Adaptive evolution among cytoplasmic piRNA proteins leads to decreased genomic auto-immunity

Luyang Wang[1], Daniel A. Barbash[2], Erin S. Kelleher[1]

1 Dept. Biology & Biochemistry, University of Houston, Houston, TX, United States of America

2 Dept. Molecular Biology & Genetics, Cornell University, Ithaca, NY, United States of America

Abstract

In metazoan germlines, the piRNA pathway acts as a genomic immune system, employing small RNA-mediated silencing to defend host DNA from the harmful effects of transposable elements (TEs). In response to dynamic changes in TE content, host genomes are proposed to alter the piRNAs that they produce in order to silence the most active TE families. Surprisingly, however, piRNA pathway proteins, which execute piRNA biogenesis and enforce silencing of targeted sequences, also evolve rapidly and adaptively in animals. If TE silencing evolves through changes in piRNAs, what necessitates changes in piRNA pathway proteins? Here we used interspecific complementation to test for functional differences between *Drosophila melanogaster* and *D. simulans* alleles of three adaptively evolving piRNA pathway proteins: Armitage, Aubergine and Spindle-E. Surprisingly, we observed interspecific divergence in the regulation of only a handful of TE families, which were more robustly silenced by the heterospecific piRNA pathway protein. This suggests that positive selection does not act on piRNA effector proteins to enhance their function in TE repression, but rather that TEs may evolve to “escape” silencing by homospecific alleles. We also discovered that *D. simulans* alleles of *aub* and *armi* exhibit enhanced off-target effects on host transcripts in a *D. melanogaster* background, suggesting the avoidance of genomic auto-immunity as a critical target of selection. Our observations suggest that piRNA effector proteins are subject to an evolutionary trade-off between defending the host genome from the harmful effect of TEs while also minimizing friendly fire against host genes.

Author Summary

Transposable elements are mobile fragments of selfish DNA that burden host genomes with deleterious mutations and incite genome instability. Host cells employ a specialized small-RNA mediated silencing pathway, the piRNA pathway, to act as a genomic immune system suppressing the mobilization of TEs. Changes in genomic TE content are met with rapid changes in the piRNA pool, thereby maintain host control over transposition. However, piRNA pathway proteins—which enact piRNA biogenesis and silence target TEs—also evolve adaptively. To isolate forces that underlie this adaptive evolution, we examined functional divergence between two *Drosophila* species for three adaptively evolving piRNA pathway proteins. To our surprise, we found very few differences in TE regulation, suggesting that evolution has not generally acted to enhance control of TE parasites. Rather, we discovered interspecific differences in the regulation of host mRNAs for two proteins, which suggested that proteins evolve to avoid off-target silencing of host transcripts. We propose that the avoidance of such “genomic autoimmunity” is an important and underappreciated force driving the adaptive evolution of piRNA proteins.

Introduction

Transposable elements (TEs) are ubiquitous mobile genetic entities, whose unrestricted propagation can cause deleterious insertional mutations and chromosome rearrangements, and is often associated with cancer and sterility [1–4]. TE regulation is therefore essential, especially in germline cells, where TE insertions and associated mutations can be transmitted to the next generation. In metazoan germlines, regulation of TE transcripts is enacted by a small RNA silencing pathway, the PIWI-interacting RNA pathway (piRNA pathway), in which piRNAs complexed with PIWI-clade Argonaute proteins target complementary TEs for post-transcriptional and transcriptional silencing [5].

Host genomes are often parasitized by multiple TE families, which change rapidly in their presence and abundance [6–9]. The control of TE transcripts by complementary piRNAs may facilitate adaptation to genomic TEs through changes in piRNA species [10,11]. Surprisingly, however, the protein components of the piRNA pathway that enact piRNA biogenesis and enforce TE silencing also evolve adaptively in diverse metazoan lineages [12–16]. Evidence for adaptive evolution of piRNA pathway proteins is particularly strong in *Drosophila* [12–15,17], which has also emerged as a work horse for uncovering the mechanisms of piRNA-mediated silencing [reviewed in 18]. For example, a recent meta analysis including both *D. melanogaster* and *D. pseudoobscura* revealed that 22 of 26 piRNA pathway proteins exhibit significant signatures of adaptive protein evolution in one or both species [14].

Adaptive evolution of piRNA effector proteins is proposed to arise from an evolutionary arms race between TEs and host silencing machinery [reviewed in 19]. In the simplest scenario, effector proteins evolve adaptively in order to restore silencing of newly invading or escaper TE families. Alternatively, if TEs “fight back” by encoding RNA or protein antagonists of host silencing machinery, piRNA pathway proteins could evolve adaptively to escape TE antagonism [20]. Finally, piRNA proteins may evolve adaptively to avoid “genomic auto-immunity” in the form of off-target silencing of host genes [19,21]. Uncovering which of these selective forces

drives the adaptive evolution of piRNA effector proteins requires elucidating the resulting functional consequences of piRNA-effector-protein divergence. For example, adaptive evolution among transcriptional silencers has led to incompatibilities between alleles of interacting proteins from different species, with dramatic consequences for piRNA production and TE control [20,22,23]. In particular, functional changes in Rhino are proposed to reflect evolutionary escape from a TE-encoded antagonist [20].

Here, we broaden our understanding of the functional consequences of adaptive evolution among *Drosophila* piRNA effector proteins by examining three additional essential piRNA pathway components that play critical roles in piRNA maturation and post-transcriptional silencing [24–30]: Armitage (Armi), Aubergine (Aub) and Spindle-E (spn-E). This work significantly extends a preliminary analysis of Aub divergence [31]. Aub is a Piwi-clade Argonaute protein which, guided by piRNAs, enacts post-transcriptional silencing of sense TE-derived mRNAs [24]. Aub cleavage also feeds forward the ping-pong amplification cycle, a core mechanism for the maturation of both sense and antisense piRNAs that also requires Spn-E [25,26,28,32]. Distinct from both Aub and Spn-E, Armi binds to antisense piRNA precursors to facilitate their sequential cleavage by the nuclease Zucchini in an alternate biogenesis mechanism referred to as “phasing” [29,33–37]. The loci encoding Aub, Spn-E and Armi all exhibit adaptive evolution along the lineage leading to *D. melanogaster, D. simulans* or both, yet the underlying evolutionary force(s) remain unknown [13,15].

To isolate diverged functions of these adaptively evolving proteins, we performed interspecific complementation, in which we compared the ability of *D. melanogaster* and *D. simulans* wild-type alleles to complement a *D. melanogaster* mutant background. While nuclear transcriptional silencers were previously demonstrated to exhibit dramatic interspecific divergence in TE regulation and piRNA production [20,23], we observed only minor allelic differences in both of these functions. Rather, we uncovered idiosyncratic differences in the regulation of a small handful of TEs suggesting potential element-specific adaptations. We also

observed that *D. simulans* alleles of *aub* and *armi* exhibit reduced efficiency of piRNA maturation in association with increased off-target regulation of host mRNAs. We propose that in contrast to nuclear transcriptional silencers, selection acts cytoplasmic piRNA proteins to maximize their specificity to piRNA production and TE transcripts, while minimizing non-functional or deleterious interactions with host mRNA.

Results

Identifying functional divergence through interspecific complementation.

Previous divergence-based analyses of Aub and Spn-E suggest that adaptive evolution is not confined to a particular functional domain but is dispersed throughout the proteins [12,15]. Consistent with these findings, we identified abundant amino-acid differences between *D. melanogaster* and *D. simulans* throughout Aub and Spn-E (Fig 1A). Armi does not exhibit strong evidence of positive selection in divergence-based tests, however, an excess of amino acid substitutions exists between *D. melanogaster* and *D. simulans*, which have likely arisen by positive selection in one or both lineages [15]. Similar to Aub and Spn-E we observe that these fixed differences are scattered throughout the protein, both inside and outside of functional domains (Fig 1A).

To isolate phenotypic differences between *D. melanogaster* and *D. simulans* alleles that result from adaptive evolution, we employed interspecific complementation, in which we compared the ability of *D. melanogaster* and *D. simulans* wild-type alleles to complement a *D. melanogaster* mutant background. For each selected piRNA protein, we generated and compared three genotypes: 1) a trans-heterozygous loss-of-function mutant, 2) a mutant with a *D. melanogaster* genomic transgene rescue, and 3) a mutant with a *D. simulans* genomic transgene rescue (S1 Fig). The transgenes include the complete genomic region from either *D. melanogaster* or *D. simulans* including upstream and downstream sequences containing potential cis-regulatory elements. Furthermore, transgenes were inserted into matched *attP*

sites by ΦC31 integrase [38], in order to avoid position effects. Phenotypes for which the *D. simulans* alleles fail to fully complement the mutant, or otherwise differ between the alleles of the two species, point to diverged functions that are potential targets of adaptive evolution.

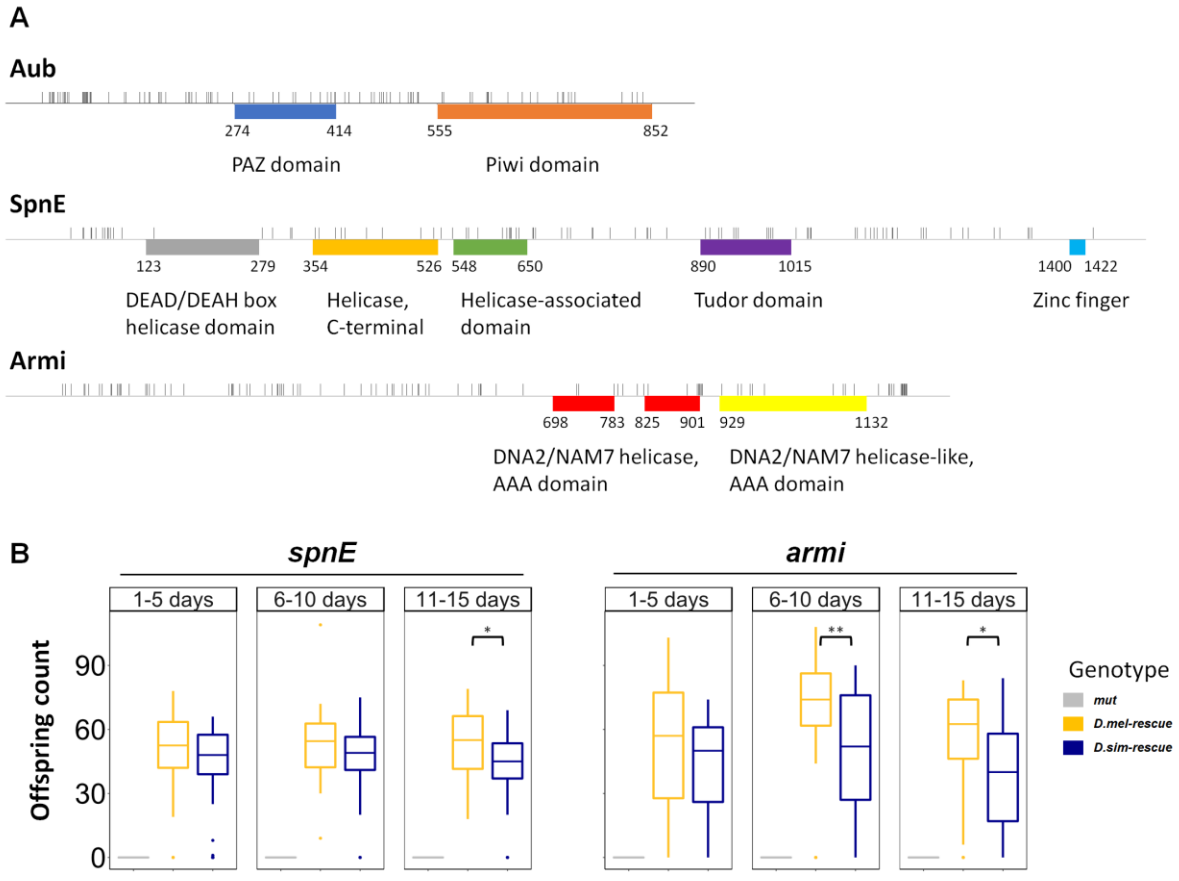


Fig 1. Functional and sequence divergence in piRNA pathway proteins. (A) Amino acid substitutions between *D. melanogaster* [40] and *D. simulans* [41] reference alleles are indicated as grey tick marks. Starting and ending amino acids for InterPro [42] annotated functional domains are indicated. (B) Female fertility for *D. melanogaster* and *D. simulans* transgenic rescues are compared for three different age classes. Females with the *D. simulans spn-E* transgene are significantly less fertile across the experiment (Repeated measures ANOVA, $F_{1,172} = 4.043$, $p < 0.05$) and at the third time point we measured (11-15 days, $t_{56} = 2.304$, $p < 0.05$). Females with the *D. simulans armi* transgene are significantly less fertile across the experiment (Repeated measures ANOVA, $F_{1,175} = 8.824$, $p < 0.01$) and

at the second time (06-10 days, t_{57} = 3.0718, $p < 0.01$) and the third time point we measured (11-15 days, t_{57} = 2.5915, $p < 0.05$). Samples sizes are 25-35 females. * denotes $p \leq 0.05$. ** denotes $p \leq 0.01$.

We first considered the effect of transgenic rescue on female fertility. Homozygosity or trans-heterozygosity for loss of function alleles in all three genes causes complete female sterility (Fig 1B, S1 Table), while heterozygotes are fertile [39]. For all three proteins, fertility is restored by transgenic rescues from the two species to different degrees, with *D. melanogaster* transgenes conferring higher fertility than their *D. simulans* counterparts (Fig 1B, S1 Table) [31]. Because *D. simulans* transgenes do not exhibit significantly reduced expression when compared to *D. melanogaster* (S2 Fig), these fertility effects most likely reflect amino acid sequence divergence in the encoded proteins.

Idiosyncratic differences in TE regulation

To uncover molecular phenotypes that relate to fertility differences, we first examined whether *D. melanogaster* and *D. simulans* alleles differed with respect to TE transcriptional control and associated piRNA production using RNA-seq and small RNA-seq (S5 Fig, S2 Table). Enhanced piRNA-mediated negative regulation of TEs is an obvious target of positive selection acting on piRNA pathway proteins. Furthermore, mutants for all three proteins exhibit dramatic decreases in piRNA abundance for the majority of TE families, and corresponding increases in target TE transcripts, when compared to wild-type flies [26,32,43].

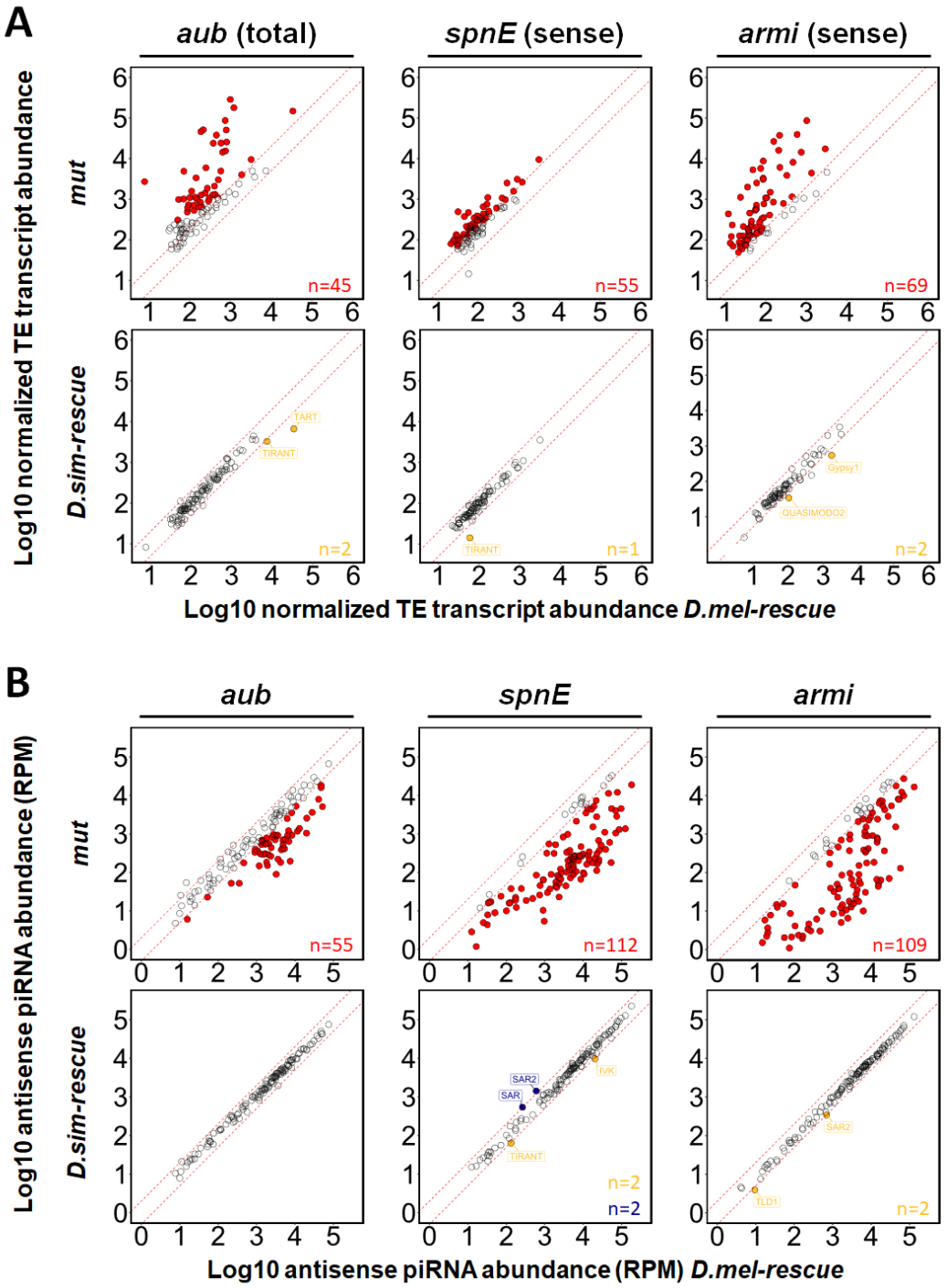
We compared TE transcript abundance between mutants and transgenic rescues using mRNA-seq (*aub*) and stranded total RNA-seq (*spn-E* and *armi*) (S3 Table). For total RNA-seq data, we focused on the regulation of sense TE transcripts that code for protein and act as transposition intermediates for retrotransposons. For all three proteins, *D. melanogaster* transgenic rescues greatly reduce transcript abundance for the majority of TE families when

compared to the corresponding mutant (Fig 2A). Surprisingly, however, in all three cases the majority of TEs are not differentially expressed between the transgenic rescues, implying that negative regulation of TEs is fairly conserved between species (Fig 2A). Importantly, we did not observe any systematic differences in expression for germline or soma-specific protein-coding genes between the transgenic rescues, strongly suggesting that the germline-to-soma ratio is consistent between transgenes (S2 Fig).

Despite an overall conservation of TE repression, we discovered idiosyncratic differences in regulation, in which individual TE families are more robustly silenced by the *D. melanogaster* or *D. simulans* allele. While we cannot rule out that some differences in TE regulation could reflect residual background effects, 5 out of 5 TE families whose transcript abundance differs between transgenic rescues for one of the three proteins are more robustly silenced by the *D. simulans* allele. This bias towards enhanced negative regulation by *D. simulans* is not consistent with the random segregation of TE copies during backcrossing. Furthermore, the *tirant* LTR retrotransposon is more robustly silenced by the *D. simulans* allele of both *aub* and *spn-E.* Differences in silencing of individual TE families suggest lineage-specific coevolution between host-regulatory machinery and individual TE families.

Divergence in TE regulation between *D. melanogaster* and *D. simulans* alleles could arise from differential production of corresponding antisense piRNAs, or from altered interaction of protein alleles with target transcripts. To differentiate between these alternatives, we compared antisense TE-derived piRNAs between mutants and transgenic rescues for all three proteins (Fig 2B, S4 Table). Notably, when compared to mutants, *D. melanogaster* transgenic rescues increased piRNA abundance similarly to heterozygotes, which are generally considered wild-type with respect to piRNA production and silencing (S3A Fig). *Drosophila melanogaster* transgenic rescues also establish similar profiles of ping-pong and phasing piRNA biogenesis as compared to heterozygotes (S3B Fig). Transgenically expressed proteins are therefore fully functional with respect to piRNA production.

Of 5 TE families that are differentially regulated by the transgenic rescues of the two species (Fig 2A), only *tirant* exhibited a corresponding change in antisense piRNA abundance (S4 Fig). Furthermore, similar to their mRNA targets, *tirant* antisense piRNAs were increased in the *D. melanogaster* transgene, which is not consistent with piRNA loss as a cause of TE derepression. Differences in TE regulation between alleles therefore occur independently of piRNA abundance, and may reflect direct interactions between proteins and TE transcripts. Furthermore, *D. simulans* alleles have very similar effects on the overall pool of TE-derived piRNAs as *D. melanogaster*, with only five repeat classes differentially abundant between any of the three pairs of transgenic rescues, two of which are satellite repeats (SAR and SAR2, Fig 2B, S4 Fig). Interspecific divergence in piRNA production is therefore modest between alleles, with no detectable impact on the regulation of genomic TEs.



1 **Fig 2. Minimal differences in TE regulation and piRNA production between alleles.** TE transcript
2 abundance (A) and TE-derived antisense piRNA abundance (B) is compared between *D. melanogaster*
3 rescues and trans-heterozygous mutants ("mut", upper row) or *D. simulans* rescues (lower row) for *aub*,
4 *spn-E* and *armi*. Red dashed lines indicate the 2 fold-change threshold. TE families whose abundance

differs significantly between mutants and *D. melanogaster* rescues are indicated in red ($p > 0.05$ for TE transcripts). *P*-values were not considered for small RNA analysis because the small number of TE families in the analysis (<130 families, S4 Table) is unlikely to provide a sufficiently large sample size for the statistical model implemented in DESeq2 [44]. TE families whose abundance is higher in *D. melanogaster* rescues than in *D. simulans* rescues are in yellow, whereas the reciprocal is in blue. mRNA abundance is based on one biological replicate (*aub*), sense total RNA abundance is based on three biological replicates (*spn-E* and *armi*). TE-derived piRNA abundance is based on two biological replicates for *aub* and three biological replicates for *spn-E* and *armi*, and was normalized to the total number of sequenced miRNAs in the same library.

***Drosophila simulans* alleles exhibit reduced piRNA biogenesis.**

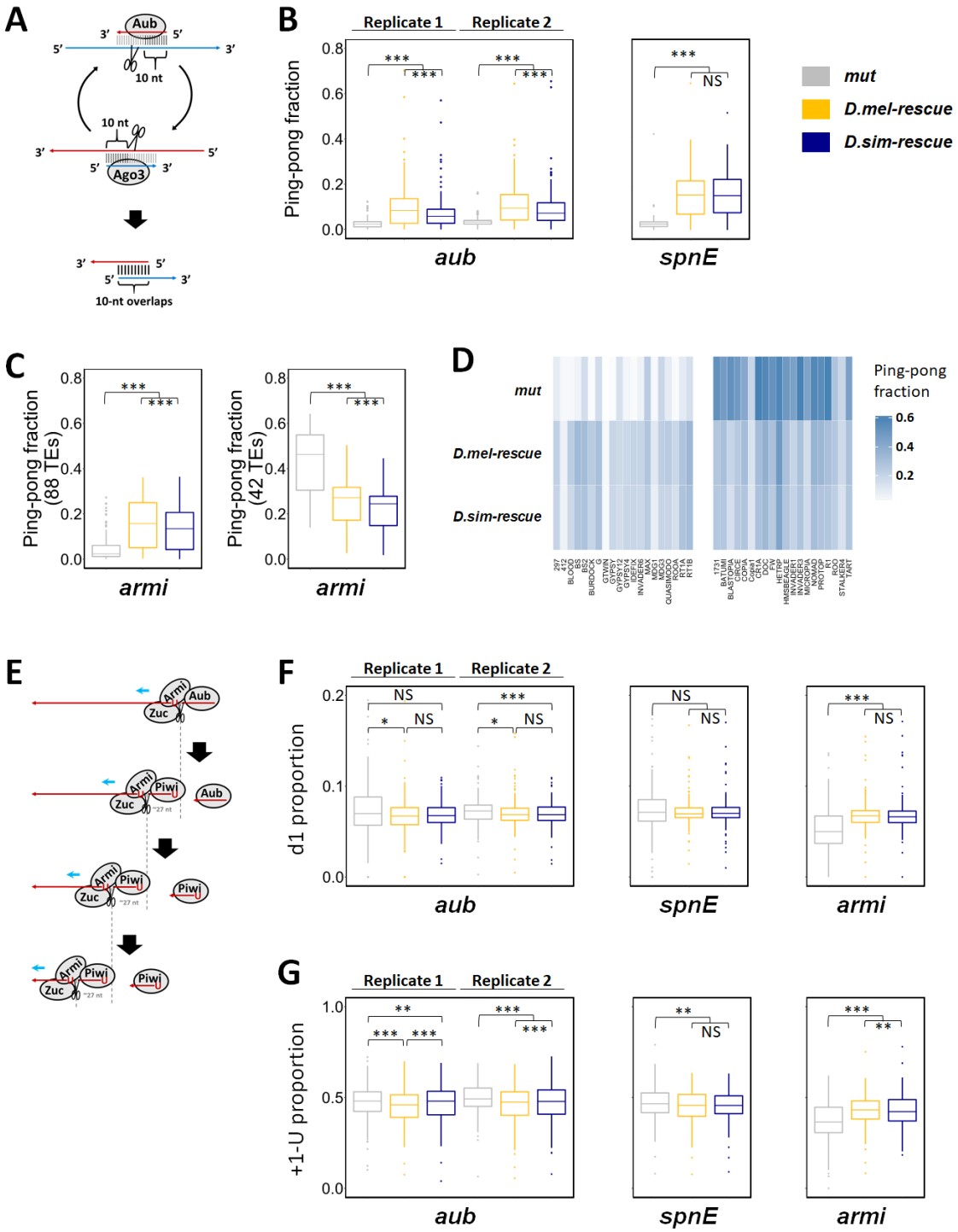
Despite the absence of large-scale interspecific differences in antisense piRNAs that regulate TEs (Fig 2B), we interrogated piRNA pools associated with each of the transgenic rescues for evidence of underlying differences in piRNA biogenesis. We examined molecular signatures of the two major mechanisms of piRNA biogenesis: ping-pong and phasing. Ping-pong biogenesis produces piRNAs through reciprocal cleavage of complementary precursors (Fig 3A) [24,25]. The frequency of ping-pong amplification is therefore estimated by the fraction of piRNAs occuring on opposite strands of the TE consensus whose sequences overlap by 10 bp, a reflection of the cleavage-site preference of the key ping-pong cycle factors Aub and Argonaute-3 (Ago-3, Fig 3A-D, S5 Table) [24–26]. In contrast, phasing biogenesis occurs through sequential cleavage of a single RNA strand, which is usually antisense [33,34]. Phasing is detected from the fraction of piRNAs whose 3' ends are immediately followed by a uracil residue (+1-U), as well as the frequency of piRNAs from the same strand that are separated by a distance of a single nucleotide (d1), both of which are diagnostic of cleavage by the phasing nuclease Zucchini (Fig 3E-G, S6 and S7 Table) [33,34]. In general, ping-pong and phasing are

inversely correlated in mutant piRNA pools, because reducing the frequency of one leads to a proportional increase in the other [33,34].

Aub plays a direct role in ping-pong amplification by cleaving piRNA precursors (Fig 3A) [24–26], and *spn-E* is required for the localization of Aub into the perinuclear nuage, where ping-pong occurs [28]. Mutations in either gene therefore cause a complete collapse of ping-pong amplification (Fig 3B, S6A and S6B Fig, S5 Table) [26,32]. Both *D. melanogaster* and *D. simulans aub* and *spn-E* alleles exhibited a dramatic increase in the ping-pong fraction, indicating a conserved role in ping-pong biogenesis (Fig 3B, S6A and S6B Fig). However in the case of *aub,* ping-pong fractions associated with the *D. simulans* transgenic rescue were modestly yet significantly lower than *D. melanogaster,* and there was a corresponding proportional increase in phased piRNA biogenesis (Fig 3B, 3F and 3G, S7 Fig, S6 and S7 Table), suggesting reduced efficiency of ping-pong. By contrast, *D. simulans spn-E* allele did not reduce ping-pong (Fig 3B, S6B Fig, S5 Table), yet there was a modest but significant increase in the d1 proportion with the *D. simulans spn-E* rescue (Fig 3F, S6 Table), potentially suggesting increased phasing.

Armi promotes the production of phased piRNAs by binding to antisense piRNA intermediates and facilitating their cleavage by the nuclease Zucchini (Fig 3E) [29,33,34]. Both d1 and +1-U are therefore significantly reduced in *armi* mutants (Fig 3F and 3G, S7 Fig, S6 and S7 Table). While Armi is not involved in ping-pong, phasing produces Aub-bound antisense piRNAs, which are required for ping-pong biogenesis for some TE families [26,45]. Ping-pong fractions are therefore decreased in *armi* mutants for some TE families (Fig 3C and 3D, S6C Fig, S5 Table). By contrast, for TE families that do not rely on phased piRNA production for ping-pong, ping-pong-derived piRNAs proportionally increase in *armi* mutants, owing to the loss of phased piRNAs (Fig 3C and 3D, S6C Fig, S5 Table). Although exhibiting piRNA production similar to the *D. melanogaster* allele (Fig 2B), the *D. simulans armi* rescue exhibited modestly but significantly reduced +1-U proportion, indicating reduced phasing (Fig 3G, S7B Fig, S7

Table). However, the more dramatic and statistically significant allelic effect is on ping-pong biogenesis, which is reduced for most TE families by the *D. simulans armi* rescue when compared to *D. melanogaster* (Fig 3C and 3D, S6C Fig, S5 Table). Importantly, this reduction occurs regardless of whether *armi* function enhances or represses ping-pong biogenesis, revealing a global inhibitory effect imposed by *D. simulans armi*. Indeed, although the differential abundance of TE and repeat-derived piRNAs between transgenic rescues rarely exceeded two-fold, significantly more TE families were more abundant in the presence of the *D. melanogaster* armi rescue compared to the *D. simulans armi* rescue (118 out of 131 TE families, Sign-test, *P*-value $< 10^{-15}$). Therefore, the modest reductions in ping-pong and phasing biogenesis exhibited by the *D. simulans armi* allele lead to a similarly modest reduction in the abundance of TE and repeat-derived piRNAs.



1

2 **Fig 3. *Drosophila simulans* alleles reduce ping-pong biogenesis and phasing biogenesis.** (A) Simplified

3 diagram of ping-pong amplification loop. (B) Ping-pong fractions of TE-derived piRNAs from 142 TE

4 families are compared between trans-heterozygous mutants and transgenic rescues for *aub* and *spn-E.*

(C) Ping-pong fractions of TE-derived piRNAs are compared between trans-heterozygous mutants and transgenic rescues for *armi*. Comparison for 92 and 43 TE families whose ping-pong fractions are decreased (left) or increased (right), respectively, in *armi* mutants as compared to the *D. melanogaster* transgenic rescue. (D) Left: ping-pong fraction heat map for the 20 most piRNA-abundant TE families from panel C left. Right: ping-pong fraction heat map for 20 most piRNA-abundant TE families from panel C right. (E) Zuc-dependent phased piRNA biogenesis. (F) Proportions of 1 nt distance between adjacent piRNAs (d1) mapped to the TE consensus sequences are compared between each genotype of each gene. (G) Proportions of uridine residue immediately after the 3′ ends of piRNAs (+1-U) mapped to the TE consensus sequences are compared between each genotype of each gene. Statistical significance was assessed by the Wilcoxon signed-rank test. For *aub*, two biological replicates of each genotype generated at different times are shown separately. For *spn-E* and *armi*, average of three biological replicates of each genotype generated at the same time are shown. NS denotes $p > 0.05$. *, **, and *** denote $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, respectively.

Increased off-target effects of *D. simulans* alleles suggest genomic auto-immunity.

While effective negative regulation of TE transcripts is a critical function of piRNA pathway proteins, it is equally important that they avoid off-target effects that interfere with the function of host genes. [19,21]. Aub, Spn-E and Armi are all RNA binding proteins that must specifically interact with piRNAs, piRNA precursors, and target transcripts, while avoiding interactions with cytoplasmic mRNAs. We therefore considered whether off-target effects differ between *D. melanogaster* and *D. simulans* alleles, predicting that *D. simulans* alleles may produce more off-target effects as they are not adapted to avoid interactions with *D. melanogaster* transcripts.

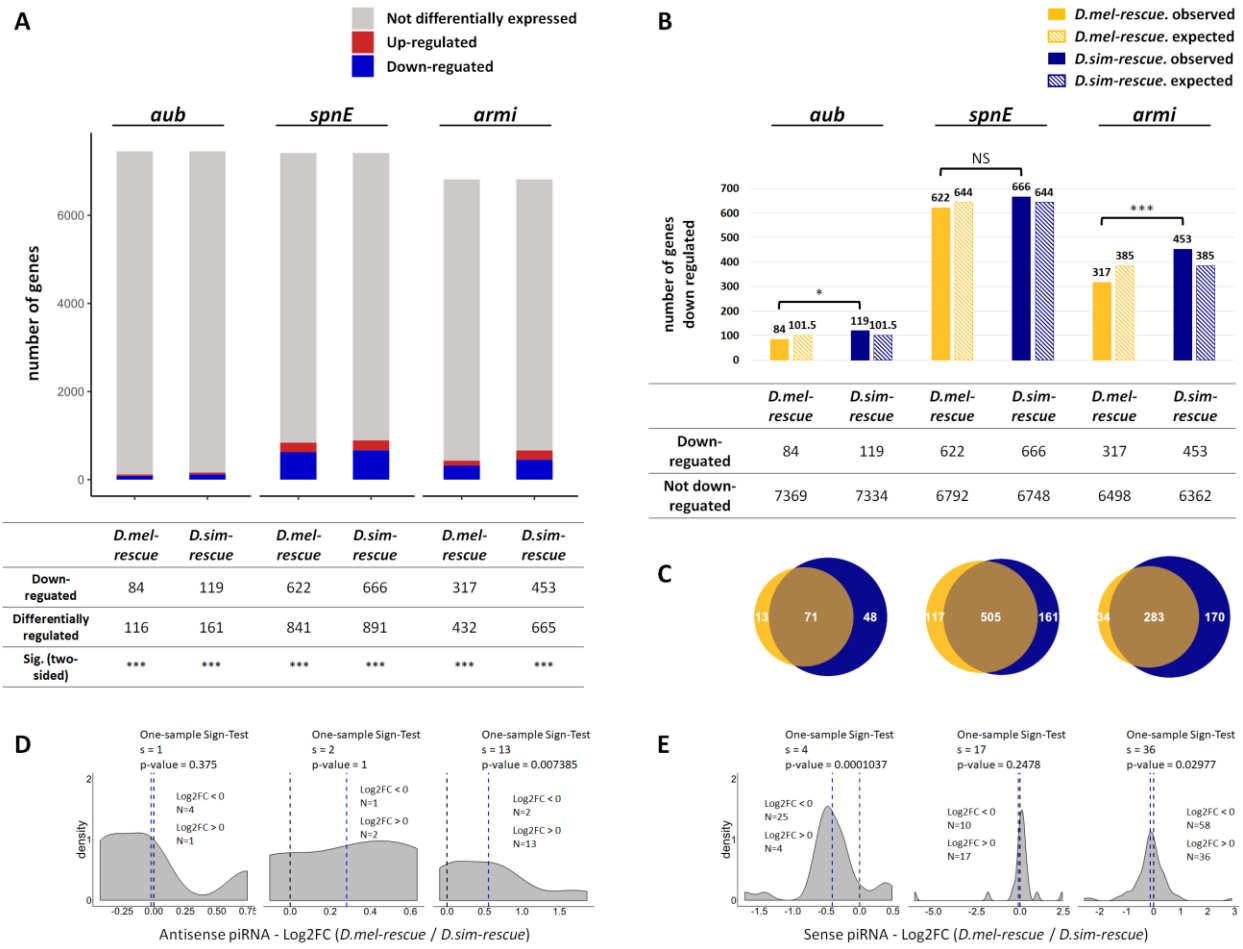
To test this prediction, we first identified protein-coding genes that are negatively regulated by piRNA pathway proteins by comparing their expression levels between mutants and transgenic rescues (S8 Table). Protein-coding genes whose expression is significantly

reduced in transgenic rescues (>1.5 fold) are candidates for off-target effects of piRNA-mediated silencing. We observed that for all three proteins, significantly more genes decreased than increased in expression in transgenic rescues as compared to mutants (Fig 4A), suggesting that piRNA pathway proteins tend to reduce the expression of protein-coding genes. Furthermore, the majority of protein-coding genes that are negatively regulated by *D. melanogaster* rescues are also repressed by *D. simulans* rescues, suggesting a shared impact on the expression of many protein-coding genes (Fig 4B and 4C). However, for all three proteins, we observed more protein-coding genes whose expression significantly is reduced by the *D. simulans* rescue than the *D. melanogaster* rescue, with the difference being significant for both *aub* and *armi* (Fig 4B). Together, these observations suggest that *D. simulans* alleles exhibit expanded off-target regulation of protein-coding genes.

Increased off-target effects of *D. simulans* alleles could be explained by increased production of antisense genic piRNAs that target host transcripts, or by piRNA-independent interactions between proteins and mRNAs. Recent analyses of off-target interactions between host mRNAs and Piwi (closely related to Aub) or Armi support the latter scenario, suggesting that while the binding of host mRNAs by piRNA proteins reduces mRNA abundance, it does not result from enhanced antisense genic piRNA production [30,46,47]. We therefore compared the abundance of antisense genic piRNAs, which target the silencing of sense transcripts, between transgenic rescues (S9 Table). Of 48 transcripts that are negatively regulated by *D. simulans aub*, but not *D. melanogaster aub*, only 5 are targeted by antisense piRNAs. Furthermore, no significant bias towards increased antisense piRNA production in the *D. simulans* allele was observed for these genes (Fig 4D). Similarly of 170 transcripts that are negatively regulated by *D. simulans armi*, but not *D. melanogaster armi*, only 15 are targeted by antisense piRNAs. Furthermore, antisense piRNAs corresponding to these genes are more abundant in the presence of the *D. melanogaster* allele, which is the opposite of what is

expected if enhanced negative regulation by *D. simulans* results from accumulating antisense piRNAs (Fig 4D). Finally, *D. simulans* transgenic rescues do not exhibit expanded production of genic piRNAs for any of the three proteins(S8 Fig). Therefore, enhanced negative regulation of host mRNAs by *D. simulans* Aub and Armi appears to be independent of antisense piRNA production, and may arise through promiscuous binding of *D. simulans* proteins to *D. melanogaster* transcripts.

If *D. simulans* Aub and Armi proteins bind promiscuously to *D. melanogaster* mRNAs, they could introduce them into the piRNA pool by treating them as substrates for piRNA maturation. Consistent with this model, we observed that sense piRNAs derived from genic transcripts that are uniquely negative regulated by *D. simulans* are significantly more likely to be more abundant in the presence of *D. simulans* alleles of *aub* and *armi* when compared to *D. melanogaster* (Fig 4E). Therefore, while off-target regulation occurs independently of antisense piRNAs, it leads to a consequential increase in sense piRNA production.



	D.mel-rescue	D.sim-rescue	D.mel-rescue	D.sim-rescue	D.mel-rescue	D.sim-rescue
Down-reguated	84	119	622	666	317	453
Differentially regulated	116	161	841	891	432	665
Sig. (two-sided)	***	***	***	***	***	***

	D.mel-rescue	D.sim-rescue	D.mel-rescue	D.sim-rescue	D.mel-rescue	D.sim-rescue
Down-reguated	84	119	622	666	317	453
Not down-reguated	7369	7334	6792	6748	6498	6362

Fig 4. Negative regulation of protein-coding genes suggests increased genomic auto-immunity of *D. simulans* alleles. (A) The number of genes whose expression levels are decreased/increased (>1.5 fold, blue/red) in the presence of each transgene as compared to the corresponding mutant. Statistical significance was assessed by the binomial test evaluating the probability of the observed proportion of down-regulated genes as compared to upregulated genes under the null hypothesis that the two probabilities are equal. (B) The number of genes whose expression levels are decreased (>1.5 fold) in the presence of each transgene as compared to the corresponding mutant. Contingency tables are shown below. Statistical significance was assessed by the Pearson's Chi-squared Test of Independence. Expected counts were calculated based on the assumption of statistical independence between transgenic rescue allele and effect of corresponding regulation (i.e. genes are equally likely to down-

regulated by either allele). (C) Overlap of genes whose expression levels are decreased (>1.5 fold) in the presence of each transgene as compared to the corresponding mutant for *aub* (left Venn diagram), *spn-E* (middle Venn diagram) and *armi* (right Venn diagram). (D, E) Probability density plots for log2 fold-change of antisense (D) / sense (E) piRNA abundance between rescues by *D. melanogaster* and *D. simulans* alleles, for genes whose expression levels are down-regulated by *D. simulans* alleles only. Log2 fold-change values were based on one biological replicate for *aub* and three biological replicates for *spn-E* and *armi*, and were obtained from a DESeq analysis for *aub* and a DESeq2 analysis for *spn-E* and *armi* (adjusted $p < 0.05$). NS denotes $p > 0.05$. * denotes $p \leq 0.05$. *** denotes $p \leq 0.001$.

Discussion

Despite pervasive adaptive evolution and gene duplication among piRNA pathway proteins in both insect and vertebrate lineages [12–15,48,49], the underlying forces that drive these evolutionary dynamics remain unclear. By performing interspecific complementation on three adaptively evolving piRNA pathway genes, we revealed diverged functions that may have arisen through positive selection. For all three proteins we observed idiosyncratic differences in TE regulation between *D. melanogaster* and *D. simulans* alleles, which is consistent with genetic conflict between host and parasite. However, we also revealed more extensive off-target effects and reduced efficiency of piRNA maturation associated with *D. simulans* alleles of both *aub* and *armi*, suggesting that selection acts to maximize biogenesis function while minimizing friendly fire on host transcripts. Taken together, our results suggest that positive selection acts at multiple molecular and functional interfaces within the piRNA pathway.

The simplest explanation for the adaptive evolution of piRNA proteins is that selection acts to maximize host control of TE transposition. At face value, TEs that are differentially expressed between transgenes from the two species seem to support this model (Fig 2A). However, all 5 of these TE families were more robustly silenced by the heterospecific *D. simulans* proteins. This suggests that rather than conspecific piRNA proteins being well-adapted to silence their genomic TEs, active genomic TEs may be well-adapted to evade silencing by their host regulators. Indeed the *tirant* element, which is more robustly regulated by *D. simulans* alleles of both *armi* and *spn-E*, is unusually active in *D. melanogaster* but is actively lost from *D. simulans* [50–52]. We propose that the differential evolutionary dynamics of *tirant* in these two lineages may in part reflect the differences in host-control that we have uncovered.

Genomic auto-immunity was recently proposed as an additional target of positive selection among piRNA proteins [19]. The deliberate non-specificity of piRNA pathway proteins that allows them to target any sequence represented among piRNAs for silencing presents a huge liability for host-gene regulation: how can piRNA proteins avoid deleterious interactions

with host transcripts? Furthermore, RNA-immunoprecipitation (RIP) and cross-linking immunoprecipitation (CLIP) of Piwi suggest that piRNA proteins may also negatively regulate host mRNAs by binding them directly in a piRNA-independent manner [46,47]. Similar observations have been made with Armi protein, with the ATP-ase domain being required to disassociate Armi from host mRNAs in the cytoplasm [30]. We observed that *D. simulans armi* and *aub* alleles are characterized by expanded negative regulation of host genes (Fig 4A-C), which is accompanied by reduced efficiency of TE-derived piRNA production (Fig 3). Importantly, this observation is not consistent with a subtle difference in protein abundance between transgenic rescues, since this would reduce both piRNA biogenesis and off-target effects. Rather our observations are fully consistent with *D. simulans* alleles binding promiscuously to host mRNAs, thereby reducing host-gene expression and depleting the pool of protein available to enact piRNA maturation. Nevertheless, we cannot discount an equally intriguing alternative explanation: that *D. simulans* alleles have decreased affinity for interacting protein partners that mediate piRNA biogenesis, which liberates them to bind more frequently to host mRNAs. Future comparisons of molecular interactions involved *D. melanogaster* and *D. simulans* alleles could differentiate between these alternatives.

Our observations considerably expand our understanding of the enigmatic forces that drive adaptive evolution across the piRNA pathway. The three proteins we studied here, which are cytoplasmic factors involved in piRNA maturation and post-transcriptional silencing, provide an informative contrast to similar studies of three adaptively evolving nuclear transcriptional silencing factors: Rhino, Deadlock and Cutoff (Parhad et al. 2017; Yu et al. 2018; Parhad et al. 2019). In comparison to the modest functional differences we observed between *D. melanogaster* and *D. simulans* alleles, nuclear factors are characterized by dramatic interspecific divergence, with *D. simulans* alleles behaving similarly to loss of function or dominant negative mutations (Parhad et al. 2017; Yu et al. 2018; Parhad et al. 2019). Furthermore, adaptive evolution among transcriptional silencers has resulted in interspecific

incompatibilities between interacting proteins, as opposed to the divergence in protein-RNA interactions that our data suggest. These observations logically reflect differences in the molecular functions of the two classes of proteins, with transcriptional regulation relying on suites of proteins that modify chromatin or regulate RNA-polymerase, while piRNA maturation and post-transcriptional silencing relies more on interactions between proteins and RNA.

Interestingly, a recent meta-analysis of piRNA protein evolution in insects revealed that while positive selection is pervasive throughout the piRNA pathway, signatures of adaptive evolution are significantly stronger among nuclear transcriptional regulators, as compared to the cytoplasmic factors we studied here (Palmer et al. 2018). Thus, our functional analysis recapitulates an evolutionary signature in sequence data. Why might nuclear transcriptional regulators diverge more rapidly or dramatically than their cytoplasmic counterparts? Enhanced positive selection on nuclear factors may reflect their greater potential to fully suppress the expression of host genes through disrupted chromatin state, as opposed to merely depleting host transcripts through non-productive binding (Blumenstiel et al. 2016). We therefore propose that off-target effects may play an under-appreciated role in the evolution of both nuclear and cytoplasmic piRNA proteins, with the strength and consequences of positive selection depending on the mechanisms of—and costs to—host gene regulation.

Materials and methods

Fly strains and crosses

All *Drosophila* strains were reared at room temperature on standard cornmeal media.

For the studies of *aubergine* (*aub*), the following *D. melanogaster* strains were used: *w; aub^{N11} bw^{1}/CyO*, *yw; aub^{HN} bw^{1}/CyO*, *yw; aub^{HN} bw^{1}/CyO; ΦP{D. melanogaster aub}*, *yw; aub^{HN} bw^{1}/CyO; ΦP{D. simulans aub}*. *w; aub^{N11} bw^{1}/CyO*, was a gift from Paul MacDonald. *yw; aub^{HN} bw^{1}/CyO* was obtained by extracting *yw* into *aub^{HN} bw^{1}/CyO* (Bloomington *Drosophila* Stock

Center #8517). *yw; aub^{HN} bw^{1}/CyO; ΦP{D. melanogaster aub}* and *yw; aub^{HN} bw^{1}/CyO; ΦP{D. simulans aub}*, originally generated in Kelleher et al [31], were backcrossed for 6 generations in *yw; aub^{HN} bw^{1}/CyO* to minimize background effects that could lead to differences between transgenic stocks that were unrelated to the transgenes.

For the studies of *spindle-E* (*spn-E*), the following *D. melanogaster* strains were used: *yw; spn-E^{1}/TM6*, *yw; spn-E$^{hls-03987}$/TM6*, *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. melanogaster spn-E}*, *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. simulans spn-E}*. *yw; spn-E^{1}/TM6* and *yw; spn-E$^{hls-03987}$/TM6* were obtained by crossing *spn-E^{1}/TM3* and *spn-E$^{hls-03987}$/TM3* (gifts from Celeste Berg) to *yw; TM3/TM6*. To generate *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. melanogaster spn-E}* and *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. simulans spn-E}, w^{1118}; ΦP{D. melanogaster spn-E}* and *w^{1118}; ΦP{D. simulans spn-E}* were first crossed to *yw: TM3/TM6. +/TM6; ΦP{D. melanogaster spn-E}*/+ and *+/TM6; ΦP{D. simulans spn-E}*/+ offspring were then crossed to *yw; spn-E$^{hls-03987}$/TM3.* Finally, *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. melanogaster spn-E}*/+ and *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. simulans spn-E}*/+ offspring were backcrossed into *yw; spn-E$^{hls-03987}$/TM6* for 6 generations, and subsequently homozygosed for the transgene, to minimize background effects.

For the studies of *armitage* (*armi*), the following *D. melanogaster* strains were used: *yw; armi^{1}/TM6*, *w; armi$^{72.1}$/TM6*, *w; armi$^{72.1}$/TM6; ΦP{D. melanogaster armi}*, *w; armi$^{72.1}$/TM6*; *ΦP{D. simulans armi}*. *yw; armi^{1}/TM6* was obtained by crossing *yw; armi^{1}/TM3* (Bloomington *Drosophila* Stock Center #8513) to *yw; TM3/TM6*. *w; armi$^{72.1}$/TM6* was obtained from Bloomington *Drosophila* Stock Center (#8544). To generate *w; armi$^{72.1}$/TM6; ΦP{D. melanogaster armi}* and *w; armi$^{72.1}$/TM6*; *ΦP{D. simulans armi}*, *w^{1118}; ΦP{D. melanogaster armi}* and *w^{1118}; ΦP{D. simulans armi}* were first crossed to *yw; TM3/TM6*. *+/TM3; ΦP{D. melanogaster armi}*/+ and *+/TM3; ΦP{D. simulans armi}*/+ offspring were then crossed to *w; armi$^{72.1}$/TM6*. Finally, *w; armi$^{72.1}$/TM3; ΦP{D. melanogaster armi}*/+ and *w; armi$^{72.1}$/TM3*; *ΦP{D. simulans armi}*/+ were backcrossed into *w; armi$^{72.1}$/TM6* for 6 generations, and subsequently homozygosed for the transgene, to minimize background effects.

Experimental genotypes were obtained from the following crosses. For studies of *aub*, virgin females *w; aub^{N11} bw^{1}/CyO* were crossed to (1) *yw; aub^{HN} bw^{1}/CyO*, (2) *yw; aub^{HN} bw^{1}/CyO; ΦP{D. melanogaster aub}* or (3) *yw; aub^{HN} bw^{1}/CyO; ΦP{D. simulans aub}* males. For studies of *spn-E,* virgin females *yw; spn-E^{1}/TM6* were crossed to (1) *yw; spn-E$^{hls-03987}$/TM6*, (2) *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. melanogaster spn-E}* or (3) *yw; spn-E$^{hls-03987}$/TM6; ΦP{D. simulans spn-E}* males. For studies of *armi,* virgin females *yw; armi^{1}/TM6* were crossed to (1) *w; armi$^{72.1}$/TM6*, (2) *w; armi$^{72.1}$/TM6; ΦP{D. melanogaster armi}* or (3) *w; armi$^{72.1}$/TM6*; *ΦP{D. simulans armi}* males. Crosses were maintained at 25ºC on standard cornmeal media.

Generation of transgenic lines

To introduce *D. melanogaster* and *D. simulans* alleles into *D. melanogaster*, we used *Φ*C31 integrase-mediated transgenesis system [38], which allows for site-specific integration. To generate transgenes for site specific integration, the gene and flanking regulatory regions of *spn-E* (~9.7Kb, *D. melanogaster* Release 6, 3R:15835349..15845065; *D. simulans* Release 2, 3R:9575537..9585081) [53,54] and *armi* (~6Kb, *D. melanogaster* Release 6, 3L:3460305..3466368; *D. simulans* Release 2, 3L:3357002..3363099) [53,54] were PCR-amplified by using corresponding primers (below) and iProof high-fidelity taq DNA polymerase (Bio-Rad).

D.mel/D.sim-spn-E forward primer: ATTGAACGCCGTCTATGCCAAGC

D.mel/D.sim-spn-E reverse primer: ACTGTTCGCCATTGCCACAGATTG

D.mel/D.sim-armi forward primer: CACCGCTGAAAGATACGCACACG

D.mel-armi reverse primer: GCTAGCCTGCGCTTGGGAGTGTTACCATTCG

D.sim-armi reverse primer: GCTAGCCTGACCTCGGGAGTGTTACCACTTC

The PCR products were cloned into pCR-Blunt-II-Topo according to manufacturer instructions (Invitrogen). Mutation-free clones were verified by sequencing.

attB containing constructs used for site-specific integration were generated by subcloning the NotI/BamHI fragment of each *spn-E* TOPO plasmid, and the NotI/NheI fragment of each *armi* TOPO plasmid into NotI/BamHI and NotI/XbaI-linearized pCasper4/attB, respectively. *spn-E* and *armi* transgenic constructs were introduced into *D. melanogaster* at the P{CaryP}attP40 site, and site-specific integration of transgenes was confirmed by PCR [55]. The resulting transgenes were made homozygous in *D. melanogaster* w^{1118}.

Female fertility

25-35 individual virgin females of each experimental genotype were crossed to two *ywF10* males on standard cornmeal media at 25°C. Fresh media and new males were provided every 5 days. The number of progeny from each 5-day period was quantified.

Small RNA-seq

3-6-day old female ovaries were dissected from each experimental genotype and placed directly in Trizol reagent (Invitrogen), and homogenized. For *aub* genotypes, Illumina small RNA libraries were prepared by Fasteris according to a proprietary protocol that depletes for 2S-RNA. Because two biological replicates prepared at different time points (5/13 and 7/13), they are analyzed separately. Small RNA libraries for *spn-E* and *armi* genotypes were prepared as described in [56]. In brief, total RNAs were extracted according to the manufacturer's instructions, and size fractionated on a 12% polyacrylamide/urea gel to select for 18-30 nt small RNAs. Small RNAs were treated with 2S Block oligo (5'-TAC AAC CCT CAA CCA TAT GTA GTC CAA GCA/3SpC3/-3'), and were subsequently ligated to 3' and 5' adaptors, reverse transcribed and PCR amplified using NEBNext Multiplex Small RNA Library Prep Set for Illumina. Small RNA libraries were further purified from a 2% agarose gel and sequenced on a Illumina NextSeq 500 at the University of Houston Seq-N-Edit Core.

RNA-seq

RNA-seq libraries for the studies of *aub* were generated by Weill Cornell Epigenomics Core according to the protocol of [57]. Briefly, total RNA was extracted from the same ovaries as above, and mRNAs were isolated using poly-T Dynabeads (Invitrogen) according to the manufacturer's instructions. Isolated mRNAs were further fragmented using fragmentation buffer (Ambion), ethanol precipitated, and reverse transcribed using Superscript II (Invitrogen) and random hexamer primers. Second-strand synthesis was performed using DNA polymerase I (Promega). cDNA was purified on a MinElute column (Qiagen), repaired with End-IT DNA repair kit (Epicentre), A-tailed with Klenow enzyme (New England Biolabs), and ligated to Illumina adaptors. Ligated cDNA was gel purified with the MinElute gel purification kit (Qiagen), PCR amplified, and gel purified again to make libraries.

RNA-seq libraries for the studies of *spn-E* and *armi* were prepared by using TruSeq Stranded Total RNA Library Preparation Kit for Illumina. 50 bp reads from each library were sequenced on a HiSeq 2000 (Aub and spn-E) and a HiSeq 2500 (Armi) by the Weill-Cornell Epigenomics Core. RNA-seq and small RNA-seq data sets are deposited under PRJNA494103.

Bioinformatic analysis of small RNA-seq libraries

3' Illumina adaptors were removed from sequencing reads by Cutadapt [58]. Sequence alignments were made by Bowtie [59]. Contaminating ribosomal RNAs were identified and removed by mapping sequencing reads to annotated ribosomal RNAs from flybase [60]. TE-derived piRNAs and genic piRNAs were identified by aligning sequencing reads ranging from 23-30 nucleotides (nt) to Repbase [61]Flybase [60], respectively, allowing for up to 2 mismatches. The number of reads mapped to each TE family or gene were counted using a Linux shell script. Redundant TE families in Repbase were identified by checking sequence identity (those consensus sequences that were >90% identical across >90% of their length were categorized as the same TE family), and reads mapped to multiple redundant TE families were

counted only once. Reads mapped to multiple non-redundant TE families were discarded. To identify miRNAs sequencing reads ranging from 18-22 nt were aligned to a miRNA reference sequence from Flybase [60]. TE families or genes with low read count (< 50 on average) in every genotype libraries were discarded. piRNA counts for each TE family or gene were normalized to the total number of sequenced miRNAs from each library. Normalized values were used for comparisons of the abundance of piRNAs between libraries.

Bioinformatic analysis of RNA-seq libraries

Removal of ribosomal RNAs, and identification of TE-derived reads was performed as for small RNA libraries (above) except that 3 mismatches were permitted between sequencing reads and TE consensus sequences. Non TE-derived reads were aligned to flybase annotated transcripts in the *D. melanogaster* reference genome (*D. melanogaster* Release 6) [53,60] by TopHat [62], requiring unique mapping. The number of reads from each protein-coding gene were counted using HTseq-count [63]. TE families or genes with low read count (< 50 on average) in every genotype were discarded. Differential expression was estimated concurrently for TEs and protein-coding genes by DESeq for *aub* [64] and DESeq2 for *spn-E* and *armi* [44]. TEs or protein-coding genes were considered differentially expressed if they exhibited an adjusted *P*-value < 0.05 and a fold-change > 2 for TEs and > 1.5 for protein-coding genes.

Ping-pong fraction

Ping-pong fraction was calculated as described in [65]. In brief, small RNA sequencing reads ranging from 23-30 nt were aligned to TE consensus sequences from Repbase [61], and redundant TE families in Repbase were identified as described above. For each piRNA, the proportion of overlapping antisense binding partners whose 5’ end occur on the 10th nucleotide was determined. This fraction was subsequently summed across all piRNAs from a given TE family, while incorporating the difference in sampling frequency between individual piRNAs.

Finally, this sum was divided by the total number of piRNAs aligned to the TE family of interest. For multi-mappers, reads were apportioned by the number of times they can be aligned to the reference.

Phasing analysis

Small RNA sequencing reads ranging from 23-30 nt were aligned to the Repbase [61], and redundant TE families in Repbase were identified as described above. To calculate the d1 proportion [34], the number of piRNAs whose 5' end was 1-22 nt downstream piRNA was determined for every TE-derived piRNA. The fraction of distances corresponding to 1 nt was then calculated. To calculate the +1-U proportion [34], the nucleotide after the 3' end of each piRNA was determined based on alignment to the Repbase [61]. The frequency of each nucleotide at the +1 position was subsequently summed across all piRNAs from a given TE family, and the proportion of uridine was calculated. For both analyses, multiply-mapping reads were apportioned by the number of times they aligned to the reference.

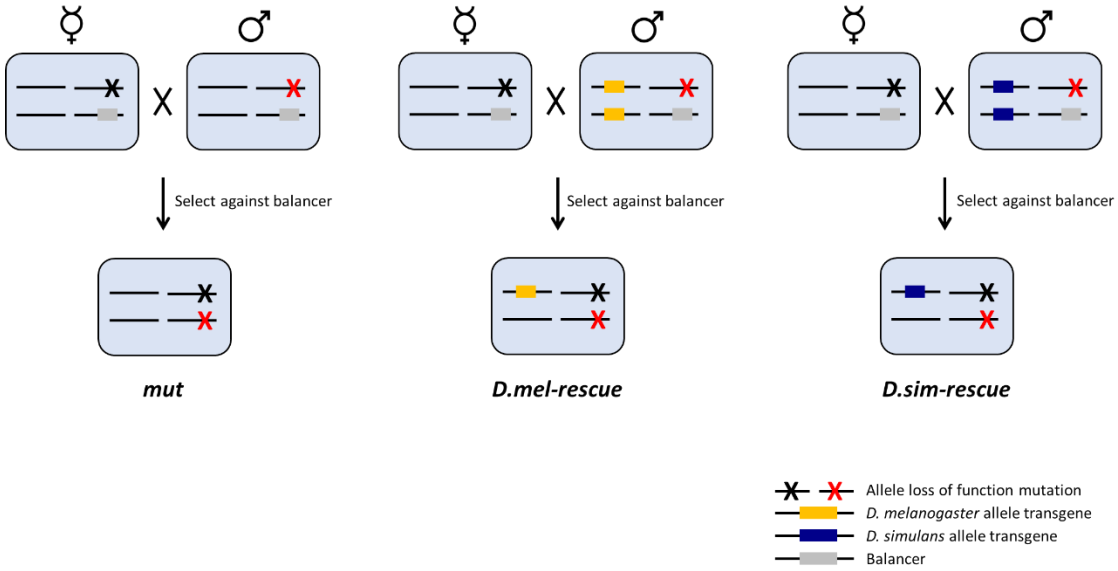
Acknowledgments

This research was supported by the Cornell Center for Comparative and Population Genomics and the University of Houston Division of Research. E.S.K. was supported by the Cornell Center for Comparative and Population Genomics, an NIH National Research Service Award (F32GM090567-01), and NSF-DEB 1457800 (to E.S.K.). Luyang Wang was supported by NSF-DEB 1457800 (to E.S.K.). We are grateful to Shuqing Ji for assistance with cloning of genomic transgenes, and Paul MacDonald and Celeste Berg for providing *Drosophila* strains. D.A.B. was supported by NIGMS R01GM074737. We are grateful to four anonymous reviewers for their insights and comments about the previous version of this manuscript.

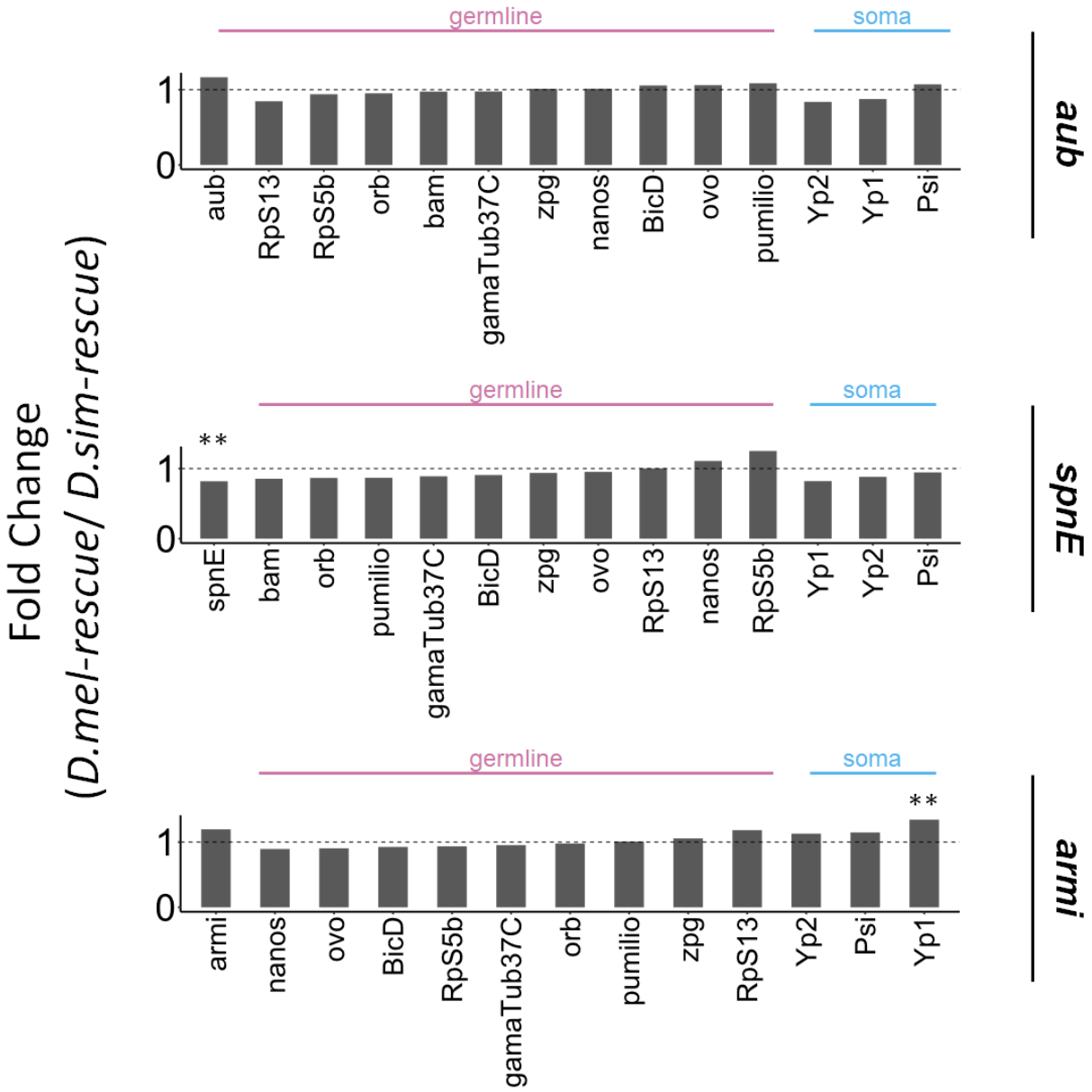
Author Contributions

E.S.K. and D.A.B. designed research. E.S.K. and L.W. performed experiments and analyzed data. E.S.K., L.W. and D.A.B. wrote paper.

Supporting information

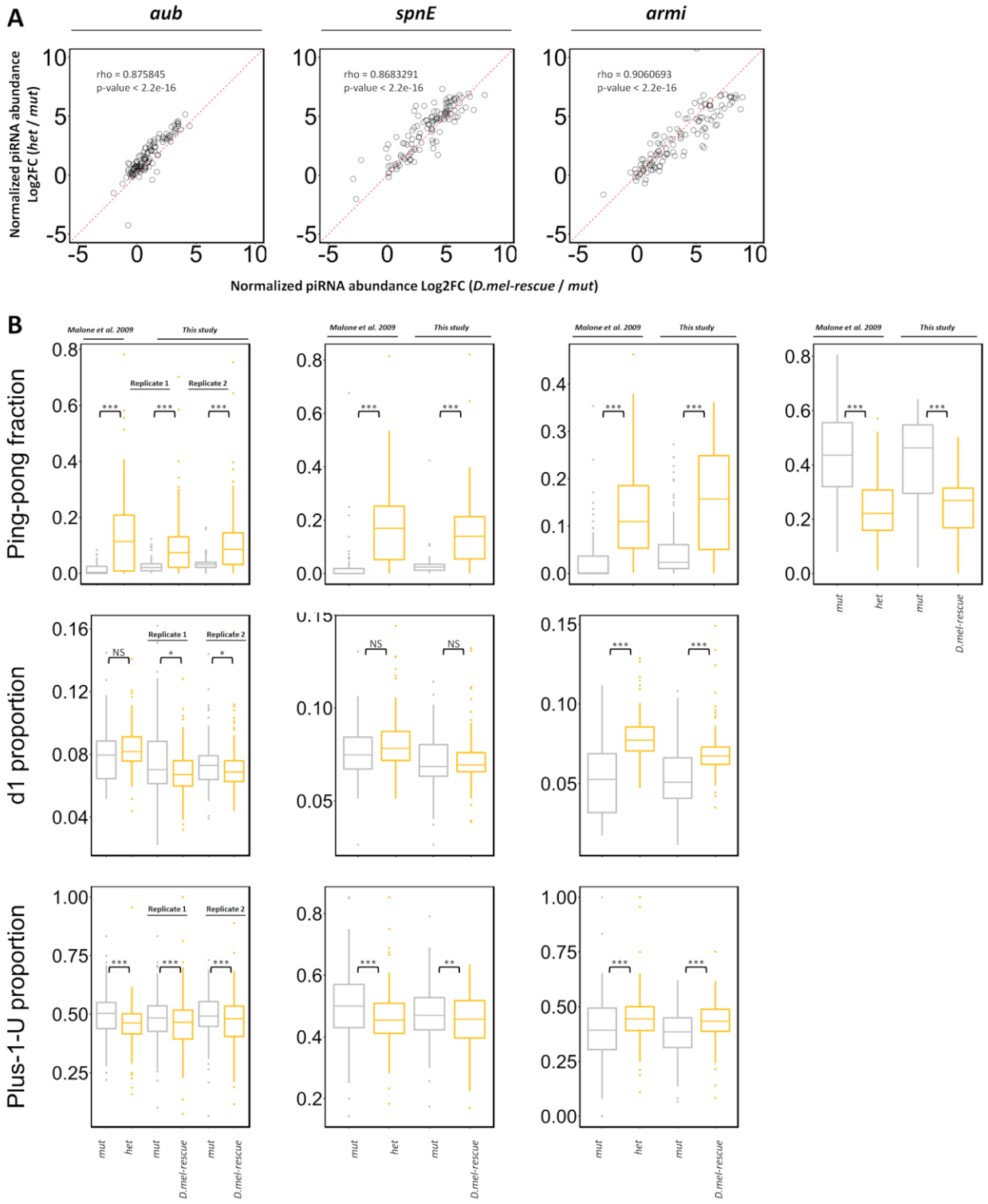


S1 Fig. *Drosophila melanogaster* genotypes and cross scheme.

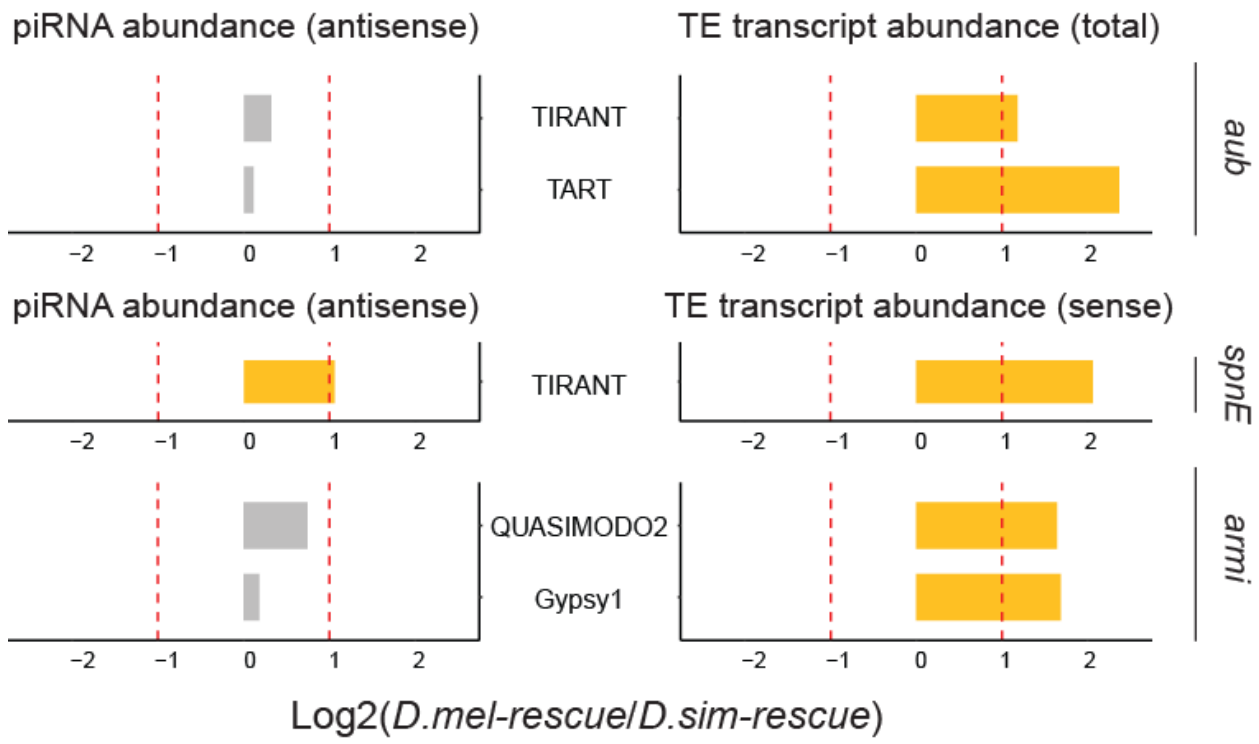


S2 Fig. Similar expression level of *aub*, *spn-E* and *armi* transgenes, as well as germline and soma specific genes between *D. melanogaster* transgenic rescue and *D. simulans* transgenic rescue. Fold-change of expression level of *aub*, *spn-E*, *armi*, germline-specific genes and soma-specific genes between *D. melanogaster* transgenic rescue and *D. simulans* transgenic rescue are shown. Fold-change values are based on one biological replicate for *aub* and three biological replicates for *spn-E* and *armi*, and were obtained from a DESeq analysis for *aub* and a DESeq2 analysis

for *spn-E* and *armi*. ** denotes $p \leq 0.01$. NS if not labeled.



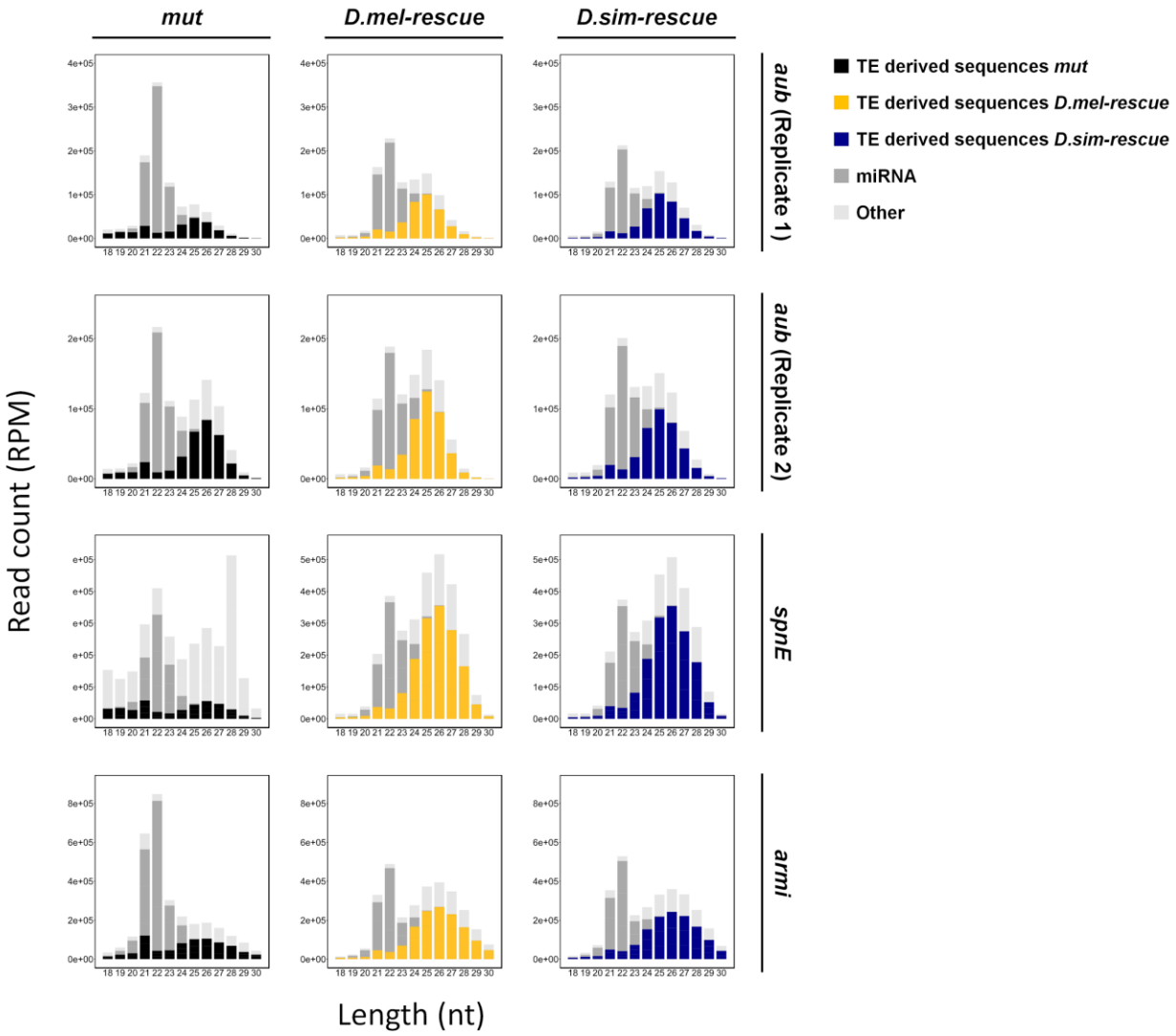
1 **S3 Fig. *D. melanogaster* transgenes exhibit similar profiles of piRNA biogenesis to**
2 **heterozygotes.** *D. melanogaster* transgenes and heterozygotes are compared to
3 transheterozygous mutants with respect to piRNA abundance (A), ping-pong and phasing
4 biogenesis (B). Statistical significance was assessed by the Wilcoxon signed-rank test. For *aub*,
5 two biological replicates of each genotype generated at different times are shown separately.
6 For *spn-E* and *armi*, average of three biological replicates of each genotype generated at the
7 same time are shown. NS denotes $p > 0.05$. *, **, and *** denote $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$,
8 respectively.
9



10
11 **S4 Fig. Decoupling between changes in TE transcript abundance and changes in TE-derived**
12 **piRNA abundance.** Log2 fold-change TE transcript abundance and TE-derived antisense piRNA
13 abundance between two transgenic rescues for the TE families whose TE transcript abundance

1 is substantively different (> 2 fold) between two rescues from Fig 2A. Red dashed lines indicate

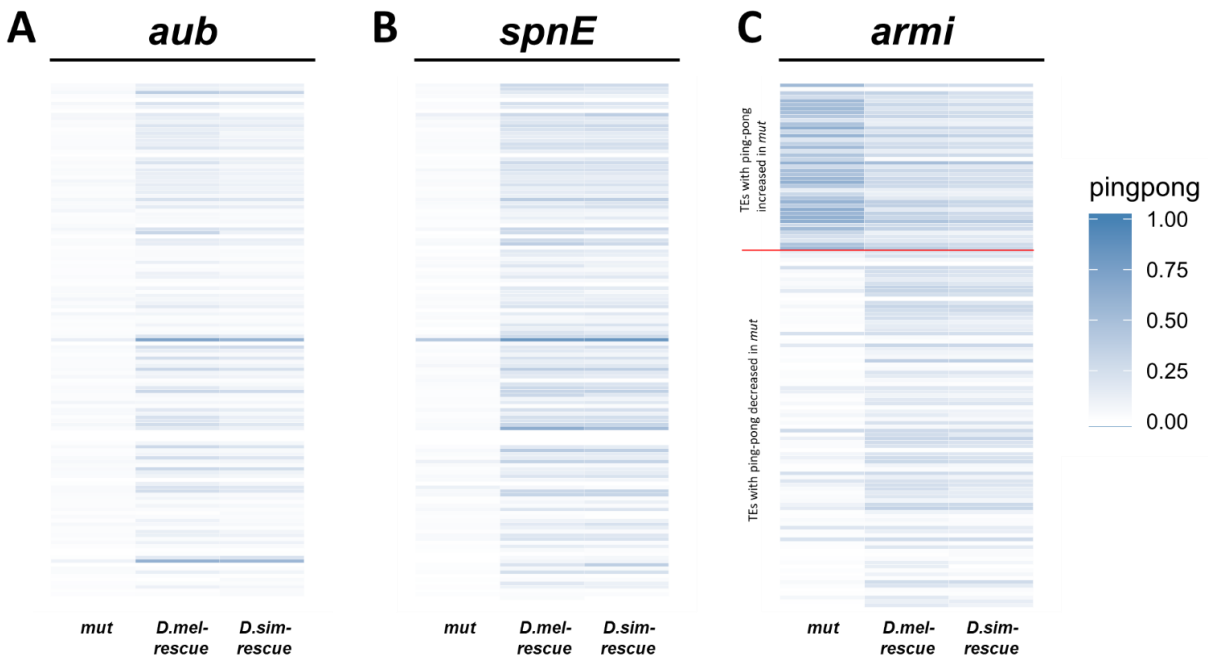
2 the 2 fold-change threshold.



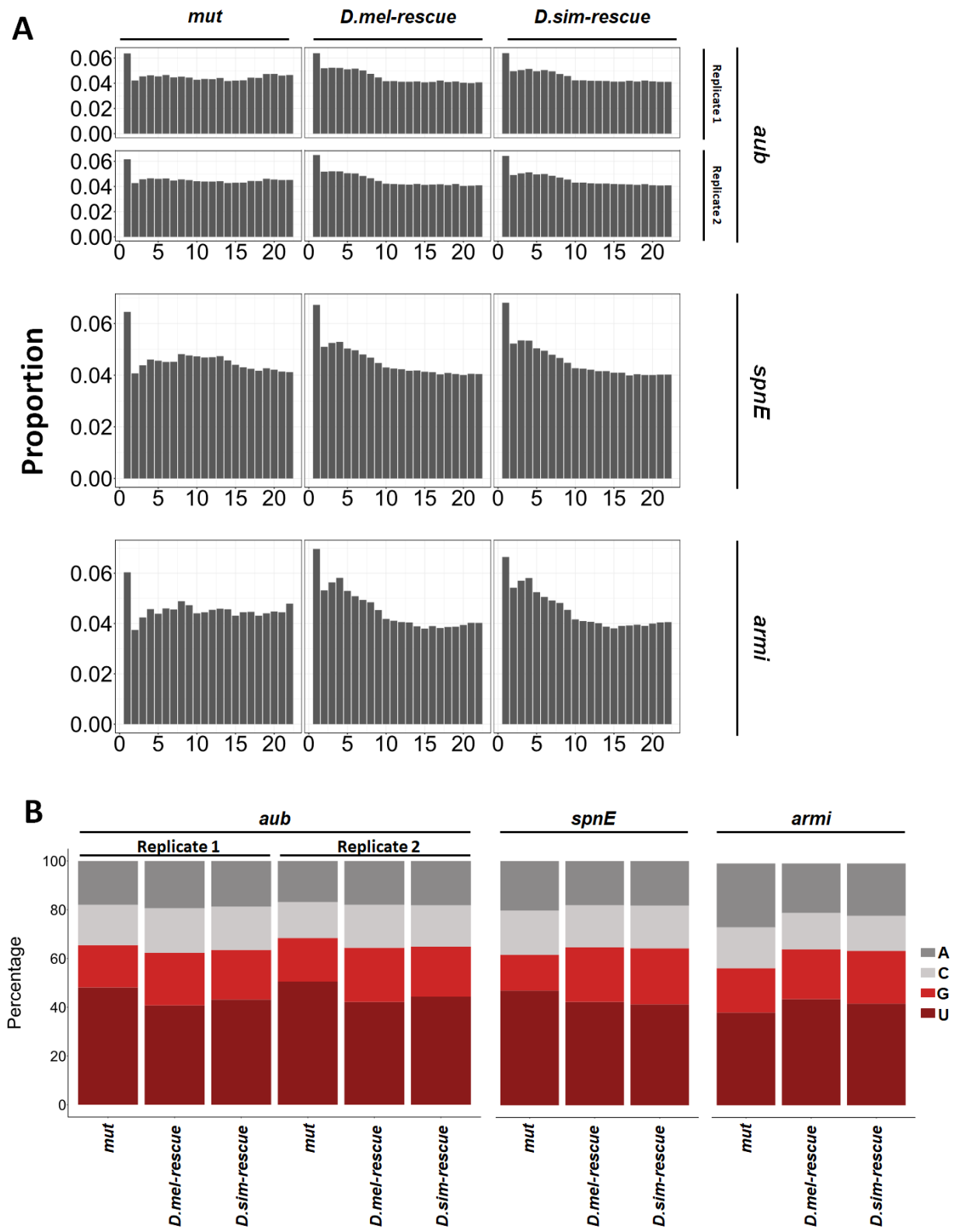
3

4 S5 Fig. **Composition of the small RNA pool for each protein studied.**

5



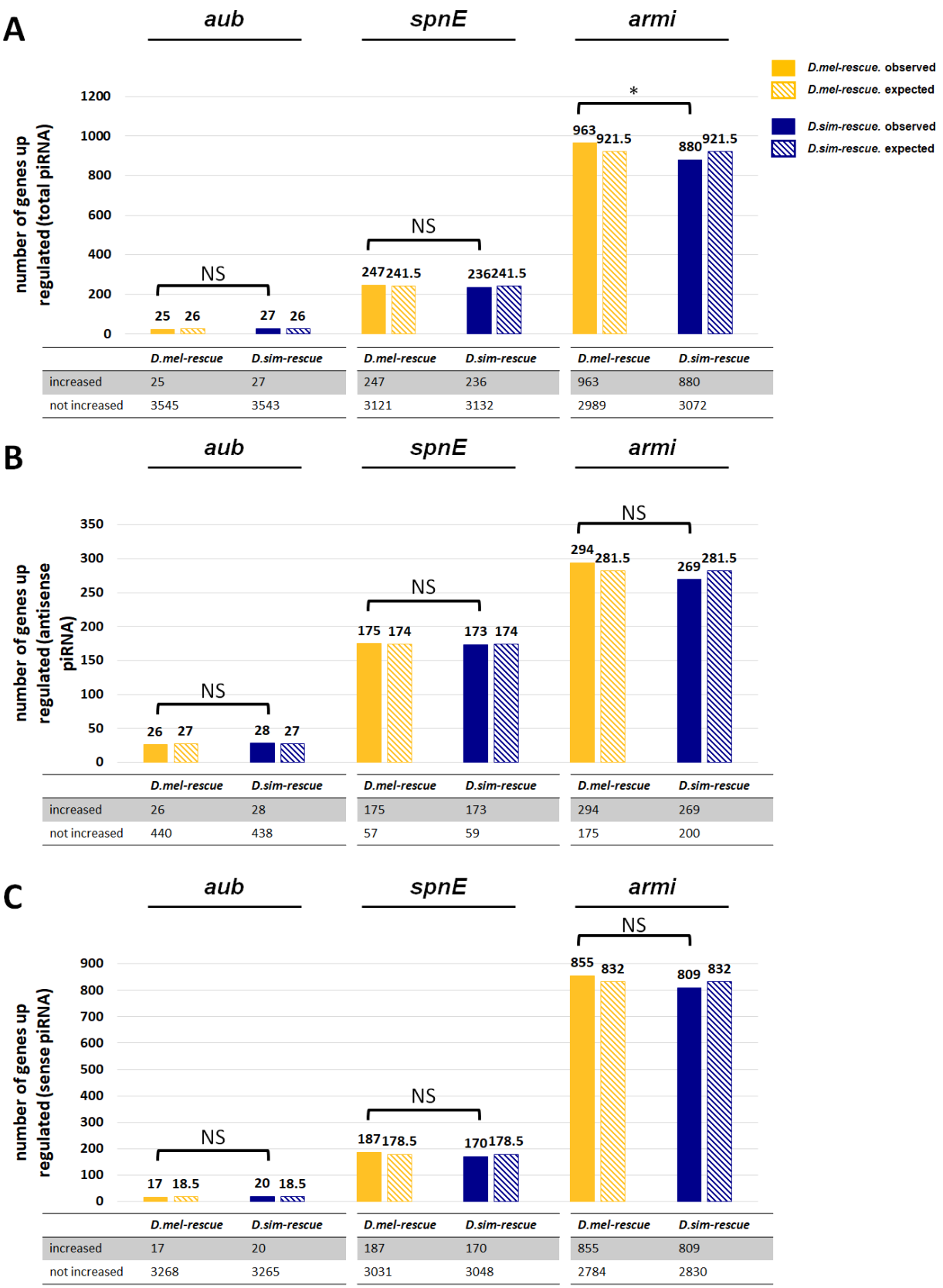
S6 Fig. Ping-pong fraction heat map for each protein studied. (A) *aub*, (B) *spn-E* and (C) *armi*. Among (C), 92 and 43 TE families whose ping-pong fractions are decreased (below red line) or increased (above red line), respectively, in *armi* mutant as compared to those in *D. melanogaster* transgenic rescue are shown.



1

2 **S7 Fig. Observed peaks of 1nt distance (A) and +1-U bias (B) among each genotype for each**

1 **protein studied.**


A
aub
spnE
armi
D.mel-rescue. observed
D.mel-rescue. expected
D.sim-rescue. observed
D.sim-rescue. expected
number of genes up regulated (total piRNA)
1200
1000
800
600
400
200
0
NS
NS
*
25 26
27 26
247 241.5
236 241.5
963 921.5
880 921.5
D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue
increased 25 27 247 236 963 880
not increased 3545 3543 3121 3132 2989 3072
B
aub
spnE
armi
number of genes up regulated (antisense piRNA)
350
300
250
200
150
100
50
0
NS
NS
NS
26 27
28 27
175 174
173 174
294 281.5
269 281.5
D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue
increased 26 28 175 173 294 269
not increased 440 438 57 59 175 200
C
aub
spnE
armi
number of genes up regulated (sense piRNA)
900
800
700
600
500
400
300
200
100
0
NS
NS
NS
17 18.5
20 18.5
187 178.5
170 178.5
855 832
809 832
D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue D.mel-rescue D.sim-rescue
increased 17 20 187 170 855 809
not increased 3268 3265 3031 3048 2784 2830

1

S8 Fig. Auto-immunity analysis for the genic piRNA profile. The number of genes whose corresponding total (A) / anti-sense (B) / sense (C) piRNA abundance is increased (>1.5 fold) in the presence of each transgene as compared to the mutant. Contingency tables are shown below. Log2 fold-change values were based on two biological replicates for *aub* and three biological replicates for *spn-E* and *armi*, and were obtained from a DESeq2 analysis (adjusted $p < 0.05$). Statistical significance was assessed by the Pearson's Chi-squared test. NS denotes $p > 0.05$. * denotes $p \leq 0.05$.

S1 Table. Offspring count from the female fertility test.

S2 Table. RNA-seq and small RNA-seq library statistics.

S3 Table. Normalized abundance and differential expression of TE transcripts.

S4 Table. Normalized abundance and differential expression of TE-derived piRNAs.

S5 Table. piRNA ping-pong biogenesis signature for TE-derived piRNAs.

S6 Table. piRNA phasing biogenesis signature (d1 proportion) for TE-derived piRNAs.

S7 Table. piRNA phasing biogenesis signature (+1-U proportion) for TE-derived piRNAs.

S8 Table. Protein-coding genes that are differentially regulated by transgenes as compared to the mutant.

S9 Table. Log2 fold-change of piRNA abundance between D. mel rescue vs D. sim rescue, for genes whose expression levels are down-regulated by D. simulans alleles only.

References

1. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. Oncogene. 2004;23: 8841–8846.

2. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene. 2008;27: 404–408.

3. Kidwell MG, Novy JB. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: Sterility Resulting from Gonadal Dysgenesis in the P-M System. Genetics. 1979;92: 1127–1140.

4. Vilà MR, Gelpí C, Nicolás A, Morote J, Schwartz S, Schwartz S, et al. Higher processing rates of Alu-containing sequences in kidney tumors and cell lines with overexpressed Alu-mRNAs. Oncol Rep. 2003;10: 1903–1909.

5. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41: 324–337.

6. Anxolabéhère D, Kidwell MG, Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol Biol Evol. 1988;5: 252–269.

7. Yang H-P, Barbash DA. Abundant and species-specific DINE-1 transposable elements in 12 Drosophila genomes. Genome Biol. 2008;9: R39.

8. de la Chaux N, Wagner A. BEL/Pao retrotransposons in metazoan genomes. BMC Evol Biol. 2011;11: 154.

9. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol Biol Evol. 2012;29: 1005–1017.

10. Lu J, Clark AG. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res. 2010;20: 212–227.

11. Malone CD, Hannon GJ. Molecular evolution of piRNA and transposon control pathways in Drosophila. Cold Spring Harb Symp Quant Biol. 2009;74: 225–234.

12. Kolaczkowski B, Hupalo DN, Kern AD. Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol Biol Evol. 2011;28: 1033–1042.

13. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci. 2009;364: 99–115.

14. Palmer WH, Hadfield JD, Obbard DJ. RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics. 2018;208: 1585–1599.

15. Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD. Recurrent and recent selective sweeps in the piRNA pathway. Evolution. 2013;67: 1081–1090.

16. Yi M, Chen F, Luo M, Cheng Y, Zhao H, Cheng H, et al. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. Genome Biol Evol. 2014;6: 1393–1407.

17. Vermaak D, Henikoff S, Malik HS. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet. 2005;1: 96–108.

18. Senti K-A, Brennecke J. The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. 2010;26: 499–509.

19. Blumenstiel JP, Erwin AA, Hemmer LW. What Drives Positive Selection in the Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis. Yale J Biol Med. 2016;89: 499–512.

20. Parhad SS, Tu S, Weng Z, Theurkauf WE. Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev Cell. 2017;43: 60–70.e5.

21. Castillo DM, Mell JC, Box KS, Blumenstiel JP. Molecular evolution under increasing transposable element burden in Drosophila: a speed limit on the evolutionary arms race. BMC Evol Biol. 2011;11: 258.

22. Yu B, Lin YA, Parhad SS, Jin Z, Ma J, Theurkauf WE, et al. Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. EMBO Rep. 2018;19. doi:10.15252/embr.201745418

23. Parhad SS, Yu T, Zhang G, Rice NP, Weng Z, Theurkauf WE. Adaptive evolution targets a piRNA precursor transcription network. bioRxiv. 2019. p. 678227. doi:10.1101/678227

24. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 2007;315: 1587–1590.

25. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128: 1089–1103.

26. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137: 522–535.

27. Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, et al. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes Dev. 2010;24: 2499–2504.

28. Andress A, Bei Y, Fonslow BR, Giri R, Wu Y, Yates JR 3rd, et al. Spindle-E cycling between nuage and cytoplasm is controlled by Qin and PIWI proteins. J Cell Biol. 2016;213: 201–211.

29. Pandey RR, Homolka D, Chen K-M, Sachidanandam R, Fauvarque M-O, Pillai RS. Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet. 2017;13: e1006956.

30. Ge DT, Wang W, Tipping C, Gainetdinov I, Weng Z, Zamore PD. The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria. Mol Cell. 2019;74: 982–995.e6.

31. Kelleher ES, Edelman NB, Barbash DA. Drosophila Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. PLoS Biol. 2012;10. doi:10.1371/journal.pbio.1001428

32. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137: 509–521.

33. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348: 812–817.

34. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348: 817–821.

35. Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, et al. Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell. 2015;59: 819–830.

36. Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, et al. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. 2015;29: 617–629.

37. Ishizu H, Kinoshita T, Hirakata S, Komatsuzaki C, Siomi MC. Distinct and Collaborative Functions of Yb and Armitage in Transposon-Targeting piRNA Biogenesis. Cell Rep. 2019;27: 1822–1835.e8.

38. Groth AC, Fish M, Nusse R, Calos MP. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics. 2004;166: 1775–1782.

39. Schüpbach T, Wieschaus E. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics. 1991;129: 1119–1136.

40. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of Drosophila melanogaster. Science. 2000;287: 2185–2195.

41. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007;450: 203–218.

42. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019;47: D351–D360.

43. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, et al. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell. 2009;138: 1137–1149.

44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

45. Olivieri D, Senti K-A, Subramanian S, Sachidanandam R, Brennecke J. The Cochaperone Shutdown Defines a Group of Biogenesis Factors Essential for All piRNA Populations in Drosophila. Mol Cell. 2012. doi:10.1016/j.molcel.2012.07.021

46. Post C, Clark JP, Sytnikova YA, Chirn G-W, Lau NC. The capacity of target silencing byDrosophilaPIWI and piRNAs. RNA. 2014. pp. 1977–1986. doi:10.1261/rna.046300.114

47. Sytnikova YA, Rahman R, Chirn G-W, Clark JP, Lau NC. Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in Drosophila ovarian cell cultures. Genome Res. 2014;24: 1977–1990.

48. Levine MT, McCoy C, Vermaak D, Lee YCG, Hiatt MA, Matsen FA, et al. Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. PLoS Genet. 2012;8: e1002729.

49. Levine MT, Vander Wende HM, Hsieh E, Baker EP, Malik HS. Recurrent Gene Duplication Diversifies Genome Defense Repertoire in Drosophila. Mol Biol Evol. 2016;33: 1641–1653.

50. Cañizares J, Grau M, Paricio N, Moltó MD. Tirant is a new member of the gypsy family of retrotransposons in Drosophila melanogaster. Genome. 2000;43: 9–14.

51. Fablet M, McDonald JF, Biémont C, Vieira C. Ongoing loss of the tirant transposable element in natural populations of Drosophila simulans. Gene. 2006;375: 54–62.

52. Kofler R, Nolte V, Schlötterer C. Tempo and Mode of Transposable Element Activity in Drosophila. PLoS Genet. 2015;11: e1005406.

53. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. 2015;25: 445–458.

54. Hu TT, Eisen MB, Thornton KR, Andolfatto P. A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Res. 2013;23: 89–98.

55. Venken KJT, He Y, Hoskins RA, Bellen HJ. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science. 2006;314: 1747–1751.

56. Wickersheim ML, Blumenstiel JP. Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. Biotechniques. 2013;55: 269–272.

57. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18: 1509–1517.

58. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12.

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

60. Gramates LS, Marygold SJ, Santos G dos, Urbano J-M, Antonazzo G, Matthews BB, et al.

FlyBase at 25: looking to the future. Nucleic Acids Res. 2017;45: D663–D671.

61. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 11.

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

63. Anders S, Pyl PT, Huber W. HTSeq - A Python framework to work with high-throughput sequencing data. bioRxiv. 2014. p. 002824. doi:10.1101/002824

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

65. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science. 2008;322: 1387–1392.