

Rewired RNAi-Mediated Genome Surveillance in House Dust Mites

Mondal MH¹, Ortolano NA¹, Scott KE¹, Taylor CB¹, Klimov PB^{2,3}, Flynt AS¹.

¹ University of Southern Mississippi, Hattiesburg, MS, USA

² University of Michigan, Ann Arbor, MI, USA

³ Tyumen State University, Tyumen, Russia

Abstract

House dust mites are common pests with an unusual evolutionary history, being descendants of a parasitic ancestor. Transition to parasitism is frequently accompanied by genome rearrangements, possibly to accommodate the genetic change needed to access new ecology. Transposable element (TE) activity is a source of genomic instability that can trigger large-scale genomic alterations. Eukaryotes have multiple transposon control mechanisms, one of which is RNA interference (RNAi). Investigation of the dust mite genome failed to identify a major RNAi pathway: the Piwi-associated RNA (piRNA) pathway, which has been replaced by a novel small-interfering RNAs (siRNAs)-like pathway. Co-opting of piRNA function by dust mite siRNAs is extensive, including establishment of TE control master loci that produce siRNAs. Interestingly, other members of the Acari have piRNAs indicating loss of this mechanism in dust mites is a recent event. Flux of RNAi-mediated control of TEs provides a mechanism for unusual arc of dust mite evolution.

House dust mites are ubiquitous inhabitants of human dwellings, and are the primary cause of indoor allergy (Arlan, 1991). Dust mites have an unusual evolutionary history, descending from a parasitic ancestor (Klimov and B, 2013). Parasite genomes are typically highly modified; possibly to accommodate genetic novelty needed to productively interact with a host (Brookfield, 2011; Poulin and Randhawa, 2015). The sequence of events leading to adoption of a parasitic lifestyle may require genomic crisis to yield the rewired parasite genome. Dust mites represent an extreme case potentially experiencing a second round of genomic change to reacquire a free-living ecology.

Transposable element (TE) activity is a major source of genome instability (Fedoroff, 2012; Hedges and Deininger, 2007). RNA interference (RNAi), a process that employs small RNAs associated with Argonaute/Piwi (Ago/Piwi) proteins, directs TE silencing in multicellular organisms (Buchon and Vaury, 2006). In many animals the Piwi-associated RNA (piRNA) pathway is the primary RNAi-based defense (Crichton et al., 2014; Senti et al., 2015). In arthropods and vertebrates piRNAs are 25-30 nt long, and unlike other small RNAs, such as microRNAs (miRNAs) and small-interfering RNAs (siRNAs), they are not excised from double-stranded RNA (dsRNA) precursors by the RNase III enzyme Dicer (Weick and Miska, 2014). piRNAs are generated in two collaborative pathways: primary processing and the “ping-pong” amplification cycle (Czech and Hannon, 2016). Primary piRNAs are derived from designated single-stranded transcripts, which initiate the ping-pong cycle where Piwi proteins collaborate to capture fragments of TEs and convert them to new piRNAs (Iwasaki et al., 2015; Siomi et al., 2011). Ago/Piwi proteins may possess “slicer” activity where transcripts base-paired with a small RNA are cut 10 nucleotides (nt) from the 5’ end of the small RNA (Huang et al., 2014). TE transcripts processed into ping-pong piRNAs are produced by slicing, and therefore exhibit 10 nt 5’ overlaps with cognate, antisense piRNAs (Iwasaki et al., 2015). piRNA cluster master loci that generate primary piRNAs are comprised of TE fragments, and serve as catalogs of restricted transcripts (Malone et al., 2009). Loss of master loci integrity compromises TE repression and causes sterility.

While piRNA regulation of TEs is common in animals, it has been lost in several nematodes (Sarkies et al., 2015). In these species alternative mechanisms restrict TE mobilization that involve Rdrp (RNA dependent RNA polymerase) and Dicer. Conversion of TE transcripts by Rdrp into dsRNA substrates of Dicer results in siRNA generation. Ago proteins then associate with nascent TE transcripts, recruiting chromatin modulators including DNA methyltransferase. This process, RNA-induced transcriptional silencing (RITS) is common in plants and fungi (Klenov and Gvozdev, 2005; Pikaard, 2006; Verdel et al., 2009). Outside nematodes clades, RITS has not been observed in vertebrates or other ecdysozoans—potentially due to absence of Rdrp (Tomoyasu et al., 2008). One possible exception is chelicerata arthropods, a lineage where dust mites belong, which possess Rdrp proteins. RNAi pathways in chelicerates appear complex as they have both Rdrp and Piwi class Argonaute proteins (Grbic et al., 2011; Kurscheid et al., 2009). Here we investigate the status of small RNA pathways in the dust mite to shed light on TE control mechanisms in this highly derived organism.

We obtained a genome sequence for the American house dust mite *Dermatophagoides farinae* using Illumina and PacBio platforms (Supp Methods). Ago/Piwi proteins were identified in the *D. farinae* genome using amino acid sequences of seven Ago and six Piwi proteins from *Tetranychus urticae*—the closest relative of *D. farinae* with a high quality annotated genome (Grbic et al., 2011). Eight confident Ago homologs were found and annotated with RNA-Seq data (Supp. Fig. 1A)(Chan et al., 2015). Ago proteins from *T. urticae*, *D. melanogaster*, *C. elegans*, and *Ascaris suum* were compared to *D. farinae* Agos using amino acid sequences of Paz, Mid, and Piwi domains. Our phylogenetic analysis recovered two Ago family members likely involved in miRNA and siRNA pathways (Fig. 1A)(Meister, 2013). The remainder belongs to a divergent clade specific to dust mites (Fig. 1A). Surprisingly, none of the Agos from *D. farinae* belong to the Piwi clade. We examined *D. farinae* Agos for the presence of slicer activity. The DEDH slicer motif, which is common in metazoan Ago and Piwi proteins, were found in the *D. farinae* Ago1(miRNA) and Ago2 (siRNA). The divergent Agos have an uncommon DEDD catalytic motif (Fig. 1B). Orthologs containing a DEDD motif could be found in scabies (*S. scabiei*), social spiders (*S. mimosarum*) and *C. elegans*

emphasizing the unusual nature of this Ago clade (Rider et al., 2015; Sanggaard et al., 2014; Tolia and Joshua-Tor, 2007).

Next, a dust mite small RNAs library was analyzed for presence of piRNA-class transcripts. An algorithm that determines overlap probabilities was used to investigate whether the 10 nt overlap characteristic of ping-pong piRNAs was present in larger small RNAs (24-31nt) (Antoniewski, 2014). For fruit fly RNAs, we observed a peak at 10 nt overlap length, but not for dust mites (Fig. 1C). Investigation of overhangs revealed that 24-31 nt dust mite small RNAs possess a 2 nt overhang, a characteristic of Dicer cleavage (Fig. 1D). This indicates 24-31 nt small RNAs in dust mite are siRNAs produced by Dicer, and not piRNAs (Wen et al., 2014).

To further investigate the biology of dust mite siRNAs, we identified small RNA producing loci having >1000 read density that are longer than 200 nt. We found ~400 loci that produced 75% of the small RNAs mapping to the *D. farinae* genome (Supp. Table 1&2). The identities of the regions were determined using blast2go (Conesa et al., 2005). Nearly a third of the loci were rRNA or mRNA. The remainder showed homology to TEs, along with a large cohort of loci that lacked similarity to known sequences. These unknown elements produce over half of the reads mapping to longer loci. The remaining 21% of small RNAs align to TEs. To determine whether any of the TE loci might be processed into piRNAs we investigated if strand bias was present. Inspection of loci revealed reads emanating from both strands when considering all mapping events or only unique hits (Supp. Table 2). This indicates that like the ping-pong pathway, primary piRNAs are also absent from dust mites.

To investigate the role of dust mite small RNAs in genome surveillance we sought to characterize biogenesis of small RNAs across all TEs. Nearly 700 TE loci were annotated using RepeatMasker (Smith et al., 2007). Compared to the bimodal distribution of TE small RNAs in *Drosophila* we found dust mite small RNAs that mapped to TE loci were unimodal with a small portion in the 26-31 nt size range (Supp. Fig. 3). The dust mite small RNAs also did not exhibit a strong 1U bias typical for piRNAs in *Drosophila* (Fig. 2A). Instead we observe an equal distribution of A and U at the first position. Furthermore, the TE derived small RNA show a 2 nt overhang bias

(Fig. 2B). Together this suggests siRNAs are the main mode for controlling TEs in dust mites, accommodating the apparent loss of piRNAs.

To verify whether TEs are controlled by siRNAs we sought to elicit RNAi against dust mite Dicer proteins by delivery of dsRNA trigger. Dust mites tolerate being soaked in dsRNA; after which, knockdown of target gene expression can be observed (Supp. Fig. 4). In the *D. farinae* genome we found three Dicers (Supp. Fig. 1B). Depletion by RNAi of Dcr1 and Dcr2 but not Dcr3 caused derepression of a Gypsy element, a known target of piRNAs in flies (Fig. 2C)(Brennecke et al., 2007). We identified moderate to strong nuclear localization signal (NLS) in Dcr1 and Dcr2, but not in Dcr3 (Supp. Table 3). Several of the Ago proteins and an Rdrp homolog also possess NLS domains. This suggests that dust mites may be repressing TE activity through a nuclear RITS-type mechanism acting on nascent transcripts.

Investigation of RNAi in dust mites revealed loss of the piRNA pathway and replacement by siRNAs. This is similar to observations in nematodes where piRNAs were lost multiple times during this group's radiation (Sarkies et al., 2015). The loss of piRNA activity in dust mites and nematodes is likely to be tolerated due to compensation by amplifying siRNAs produced by Rdrp (Sarkies et al., 2015). The collective function of dust mite Rdrps, however, appears to be distinct from nematodes, as only processive versions are present (Supp. Fig. 5). This suggests the *de novo* siRNA pathway is not present in mites, which is consistent with our observation that small RNAs in dust mites are Dicer products (Figure 2). Substantial Rdrp activity does appear to be present in dust mites; dsRNA soaking causes elevation of target mRNA that accompanies decrease in protein accumulation (Supp. Fig. 4B).

Dust mites also differ from nematodes that lack piRNAs in the organization of siRNA producing loci. A key feature of piRNA biology is the cataloging of restricted sequences into master loci. In the nematode lineages lacking piRNAs master loci also appear to be absent (Sarkies et al., 2015). This is not the case in dust mites (Fig. 3A). Three loci were discovered that span 62 kb, contain sequences from multiple varieties of TE, and exhibit homology to 70% of TE mapped small RNAs. Similar regions could not be found in the *S. scabiei* genome (Rider et al., 2015). Poor conservation is a

characteristic of piRNA master loci (Shi et al., 2013). The dust mite loci generate siRNAs, arising from an apparent dsRNA precursor as both strands of the loci show similar rates of read mapping (Fig 3A). It is unclear whether these regions are piRNA master loci that converted to siRNA production or are the result of genome rearrangements that occurred following the loss of the piRNA pathway function.

Nematodes ultimately use siRNAs to drive DNA methylation (Sarkies et al., 2015). Dust mites further differ from nematodes in this regard. We found a single DNA methyltransferase in the *D. farinae* genome, a Dnmt1 homolog. It is likely a pseudogene as it appears to be truncated and shows little evidence of expression (Fig 3B,C). Furthermore, substantial methylation was not observed in dust mites using an immunoblotting method (Supp. Fig. 5). This is consistent with loss of DNA methylation of TEs in arthropods, and highlight the distinct, derived nature of small RNA-mediated genome surveillance in dust mites (Yan et al., 2015).

This work provides insight into the elaborate nature of RNAi in chelicerates. Most members of this clade appear to have both Piwi proteins and Rdtps (Grbic et al., 2011; Kurscheid et al., 2009; Sanggaard et al., 2014). However, there doesn't appear to be much benefit to having both pathways as many chelicerate genomes are heavily colonized by TEs (Gulia-Nuss et al., 2016). The loss of the piRNA pathway in dust mites probably occurred in the parasitic ancestor. Inspection of the scabies genome failed to uncover piwi homologs; however, without transcriptome information systematic annotation of scabies Ago proteins was not possible (Rider et al., 2015). It is not clear whether loss of piRNAs accompanied transition to parasite ecology or occurred later in the group's radiation.

Flux of small RNA-mediated TE surveillance correlates with evolutionary innovation; for example, higher arthropods lost Rdtp in favor of piRNA control of TE (Maida and Masutomi, 2011). This also occurred when vertebrates diverged from basal chordates (Putnam et al., 2008). In both cases loss of Rdtp accompanied innovation in body plan and sensory organs. In vertebrates whole genome duplication occurred twice following descent from an Rdtp expressing ancestor affirming a period of genome instability (Putnam et al., 2008). TE activity may be fortuitous for adaptation, and

dramatic evolutionary changes require extreme events such as perturbation of surveillance mechanisms.

References

- Antoniewski, C. (2014). Computing siRNA and piRNA overlap signatures. *Methods Mol Biol* 1173, 135-146.
- Arlan, L.G. (1991). House-dust-mite allergens: A review. *Experimental & Applied Acarology* 10, 167-186.
- 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.
- Brookfield, J.F. (2011). Host-parasite relationships in the genome. *BMC Biol* 9, 67.
- Buchon, N., and Vaury, C. (2006). RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity (Edinb)* 96, 195-202.
- Chan, T.F., Ji, K.M., Yim, A.K., Liu, X.Y., Zhou, J.W., Li, R.Q., Yang, K.Y., Li, J., Li, M., Law, P.T., *et al.* (2015). The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 135, 539-548.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., and Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol Life Sci* 71, 1581-1605.
- Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem Sci*.
- Fedoroff, N.V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 338, 758-767.
- Grbic, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouze, P., Grbic, V., Osborne, E.J., Dermauw, W., Ngoc, P.C., Ortego, F., *et al.* (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487-492.
- Gulia-Nuss, M., Nuss, A.B., Meyer, J.M., Sonenshine, D.E., Roe, R.M., Waterhouse, R.M., Sattelle, D.B., de la Fuente, J., Ribeiro, J.M., Megy, K., *et al.* (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun* 7, 10507.
- Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 616, 46-59.
- 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.
- Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* 84, 405-433.

Klenov, M.S., and Gvozdev, V.A. (2005). Heterochromatin formation: role of short RNAs and DNA methylation. *Biochemistry (Mosc)* *70*, 1187-1198.

Klimov, P.B., and B, O.C. (2013). Is permanent parasitism reversible?--critical evidence from early evolution of house dust mites. *Syst Biol* *62*, 411-423.

Kurscheid, S., Lew-Tabor, A.E., Rodriguez Valle, M., Bruyeres, A.G., Doogan, V.J., Munderloh, U.G., Guerrero, F.D., Barrero, R.A., and Bellgard, M.I. (2009). Evidence of a tick RNAi pathway by comparative genomics and reverse genetics screen of targets with known loss-of-function phenotypes in *Drosophila*. *BMC Mol Biol* *10*, 26.

Maida, Y., and Masutomi, K. (2011). RNA-dependent RNA polymerases in RNA silencing. *Biol Chem* *392*, 299-304.

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.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* *14*, 447-459.

Pikaard, C.S. (2006). Cell biology of the Arabidopsis nuclear siRNA pathway for RNA-directed chromatin modification. *Cold Spring Harb Symp Quant Biol* *71*, 473-480.

Poulin, R., and Randhawa, H.S. (2015). Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology* *142 Suppl 1*, S6-S15.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., *et al.* (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* *453*, 1064-1071.

Rider, S.D., Jr., Morgan, M.S., and Arlian, L.G. (2015). Draft genome of the scabies mite. *Parasit Vectors* *8*, 585.

Sanggaard, K.W., Bechsgaard, J.S., Fang, X., Duan, J., Dyrland, T.F., Gupta, V., Jiang, X., Cheng, L., Fan, D., Feng, Y., *et al.* (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun* *5*, 3765.

Sarkies, P., Selkirk, M.E., Jones, J.T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N.M., Schiffer, P.M., *et al.* (2015). Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biol* *13*, e1002061.

Senti, K.A., Jurczak, D., Sachidanandam, R., and Brennecke, J. (2015). piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev* *29*, 1747-1762.

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome Research* *23*, 497-508.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* *12*, 246-258.

Smith, C.D., Edgar, R.C., Yandell, M.D., Smith, D.R., Celniker, S.E., Myers, E.W., and Karpen, G.H. (2007). Improved repeat identification and masking in Diptera. *Gene* *389*, 1-9.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the argonautes. *Nat Chem Biol* *3*, 36-43.

Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, M., Grossmann, D., and Bucher, G.

(2008). Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. *Genome Biol* 9, R10.

Verdel, A., Vavasseur, A., Le Gorrec, M., and Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. *Int J Dev Biol* 53, 245-257.

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

Wen, J., Mohammed, J., Bortolamiol-Becet, D., Tsai, H., Robine, N., Westholm, J.O., Ladewig, E., Dai, Q., Okamura, K., Flynt, A.S., *et al.* (2014). Diversity of miRNAs, siRNAs, and piRNAs across 25 *Drosophila* cell lines. *Genome Res* 24, 1236-1250.

Yan, H., Bonasio, R., Simola, D.F., Liebig, J., Berger, S.L., and Reinberg, D. (2015). DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu Rev Entomol* 60, 435-452.

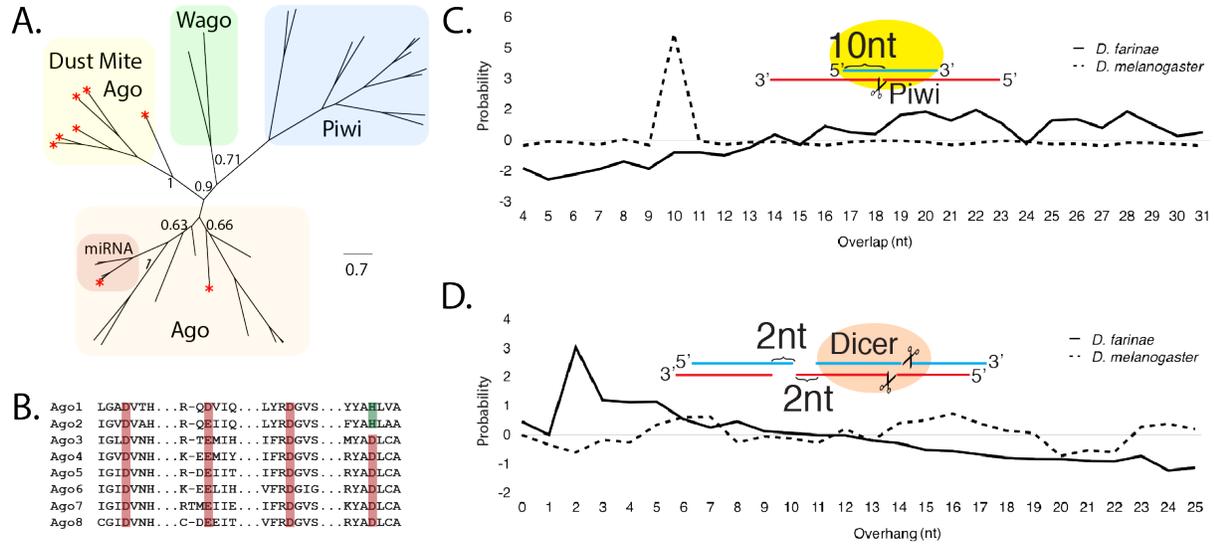


Figure 1: Absence of Piwi/piRNA pathway in dust mites. **A.** Relationship of Ago/Piwi proteins from *D. farinae*, *Drosophila*, *C. elegans*, and *A. suum* using conserved Paz, Mid and Piwi domains. Asterisk indicates dust mite proteins. Only two Wago proteins included on tree for simplicity. Bootstrap values for major nodes indicated. **B.** Alignment of dust mite Ago "slicer" DEDH/D motif. Active site residues highlighted in red or green. **C.** Overlap probability for nucleotide overlaps for 24-31nt RNAs from *D. farinae* and *D. melanogaster*. **D.** Probability associated with overhangs of small RNA populations shown in C. Overhangs calculated by subtracting overlap length from read length for every size. Diagrams in C and D showing overlap and overhang characteristic of piRNA and siRNA biogenesis, respectively.

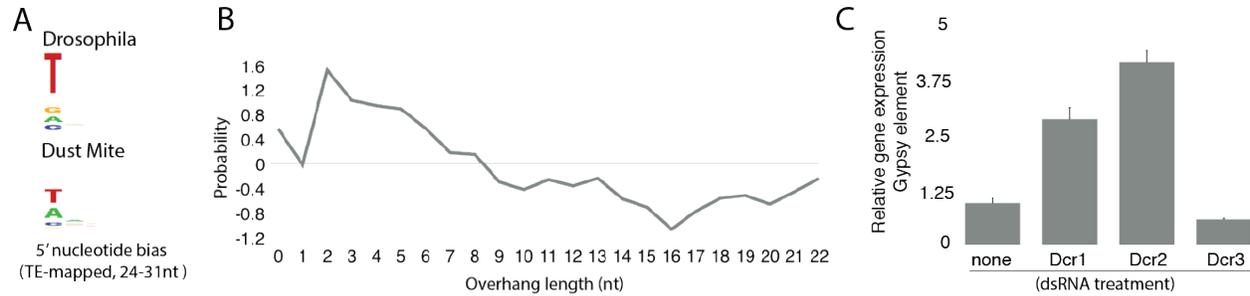


Figure 2: TE derived small RNAs are Dicer dependent. **A.** Seqlogo showing nucleotide bias at the 5' end of 24-31nt small RNA that map to TEs in *Drosophila* and dust mites **B.** Probability of overhangs associated with small RNAs mapping to annotated TEs in the *D. farinae* genome. **C.** Expression of a gypsy element following dsRNA soaking against the three dust mite Dicers. Error bars represent SEM.

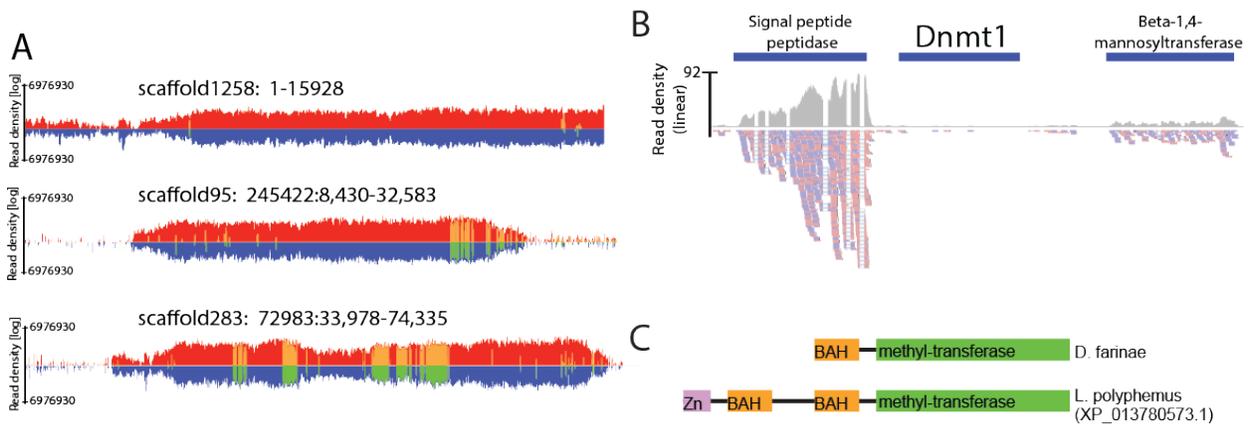


Figure 3: Characteristics of small RNA-mediated TE control in dust mites. A. siRNA producing TE-control master loci. Read density of all mapping events to the positive strand in red, negative strand blue. Density of uniquely mapping reads in yellow for positive strand and green for negative strand. **B.** Expression of dust mite DNA (cytosine-5)-methyltransferase 1 (Dnmt1) in mixed stage RNA-Seq data. Blue bar represents dust mite Dnmt1 locus in the scaffold. Read density in region shown as grey plot. Reads mapping below. Plus strand mapping in red, minus strand mapping in blue. **(C)** Domain structure of truncated *D. farinae* Dnmt1 and an intact ortholog from *Limulus polyphemus*.