation appears modest when compared with the unrescued mutant, impacting the expression of only 12 genes (figs. 1B and 5A, 5B). However, it is revealed as quite dramatic when the gene expression profiles of the *D. simulans* and *D. melanogaster* transgenic rescues are compared with each other (fig. 5C). In total, 159 genes are differentially expressed between the transgenic rescues, 141 of which exhibit higher expression in the presence of *D. simulans cuff*. The two transgenes exhibit opposing effects on the expression of these 141 genes, with expression values being higher in *D. simulans* rescues than mutants, but lower in *D. melanogaster* rescues (fig. 5D).

The disparate impacts of *D. simulans* and *D. melanogaster cuff* on gene expression are potentially explained by the former's sequestration of other *D. melanogaster* proteins into nonfunctional complexes, including the conserved transcriptional coregulator C-terminal-Binding Protein (CtBP; Parhad et al. 2020). Although physical interaction between Cuff and CtBP is required to suppress mRNA transcription at piRNA clusters, sequestration of CtBP could affect its function in regulating protein-coding genes (fig. 1A; Phippen et al. 2000; Fang et al. 2006). To evaluate this possibility, I compared gene expression changes resulting from ovarian *CtBP* knock-down (Parhad et al. 2020; supplementary table S4, Supplementary Material online), to those arising from *D. simulans cuff*. For *CtBP*-KD1 (*P{KK108401}VIE-260B*), upregulated genes were unrelated to those upregulated by *D. simulans cuff* as compared with *D. melanogaster cuff* (Fisher's exact test *P* value = 0.522, fig. 5E). However, for *CtBP*-KD2 (*P{GD4268}v37609*) upregulated genes are highly significantly enriched for those upregulated by *D. simulans cuff*, with 41 genes commonly upregulated in both genotypes ($\chi^2 = 451.05$, *df* = 1, *P* value < 10^{-15} , fig. 5F). Although differences in gene regulation between the two knockdowns could be explained by an off-target effect of one or both constructs, it is interesting that *CtBP*-KD1 has a stronger impact on *CtBP* expression than *CtBP*-KD2 (74% as compared with 30% decrease in expression [Parhad et al. 2020]). Thus, an intriguing alternative explanation for this inconsistency is that CtBP's effects on genic targets of transcriptional repression are highly dosage-dependent and that *D. simulans cuff* is more similar to a mild reduction in *CtBP* function.

Although CtBP is often considered a transcriptional corepressor, it can also act as a transcriptional coactivator (Fang et al. 2006; Bhambhani et al. 2011). I therefore also compared genes that were downregulated by *D. simulans cuff* and *CtBP* KD, whose activated expression might depend on *CtBP*. Although only 18 genes were downregulated by *D. simulans cuff*, these were significantly enriched for genes downregulated by both *CtBP* knockdowns (*CtBP*-KD1: $\chi^2 = 188.15$, *df* = 1, *P* value < 10^{-15} , *CtBP*-KD2: Fisher's Exact Test *P* value = 0.006641, fig. 5G–H). Taken together, my observations suggest that the considerable impact of *D. simulans cuff* on host

gene regulation may be partly explained by its sequestration of CtBP. Genes regulated by *D. simulans cuff* but not *CtBP* could be targets of other sequestered proteins, such as TRF2 (Parhad et al. 2020).

Discussion

Here I examined the potential for genomic autoimmunity to shape interspecific divergence in the RDC complex: a key regulator of piRNA precursor transcription. For both Rhi and Cuff, I observed expanded regulation of *D. melanogaster* genes by a *D. simulans* protein or domain, consistent with the autoimmunity model. In the case of Rhi, *D. simulans* hinge and chromo domains establish enhanced negative regulation of a single locus: the histone gene cluster. By contrast, *D. simulans cuff* promotes the increased expression of numerous protein-coding genes throughout the genome, potentially through the sequestration of the transcriptional coregulator CtBP. The unifying and novel observation from both of these analyses is that autoimmunity does not occur through the TE-dependent recruitment of piRNA machinery to genomic sites, as was originally proposed (Blumenstiel et al. 2016). Rather, I propose that shared regulatory machinery between piRNA clusters and host genes provides opportunities nonfunctional interactions that give rise to autoimmunity.

I discovered that both the *D. melanogaster* and *D. simulans* Rhi proteins, as well as all of their fusion proteins, localize to the histone gene cluster (fig. 4C). Although the histone gene cluster does not contain TEs, it employs a noncanonical transcriptional program that shares some features with piRNA transcription. Rhi may therefore be recruited to the histone gene cluster through interactions with shared regulatory factors. In particular, the transcription of both piRNA precursors and *his1* is initiated by TRF2, although in the former case TRF2 is recruited by Rhi and Del rather than direct binding to the promoter (Isogai et al. 2007; Andersen et al. 2017). Rhino's association with TRF2 may therefore recruit Rhi indirectly to *his1* promoters. Consistent with this model, the largest Rhi peak for all transgenes occurs upstream of *his1*, and *his1* is also the most strongly negatively regulated histone by the *D. melanogaster* transgene and the chromo and hinge fusion constructs (fig. 4A). However, because the association between TRF2 and Rhi is thought to be mediated through Del, TRF2 association cannot explain the recruitment of the *D. simulans* and chromo shadow fusion Rhi proteins to the histone gene cluster (fig. 4C).

A non-mutually exclusive alternative mechanism for Rhi recruitment is the histone modification H3K9me3, which is bound by Rhi to initiate piRNA precursor transcription (Le Thomas et al. 2014; Mohn et al. 2014). The histone methyltransferase *Su(var)3-9* localizes to the histone gene cluster in salivary glands, and *Su(var)3-9* mutants exhibit increased histone expression (Ner et al. 2002). Given the role of *Su(var)3-9* in depositing H3K9me3 during oogenesis (Yoon et al. 2008),

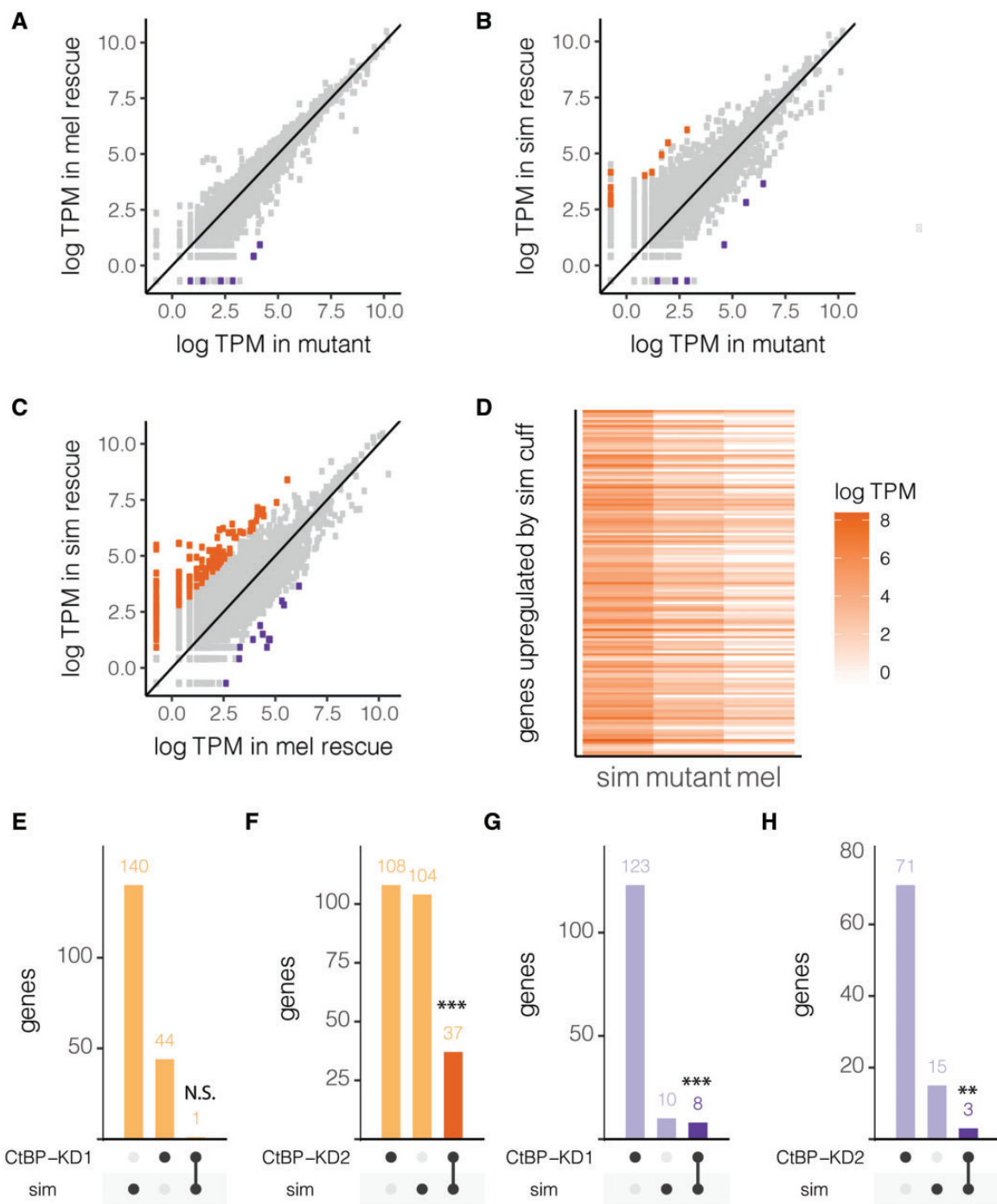


FIG. 5.—*Drosophila simulans* cuff upregulates numerous host genes, potentially by sequestering CtBP. (A–C) Correlation plots of the log-scale gene expression levels in transcripts per million (TPM) between the three *cuff* genotypes: unrescued mutant (mut), *D. melanogaster* (mel), and *D. simulans* (sim) transgenic rescues. (D) Log expression levels (TPM) are compared between all three genotypes for 141 genes upregulated by the *D. simulans* (sim) as compared with the *D. melanogaster* (mel) *cuff* transgenic rescue. (E–H) Upset plots comparing genes upregulated (E and F) and downregulated (G and H) by *D. simulans* as compared with *D. melanogaster* transgenic rescues and in CtBP knockdown as compared with control flies. CtBP-KD1:*P*{*KK108401*}*VIE-260B*, CtBP-KD2:*P*{*GD4268*}*v37609*. Results of Fisher’s exact test (CtBP-KD2, downregulated) or χ^2 test-of-independence (all others, *df* = 1) indicating the significance of overlap in upregulated or downregulated genes between *D. simulans* *cuff* transgenic rescues CtBP-KD are also indicated. N.S. denotes *P* value > 0.05, ** denotes *P* value < 0.01, and *** denotes *P* value < 0.001.

it seems likely that this heterochromatic mark also occurs at the histone gene cluster in ovaries, potentially leading to Rhi recruitment. Regardless of the mechanism, my observation suggests that Rhi-dependent autoimmunity arises through the recruitment of Rhi to nontarget sites by physical interactions between Rhi and other regulatory proteins.

In contrast, *cuff* autoimmunity impacts the expression of genes throughout the genome, with the foreign *D. simulans* protein upregulating the expression of hundreds of genes when compared with its *D. melanogaster* counterpart (fig. 5C). This is quite distinct from an autoimmunity model in which piRNA machinery is recruited to off-target sites where they establish silencing. The positive impact of *D. simulans* Cuff on many genes may arise from its enhanced affinity for CtBP and other transcriptional coregulators (Parhad et al. 2020). If Cuff traps these proteins in nonfunctional complexes, thereby reducing their availability for gene regulation (Parhad et al. 2020), the expression of target genes throughout the genome would be impacted. In support of this model, there is significant concordance between genes up or down-regulated by *D. simulans cuff* and those repressed or activated by CtBP function (fig. 5E–G). Thus, in the case of *cuff*, the foreign *D. simulans* protein may disrupt gene regulation indirectly through its affinity for transcriptional coregulators that play accessory roles in piRNA biogenesis.

Selection against genomic autoimmunity is proposed to accompany selection for genome defense in driving and adaptive evolution of piRNA proteins. Specifically, invading or escaping TEs select for expanded piRNA-mediated regulation, and subsequently, compensatory mutations may arise that decrease off-target effects (Blumenstiel et al. 2016). The phenotype of *D. simulans cuff* is consistent with this model. Enhanced affinity for cofactors such as CtBP might have facilitated expanded TE silencing in the *D. simulans* lineage. Subsequently, regulatory changes elsewhere in the system, such as increased abundance of CtBP and shared cofactors, would resolve impacts on gene regulation.

My observations with Rhi also support a tension between the robustness and specificity of defense. The histone array is occupied by all Rhi proteins (fig. 4C), but expanded negative regulation of histone genes is associated with both the hinge and chromo domains of *D. simulans* Rhi (fig. 4B). Selection may therefore have acted to reduce histone repression by the *D. melanogaster* protein. In support of this, while fitness effects of the chromo and hinge fusion constructs were not extensively studied (fig. 2B; Parhad et al. 2017), repression of histone transcripts can have serious fitness consequences because maternally transmitted histone mRNAs are required for early zygotic cell divisions (Sullivan et al. 2001). Interestingly, all three rhino domains contain amino acids that have evolved adaptively across the *melanogaster* group, and the hinge and chromo shadow domains in particular show excess amino acid substitution between *D. melanogaster* and *D. simulans* (Vermaak et al. 2005). I propose that while the interface with

Del may explain selection in the chromo shadow domain as suggested previously (fig. 2A; Parhad et al. 2017), and that selection against histone autoimmunity may be promoting divergence of the hinge and chromo domains.

Materials and Methods

Data Sets and Quality Control

rhi and *del* ovarian ribo-depleted and stranded RNA-seq, small RNA-seq, and ChIP seq (*rhi* only) data sets are from Parhad et al. (2017). *cuff* and *CtBP* ovarian RNA-seq data were from Parhad et al. (2020). All Illumina libraries downloaded and analyzed are described in [supplementary table S1, Supplementary Material](#) online. Data were downloaded from the NCBI Sequenced Read Archive. Adaptors were removed and low-quality bases were trimmed from all raw-reads using trim-galore (Krueger 2015).

RNA-Seq Analysis

RNA-seq reads were aligned to release 6.33 of the *D. melanogaster* transcriptome using Kallisto (Bray et al. 2016), in order to estimate the abundance of each transcript. The estimated number of reads was then summed across all transcripts from the same gene to obtain the estimated read count for each gene. For histone gene copies, the estimated number of reads was further summed across all copies of the gene, because reads cannot be reliably assigned to individual copies. Genic read counts were then used to estimate differential expression.

For *del* and *cuff* mutants and transgenic rescues, as well as for *CtBP* knockdown and control flies (*white* knockdown) only one biological replicate was available for each genotype. We therefore used DESeq to estimate differential expression (method="blind," sharingMode="fit-only"; Anders and Huber 2010), and significant differences were detected using a negative binomial test. Genes with fewer than 50 reads in all samples were excluded. For *rhi* mutants and transgenic rescues, two biological replicates were available for each genotype. We therefore estimated differential expression and detected statistical significance with DESeq2 (Love et al. 2014). Genes with fewer than 50 average reads across samples were excluded. Regardless of the analysis package, a gene was considered differentially expressed if the adjusted *P* value was less than 0.05.

ChIP-Seq Analysis

ChIP Seq reads were aligned to a reference genome using BWA (Li and Durbin 2009). To avoid complications of multiply-mapping reads in the histone gene cluster, reads were aligned to a custom version of the dm6 reference genome in which the histone array (2L: 21,403,672–21,543,688) was replaced with a single representative copy

of the histone repeat (2L: 21,482,367–21,487,518), as in McKay et al. (2015). Peaks of Rhi occupancy were detected using MACS2 with broad-peaks settings at a significance cut-off of 0.1 (Zhang et al. 2008; Feng et al. 2012). A gene was considered occupied by Rhi if a peak occurred within 1,000 nt of a transcription start site, or anywhere in the transcript body inclusive of introns, based on flybase annotated transcripts.

To generate sliding window analyses of Rhi occupancy of the histone gene cluster, I first extracted read alignments overlapping the cluster via samtools (Li et al. 2009). I then calculated the nucleotide coverage, normalized to the number of aligned sequencing reads, using bedtools genomecov (Quinlan and Hall 2010). Sliding window estimates of mean coverage, relative to input were then calculated using the rollapply function from the zoo package (Zeileis and Grothendieck 2005) in R version 3.6.1 (R Development Core Team 2008).

small-RNA Analysis

Adapters were trimmed from small RNAs and putative miRNAs (18–22 nt) and piRNAs (23–32 nt) were identified using trim galore (Krueger 2015). Putative miRNAs were then aligned to all annotated miRNAs in the (dm6) reference assembly, whereas piRNAs were aligned to the custom reference with a single copy of the histone array. Sliding window analyses of piRNA abundance across the histone gene cluster were performed as with the ChIP-seq data except the coverage was normalized to the number of reads aligning to miRNAs from the same library. Similarly, differential piRNA abundance of individual histone genes was generated by counting the number of reads overlapping each annotated transcript using samtools (Li and Durbin 2009) and normalizing to the number of reads aligning to miRNAs from the same library.

Statistics and Data Visualization

Statistical testing was performed in R version 3.6.1 (R Development Core Team 2008). Data were wrangled using tidyverse (Wickham et al. 2019) and represented using UpSetR (Conway et al. 2017) and ggplot2 (Wickham 2011).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

I am grateful to the members of my lab group, Justin Blumenstiel, and one anonymous reviewer for thoughtful comments on the earlier version of this manuscript. I was supported by NIH R35-GM138112. I was supported by R35-GM138112 from the National Institute of Health.

Data Availability

All Illumina libraries analyzed in this work are previously published and were downloaded from the NCBI Sequenced Read Archive. Analyzed libraries are described in [supplementary table S1, Supplementary Material](#) online.

Literature Cited

- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11(10):R106.
- Andersen PR, Tirian L, Vunjak M, Brennecke J. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549(7670):54–59.
- Aruna S, Flores HA, Barbash DA. 2009. Reduced fertility of *Drosophila melanogaster* hybrid male rescue (Hmr) mutant females is partially complemented by Hmr orthologs from sibling species. *Genetics* 181(4):1437–1450.
- Bhambhani C, Chang JL, Akey DL, Cadigan KM. 2011. The oligomeric state of CtBP determines its role as a transcriptional co-activator and co-repressor of Wingless targets. *EMBO J.* 30:2031–2043.
- Blumenstiel JP, Erwin AA, Hemmer LW. 2016. What drives positive selection in the *Drosophila* piRNA machinery? The genomic autoimmunity hypothesis. *Yale J Biol Med.* 89(4):499–512.
- Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Curr Biol.* 28(8):1289–1295.e4.
- Bray N, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal RNA-Seq quantification with kallisto. *Nat Biotechnol.* 34(5):525–527.
- Chen Y-CA, et al. 2016. Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. *Mol Cell.* 63(1):97–109.
- Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938–2940.
- Crysnanto D, Obbard DJ. 2019. Widespread gene duplication and adaptive evolution in the RNA interference pathways of the *Drosophila obscura* group. *BMC Evol Biol.* 19: 99.
- Czech B, et al. 2018. piRNA-guided genome defense: from biogenesis to silencing. *Annu Rev Genet.* 52:131–157.
- El Baidouri M, Panaud O. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol Evol.* 5(5):954–965.
- Fang M, et al. 2006. C-terminal-binding protein directly activates and represses Wnt transcriptional targets in *Drosophila*. *EMBO J.* 25(12):2735–2745.
- Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* 7(9):1728–1740.
- Flores HA, Bubnell JE, Aquadro CF, Barbash DA. 2015. The *Drosophila* bag of marbles gene interacts genetically with *Wolbachia* and shows female-specific effects of divergence. *PLoS Genet.* 11:e1005453.
- Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464(7293):1347–1350.
- Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R. 2007. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* 21(22):2936–2949.
- Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 80(6):1655–1659.
- Kolaczowski B, Hupaló DN, Kern AD. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol Biol Evol.* 28(2):1033–1042.

- Krueger F. 2015. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. *Trim Galore*. 516(517):
- Le Thomas A, et al. 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev*. 28(15):1667–1680.
- Lewis SH, Salmela H, Obbard DJ. 2016. Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine. *Genome Biol Evol*. 8(3):507–518.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, et al.; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15(12):550.
- McKay DJ, et al. 2015. Interrogating the function of metazoan histones using engineered gene clusters. *Dev Cell*. 32(3):373–386.
- Mohn F, Sienski G, Handler D, Brennecke J. 2014. The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157(6):1364–1379.
- Naito K, et al. 2006. Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci U S A*. 103(47):17620–17625.
- Ner SS, Harrington MJ, Grigliatti TA. 2002. A role for the *Drosophila* SU(VAR)3-9 protein in chromatin organization at the histone gene cluster and in suppression of position-effect variegation. *Genetics* 162(4):1763–1774.
- Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci*. 364(1513):99–115.
- Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*. 20(2):89–108.
- Palmer WH, Hadfield JD, Obbard DJ. 2018. RNA-interference pathways display high rates of adaptive protein evolution in multiple invertebrates. *Genetics* 208(4):1585–1599.
- Parhad SS, et al. 2020. Adaptive evolution targets a piRNA precursor transcription network. *Cell Rep*. 30(8):2672–2685.e5.
- Parhad SS, Tu S, Weng Z, Theurkauf WE. 2017. Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. *Dev Cell*. 43(1):60–70.e5.
- Phippen TM, et al. 2000. *Drosophila* C-terminal binding protein functions as a context-dependent transcriptional co-factor and interferes with both mad and groucho transcriptional repression. *J Biol Chem*. 275(48):37628–37637.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- R Development Core Team. 2008. R: a language and environment for statistical computing. Available from: <http://www.r-project.org>
- Reiss D, et al. 2019. Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. *PLoS Genet*. 15(2):e1007965.
- Senti K-A, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet*. 26(12):499–509.
- Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD. 2013. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution* 67(4):1081–1090.
- Smothers JF, Henikoff S. 2001. The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. *Mol Cell Biol*. 21(7):2555–2569.
- Sullivan E, et al. 2001. *Drosophila* stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes Dev*. 15:173–187.
- Vermaak D, Henikoff S, Malik HS. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet*. 1(1):96–108.
- Vermaak D, Malik HS. 2009. Multiple roles for heterochromatin protein 1 genes in *Drosophila*. *Annu Rev Genet*. 43:467–492.
- Wang L, Barbash DA, Kelleher ES. 2020. Adaptive evolution among cytoplasmic piRNA proteins leads to decreased genomic auto-immunity. *PLoS Genet*. 16(6):e1008861.
- Wickham H. 2011. ggplot2. *WIREs Comp Stat*. 3(2):180–185.
- Wickham H, et al. 2019. Welcome to the Tidyverse. *J Open Source Softw*. 4(43):1686.
- Yang H-P, Barbash DA. 2008. Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. *Genome Biol*. 9(2):R39.
- Yi M, et al. 2014. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. *Genome Biol Evol*. 6(6):1393–1407.
- Yoon J, et al. 2008. dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in *Drosophila melanogaster*. *PLoS One*. 3(5):e2234.
- Yu B, et al. 2015. Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell Res*. 25(4):525–528.
- Yu B, et al. 2018. Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. *EMBO Rep*. 19(7):. doi:10.15252/embr.201745418.
- Zeileis A, Grothendieck G. 2005. zoo: S3 infrastructure for regular and irregular time series. *J Stat Softw*. 14:1–27.
- Zhang Y, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 9(9):R137.
- Zhang Z, et al. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157(6):1353–1363.

Associate editor: Esther Betran