Divergence of piRNA pathway proteins affects piRNA biogenesis and off-target effects, but not TE transcripts, revealing a hidden robustness to piRNA silencing

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 found very few differences in TE transcript levels, but global effects on piRNA biogenesis, particularly for Armitage. Adaptive divergence in piRNA biogenesis in the absence of clear downstream effects on TE targets reveals a surprising degree of robustness that suggests that piRNA-mediated silencing is an “all-or-nothing” threshold trait. These results lead us to propose that the feature of importance for adaptation is piRNA production rather than piRNA abundance, and we further suggest that piRNA biogenesis evolves rapidly to establish control over invader and escaper TEs or in response to TE antagonism. We also discovered that *D. simulans* alleles exhibit enhanced off-target effects on host transcripts in a *D. melanogaster* background, suggesting the avoidance of genomic auto-immunity as an additional target of selection. Taken together, we propose 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.

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, 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, a third non-mutually exclusive model proposes that piRNA proteins 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 in the piRNA-mediated transcriptional silencer *rhino* along the lineage leading to *D. simulans* alters the interaction between Rhino protein and its cofactor Deadlock [20]. The authors propose that this functional change in Rhino reflects evolutionary escape from a TE encoded antagonist, which mimics Deadlock binding to Rhino.

Here, we broaden our understanding of the functional consequences of adaptive evolution among *Drosophila* piRNA effector proteins by examining three essential piRNA pathway components, Armitage (Armi), Aubergine (Aub) and Spindle-E (SpnE); this work significantly extends a preliminary analysis of Aub divergence [22]. All three proteins exhibit adaptive evolution along the lineage leading to *D. melanogaster, D. simulans* or both [13,15]. These three proteins play critical and diverse roles in piRNA biogenesis and TE silencing. Aub is a Piwi-clade Argonaute protein which, guided by piRNAs, enacts post-transcriptional silencing of TE-derived mRNAs [23]. Aub cleavage of TE-derived mRNAs also feeds forward the ping-pong amplification cycle, a core mechanism for the production of both sense and antisense piRNAs from complementary precursors [24–26]. SpnE also contributes to ping-pong biogenesis by promoting the localization of Aub into the perinuclear nuage, where ping-pong occurs [27]. Distinct from both Aub and SpnE, Armi binds to antisense piRNA precursors to facilitate their sequential cleavage by the nuclease Zucchini in an alternate biogenesis mechanism referred to as “phasing” [28–31]. In contrast to ping-pong derived piRNAs, phased piRNAs play a major role in transcriptional silencing of TE insertions [32].

To isolate diverged functions of adaptive 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. We observe that the most dramatic functional differences between alleles impact their roles in piRNA biogenesis, revealing

a potential nexus of adaptive protein evolution. Furthermore, we observe that the piRNA losses associated with *D. simulans* alleles, which are particularly pronounced for Armi, are not associated with downstream changes in TE transcripts. The absence of effects on TE regulation suggests that TE transcript abundance is not the primary target of selection on these adaptively evolving piRNA proteins, and further reveals unexpected robustness in piRNA-mediated silencing. Finally, we reveal that *D. simulans* alleles exhibit enhanced off-target effects on the expression of host transcripts in a *D. melanogaster* background, suggesting the avoidance of genomic auto-immunity as an additional target of selection.

Results

Identifying functional divergence through interspecific complementation.

To isolate functional 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 (Figure S1). 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 [33], 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.

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 (Figure S2A) [34]. 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 (Figure S2A) [22]. Because *D. simulans* transgenes do not exhibit significantly reduced expression when compared to *D. melanogaster* (Figure S2B), these fertility effects most likely reflect amino acid sequence divergence in the encoded proteins.

Reduced piRNA abundance with *D. simulans* alleles.

To uncover molecular phenotypes that relate to fertility differences, we first examined whether *D. melanogaster* and *D. simulans* alleles differed with respect to piRNA production using small-RNA seq. piRNA pathway mutants cause a dramatic decrease in piRNA abundance for the majority of TE families when compared to the transgenic rescue by *D. melanogaster* alleles, consistent with their critical functions in piRNA biogenesis (Figure 1) [25,35–37]. While not as dramatic as the mutant, *D. simulans armi* showed a similar systematic reduction in TE-derived piRNA abundance, with 116 of 126 TE families exhibiting fewer piRNAs when compared to *D. melanogaster* (Figure 1). *Drosophila simulans armi* therefore has diverged considerably from *D. melanogaster* with respect to piRNA biogenesis. By contrast, *D. simulans aub* and *spn-E* alleles have isolated effects on individual TE families. *Drosophila simulans aub* modestly reduces piRNA biogenesis, with 12 out of 14 significantly affected TE families exhibiting reduced piRNA abundance when compared to *D. melanogaster* (Figure 1). By contrast, for *spnE*, we observed more idiosyncratic differences, with a similar number of TE families exhibiting increased piRNA abundance in the presence of the *D. simulans* allele (21 TE families) as the *D. melanogaster* allele (25 TE families, Figure 1). Importantly, we did not observe any systematic differences in expression for germline or soma-specific protein-coding genes between the transgenic rescues, indicating changes in piRNA abundance do not reflect a change in the germline-to-soma ratio (Figure S2B).

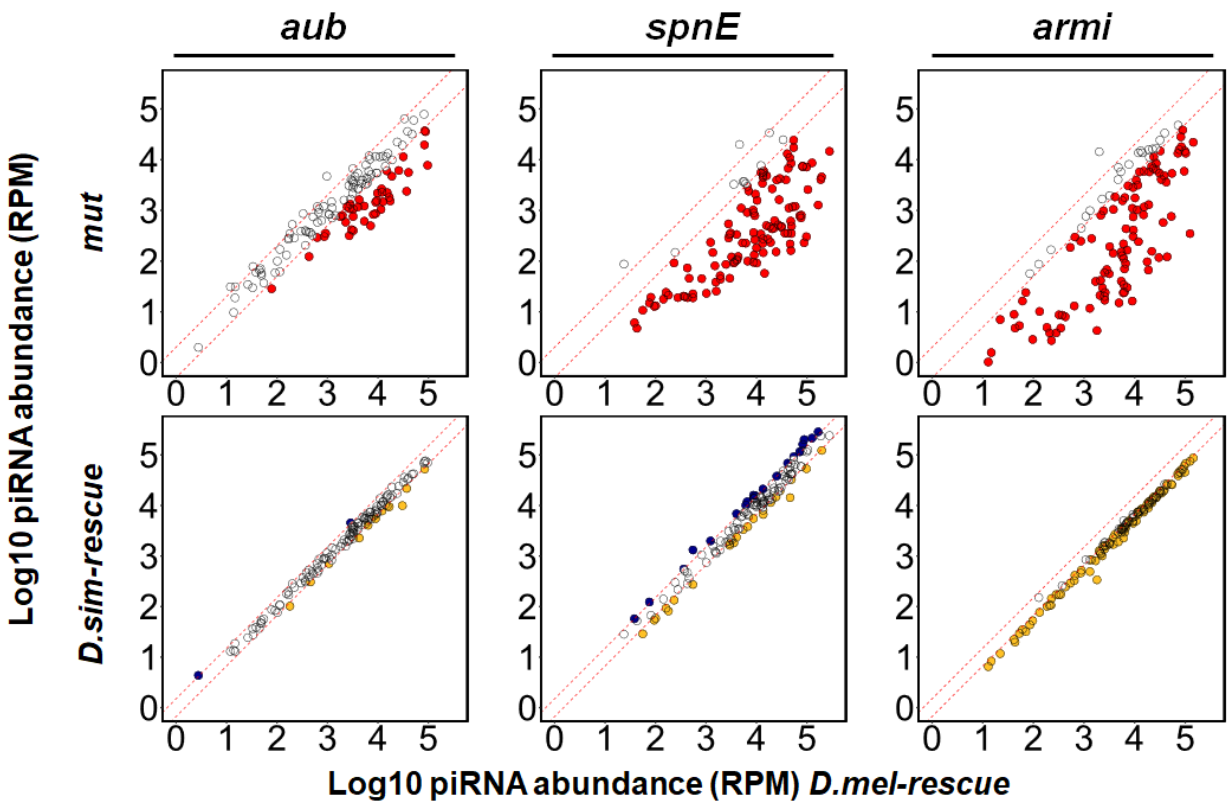


Figure 1. *Drosophila simulans armi* is characterized by a dramatic loss of TE-derived piRNAs. TE-derived piRNA abundance is compared between *D. melanogaster* rescues and trans-heterozygous mutants ("mut", upper row) or *D. simulans* rescues (lower row) for *aub*, *spnE* and *armi*. Red dashed lines indicate the 2 fold-change threshold (upper row) or 1.5 fold-change threshold (lower row). TE families whose TE-derived piRNA abundance in *D. melanogaster* transgenic rescue is 2 fold higher than that in corresponding mutant are indicated in red. TE families whose piRNA abundance differs between transgenic rescues (>1.5 fold) are indicated in yellow and blue, for TE families increased in *D. melanogaster* or *D. simulans* rescues, respectively. TE-derived piRNA abundance is based on two biological replicates for *aub* and three biological replicates for *spnE* and *armi*, and was normalized to the total number of sequenced miRNAs in the same library.

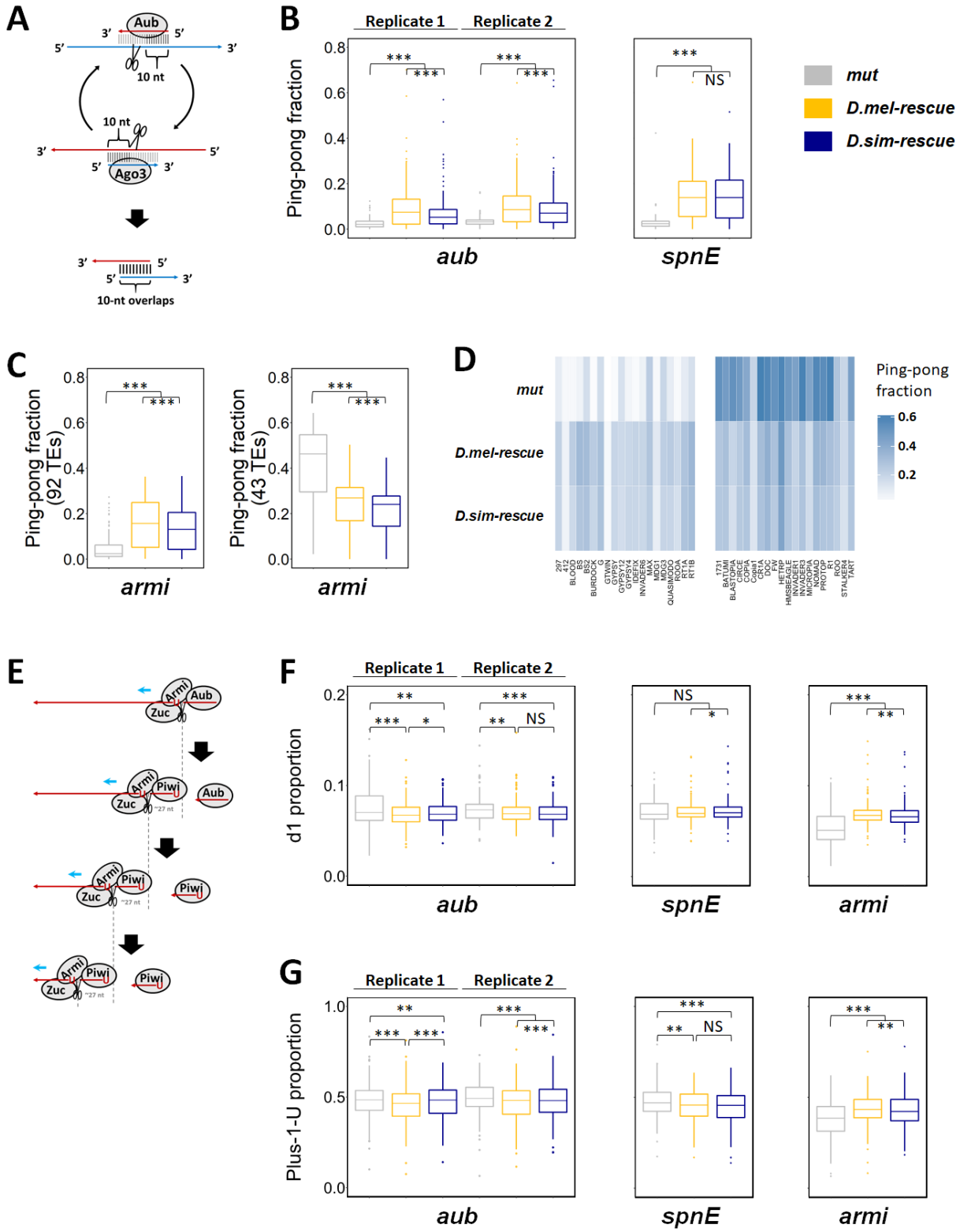
***Drosophila simulans* alleles exhibit reduced piRNA biogenesis.**

To detect underlying mechanisms that could explain differences in piRNA abundance, we examined molecular signatures of the two major mechanisms of piRNA biogenesis in our small RNA libraries: ping-pong and phasing. The frequency of ping-pong amplification was 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 factors of ping-pong cycle, Aub and Argonaute-3 (Ago-3, Figure 2A-D) [23,24,26]. Similarly, phasing biogenesis was 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 single nucleotide (d1), both of which are diagnostic of cleavage by the nuclease Zucchini (Figure 2E-G) [28,29]. 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 [28,29].

Aub plays a direct role in ping-pong amplification by cleaving piRNA precursors (Figure 2A) [23,24,26], and *spnE* is required for the localization of Aub into the perinuclear nuage, where ping-pong occurs [27]. Mutations in either gene therefore cause a complete collapse of ping-pong amplification (Figure 2B, Figure S3A and S3B) [25,26]. Both *D. melanogaster* and *D. simulans aub* and *spnE* alleles exhibited a dramatic increase in the ping-pong fraction, indicating a conserved role in ping-pong biogenesis (Figure 2B). 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 (Figure 2F and 2G, Figure S4). This reduced efficiency in ping-pong biogenesis mirrors the modest reduction in piRNA abundance we observed for this allele, with 11 out of 12 TEs whose piRNA abundance is reduced by *D. simulans aub* allele (Figure 1) showing decreased ping-pong fraction. By contrast, *D. simulans spnE* alleles do not reduce ping-pong (Figure 2B), which is consistent with the absence of a systematic loss of the *D. simulans* allele on piRNA production (Figure 1). Interestingly however, there was a modest but

significant increase in the d1 proportion with the *D. simulans spnE* rescue (Figure 2F), 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 (Figure 2E) [28,29,38]. Both d1 and +1-U are therefore significantly reduced in *armi* mutants (Figure 2F and 2G, Figure S4). 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 [26,39]. Ping-pong fractions are therefore decreased in *armi* mutants for some TE families (Figure 2C and 2D, Figure S3C). 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 (Figure 2C and 2D, Figure S3C). Consistent with reduced piRNA production (Figure 1), the *D. simulans armi* rescue exhibited modestly but significantly reduced d1 and +1-U proportion, indicating reduced phasing (Figure 2F and 2G, Figure S4). 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* (Figure 2C and 2D, Figure S3C). 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*.



1

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

Simplified diagram of ping-pong amplification loop. (B) Ping-pong fractions of TE-derived piRNAs from 142 TE families are compared between trans-heterozygous mutants and transgenic rescues for *aub* and *spnE.* (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 *spnE* 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.

piRNA biogenesis defects do not cause TE derepression.

Ultimately, the piRNA pathway represses the production of TE-derived mRNAs (reviewed in [18]). Enhanced negative regulation of TEs is therefore an obvious target of positive selection acting on piRNA pathway proteins. Furthermore, given the reduced abundance of TE-derived piRNAs in the *D. simulans armi* rescue, deregulation is expected among targeted TEs. We therefore compared TE transcript abundance between mutants and transgenic rescues using mRNA-seq (*aub*) and stranded total-RNA seq (*spnE* and *armi*).

For all three mutants, *D. melanogaster* transgenic rescues reduce transcript abundance for almost all TE families when compared to the corresponding mutant (Figure 3). Furthermore,

in all three cases, even though there are idiosyncratic differences in the expression of individual TE families, the majority of TEs are not differentially expressed between the transgenic rescues (Figure 3, Figure S5). This observation is particularly unexpected for *armi*, where defects of the *D. simulans* allele in ping-pong and phasing biogenesis leads to a systematic loss of piRNAs targeting the majority of TE families (Figure 1, 2C, 2D, 2F and 2G). Furthermore, the majority of TE families that are differentially expressed between the two transgenic rescues did not exhibit an associated difference in piRNA abundance (36 of 38 cases among all three genes, Figure S5). Interspecific differences in TE regulation, therefore, are unrelated to differences in piRNA production.

We further explored the overall quantitative relationship between TE transcript levels and piRNA abundance. When *aub* and *spnE* mutants are compared to transgenic rescues, the dramatic reduction of antisense piRNAs is associated with increased TE transcript abundance (Figure 4A). Strikingly, however, we observed no correlated changes between piRNA and mRNA pools for *aub* or *spnE* when the two transgenic rescues are compared to each other, suggesting the magnitude of changes in piRNA abundance are not sufficient to impact downstream targets (Figure 4A). *armi* mutants exhibit a distinct relationship between piRNA losses and TE transcripts, in which increased antisense TE transcript abundance is associated with minimal losses of antisense piRNAs (Figure 4B). This counterintuitive positive relationship between mRNA and piRNA pools likely reflects enhanced ping-pong biogenesis of bidirectionally transcribed TEs in *armi* mutants (Figure 4B). Nevertheless, comparisons between the two *armi* transgenic rescues again do not reveal any connections between changes in TE-derived piRNAs and mRNAs. Therefore, modest changes in piRNA production associated with different transgenes do not translate to altered TE transcription, revealing a hidden robustness to piRNA-mediated regulation.

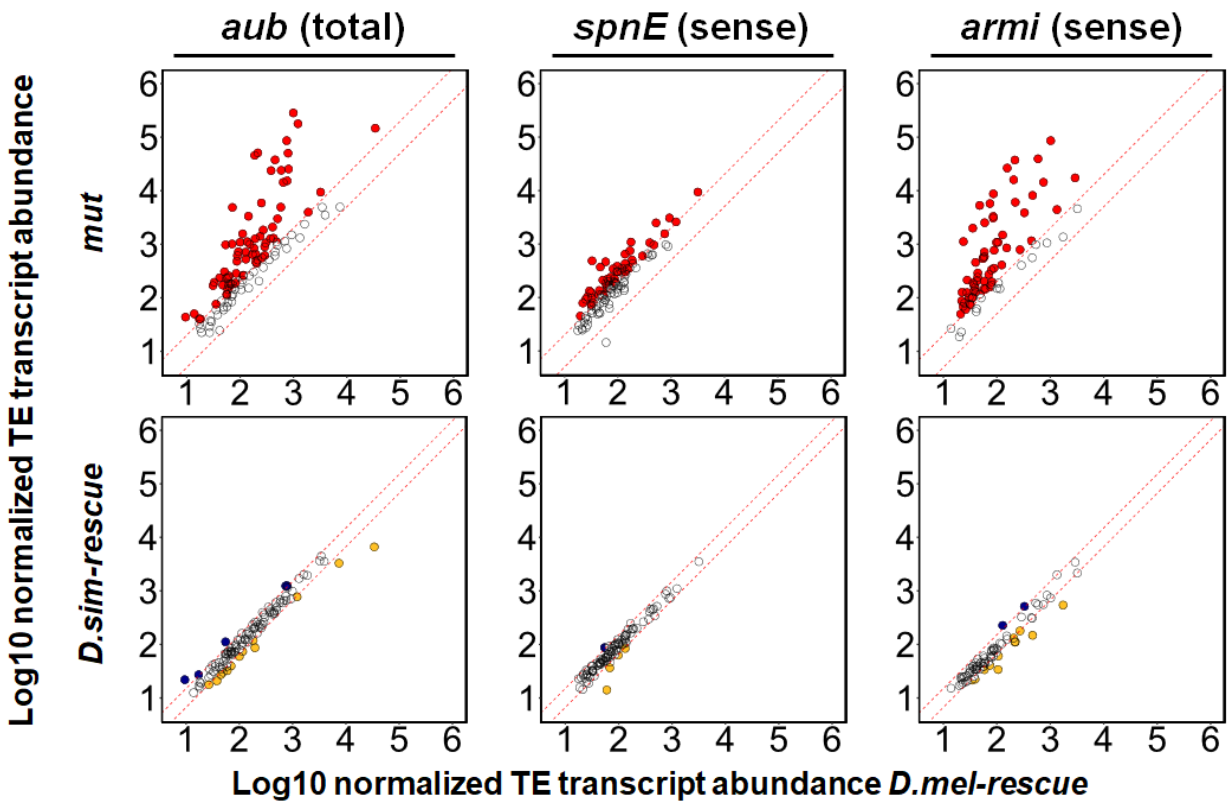
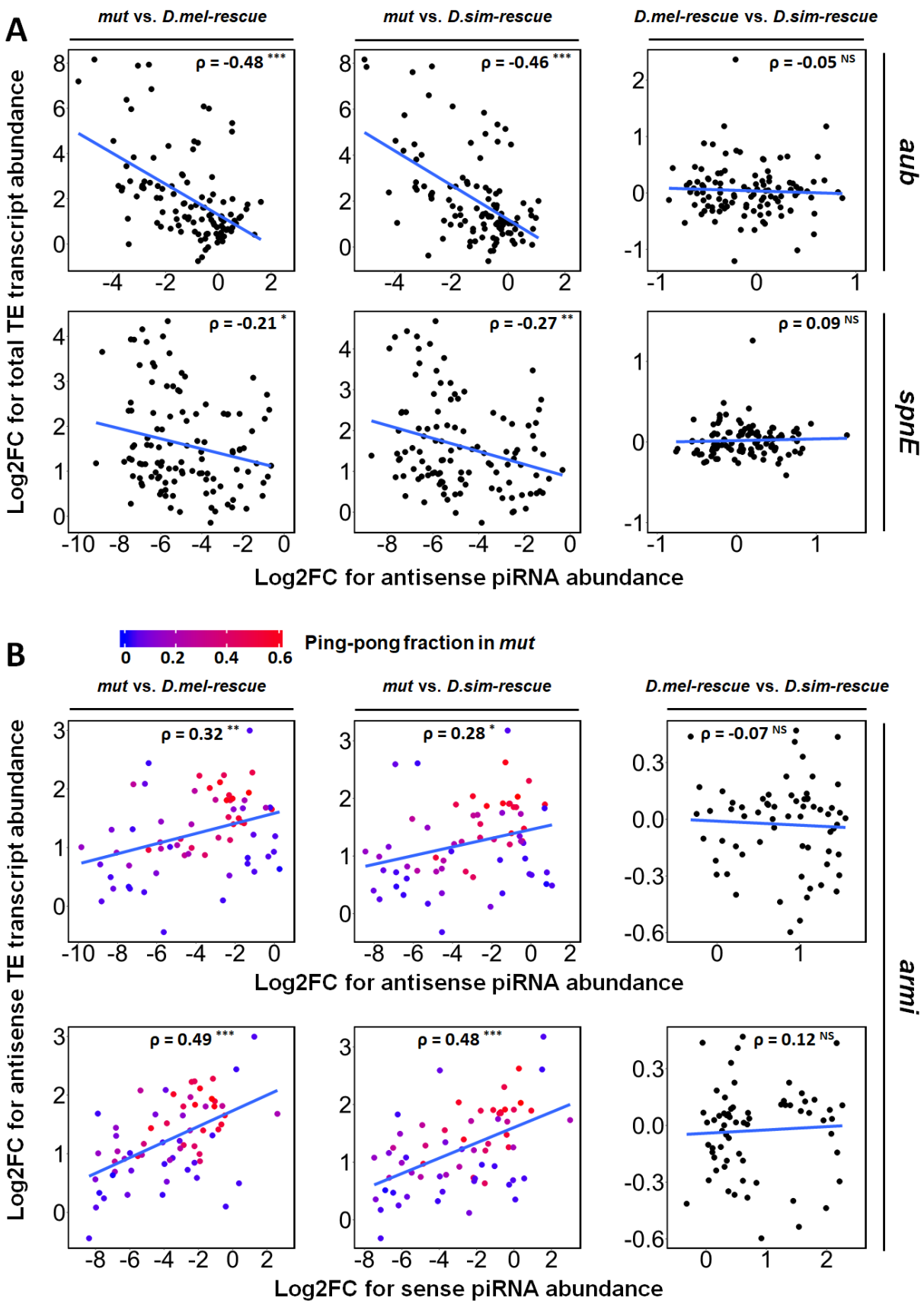


Figure 3. Negative transcriptional regulation of TEs is largely conserved between species. TE transcript abundance in *D. melanogaster* rescues is compared to that in trans-heterozygous mutants (upper row) or in *D. simulans* rescues (lower row) for *aub* (total TE transcripts), *spnE* and *armi* (sense TE transcripts). Red dashed lines indicate the 2 fold-change threshold (upper row) or 1.5 fold-change threshold (lower row). TE families whose transcript abundance in *D. melanogaster* transgenic rescue is >2 fold lower than that in corresponding mutant are indicated in red. TE families whose transcript abundance differs between transgenic rescues (>1.5 fold, adjusted *p*-value < 0.05) are indicated in yellow and blue, for TE families increased in *D. melanogaster* or *D. simulans* rescues, respectively. TE transcript abundance is based on one biological replicate for *aub* and three biological replicates for *spnE* and *armi*, and was obtained from a combined DESeq2 analysis that included both TEs and protein-coding genes.


A
mut vs. D.mel-rescue
mut vs. D.sim-rescue
D.mel-rescue vs. D.sim-rescue
ρ = -0.48 ***
ρ = -0.46 ***
ρ = -0.05 NS
ρ = -0.21 *
ρ = -0.27 **
ρ = 0.09 NS
aub
spnE
Log2FC for total TE transcript abundance
Log2FC for antisense piRNA abundance
B
0 0.2 0.4 0.6
Ping-pong fraction in mut
mut vs. D.mel-rescue
mut vs. D.sim-rescue
D.mel-rescue vs. D.sim-rescue
ρ = 0.32 **
ρ = 0.28 *
ρ = -0.07 NS
ρ = 0.49 ***
ρ = 0.48 ***
ρ = 0.12 NS
armi
Log2FC for antisense TE transcript abundance
Log2FC for antisense piRNA abundance
Log2FC for sense piRNA abundance

1

Figure 4. Allelic differences in piRNA production and TE transcripts are unrelated. (A). Relationship between differences in total TE transcript abundance and in antisense piRNA abundance for different *aub* and *spnE* genotypes. (B). Relationship between changes in antisense TE transcript abundance and changes in antisense or sense piRNA abundance for each comparison for *armi*. Dot color in panel B indicates the ping-pong fraction of each TE family in the *armi* mutant. ρ: Spearman's rank correlation coefficient. Blue line indicates the least-squares regression line. NS denotes $p > 0.05$. * denotes $p \leq 0.05$. ** denotes $p \leq 0.01$. *** denotes $p \leq 0.001$.

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

Avoidance of genomic auto-immunity syndrome is an alternative, non-TE centered hypothesis to explain the adaptive evolution of piRNA pathway proteins [19,21]. Under this model, piRNA pathway proteins experience selection to minimize their off-target effects on protein-coding genes, which would reduce host fitness. Selection to avoid genomic auto-immunity therefore predicts a greater number of off-target effects with the *D. simulans* rescues, which are not adapted to the *D. melanogaster* background. To test this prediction, we identified protein-coding genes that are negatively regulated by piRNA pathway proteins by comparing their expression levels between mutants and transgenic rescues. Protein-coding genes whose expression is significantly reduced in transgenic rescues (>1.5 fold) are candidate off-target effects of piRNA-mediated silencing. For all three genes, we observed more protein-coding genes whose expression is reduced by the *D. simulans* rescue than the *D. melanogaster* rescue, although the difference is significant only for *aub* and *armi* (Figure 5A). 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 (Figure 5B). Together, these observations suggest that *D. simulans* alleles have enhanced impacts on protein-coding genes, consistent with off-target effects.

Increased off-target effects of *D. simulans* alleles on protein-coding genes could be explained by increased production of genic piRNAs, or more efficient silencing of target mRNAs. To differentiate between these alternatives, we compared genic piRNAs between transgenic rescues. Surprisingly, *D. simulans* transgenic rescues of *aub* and *armi* exhibited fewer protein-coding genes with increased piRNA abundance (>1.5 fold) when compared to mutants than *D. melanogaster* transgenic rescues (Figure S6A). Furthermore, the magnitude of increased protein-coding piRNA production was not higher in *D. simulans* than *D. melanogaster* transgenic rescues when compared to the mutants (Figure S6B). Therefore, similar to our observations with TEs, allelic differences in piRNAs and target protein-coding transcripts are not necessarily related.

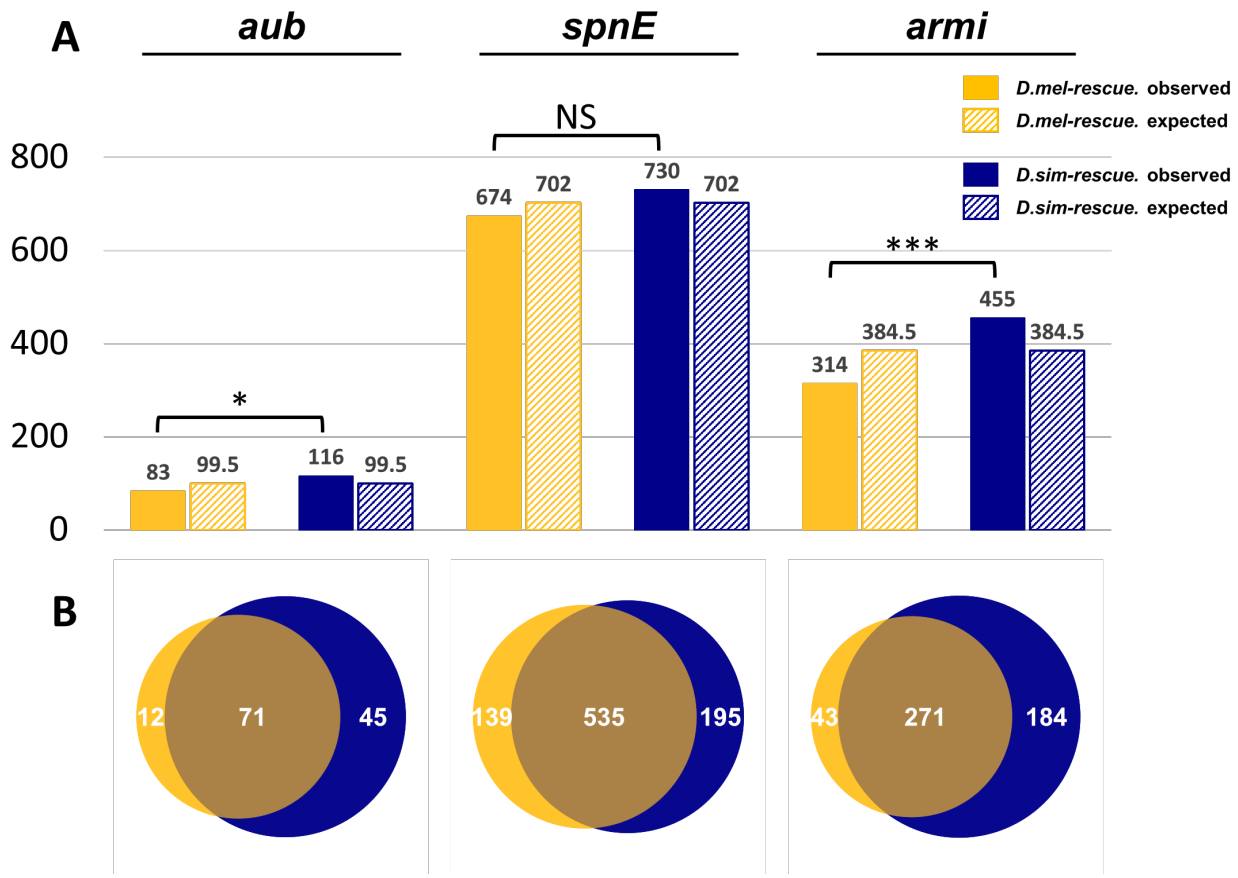


Figure 5. Negative regulation of protein-coding genes suggests increased genomic auto-immunity of *D. simulans* alleles. (A) The number of genes whose expression levels are decreased (>1.5 fold) in the

presence of each transgene as compared to the corresponding mutant. Log2 fold-change values were based on one biological replicate for *aub* and three biological replicates for *spnE* and *armi*, and were obtained from a DESeq analysis for *aub* and a DESeq2 analysis for *spnE* and *armi* (adjusted $p < 0.05$). Statistical significance was assessed by the Pearson's Chi-squared Test of Independence, where NS denotes $p > 0.05$. * denotes $p \leq 0.05$. *** denotes $p \leq 0.001$. (B) 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), *spnE* (middle Venn diagram) and *armi* (right Venn diagram).

Discussion

Despite pervasive adaptive evolution and gene duplication among piRNA pathway proteins in both insect and vertebrate lineages [12–15,40,41], 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. Our discovery of reduced piRNA biogenesis associated with *D. simulans aub* and *armi* alleles implicates piRNA production as a nexus of adaptive evolution. Furthermore, because piRNA losses associated with *D. simulans* alleles do not result in downstream transcriptional derepression of target TEs, our observations do not support TE transcript abundance as a major target of positive selection. Additionally, for both *aub* and *armi*, we observed that *D. simulans* alleles are associated with enhanced off-target effects on host gene transcripts, supporting genomic auto-immunity as a second driver of adaptive evolution. Taken together, our data suggest that positive selection acts at multiple molecular and functional interfaces within the piRNA pathway.

piRNAs direct TE silencing, and the loss of specific piRNAs in many mutant backgrounds is associated with a corresponding increase in target TE transcripts (Figure 4A) [25,35]. However, we discovered that reduced piRNA production associated with *D. simulans* alleles does not cause derepression of the corresponding target TEs (Figure 1, 3, and 4). While we cannot rule out the possibility that derepression occurs but is not detected by RNA-seq in whole ovaries because it happens at a particular stage of oogenesis [42], our observations suggest that piRNA-mediated silencing is a threshold trait: once sufficient piRNAs are produced to target a particular TE family, additional piRNAs do not increase the strength of silencing. Indeed, threshold effects have been observed in piRNA-mediated silencing of reporter genes [43], and in the “all-or-nothing” accumulation of Rhino—an effector of transcriptional silencing—at piRNA-target loci [44]. Threshold effects are further supported by the lack of correlation between piRNA abundance and transposition rate among active TE families in *D. melanogaster*,

suggesting that there is no selection for increased piRNA production once silencing is established [45].

What are the evolutionary implications of this model of threshold effects? Threshold effects may explain why we observed more dramatic allelic differences affecting piRNA biogenesis compared to TE transcript levels. If silencing is determined by a critical mass of piRNAs, then establishing silencing over invading or escaper TEs, or escaping from TE-encoded antagonists, will be arbitrated at the level of piRNA production. piRNA biogenesis is therefore expected to evolve adaptively. While it may seem counterintuitive that we did not observe any incompatibilities between *D. simulans* alleles and *D. melangoaster* TEs, in which insufficient piRNAs are produced to enact silencing, it must be considered that *D. simulans* and *D. melanogaster* alleles are largely coevolving with the same TE families as a consequence of the frequent horizontal transfer of TEs between these two species [46–49]. Therefore, the incompatibilities that emerge between species are expected to occur between components of the piRNA biogenesis machinery that have evolved independently in each lineage, rather than between piRNA machinery and target TEs.

The *armi* phenotype we discovered is particularly intriguing, where the *D. simulans* allele is characterized by reduced ping-pong and phasing biogenesis when compared to *D. melanogaster*, leading to a systematic loss of piRNAs (Figure 1, 2C, 2D, 2F and 2G). This is distinct from the phenotype of *armi* mutants, in which phasing is lost for all TE families (Figure 2F and 2G), but ping-pong is affected in a TE-dependent manner (Figure 2C and 2D, Figure S3C) [26,39]. We propose that reduced ping-pong and phasing by *D. simulans* Armi is potentially explained by enhanced interactions with Ago-3 and Aub during the transfer of piRNA precursors from ping-pong to phasing sites [50,51]. Increased stability of Armi-Aub or Armi-Ago-3 complexes would reduce the availability of Aub/Ago-3 for ping-pong biogenesis, while concurrently reducing the availability of Armi for phasing biogenesis (Figure 6). Enhanced affinity could have evolved if, for example, *D. simulans* Armi competes with a TE antagonist for

binding to Aub or Ago-3. Escape from TE-encoded antagonist proteins was recently proposed as the evolutionary mechanism responsible for an incompatibility between *D. simulans* Rhino and *D. melanogaster* Deadlock [20,52], two components of a complex that promotes piRNA cluster transcription [53,54].

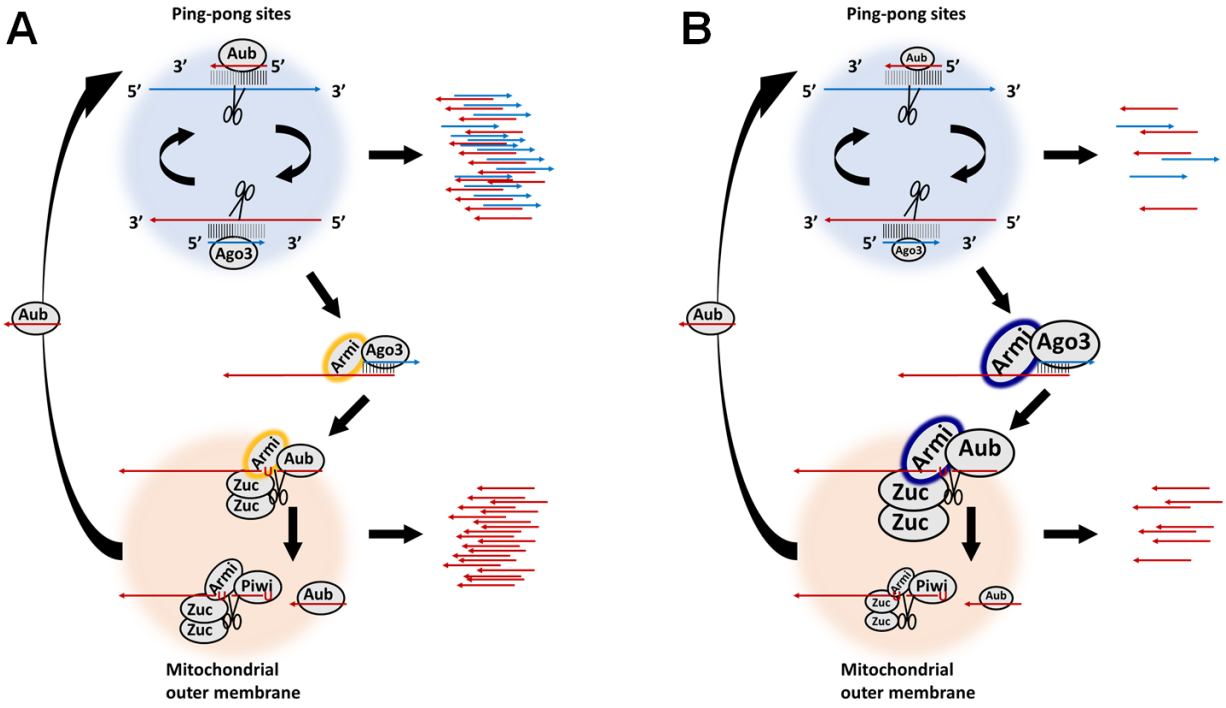


Figure 6. A model for incompatibility of *D. simulans* Armi in a *D. melanogaster* background leading to defects in piRNA biogenesis. (A) *D. melanogaster* Armi normally interacts with ping-pong factor proteins, Aub and Ago-3, allowing for normal piRNA production. (B) Incompatibility of *D. simulans* Armi in a *D. melanogaster* genetic background results in an aberrantly higher affinity to those ping-pong factor proteins, which isolates Armi and those ping-pong factors from phasing and ping-pong biogenesis, respectively, interfering with normal piRNA biogenesis.

Because piRNAs have the potential to affect any transcript including those of the host, avoidance of off-target effects on host gene expression is proposed as an alternative driver of adaptive evolution among piRNA pathway proteins [19,21]. This genomic auto-immunity model

is supported by the observation that species with a greater TE burden have longer piRNAs that favor specificity over sensitivity [19,21]. In our study, we provide direct evidence of interspecific divergence in genomic auto-immunity at the protein level, with *D. simulans* alleles of both *aub* and *armi* exhibiting enhanced repression of host genes when compared to their *D. melanogaster* counterparts (Figure 5).

Taken together, our study points to multiple diverging functions and unexpected complexity in the molecular adaptation of piRNA pathway proteins. In particular, Aub and Armi appear to have diverged functionally between *D. melanogaster* and *D. simulans* with respect to both piRNA biogenesis and their off-target effects on host genes. These findings point to a picture in which the piRNA proteins are pulled in two directions: to maintain their potency in an evolutionary arms race with TEs, but also to minimize friendly-fire by differentiating between self-and non-self when enacting genome defense.

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 [22], 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* (*spnE*), the following *D. melanogaster* strains were used: *yw; spnE^{1}/TM6*, *yw; spnE$^{hls-03987}$/TM6*, *yw; spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}*, *yw;*

1 *spnE$^{hls-03987}$/TM6; ΦP{D. simulans spnE}*. *yw; spnE^{1}/TM6* and *yw; spnE$^{hls-03987}$/TM6* were
2 obtained by crossing *spnE^{1}/TM3* and *spnE$^{hls-03987}$/TM3* (gifts from Celeste Berg) to *yw;*
3 *TM3/TM6*. To generate *yw; spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}* and *yw; spnE^{hls-}*
4 *^{03987}/TM6; ΦP{D. simulans spnE}, w^{1118}; ΦP{D. melanogaster spnE}* and *w^{1118}; ΦP{D. simulans*
5 *spnE}* were first crossed to *yw: TM3/TM6. +/TM6; ΦP{D. melanogaster spnE}/+* and *+/TM6;*
6 *ΦP{D. simulans spnE}/+* offspring were then crossed to *yw; spnE$^{hls-03987}$/TM3.* Finally, *yw;*
7 *spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}/+* and *yw; spnE$^{hls-03987}$/TM6; ΦP{D. simulans*
8 *spnE}/+* offspring were backcrossed into *yw; spnE$^{hls-03987}$/TM6* for 6 generations, and
9 subsequently homozygosed for the transgene, to minimize background effects.
10 For the studies of *armitage* (*armi*), the following *D. melanogaster* strains were used: *yw;*
11 *armi^{1}/TM6*, *w; armi$^{72.1}$/TM6*, *w; armi$^{72.1}$/TM6; ΦP{D. melanogaster armi}*, *w; armi$^{72.1}$/TM6*; *ΦP{D.*
12 *simulans armi}*. *yw; armi^{1}/TM6* was obtained by crossing *yw; armi^{1}/TM3* (Bloomington
13 *Drosophila* Stock Center #8513) to *yw; TM3/TM6*. *w; armi$^{72.1}$/TM6* was obtained from
14 Bloomington *Drosophila* Stock Center (#8544). To generate *w; armi$^{72.1}$/TM6; ΦP{D.*
15 *melanogaster armi}* and *w; armi$^{72.1}$/TM6*; *ΦP{D. simulans armi}*, *w^{1118}; ΦP{D. melanogaster*
16 *armi}* and *w^{1118}; ΦP{D. simulans armi}* were first crossed to *yw; TM3/TM6*. *+/TM3; ΦP{D.*
17 *melanogaster armi}/+* and *+/TM3; ΦP{D. simulans armi}/+* offspring were then crossed to *w;*
18 *armi$^{72.1}$/TM6*. Finally, *w; armi$^{72.1}$/TM3; ΦP{D. melanogaster armi}/+* and *w; armi$^{72.1}$/TM3*; *ΦP{D.*
19 *simulans armi}/+* were backcrossed into *w; armi$^{72.1}$/TM6* for 6 generations, and subsequently
20 homozygosed for the transgene, to minimize background effects.
21 Experimental genotypes were obtained from the following crosses. For studies of *aub*,
22 virgin females *w; aub^{N11} bw^{1}/CyO* were crossed to (1) *yw; aub^{HN} bw^{1}/CyO*, (2) *yw; aub^{HN}*
23 *bw^{1}/CyO; ΦP{D. melanogaster aub}* or (3) *yw; aub^{HN} bw^{1}/CyO; ΦP{D. simulans aub}* males. For
24 studies of *spnE,* virgin females *yw; spnE^{1}/TM6* were crossed to (1) *yw; spnE$^{hls-03987}$/TM6*, (2) *yw;*
25 *spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}* or (3) *yw; spnE$^{hls-03987}$/TM6; ΦP{D. simulans*
26 *spnE}* 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 25°C 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 [33], which allows for site-specific integration. To generate transgenes for site specific integration, the gene and flanking regulatory regions of *spnE* (~9.7Kb, *D. melanogaster* Release 6, 3R:15835349..15845065; *D. simulans* Release 2, 3R:9575537..9585081) [55,56] and *armi* (~6Kb, *D. melanogaster* Release 6, 3L:3460305..3466368; *D. simulans* Release 2, 3L:3357002..3363099) [55,56] were PCR-amplified by using corresponding primers (below) and iProof high-fidelity taq DNA polymerase (Bio-Rad).

D.mel/D.sim-spnE forward primer: ATTGAACGCCGTCTATGCCAAGC

D.mel/D.sim-spnE 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 *spnE* TOPO plasmid, and the NotI/NheI fragment of each *armi* TOPO plasmid into NotI/BamHI and NotI/XbaI-linearized pCasper4/attB, respectively. *spnE* 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 [57]. 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 *spnE* and *armi* genotypes were prepared as described in [58]. 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 [59]. 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 *spnE* 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 SpnE) 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 [60]. Sequence alignments were made by Bowtie [61]. Contaminating ribosomal RNAs were identified and removed by mapping sequencing reads to annotated ribosomal RNAs from flybase [62]. To identify TE derived piRNAs, sequencing reads ranging from 23-30 nucleotides (nt) were aligned to Repbase [63], allowing for up to 2 mismatches. The number of reads mapped to each TE family 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 [62]. TE-derived piRNA counts for each TE family 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) [55,62] by TopHat [64], requiring unique mapping. The number of reads from each protein coding gene were counted using HTseq-count [65]. Differential expression was estimated concurrently for TEs and protein-coding genes by DESeq for *aub* [66] and DESeq2 for *spnE* and *armi* [67]. TEs or protein-coding genes were considered differentially expressed if they exhibited an adjusted *p*-value < 0.05 and a fold-change > 2 when comparing transgenic rescues and mutants, or with an adjusted *p*-value < 0.05 and a fold-change > 1.5 when comparing the two transgenic rescues.

Ping-pong fraction

Ping-pong fraction was calculated as described in [68]. In brief, small RNA sequencing reads ranging from 23-30 nt were aligned to TE consensus sequences from Repbase [63], 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 [63], and redundant TE families in Repbase were identified as described above. To calculate the d1 proportion [29], 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 [29], the nucleotide after the 3' end of each piRNA was determined based on alignment to the Repbase [63]. 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.

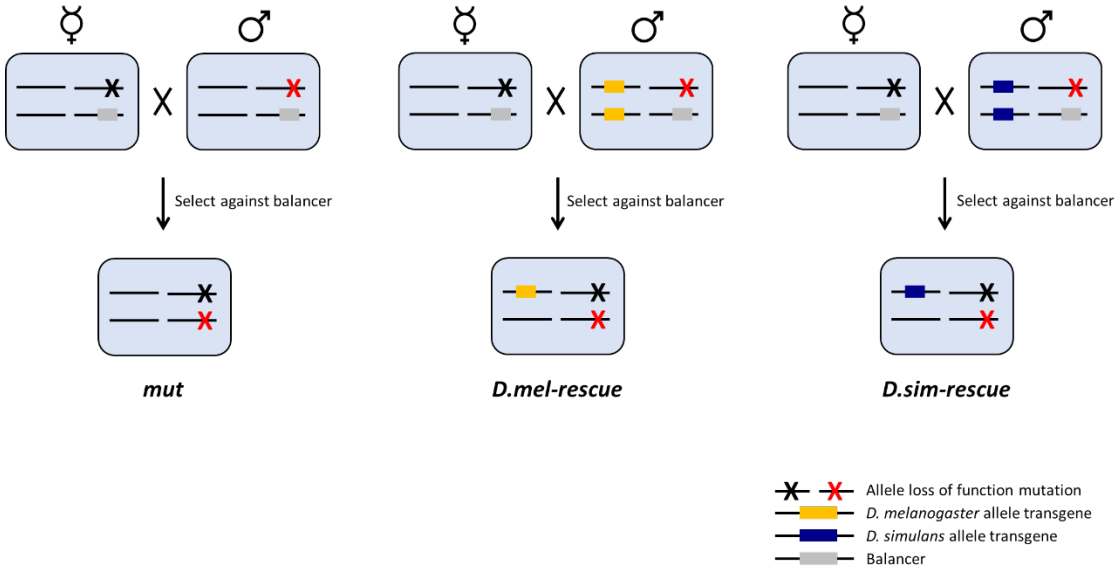
Acknowledgments

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.

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.

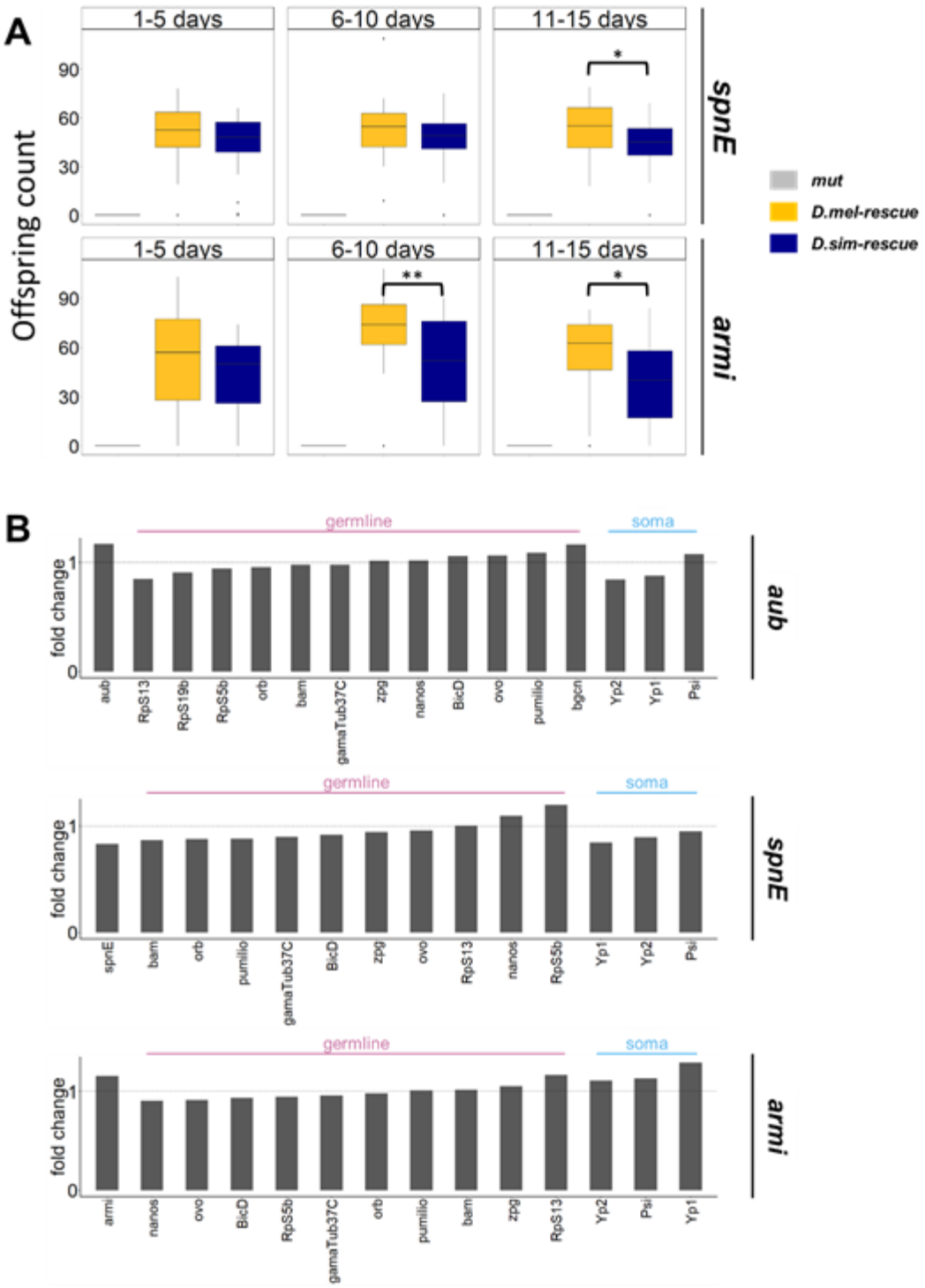
1 **Supporting information**



2

3 **Figure S1.** *Drosophila melanogaster* genotypes and cross scheme.

4



1

2 **Figure S2.** (A). Functional divergence in two piRNA pathway proteins, *spnE* and *armi*, between

3 species affects female fertility. PhiC31-mediated site-specific integrations of *D. simulans* and *D.*

4 *melanogaster* are compared for their ability to complement trans-heterozygous mutants for

female fertility across three different age classes. Females with the *D. simulans spnE* 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 $p \leq 0.01$. *** denotes $p \leq 0.001$. (B). Expression level of *aub*, *spnE* and *armi*, and the ovarian germ cell, somatic cell proportion were similar between *D. melanogaster* transgenic rescue and *D. simulans* transgenic rescue. Fold change of expression level of *aub*, *spnE*, *armi*, germline-specific genes and soma-specific genes between two transgenic rescues were shown.

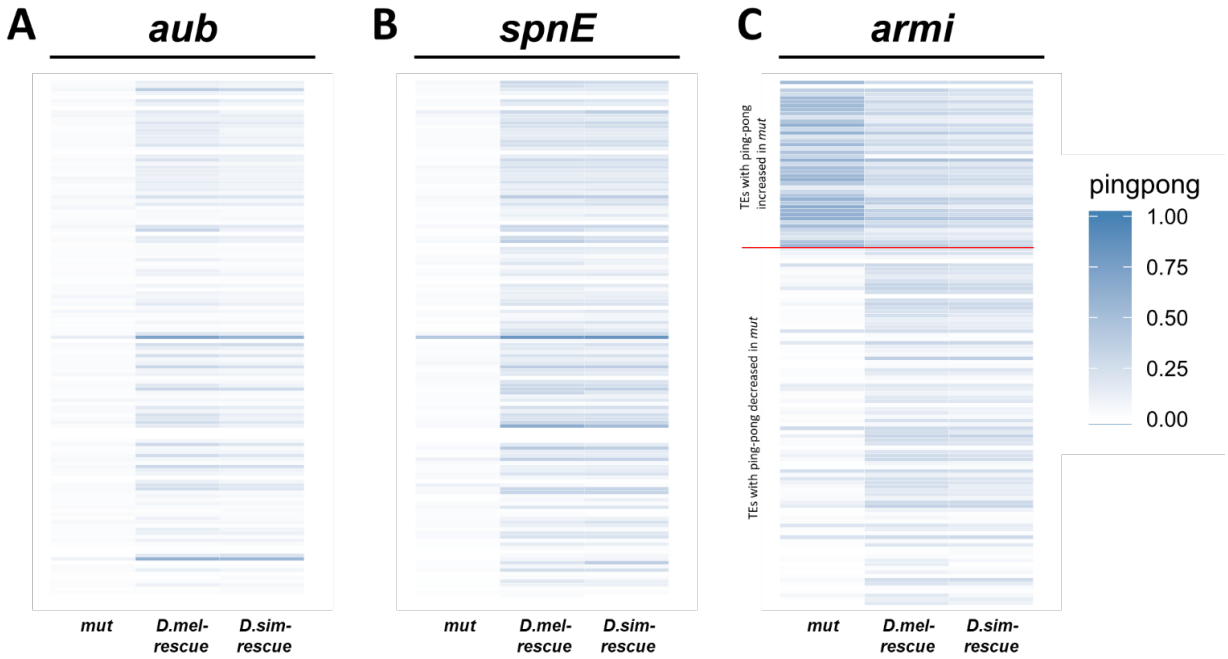


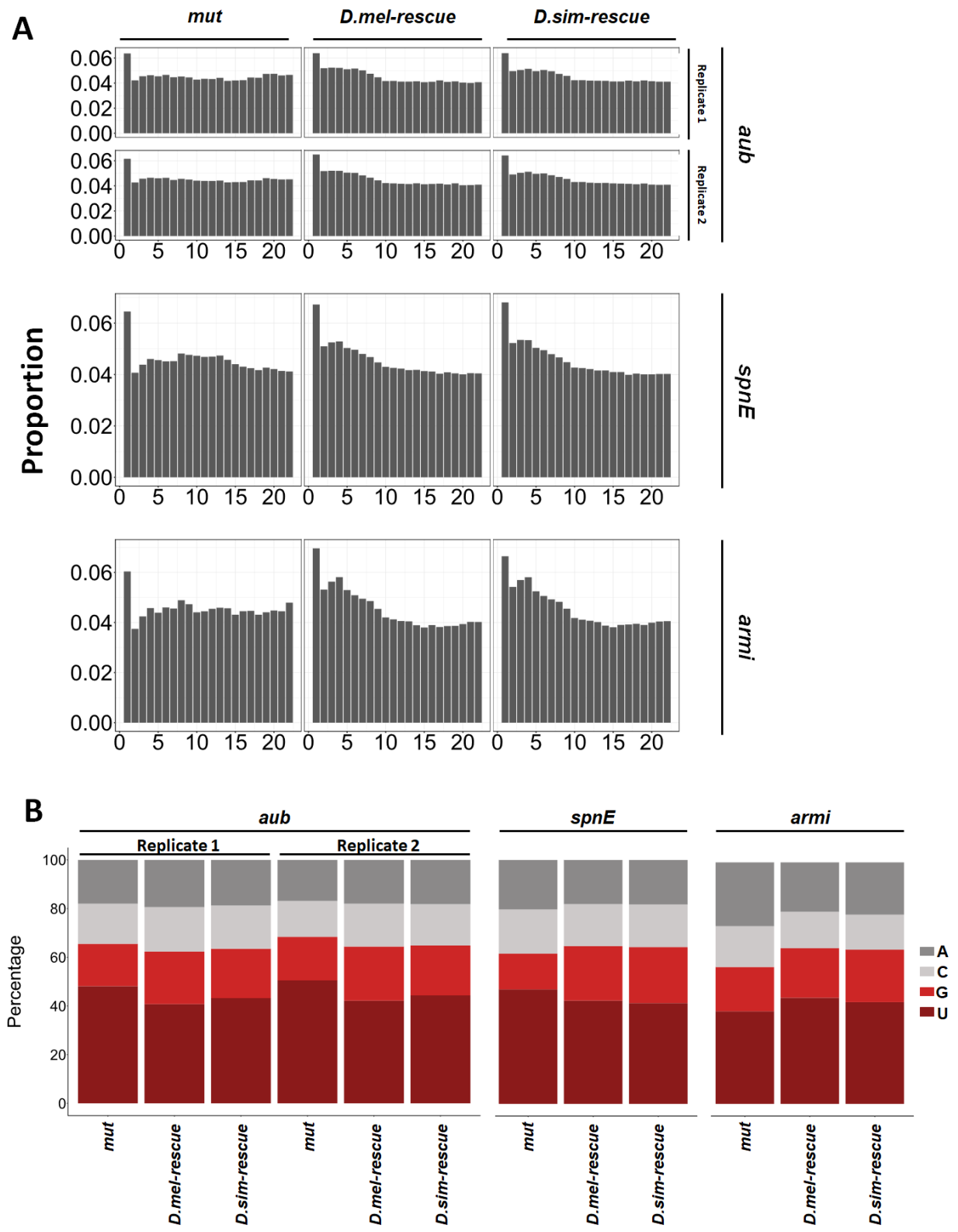
Figure S3. Ping-pong fraction heat map for *aub* (A), *spnE* (B) and *armi* (C). Among (C), 92 and 43

1 TE families whose ping-pong fractions are decreased (below red line) or increased (above red

2 line), respectively, in *armi* mutant as compared to those in *D. melanogaster* transgenic rescue

3 are shown.

4

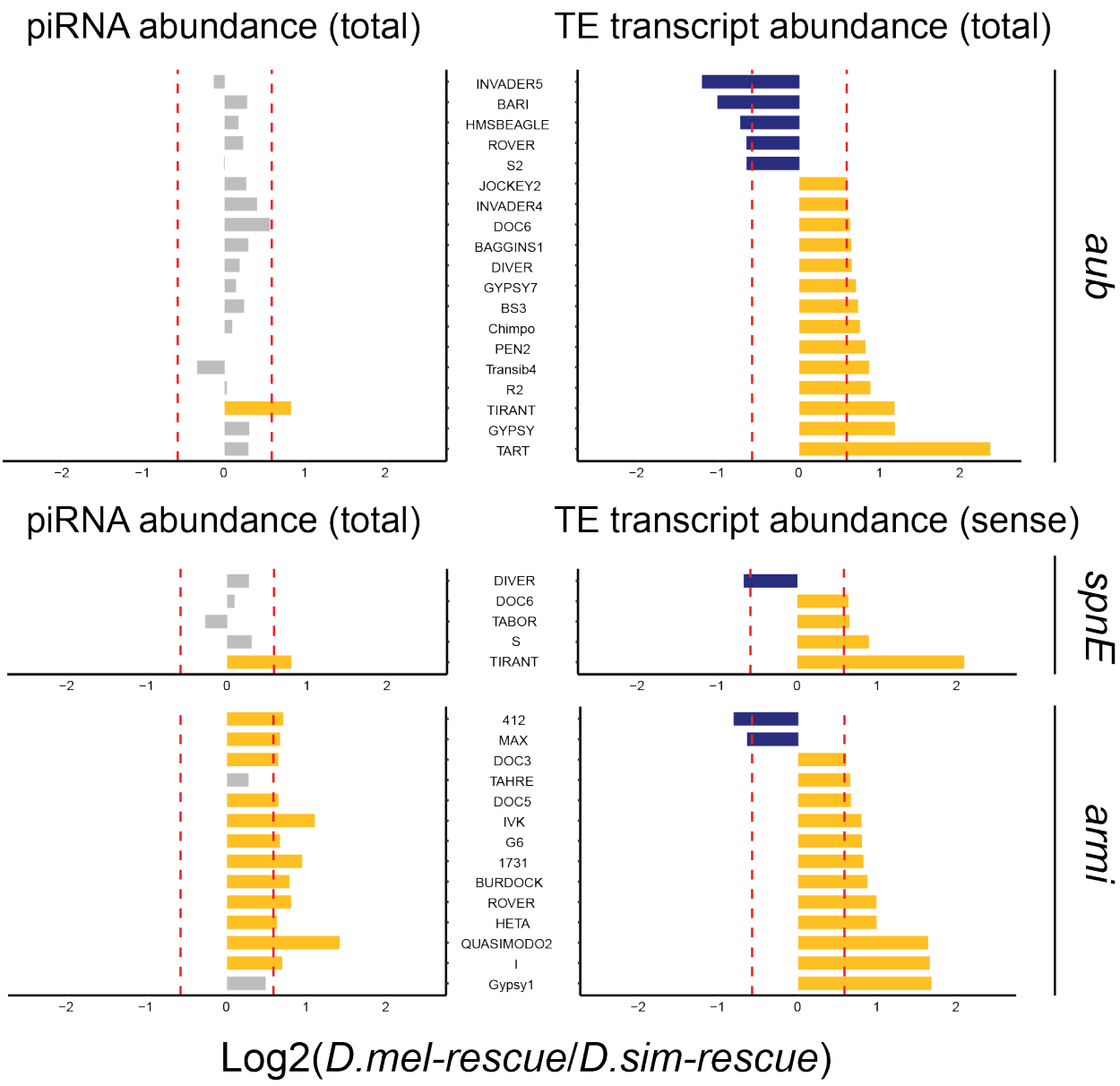


1

2 **Figure S4.** Observed peaks of 1nt distance (A) and +1 U bias (B) among each genotype for each

1 protein studied.

2



3

4 **Figure S5. Decoupling between changes in TE transcript abundance and changes in TE-derived**

5 **piRNA abundance.** This figure is showing log2-fold-changes of TE transcript abundance and TE-

6 derived piRNA abundance between two transgenic rescues for the TE families whose TE

7 transcript abundance is dramatically different (> 1.5 fold, adjusted *p*-value < 0.05) between two

rescues from Figure 3. Red dashed lines indicate the 1.5 fold-change threshold.

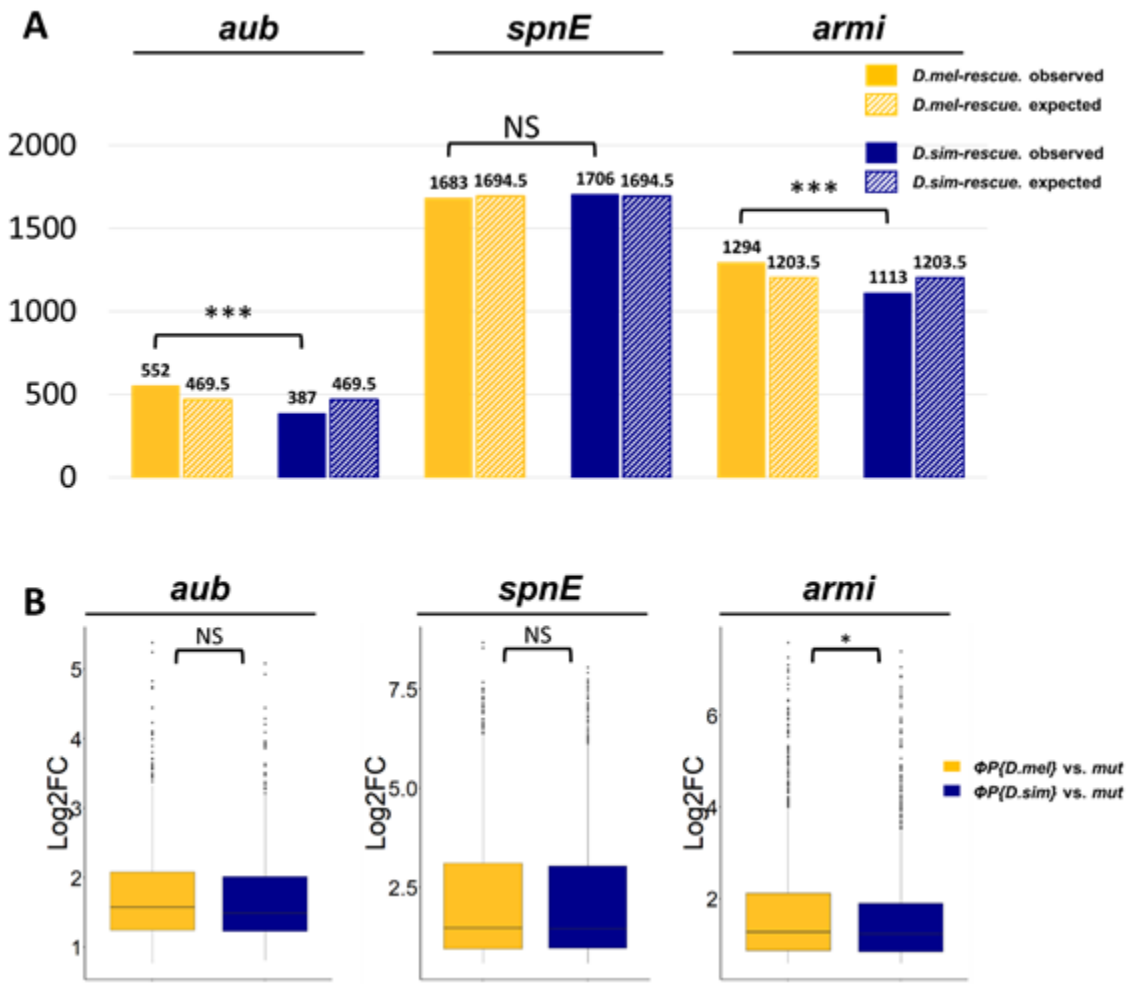


Figure S6. (A). The number of genes whose corresponding piRNA abundance is increased (>1.5 fold) in the presence of each transgene as compared to the mutant. Log2 fold-change values were based on two biological replicates for *aub* and three biological replicates for *spnE* and *armi*, and were obtained from a DESeq2 analysis (adjusted $p < 0.05$). Statistical significance was assessed by the Pearson's Chi-squared test. (B). Log2 fold-changes of abundance of piRNAs mapped to protein-coding genes between rescues and mutant from Figure S6A are compared between two transgenic rescues. Comparisons were made by a Wilcoxon rank sum test. NS

denotes $p > 0.05$. * denotes $p \leq 0.05$. *** denotes $p \leq 0.001$.

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

23. 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.

24. 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.

25. 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.

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. 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.

28. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. American Association for the Advancement of Science; 2015;348: 812–817.

29. 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.

30. 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. Elsevier Inc.; 2015;59: 819–830.

31. 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.

32. Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, et al. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 2011;30: 3977–3993.

33. 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.

34. 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.

35. 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.

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

37. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, et al. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell. 2012;151: 871–884.

38. 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.

39. 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

40. 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.

41. 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.

42. Dufourt J, Dennis C, Boivin A, Gueguen N, Théron E, Goriaux C, et al. Spatio-temporal requirements for transposable element piRNA-mediated silencing during Drosophila oogenesis. Nucleic Acids Res. 2014;42: 2512–2524.

43. Post C, Clark JP, Sytnikova YA, Chirn G-W, Lau NC. The capacity of target silencing byDrosophilaPIWI and piRNAs [Internet]. RNA. 2014. pp. 1977–1986.

doi:10.1261/rna.046300.114

44. Akulenko N, Ryazansky S, Morgunova V, Komarov PA, Olovnikov I, Vaury C, et al. Transcriptional and chromatin changes accompanying de novo formation of transgenic piRNA clusters. RNA. 2018;24: 574–584.

45. Kelleher ES, Barbash DA. Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. Mol Biol Evol. 2013;30. doi:10.1093/molbev/mst081

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

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

48. Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. The recent invasion of natural Drosophila simulans populations by the P-element. Proceedings of the National Academy of Sciences. 2015;112: 201500758.

49. Hill T, Schlötterer C, Betancourt AJ. Hybrid Dysgenesis in Drosophila simulans Associated with a Rapid Invasion of the P-Element. PLoS Genet. 2016;12: e1005920.

50. Huang H, Li Y, Szulwach KE, Zhang G, Jin P, Chen D. AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. J Cell Biol. 2014;206: 217–230.

51. 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 [Internet]. bioRxiv. 2018. p. 445825. doi:10.1101/445825

52. 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

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

54. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157: 1353–1363.

55. 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.

56. 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.

57. 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.

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

59. 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.

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

61. 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.

62. 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. Oxford University Press; 2017;45: D663–D671.

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

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

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

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

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

68. 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.