Divergence of piRNA pathway proteins affects piRNA biogenesis but not TE transcript level

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

[1] Dept. Biology & Biochemistry, University of Houston, Houston, TX, USA

[2] Dept. Molecular Biology & Genetics, Cornell University, Ithaca, NY, USA

SUMMARY

In metazoan germlines, the piRNA pathway acts as a genomic immune system [1,2]: employing small-RNA mediated silencing to defend host DNA from the harmful effects of transposable elements (TEs). In response to dynamic changes in genomic TE content, host genomes are proposed to alter the piRNAs that they produce in order to silence the most active TE families [3–5]. However, piRNA pathway proteins, which execute piRNA biogenesis and enforce silencing of targeted sequences, also evolve rapidly and adaptively in animals [6,7]. 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 find very few differences in TE transcript levels, but global effects on piRNA biogenesis, particularly for Armitage. Therefore, despite the fitness costs of transposition, piRNA effector proteins are not under selection to enhance TE silencing. Rather, our results suggest TE antagonism of host piRNA production. Furthermore, the absence of correlated downstream effects on TE transcripts suggest a fundamentally different relationship between piRNA abundance and TE transcript regulation between alleles. Finally, we reveal that *D. simulans* alleles exhibit enhanced off-target effects on host transcripts in a *D. melanogaster* background, suggesting the avoidance of genomic auto-immunity [8] as an additional target of selection. Taken together, our results point to multiple diverging functions, unveiling unexpected complexity in the molecular adaptation of piRNA pathway proteins.

RESULTS AND DISCUSSION

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 each allele to complement a *D. melanogaster* mutant background. For each piRNA protein, Aubergine (Aub), Spindle-E (SpnE) and Armitage (Armi), we compared

three genotypes, 1) trans-heterozygous loss-of-function mutants, 2) mutants with a *D. melanogaster* genomic transgene rescue, and 3) mutants with *D. simulans* genomic transgene rescue. 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.

Homozygosity or trans-heterozygosity for loss of function alleles in all three genes causes complete female sterility (Figure S1A) [9]. For all three proteins, fertility was only partially rescued by *D. simulans* transgene when compared to its *D. melanogaster* counterpart, indicating divergence in gene function (Figure S1A) [10]. Importantly, *D. simulans* transgenes do not exhibit significantly reduced expression when compared to *D. melanogaster* transgenes (Figure S1B), indicating that fertility effects reflect amino acid sequence divergence.

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 their functions in piRNA production. Biochemical functions of piRNA pathway proteins, including piRNA biogenesis, would be expected to evolve adaptively if TEs encode antagonists of piRNA-mediated silencing, or if TE-derived transcripts evolve to evade piRNA processing and associated silencing [8,11]. Aub and SpnE are involved in ping-pong biogenesis, which produces sense and antisense piRNAs from precursor transcripts through a homology-dependent amplification loop [3,12,13]. By contrast, *armi* alleles disrupt both ping-pong amplification and phased piRNA-biogenesis, which gives rise to piRNAs from a single strand through sequential cleavage by the nuclease Zucchini [14–17].

To determine the overall impact of adaptive protein evolution on piRNA production, we compared TE-derived piRNA abundance among mutants and transgenic rescues using small-RNA seq. For all three piRNA pathway mutants, transgenic rescue by both *D. melanogaster* and *D. simulans* alleles is associated with a dramatic increase in piRNAs for the majority of germline TE families, indicating an overall conservation of piRNA production (Figure 1A-C). Nevertheless, we uncovered allelic differences in the abundance of piRNAs from specific TE families, indicative of functional divergence between species. 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 did not reflect a change in the germline-to-soma ratio (Figure S1B).

armi exhibited the most dramatic differences in piRNA abundance between alleles, with 81

of 84 TE families exhibiting reduced piRNA abundance in the *D. simulans* rescue when compared to *D. melanogaster* (Figure 1C and 1F). *D. simulans aub* was characterized by similar, but less dramatic reductions in piRNA biogenesis: of 10 TE families whose piRNA abundance differed between *D. melanogaster* and *D. simulans* alleles, 9 exhibited lower abundance in the *D. simulans* rescue (Figure 1A and 1D). By contrast, for *spnE*, less than half (11 of 25) of differentially abundant TE families exhibit reduced abundance in the *D. simulans* rescue when compared to *D. melanogaster* (Figure 1B and 1E), suggesting functional divergence does not result in systematically reduced functionality of *D. simulans* alleles in a *D. melanogaster* background.

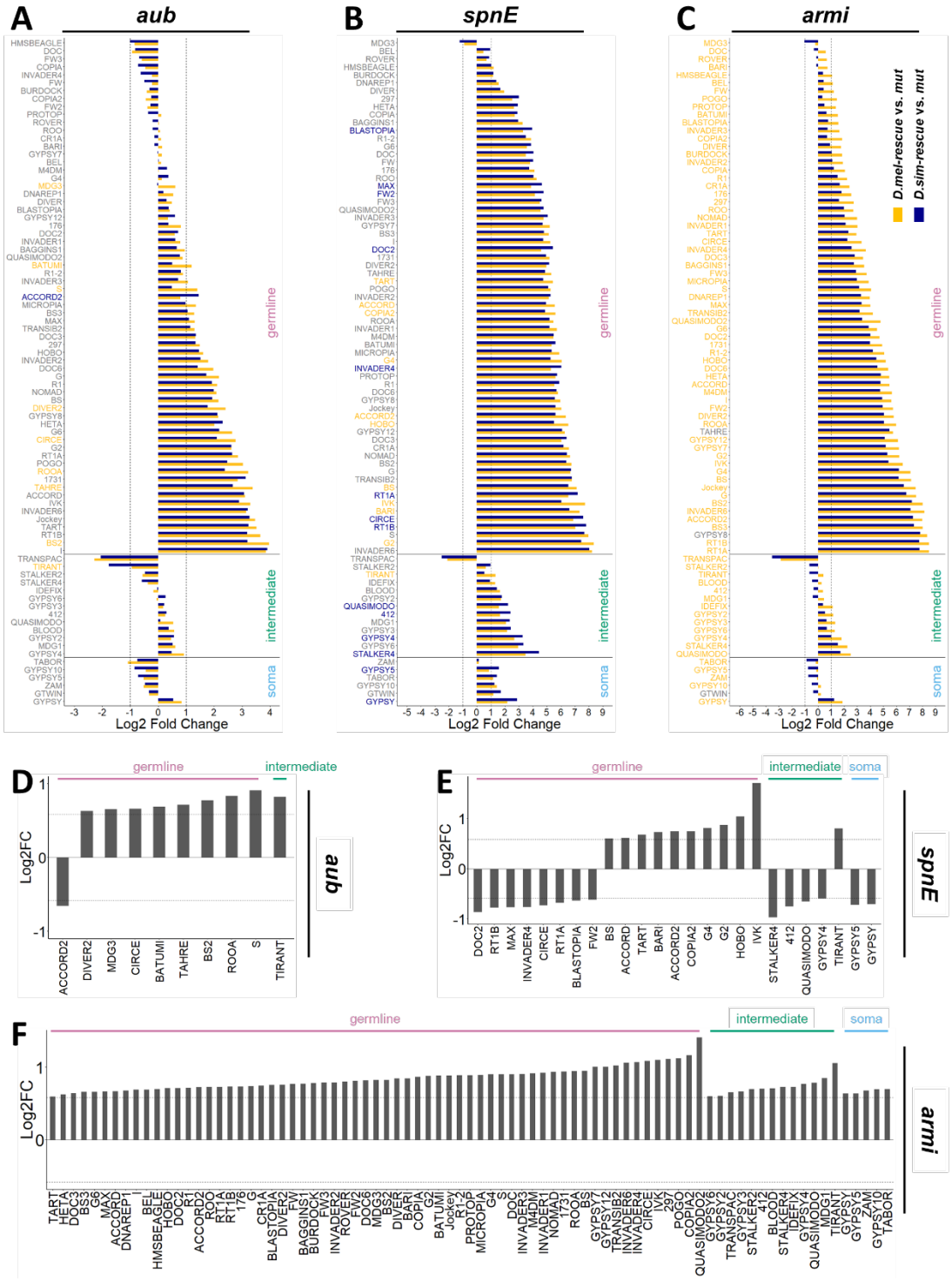


Figure 1. *D. simulans armi* is characterized by a dramatic loss of TE-derived piRNAs. (A-C) Log2 fold-change of TE-derived piRNA abundance in transgenic rescues as compared to trans-heterozygous mutants for *aub*, *spnE* and *armi*. Dashed lines indicate the 2 fold-change threshold. TE families whose TE-derived piRNA 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. (D-F) Log2 fold-change for TE families whose TE-derived piRNAs are significantly differentially abundant between transgenic rescues. Dashed lines indicate the 1.5 fold-change threshold. TE families were grouped into germline-specific, soma-specific and intermediate [3]. Log2 fold-change values are based on two biological replicates for *aub* and three biological replicates for *spnE* and *armi*. piRNA abundance for each TE family was normalized to the total number of sequenced miRNAs in the same library.

<u>*D. simulans* alleles exhibit reduced piRNA biogenesis.</u>

To detect underlying differences in piRNA biogenesis, which could give rise to interallelic differences in piRNA abundance, we determined molecular signatures of ping-pong and phased piRNA biogenesis from our small RNA data. The frequency of ping-pong amplification was estimated by the fraction of piRNAs occuring on opposite strands of the TE consensus and whose sequences overlap by 10 bp, a reflection of the cleavage-site preference of the Piwi-Argonautes Aub and Argonaute-3 (Ago-3, Figure 2A) [3,12,18]. Similarly, phased biogenesis is detected from the fraction of piRNA 3' ends 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 2F) [16,17]. In general, ping-pong and phasing are inversely correlated in mutant piRNA pools, because reducing one mechanism of biogenesis leads to a proportional increase in the other [16,17].

Aub plays a direct role in ping-pong amplification by cleaving piRNA precursors (Figure 2A) [3,12,18], and *spnE* is required for the localization of Aub into the perinuclear nuage, where ping-pong occurs [19]. Mutants for both genes exhibit a complete collapse of ping-pong amplification, as indicated by very low ping-pong fractions for all TE families (Figure 2B) [3,13]. Similar to our observations with TE-derived piRNA abundance (Figure 1A-C), both *D. melanogaster* and *D. simulans aub* and *spnE* alleles confer a dramatic rescue of ping-pong biogenesis. Ping-pong fractions in the *D. simulans aub* rescue, however, are significantly lower than *D. melanogaster* (Figure 2B), revealing that reduced piRNA abundance associated with *D. simulans aub* (Figure 1D) is accompanied by reduced *aub*-dependent piRNA production. As a consequence of reduced ping-pong biogenesis, there is a corresponding increase in phased piRNA biogenesis in the *D. simulans* rescue (Figure 2G and 2H, Figure S2). By contrast, *D. melanogaster* and *D. simulans spnE* alleles do not substantially differ with respect to ping-pong (Figure 2B), which is consistent with the absence of a systematic loss of piRNAs in the presence of the *D. simulans* allele (Figure 1E). However, these is a modest but significant increase in the d1 proportion with the *D. simulans spnE* rescue (Figure 2G), suggesting a modest increase in phased biogenesis.

Armi is believed to bind to piRNA precursors [20,21], and is required for phased piRNA biogenesis [14,15,22,23]. Consistent with a role in phasing, +1-U and d1 are reduced in *armi* mutants (Figure 2G and 2H, Figure S2). *armi* is also required for ping-pong biogenesis for some TE families, suggesting that phased piRNA production produces substrates for ping-pong [3,24]. Ping-pong fractions are therefore decreased in *armi* transgenic rescues for some TE families, and increased in others (Figure 2C and 2D). Consistent with the global reduction of TE-derived piRNAs (Figure 1C), the *D. simulans armi* rescue is characterized by modest reduction in phased biogenesis (Figure 2H, Figure S2). However, in contrast to loss-of-function mutants, ping-pong fractions are also reduced for most TE families in the *D. simulans* rescue when compared to *D. melanogaster* regardless of whether *armi* is required for ping-pong biogenesis (Figure 2C-E). Therefore, *D. simulans armi* enacts a global inhibitory effect on ping-pong. While we did not investigate the molecular basis for this inhibition, it could be mediated through Armi's physical interaction with Ago-3 and Aub [25]. If *D. simulans* Armi exhibits enhanced affinity for its binding partners, it could antagonize them from their functions in ping-pong biogenesis.

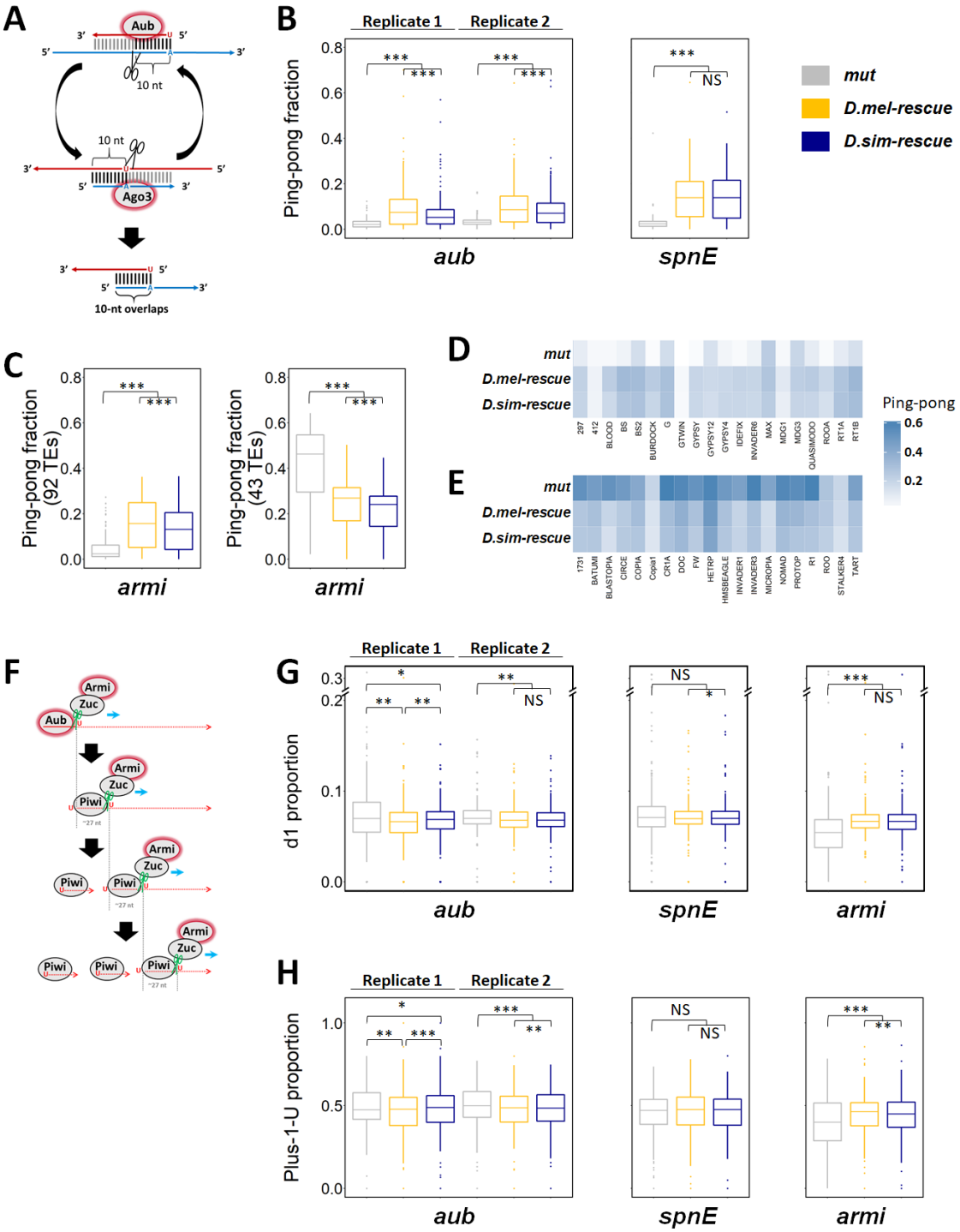


Figure 2. *D. simulans* alleles reduce ping-pong biogenesis and phased biogenesis. (A) Ping-pong amplification loop. (B) Ping-pong fractions of TE-derived piRNAs are compared between trans-heterozygous mutants and transgenic rescues for *aub* and *spnE*. (C) Ping-pong fractions of TE-derived piRNAs are 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* mutant as compared to those in *D. melanogaster* transgenic rescue are shown . (D) Ping-pong fraction heat map for 20 most piRNA-abundant TE families from panel C left. (E) Ping-pong fraction heat map for 20 most piRNA-abundant TE families from panel C right. (F) Zuc-dependent phased piRNA biogenesis. (G) Proportions of 1 nt distance between adjacent piRNAs (d1) mapped to the TE consensus sequences are compared between each genotype of each gene. Note the break in the Y-axes. (H) 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 in *D. simulans* are not associated with TE derepression.

Ultimately, the piRNA pathway represses TEs by repressing their transcripts (reviewed in [1]). Enhanced negative regulation of TE transcripts therefore presents an obvious target of positive selection acting on piRNA pathway proteins. Furthermore, given the reduced abundance of TE-derived piRNAs in the *D. simulans aub* and particularly *armi* rescues, deregulation is expected among their downstream targets. 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 proteins, both *D. melanogaster* and *D. simulans* transgenic rescues reduce transcript abundance from almost all TE families when compared to the corresponding mutant (Figure 3A-C). However, it is very striking that despite having significant differences in TE-derived piRNAs, most TE families are not differentially expressed between *D. melanogaster* and *D. simulans* transgenic rescues. Across all three proteins studied, we identified no more than 16 TE families that are differentially expressed (Figure 3D-F) [10]. Furthermore, the *D. simulans* transgene is associated with reduced rather than increased TE expression for 13 of these TE families (Figure 3D-F). This observation is particularly unexpected for *armi*, where defects of *D. simulans* alleles in ping-pong and phased biogenesis lead to a dramatic loss of TE-derived piRNAs (Figure 1C, 2B, 2D and 2E). *D. simulans* alleles therefore appear to enact more efficient piRNA-mediated silencing, conferring equivalent regulation of TE transcripts despite reduced piRNA abundance.

A fundamentally different relationship between piRNA abundance and TE transcript regulation between *D. melanogaster* and *D. simulans* alleles is further supported by comparing changes in piRNAs and TE transcripts. Because piRNAs are both produced from and regulate

TE transcripts, these two pools of RNAs covary in meaningful ways. In *aub* and *spnE* mutants, where the cleavage of TE transcripts by the ping-pong cycle is lost (Figure 2A and 2B), reduced antisense piRNAs are associated with increased TE transcript abundance (Figure S3). By contrast, in *armi* mutants, increased antisense TE transcription correlates with minimal loss of piRNAs, most likely because the presence of transcripts from both strands drives forward ping-pong biogenesis (Figure 2A, Figure S3). Strikingly, however, despite the very different effects of *armi* versus *aub* and *spnE* on piRNA biogenesis, in all three cases correlated changes between piRNA pools and mRNA pools are not observed in comparisons between the two transgenic rescues (Figure S3). This indicates again that the *D. simulans* alleles have altered, not simply reduced, function. Furthermore, these functional changes in the *D. simulans* alleles on the piRNA pool are complex and do not translate to altered TE transcription in a predictable way, and *vice versa*.

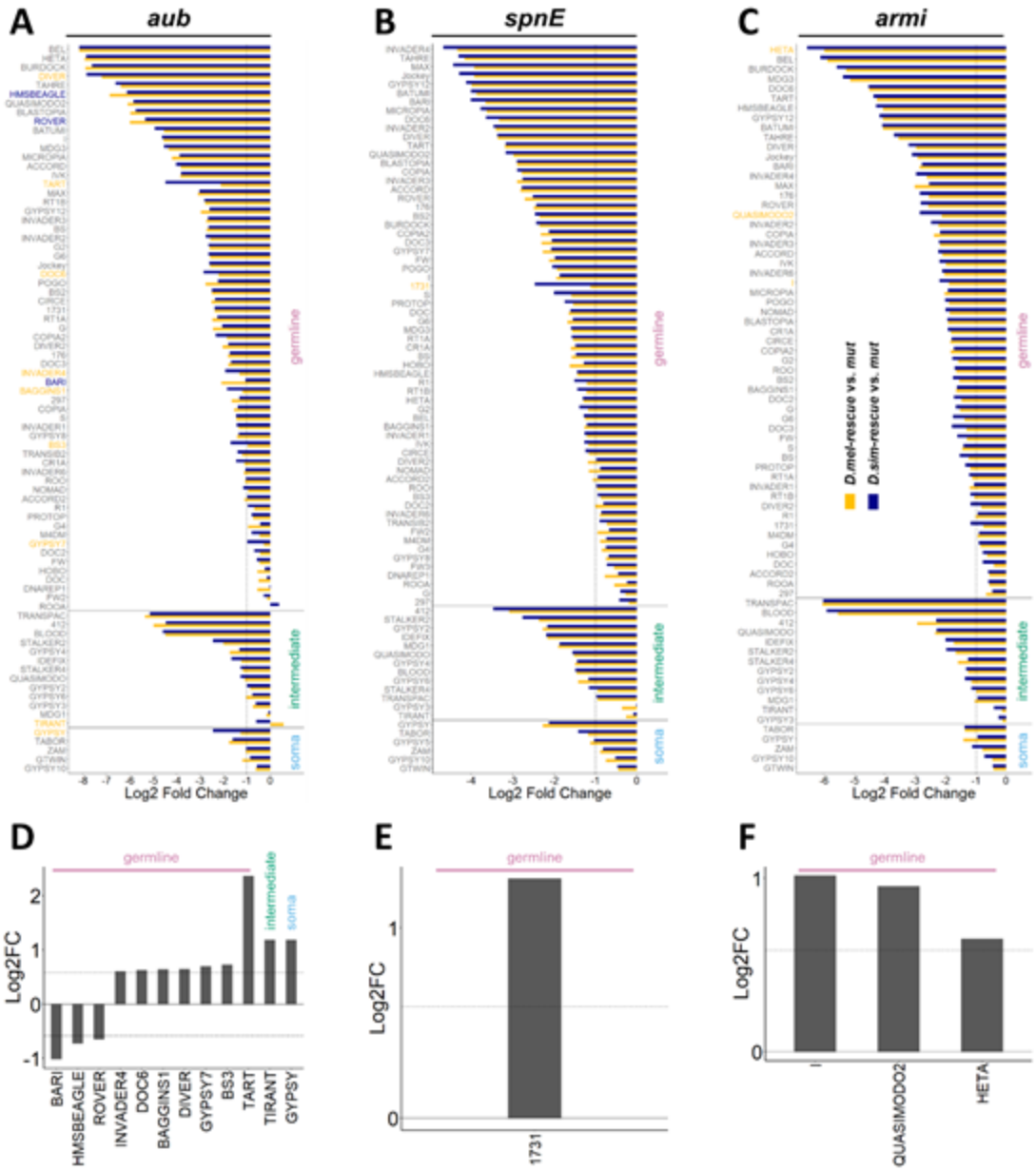


Figure 3. Negative transcriptional regulation of TEs is mostly conserved between species. (A-C) Log2 fold-change of TE transcript abundance in transgenic rescue as compared to trans-heterozygous mutants for *aub*, *spnE* and *armi*. Dashed lines indicate the log2 transformed fold-change threshold of 2. TE families whose TE transcript abundance significantly 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. (D-F) Log2 fold-change for TE families that are differentially abundant between transgenic rescues. Dashed lines indicate the 1.5 fold-change threshold. TE families were grouped into germline-specific, soma-specific and intermediate [3]. 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 combined DESeq2 analysis that included both TEs and protein-coding genes.

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

piRNA-mediated silencing can potentially be deleterious if off-target effects silence host protein-coding genes. Avoidance of this genomic auto-immunity syndrome is an alternative, non-TE centered hypothesis to explain the adaptive evolution of piRNA pathway proteins [8,26]. If piRNA proteins experience selection to minimize auto-immunity, a greater number of off-target effects are predicted to occur with the *D. simulans* rescue, which is not adapted to the *D. melanogaster* background. We therefore examined the number of protein-coding genes that are negatively regulated by piRNA pathway proteins, by comparing their expression levels in mutants and transgenic rescues. Protein-coding genes whose expression is significantly reduced in transgenic rescues (>1.5 fold) indicate candidate off-target effects of piRNA-mediated silencing. We observed that for *aub* and *armi*, a significantly greater number of protein-coding genes have reduced expression in the presence of the *D. simulans* transgene than the *D. melanogaster* transgene (Figure 4). *Drosophila simulans spnE* alleles are also associated with a greater number of negatively regulated protein-coding genes, but the difference is not statistically significant (p = 0.11). These observations suggest that *D. simulans* alleles are poorly adapted to avoid off-target effects in a *D. melanogaster* background, consistent with the auto-immunity hypothesis.

Increased off-target effects of *D. simulans* alleles on protein-coding genes could be explained by increased production of protein-coding derived piRNAs, or enhanced silencing of target mRNAs. To differentiate between these alternatives, we compared protein-coding derived piRNAs between transgenic rescues. *D. simulans* transgenic rescues of *aub* and *armi* exhibit fewer protein-coding genes with increased piRNA abundance (>1.5 fold) when compared to mutants than *D. melanogaster* transgenic rescues (Figure S4A). Furthermore, the magnitude of increased protein-coding piRNA production is not higher in *D. simulans* than *D. melanogaster* transgenic rescues (Figure S4B). Therefore, similar to our observations with TEs, enhanced negative regulation by *D. simulans* alleles does not reflect increased production of piRNAs.

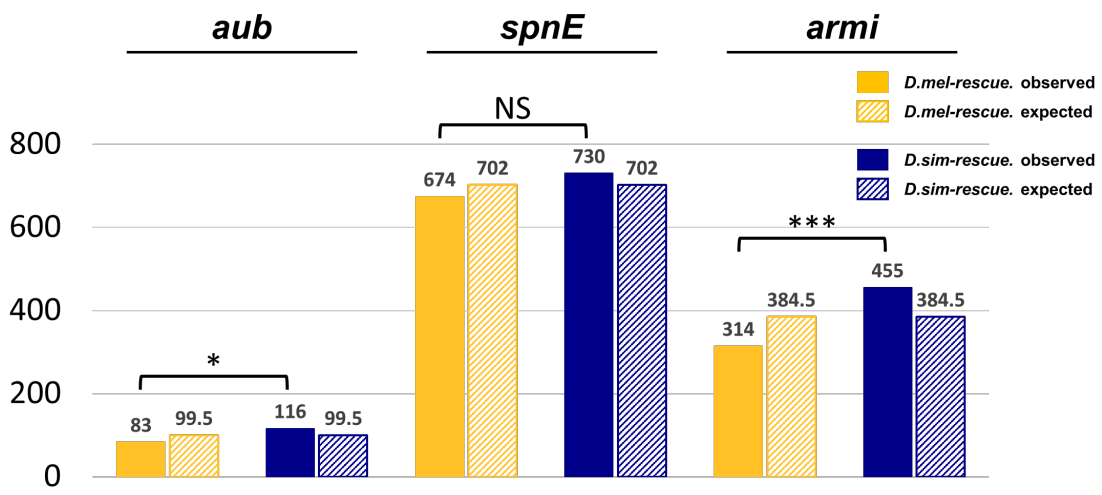


Figure 4. Negative regulation of protein-coding genes suggests increased genomic auto-immunity of *D. simulans* alleles. The number of genes whose expression levels are decreased (>1.5 fold) in the presence of each transgene as compared to the corresponding mutant. Log2 fold-change values were based 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 $p \leq 0.001$.

Conclusion

Despite pervasive adaptive evolution and gene duplication among piRNA pathway proteins in both insect and vertebrate lineages [6,7,27–30], the underlying forces that drive these evolutionary dynamics remain unclear. Transposable elements are potentially implicated in positive selection on piRNA pathway proteins in two ways [8,30]. First, changes in genomic TE activity or composition could select for enhanced TE silencing. Second, TE-encoded antagonists of silencing could reduce piRNA pathway protein function, selecting for novel protein variants that escape antagonism. Escape from TE-encoded antagonist proteins was recently proposed as a driver of adaptive evolution in Rhino, a heterochromatic protein that defines piRNA producing sites in *Drosophila* genomes [11,31–33]. While the TE-encoded antagonist remains to be elucidated, adaptive evolution in *rhino* has resulted in an incompatibility between *D. simulans* Rhino and *D. melanogaster* allele of its interacting partner Deadlock [11,34]. We propose that a similar phenomenon may impact the functions of *D. simulans* Aub and particularly Armi in a *D. melanogaster* background, impacting the physical interactions that mediate piRNA biogenesis. In addition to divergence with respect to piRNA biogenesis, we provide the first evidence of interspecific divergence in genomic auto-immunity [8,26], with *D. simulans* alleles causing enhanced off-target effects compared to their *D.*

melanogaster counterparts.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Generation of Transgenic Lines
  - Complementation assay of female fertility
  - Small RNA-Seq
  - RNA-Seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic analysis of small RNA-Seq libraries
  - Bioinformatic analysis of RNA-Seq libraries
  - Ping-pong analysis
  - Phasing analysis

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
iProof high-fidelity taq DNA polymerase	Bio-Rad	#1725301
Trizol reagent	Invitrogen	#15596026
poly-T Dynabeads	Invitrogen	#61002
fragmentation buffer	Ambion	#AM8740
Superscript II	Invitrogen	#18064014

DNA polymerase I	Promega	#M2051
Klenow enzyme	NEB	#M0212S
Critical Commercial Assays		
NEBNext Multiplex Small RNA Library Prep Set for Illumina	NEBNext	#E7300S
TruSeq® Stranded Total RNA Library Prep	Illumina	#20020596
MinElute gel purification kit	Qiagen	#28604
End-IT DNA repair kit	Epicentre	#ER0720
Experimental Models: Organisms/Strains		
D. melanogaster: *w/yw; aub^{N11} bw^{1}/aub^{HN} bw^{1}*	[10]	N/A
D. melanogaster: *w/yw; aub^{N11} bw^{1}/aub^{HN} bw^{1}; ΦP{D. melanogaster aub}/+*	[10]	N/A
D. melanogaster: *w/yw; aub^{N11} bw^{1}/aub^{HN} bw^{1}; ΦP{D. simulans aub}/+*	[10]	N/A
D. melanogaster: *yw; spnE^{1}/spnE$^{hls-03987}$*	This paper	N/A
D. melanogaster: *yw; spnE^{1}/spnE$^{hls-03987}$; ΦP{D. melanogaster spnE}/+*	This paper	N/A
D. melanogaster: *yw; spnE^{1}/spnE$^{hls-03987}$; ΦP{D. simulans spnE}/+*	This paper	N/A

D. melanogaster: *w/yw; armi^{1}/armi$^{72.1}$*	This paper	N/A
D. melanogaster: *w/yw; armi^{1}/armi$^{72.1}$; ΦP{D. melanogaster armi}/+*	This paper	N/A
D. melanogaster: *w/yw; armi^{1}/armi$^{72.1}$; ΦP{D. simulans armi}/+*	This paper	N/A
Oligonucleotides		
spnE-forward primer: ATTGAACGCCGTCTATGCCAAGC	This paper	N/A
spnE-reverse primer- *D.mel/ D.sim*: ACTGTTCGCCATTGCCACAGATTG	This paper	N/A
armi-forward primer: CACCGCTGAAAGATACGCACACG	This paper	N/A
armi-reverse primer-*D.mel*: GCTAGCCTGCGCTTGGGAGTGTTACCATTCG	This paper	N/A
armi-reverse primer-*D.sim*: GCTAGCCTGACCTCGGGAGTGTTACCACTTC	This paper	N/A
Recombinant DNA		
pCR-Blunt-II-Topo	Invitrogen	K280002
pCasper4/attB	[35]	

Software and Algorithms		
Cutadapt	[36]	https://cutadapt.readthedocs.io/en/stable/index.html
Bowtie	[37]	http://bowtie-bio.sourceforge.net/index.shtml
Bowtie2	[38]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
DESeq	[39]	https://bioconductor.org/packages/release/bioc/html/DESeq.html
DESeq2	[40]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
TopHat	[41]	http://ccb.jhu.edu/software/tophat/index.shtml
HTseq-count	[42]	https://htseq.readthedocs.io/en/master/count.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Erin S. Kelleher (eskelleher@uh.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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 [10], 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; spnE$^{hls-03987}$/TM6; ΦP{D. simulans spnE}*. *yw; spnE^{1}/TM6* and *yw; spnE$^{hls-03987}$/TM6* were obtained by crossing *spnE^{1}/TM3* and *spnE$^{hls-03987}$/TM3* (gifts from Celeste Berg) to *yw; TM3/TM6*. To generate *yw; spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}* and *yw; spnE$^{hls-03987}$/TM6; ΦP{D. simulans spnE}, w^{1118}; ΦP{D. melanogaster spnE}* and *w^{1118}; ΦP{D. simulans spnE}* were first crossed to *yw: TM3/TM6. +/TM6; ΦP{D. melanogaster spnE}/+* and *+/TM6; ΦP{D. simulans spnE}/+* offspring were then crossed to *yw; spnE$^{hls-03987}$/TM3.* Finally, *yw; spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}/+* and *yw; spnE$^{hls-03987}$/TM6; ΦP{D. simulans spnE}/+* offspring were backcrossed into *yw; spnE$^{hls-03987}$/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 *spnE,* virgin females *yw; spnE^{1}/TM6* were crossed to (1) *yw; spnE$^{hls-03987}$/TM6*, (2) *yw; spnE$^{hls-03987}$/TM6; ΦP{D. melanogaster spnE}* or (3) *yw; spnE$^{hls-03987}$/TM6; ΦP{D. simulans 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.

METHOD DETAILS

Generation of Transgenic Lines

To introduce *D. melanogaster* and *D. simulans* alleles into *D. melanogaster*, we used *Φ*C31 integrase-mediated transgenesis system [43], 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) [44,45] and *armi* (~6Kb, *D. melanogaster* Release 6, 3L:3460305..3466368; *D. simulans* Release 2, 3L:3357002..3363099) [44,45] were PCR-amplified by using corresponding primers (see KEY RESOURCES TABLE) and iProof high-fidelity taq DNA polymerase (Bio-Rad). 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 [46]. 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 [47]. 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 [48]. 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.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic analysis of small RNA-Seq libraries

3' Illumina adaptors were removed from sequencing reads by Cutadapt [36]. Sequence alignments were made by Bowtie [37]. Contaminating ribosomal RNAs were identified and removed by mapping sequencing reads to annotated ribosomal RNAs from flybase [49]. To identify TE derived piRNAs, sequencing reads ranging from 23-30 nucleotides (nt) were aligned to Repbase [50], 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 [49]. 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) [44,49] by TopHat [41], requiring unique mapping. The number of reads from each protein coding gene were counted using HTseq-count [42]. Differential expression was estimated concurrently for TEs and protein-coding genes by DESeq for *aub* [39] and DESeq2 for *spnE* and *armi* [40]. 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 [51]. In brief, small RNA sequencing reads ranging from 23-30 nt were aligned to TE consensus sequences from Repbase [50], 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 Repbase [50], and redundant TE families in Repbase were identified as described above. To calculate the d1 proportion [16], 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 [16], the nucleotide after the 3' end of each piRNA was determined based on alignment to the reference genome (*D. melanogaster* Release 6) [44]. 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.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at …

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.

REFERENCES

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

2. Klattenhoff, C., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development *135*, 3–9.

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

4. Lu, J., and Clark, A.G. (2010). Population dynamics of PIWI-interacting RNAs (piRNAs) and

their targets in Drosophila. Genome Res. *20*, 212–227.

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

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

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

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

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

10. Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). Drosophila Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. PLoS Biol. *10*. Available at: http://dx.doi.org/10.1371/journal.pbio.1001428.

11. Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev. Cell *43*, 60–70.e5.

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

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

14. Haase, A.D., Fenoglio, S., Muerdter, F., Guzzardo, P.M., Czech, B., Pappin, D.J., Chen, C., Gordon, A., and Hannon, G.J. (2010). Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes Dev. *24*, 2499–2504.

15. Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. *29*, 3301–3317.

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

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

18. Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi,

H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587–1590.

19. Andress, A., Bei, Y., Fonslow, B.R., Giri, R., Wu, Y., Yates, J.R., 3rd, and Carthew, R.W. (2016). Spindle-E cycling between nuage and cytoplasm is controlled by Qin and PIWI proteins. J. Cell Biol. *213*, 201–211.

20. Vourekas, A., Zheng, K., Fu, Q., Maragkakis, M., Alexiou, P., Ma, J., Pillai, R.S., Mourelatos, Z., and Wang, P.J. (2015). The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. *29*, 617–629.

21. Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem. Sci. *41*, 324–337.

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

23. Pandey, R.R., Homolka, D., Chen, K.-M., Sachidanandam, R., Fauvarque, M.-O., and Pillai, R.S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet. *13*, e1006956.

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

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

26. Castillo, D.M., Mell, J.C., Box, K.S., and Blumenstiel, J.P. (2011). Molecular evolution under increasing transposable element burden in Drosophila: a speed limit on the evolutionary arms race. BMC Evol. Biol. *11*, 258.

27. Kolaczkowski, B., Hupalo, D.N., and Kern, A.D. (2011). Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol. Biol. Evol. *28*, 1033–1042.

28. Levine, M.T., McCoy, C., Vermaak, D., Lee, Y.C.G., Hiatt, M.A., Matsen, F.A., and Malik, H.S. (2012). Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. PLoS Genet. *8*, e1002729.

29. Levine, M.T., Vander Wende, H.M., Hsieh, E., Baker, E.P., and Malik, H.S. (2016). Recurrent Gene Duplication Diversifies Genome Defense Repertoire in Drosophila. Mol. Biol. Evol. *33*, 1641–1653.

30. Palmer, W.H., Hadfield, J.D., and Obbard, D.J. (2018). RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics *208*, 1585–1599.

31. Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive selection drives the evolution of

rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet. *1*, 96–108.

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

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

34. Yu, B., Lin, Y.A., Parhad, S.S., Jin, Z., Ma, J., Theurkauf, W.E., Zhang, Z.Z., and Huang, Y. (2018). Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. EMBO Rep. *19*. Available at: http://dx.doi.org/10.15252/embr.201745418.

35. Maheshwari, S., and Barbash, D.A. (2012). Cis-by-Trans Regulatory Divergence Causes the Asymmetric Lethal Effects of an Ancestral Hybrid Incompatibility Gene. PLoS Genet. *8*, e1002597.

36. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

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

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

39. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

40. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

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

42. Anders, S., Pyl, P.T., and Huber, W. (2014). HTSeq - A Python framework to work with high-throughput sequencing data. bioRxiv, 002824. Available at: https://www.biorxiv.org/content/early/2014/08/19/002824 [Accessed January 8, 2019].

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

44. Hoskins, R.A., Carlson, J.W., Wan, K.H., Park, S., Mendez, I., Galle, S.E., Booth, B.W., Pfeiffer, B.D., George, R.A., Svirskas, R., *et al.* (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. *25*, 445–458.

45. Hu, T.T., Eisen, M.B., Thornton, K.R., and Andolfatto, P. (2013). A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Res. *23*, 89–98.

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

47. Wickersheim, M.L., and Blumenstiel, J.P. (2013). Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. Biotechniques *55*, 269–272.

48. Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

49. Gramates, L.S., Marygold, S.J., Santos, G. dos, Urbano, J.-M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., *et al.* (2017). FlyBase at 25: looking to the future. Nucleic Acids Res. *45*, D663–D671.

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

51. Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387–1392.

SUPPLEMENTARY MATERIALS

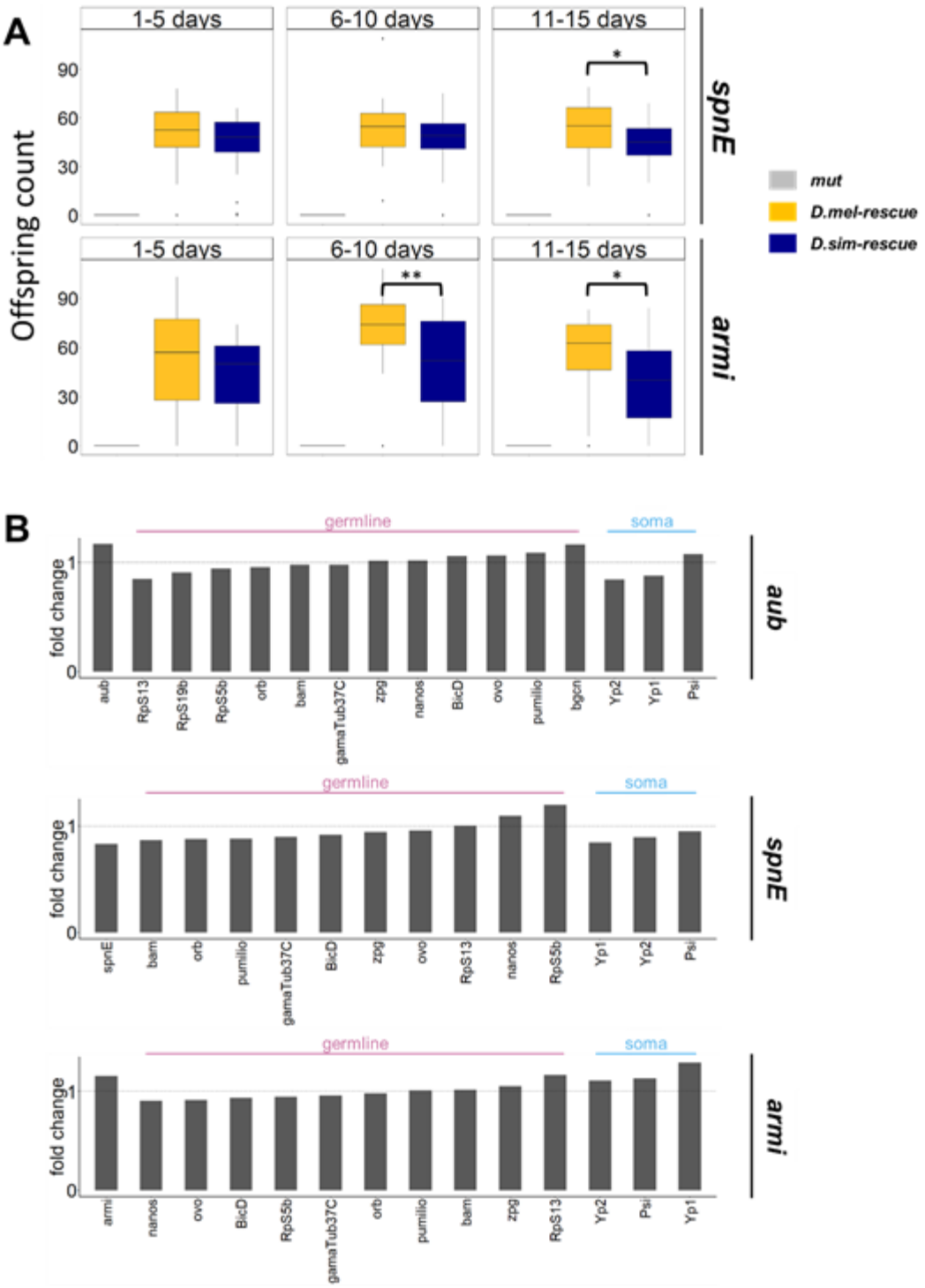


Figure S1. (A). Functional divergence in two piRNA pathway proteins, *spnE* and *armi*, between species affects female fertility. PhiC31-mediated site-specific integrations of *D. simulans* and *D. melanogaster* are compared for their ability to complement trans-heterozygous mutants for female fertility 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.

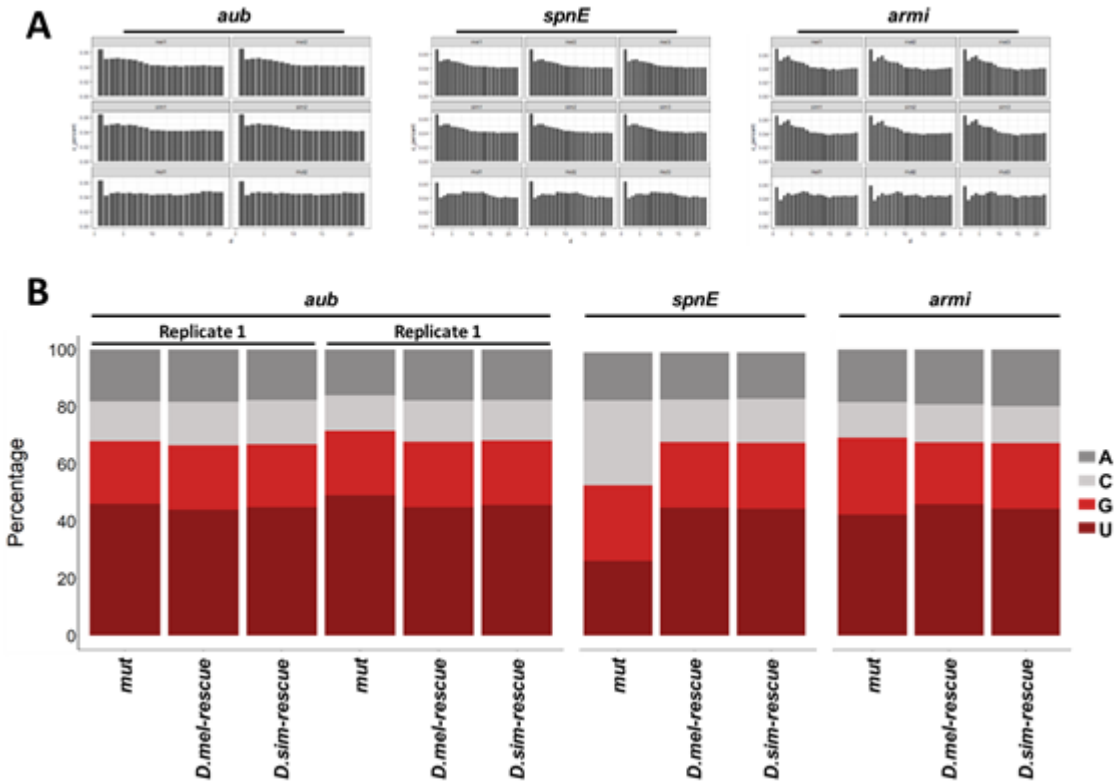
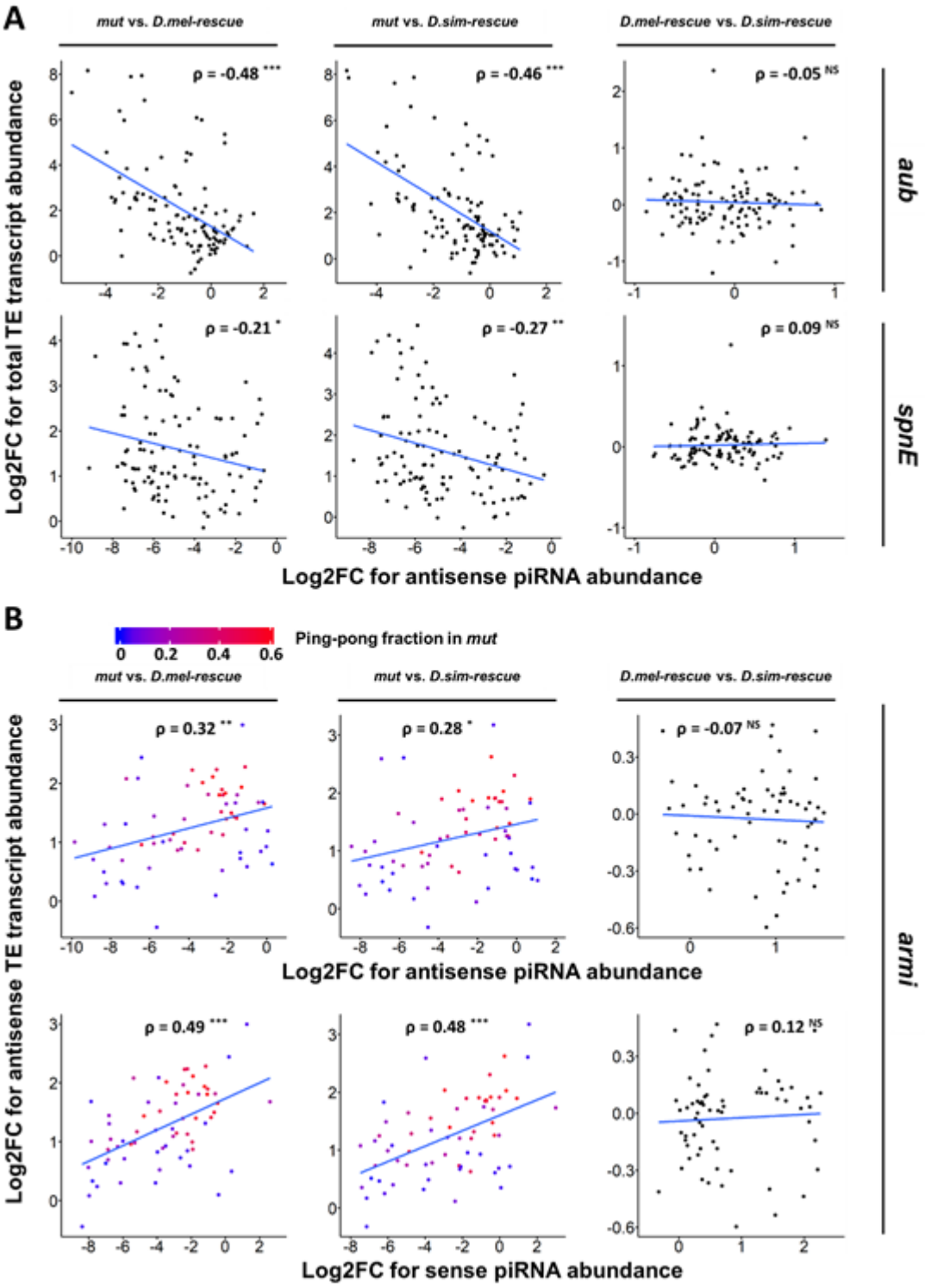


Figure S2. Observed peaks of 1nt distance (A) and +1 U bias (B) among each genotype for each protein studied.

1



2

3 **Figure S3.** Relationship of changes in TE transcript abundance and piRNA abundance similarly

4 observed between mutant and rescues is lost when comparing two rescues.

5 (A). Relationship between changes in total TE transcript abundance and changes in antisense

piRNA abundance for each comparison for *aub* and *spnE*. Negative correlations suggest that loss of piRNAs explains TE transcriptional derepression in *aub* and *spnE* mutants, but is unrelated to differences in TE transcription between transgenic rescues.

(B). Relationship between changes in antisense TE transcript abundance and changes in antisense or sense piRNA abundance for each comparison for *armi*. Positive correlations suggest that antisense TE transcription offsets piRNA losses in *armi* mutant, but is unrelated to differences in piRNA production between rescues. Dot color indicates the ping-pong fraction of each TE family in the *armi* mutant, and demonstrates that more robust ping-pong in the mutant is associated with higher levels of anti-sense transcription and reduced piRNA loss. 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 $p \leq 0.001$.

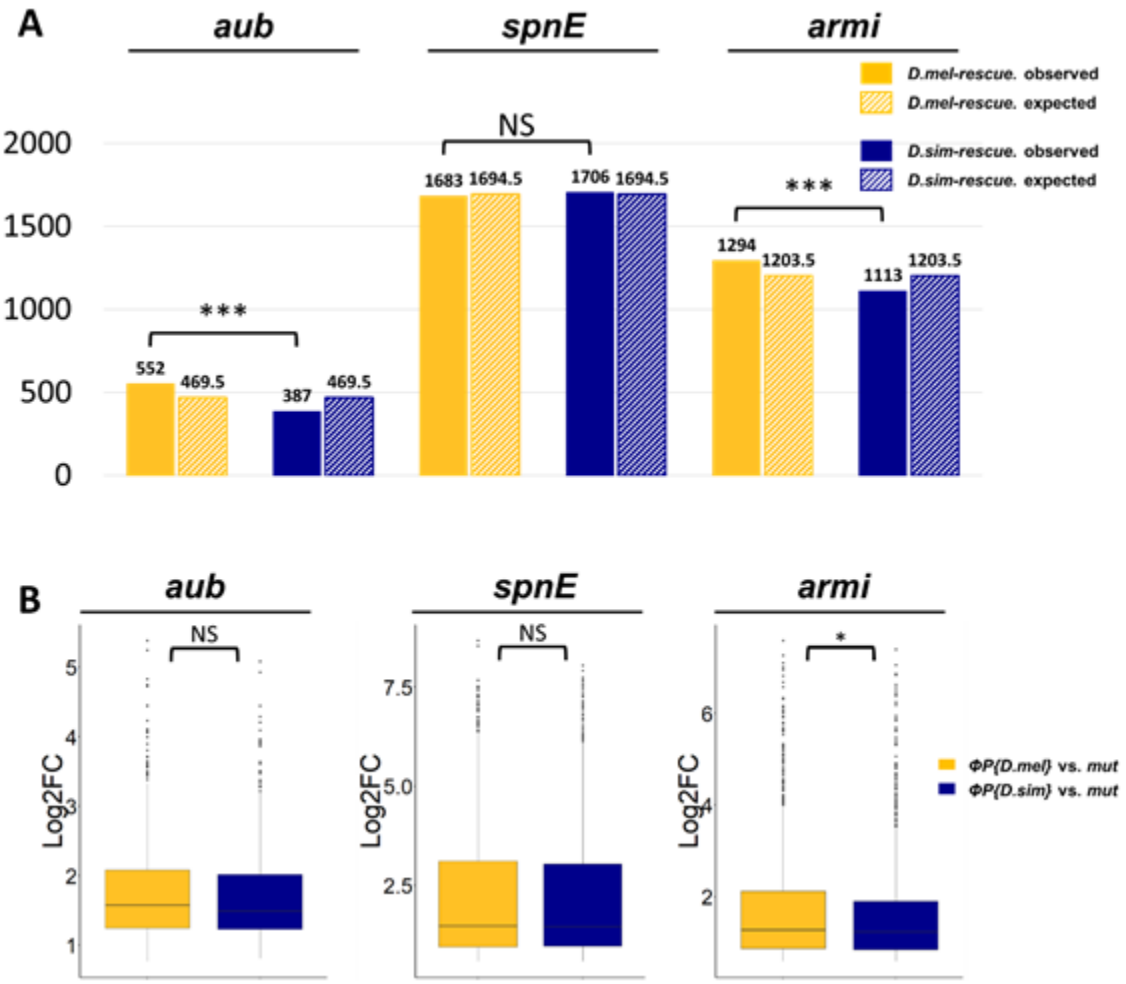


Figure S4. (A). The number of genes whose corresponding piRNA abundance is increased (>1.5 fold) in the presence of each transgene as compared to the mutant. Log2 fold-change values were based on two biological replicates for *aub* and three biological replicates for *spnE* and *armi*, and were obtained 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 S4A 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 $p \leq 0.001$.