Trends in Genetics

CellPress

Review

Evolution of piRNA-guided defense against transposable elements

Shashank Pritam[1] and Sarah Signor[1,*]

Transposable elements (TEs) shape every aspect of genome biology, influencing genome stability, size, and organismal fitness. Following the 2007 discovery of the piRNA defense system, researchers have made numerous findings about organisms' defenses against these genomic invaders. TEs are suppressed by a 'genomic immune system', where TE insertions within specialized regions called PIWI-interacting RNA (piRNA) clusters produce small RNAs responsible for their suppression. The evolution of piRNA clusters and the piRNA system is only now being understood, largely because most research has been conducted in developmental biology labs using only one to two genotypes of *Drosophila melanogaster*. While piRNAs themselves were identified simultaneously in various organisms (flies, mice, rats, and zebrafish) in 2006–2007, detailed work on piRNA clusters has only recently expanded beyond *D. melanogaster*. By studying piRNA cluster evolution in various organisms from an evolutionary perspective, we are beginning to understand more about TE suppression mechanisms and organism–TE coevolution.

Highlights

piRNA clusters produce the small RNA that silences transposable elements (TEs) yet they are very evolutionarily labile.

We do not yet know how silencing of a new TE invasion is initiated, though there are several hypotheses including the 'trap' model and initiation by siRNA.

piRNA clusters are typically expressed in either the germline or soma (some in both), and the authors hypothesize that frequent transitions occur from somatic piRNA clusters to germline.

The authors also hypothesize that new somatic piRNA clusters evolve when a promoter is co-opted to express piRNA from a neighboring TE insertion.

The suppression of transposable elements

TEs (see Glossary) are an evolutionarily important force in all eukaryotic genomes. While the exact percentage of the genome made up of transposons varies across species (up to 80% in maize, or 50% in humans), it is always a significant fraction. This percentage is made up of both young active transposons and the degraded remnants of old invasions. In general the individual insertions of transposons are thought to be either largely neutral or deleterious – there are examples of adaptive TE insertions but these appear to be an exception to the rule (reviewed in [1]). Whether individual insertions are neutral or deleterious, the presence of TEs overall is likely deleterious as the host has dedicated machinery for suppressing their transposition. In mammals and insects a mechanism of TE suppression is based on small RNAs, the specific subset of which is referred to as PIWI-interacting RNA (**piRNA**).

These piRNAs interact with PIWI clade proteins to silence TEs transcriptionally and post-transcriptionally [2–5]. The majority of piRNA is derived from the processing of noncoding transcripts that originate from large genomic regions termed **piRNA clusters** (Figure 1). These clusters contain TEs in various states of decomposition, from full length recent insertions to the remnants of ancient invasions. The TEs themselves serve as the template for their own **silencing**, producing transcripts cognate to active transposons that can be used to target them for degradation. For example, in *Drosophila* post-transcriptional silencing of TEs occurs through piRNA interacting with proteins Aub and AGO3, which cleave TE transcripts in the cytoplasm (Figure 2) [2,3,6,7]. TEs are transcriptionally silenced in the nucleus when piRNAs guide the protein PIWI to transposon insertions and deposit chromatin marks [4,5,8].

Silencing of TEs by small RNAs likely evolved deep in the history of eukaryotic evolution as it is shared across many domains of life. However, even on microevolutionary timescales there is

[1]Biological Sciences, North Dakota State University, Fargo, ND, USA

*Correspondence: sarah.signor@ndsu.edu (S. Signor).

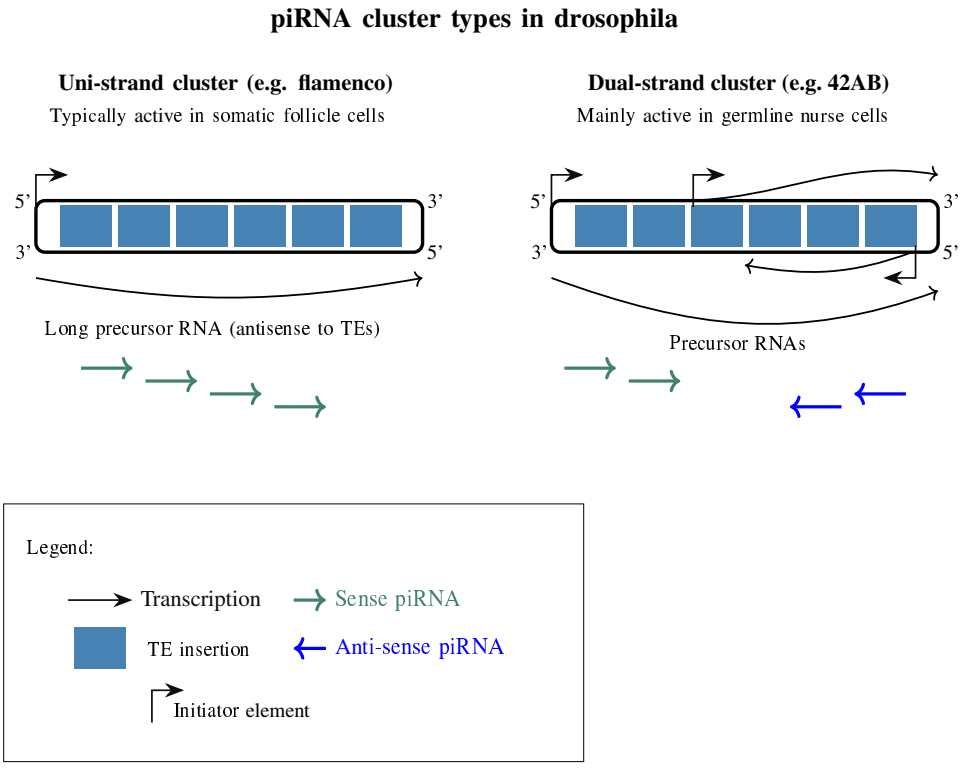

Figure 1. Comparison of unistrand and dual-strand piRNA clusters in *Drosophila melanogaster*. Abbreviations: piRNA, PIWI-interacting RNA; TE, transposable element.

considerable variation in the genes involved in the piRNA pathway, the loci which produce piRNA, and the exact TEs that are being targeted. In this review we aim to summarize our current understanding of the recent evolution of piRNA suppression, particularly with regard to the piRNA clusters which produce primary piRNA. How these loci evolve may shed light on the function and origin of piRNA clusters. We know very little about the piRNA pathway outside of model species such as *Drosophila melanogaster*. The last few years have produced a wealth of research on the evolution of the piRNA which is shaping our understanding of its origin and function.

piRNA production

piRNA are 23–32 bp small RNA produced from piRNA clusters, regions of the genome dense with TE insertions (Figure 1). In most metazoan genomes up to a few hundred piRNA clusters can be identified. Their size can vary widely, from a few thousand base pairs to hundreds of kilobases. Typically they occupy 5% or less of eukaryotic genomes. piRNA clusters are dispersed in animal genomes and the region which they occupy is not conserved in different phylogenetic groups. For example, in *Drosophila* they are located primarily in pericentric heterochromatin, while in mice they are euchromatic [2]. piRNA production from piRNA clusters can be unistrand or double stranded, and the tissues and exact mechanisms where piRNA is produced varies between species.

Germline production of primary piRNA

In the germline cells of insects an entirely distinct mode of transcription occurs to produce piRNA transcripts from piRNA clusters (Figures 1 and 2). piRNA clusters are marked by H3K9me3

Glossary

Ping-pong cycle: amplification process where sense and antisense transcripts from active TEs feed into the piRNA pathway, generating more piRNAs and enhancing silencing.
piRNA: 23–31 bp small RNA bound by PIWI proteins, derived from long precursor transcripts.
piRNA cluster: genomic region that produces piRNA precursors.
PIWI proteins: argonaute family proteins that bind piRNAs and mediate their functions in transposon silencing.
Silencing: repression of TE activity through piRNA-mediated mechanisms, occurring at both transcriptional and post-transcriptional levels.
Transposable element (TE): diverse mobile DNA elements that can transpose in animal genomes, primary targets of piRNA-mediated silencing.

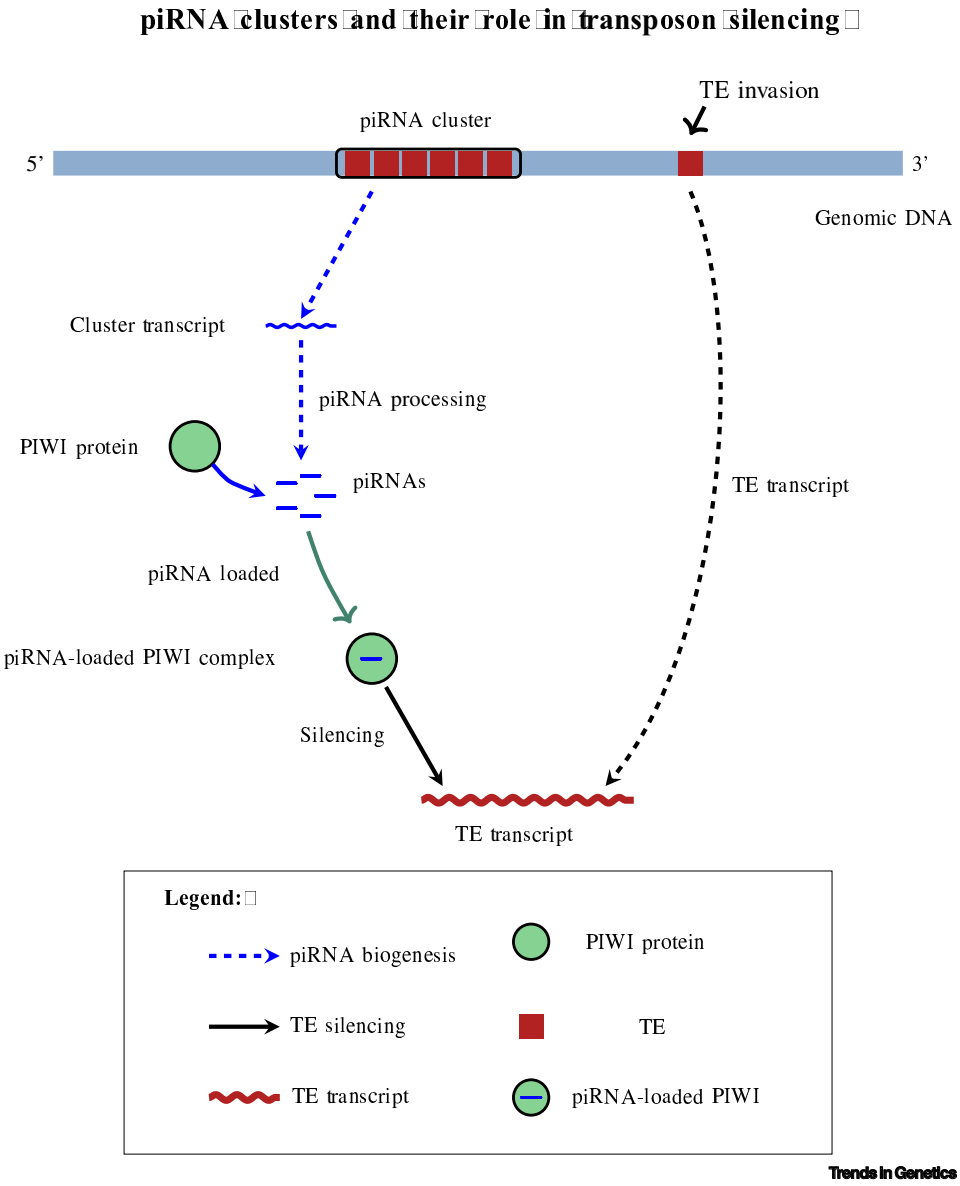

Figure 2. piRNA and the silencing of active TEs. piRNA clusters are distinct regions of the genome occupied by recent TE insertions and the degraded remains of old TEs. They are transcribed into precursor RNA which is post-transcriptionally processed into mature piRNA. These mature piRNA are then bound by **PIWI proteins**. Along with several cofactors these piRNA-PIWI complexes silence TEs both transcriptionally and post-transcriptionally. Abbreviations: piRNA, PIWI-interacting RNA; TE, transposable element.

methylation, an epigenetic mark that initiates heterochromatin formation [9]. H3K9me3 is usually associated with constitutive heterochromatin and is generally considered to suppress transcription [4,5,10–12]. The locus is then licensed for non-canonical transcription from these chromatin marks rather than from a promoter by the Rhino-Deadlock-Cutoff complex. Transcription is dual-stranded and occurs at many positions within the heterochromatic piRNA clusters [2,11,13,14]. Co-transcriptional processes such as splicing and polyadenylation are also suppressed in these regions [11,15]. Following transcription, piRNA transcripts are exported to the cytoplasm and

processed into piRNA by a protein called Zucchini [16,17]. A more detailed discussion of the production of piRNA can be found here [10].

Ovarian somatic production of piRNA

In *Drosophila* there is a separate piRNA pathway in the somatic support cells of the ovary (Figures 1 and 2). This is to suppress *gypsy*-class transposons which possess an *envelope* protein. The *envelope* protein allows TEs to move from the somatic support cells to the germline during development. For example, in the ovarian somatic cells production of primary piRNA occurs through canonical transcription from an RNA Pol II promoter rather than non-canonical transcription licensed by Rhino [5]. The locus which primarily performs this function in *Drosophila* is called *flamenco*, a ~ 400 kb region of the genome filled with antisense insertions of primarily *gypsy*-class elements. *flamenco* produces a long noncoding transcript which is processed and alternatively spliced like a gene before being exported to the cytoplasm [18]. Only PIWI is expressed in the ovarian somatic piRNA system, silencing TEs transcriptionally in the cytoplasm. *flamenco* -like loci are found across drosophilids and in mosquitoes, but this system of repression appears to be unique to arthropods [19].

The ping-pong cycle

There is a secondary pathway for the amplification of the primary piRNA signal in flies, fish, mammals, and many other organisms referred to as the **ping-pong cycle** [2,3]. The ping-pong cycle is an amplification loop in which the cleaved transcripts of TEs are fed back into the pool of piRNA. This pathway connects piRNA biogenesis with target silencing. Specifically, antisense piRNAs are bound by Aub, which cleaves the transcripts of active transposons to produce sense strand piRNAs [20–22]. These sense strand piRNAs are bound by Ago3, which cleaves antisense piRNA cluster transcripts, producing the piRNA that are then bound to Aub. The ping-pong loop is the predominant mechanism that produces piRNA precursors for PIWI [20–22]. Interruption of the ping-pong cycle causes transposon activation [23,24]. A hallmark of the ping-pong cycle is its specific signature of a 10 bp offset between sense and antisense piRNAs. The ping-pong cycle relies on a number of protein cofactors which have been reviewed extensively elsewhere [20,25]

Establishing TE silencing

There are currently three main hypotheses for the process by which a newly invading TE is silenced by piRNA. The initial identification of piRNA clusters as TE suppression loci led to the development of a theory for the silencing of TEs termed the ‘trap’ model [2]. Under the trap model, large piRNA clusters produce the majority of piRNA within the organism. A newly invading TE then integrates itself into the piRNA cluster and is subsequently repressed by piRNA produced from the region [18,26–30]. Integration of a TE into a piRNA cluster is then considered sufficient to activate the entire TE suppression pathway, including the ping-pong cycle. The trap model of TE suppression has empirical support from several studies. First, when artificial sequences are inserted into piRNA clusters they trigger piRNA production cognate to those sequences [31]. Second, when a reporter construct is inserted into a large piRNA cluster it is sufficient to silence the reporter [32,33]. However, there is also evidence from simulations to the contrary such as fewer insertions in piRNA clusters than expected, smaller piRNA clusters than expected, and the experimental ability to delete multiple large clusters without consequences [34,35].

There are two other models which have been proposed to explain how newly invading TEs become piRNA targets, the first of which is called the siRNA model. Under the siRNA model it is maternally inherited siRNA that is necessary and sufficient to activate the production of piRNA. siRNAs are produced from cleavage of dsRNA, which could be readily formed by the transcripts of TEs [33]. The second is called the ‘crank-up’ model, where instead of TE insertions

into a piRNA cluster being the critical step to initiate host silencing it is activation of the ping-pong cycle [36]. This was based on the experimental observation that the *P*-element was able to escape host control despite several insertions into piRNA clusters. Likewise, the same study found that production of siRNA from dsRNA of the *P*-element was insufficient to initiate silencing [36]. The only part of the piRNA pathway that corresponded completely with TE suppression was the activation of the ping-pong pathway. All of these models – the trap model, siRNA model, and crank-up model – remain theoretical and additional research is needed to differentiate between them. However, the evolution of piRNA clusters can help us to differentiate between these different models of TE suppression; for example, if more TE copies are routinely observed in piRNA clusters than would be predicted under the trap model, this could suggest it is not an accurate reflection of piRNA suppression.

To conclude, the piRNA based suppression of TEs is widespread in eukaryotic evolution, while the details of piRNA production and its effectors have diverged considerably among different organisms. Most organisms rely on piRNA clusters for the production of piRNA, though these clusters are made of different components, located in different regions, and transcribed differently. One conserved aspect of the pathway is the ping-pong cycle, which has been hypothesized to be the crucial step for TE suppression. Next, we will review what is known about the evolution of piRNA clusters and how that may inform our understanding of TE suppression.

Evolution of piRNA clusters

piRNA clusters have largely been described in a limited number of genotypes of model organisms, such as *D. melanogaster* or the mouse model (Box 1). However, it is unknown how conserved different aspects of these systems are outside of model organisms (Box 2). For example, do all drosophilids utilize large piRNA clusters to produce the majority of their piRNA? If large piRNA clusters are not conserved outside of *D. melanogaster*, the trap model of TE suppression may not hold. We will review here the current state of our knowledge about the evolution of piRNA clusters, focusing largely on *Drosophila* due to the preponderance of work done in this system.

Evolution of the ovarian somatic TE suppression system

In *D. melanogaster*, *flamenco* is the major unistranded piRNA cluster, likely originating 13.3–15 MYA ago [37]. It is shared with many related species from *D. melanogaster*'s closest relative *Drosophila simulans* to distant relatives such as *Drosophila biarmipes*. Because of its essential function in piRNA suppression - knockout of *flamenco* results in sterility – it was thought be a deeply conserved and necessary part of the piRNA system. However, recent studies have found that the *flamenco* locus is evolutionarily dynamic even within the *melanogaster* subgroup. For example, a recent study found that *flamenco* is duplicated in *D. simulans* and may have germ line activity in *D. simulans* and *Drosophila mauritiana* [38].

Box 1. piRNA clusters in mice

In mice, piRNAs are expressed during two phases that are characterized by divergent piRNA populations. TE-derived piRNAs are primarily early in embryogenesis and are termed pre-pachytene piRNA. These piRNA are necessary for male fertility and protect the genome during early stage spermatogenesis [55–57]. Later in development, the most abundant form of piRNA, termed pachytene piRNAs, are processed from piRNA precursor transcripts produced from large, bidirectionally transcribed regions of the genome [52,54,58]. These regions of the genome are inter-genic and largely depleted of TE sequences, though a fraction are targeted towards TEs. These loci are euchromatic, A-MYB promoter-dependent RNA Pol II transcription units [59]. The sequence of these piRNA-producing regions is not conserved, but the location is, and non-eutherian mammals share around 20 pachytene piRNA clusters [34]. Mice have three PIWI proteins, MIWI, MILI, and MIWI2, that are not direct homologs of *Drosophila PIWI* proteins but perform the same function as in *Drosophila*. *MIWI* and *MILI*-bound piRNAs silence TEs in the cytosol [60]. MIWI2 bound piRNAs silence TEs in the nucleus by guiding the deposition of epigenetic marks such as H3K9me3 and DNA methylation [61].

Box 2. Diversity in piRNA production across animals

While piRNA production is conserved across animals, there is considerable variability in how it is produced. In assassin bugs, all piRNA clusters are unistrand clusters and unlike *flamenco* in *Drosophila* the TEs are not oriented in a particular direction [62]. piRNA clusters are also primarily unistrand in *Tribolium castaneum* [63], *Anopheles gambiae* [64], slugs [65], planaria [65] and *Blattella germanica* [66]. Dual-stranded transcription does occur in these species and the ping-pong cycle is conserved, but it is much less abundant than in *D. melanogaster* [13,63]. The strandedness of piRNA clusters seems to be both evolutionarily labile and potentially not a hard division – in *Drosophila*, piRNA cluster 20A is a germline cluster which appears to be primarily unistrand, and in *D. simulans* and *D. mauritiana, flamenco* shows signals of a dual-stranded cluster. It may be that the ancestral trait is unistrand clusters even in the germline and that *Drosophila* retains remnants of this system.

In assassin bugs, as in other hemimetabolous and homometabolous insects, the ping-pong pathway is also not restricted to the germline [63,66–68]. In fact, even in the basal metazoan *Hydra* the ping-pong cycle is active in the somatic stem cells [69]. Most nematodes appear to have lost PIWI proteins and piRNAs altogether, though the model species *Caenorhabditis elegans* does contain a piRNA pathway that is distinct from flies and mice [70,71]. In *C. elegans*, piRNA are produced from single TEs and these do not appear to be related to the piRNA clusters of other organisms [72,73]. Our understanding of the piRNA pathway outside of model species is still in its infancy.

There also appears to be frequent conversion of *flamenco* between germline and ovarian somatic functions. For example, in *Drosophila ficusphila, flamenco* is now a dual-stranded cluster with a ping-pong signal, and a different unistrand cluster has evolved [37]. These unistrand to dual-stranded transitions appear to happen frequently in *Drosophila* evolution. In the *obscura* group the region syntenic to this second *flamenco*-like locus is now dual-stranded and yet another *flamenco*-like unistrand cluster has evolved [37]. *Drosophila erecta* and *Drosophila yakuba* contain clusters with both germline and ovarian somatic piRNA signals, suggesting that these clusters could also potentially serve both functions, though more evidence is needed to support this claim [37].

Exactly how *flamenco* evolved to be a unistrand cluster is unknown, as is how additional unistrand clusters evolved in other species. However, recent work on the ovarian somatic suppression system in *D. melanogaster* highlights how ovarian somatic clusters could potentially evolve (Figure 3) [39]. In *D. melanogaster 17.6* is not suppressed by *flamenco*. Why a *flamenco* insertion has not arisen is unclear, however a small unistrand cluster appears to have evolved to suppress *17.6*. This new cluster has a partial copy of an old *quasimodo*-related element with an intact long terminal repeat (LTR) and 5′ untranslated region (UTR) neighbors insertions of *17.6* and *idefix* [39]. This *quasimodo*-like fragment exists in *D. simulans* and *Drosophila sechellia* but does not have any neighboring TE insertions. In *D. melanogaster*, the initiator motif in the first LTR of the *quasimodo*-like insertions has been co-opted to be the start of a unistrand piRNA cluster (Figure 3) [39]. Transcripts are produced from these TE insertions which are processed into piRNA, silencing *17.6* in ovarian somatic cells [39]. This is likely how other unistrand clusters evolve – an initiator motif from a TE or other source becomes associated with a TE and begins to act as a cluster, which in turn favors the insertion of additional TEs causing the locus to grow. Exactly how or why this cluster evolves to be unistrand as opposed to joining the germline suppression system is unknown and would be an interesting area of further research.

In the *obscura* species group a gene essential for processing transcripts from *flamenco* has been lost (*fs(1)Yb*), yet abundant ovarian somatic piRNA are produced, suggesting that a different gene or process has evolved to serve this function [40]. There is also evidence that mosquitoes contain unistrand piRNA clusters for suppressing somatic TEs, which implies that this approach is widespread across insects. However, the more fine-grained analyses described earlier suggests that the evolution of this system is complex, with frequent turnover between *flamenco* loci and their function in the ovarian somatic an germline pathways, as well as generation of new *flamenco* loci.

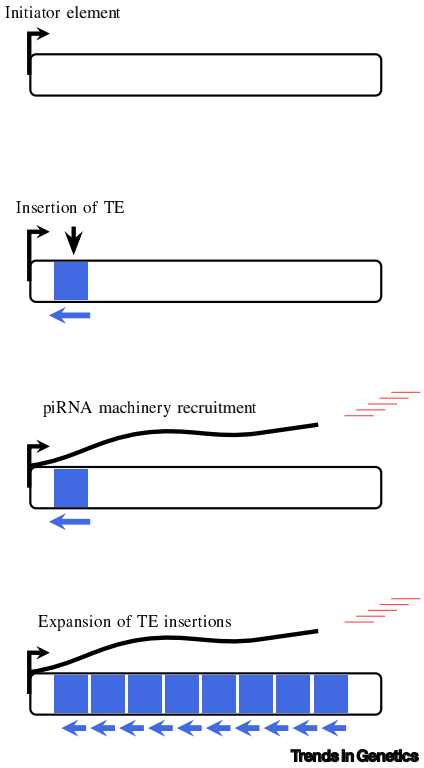

Figure 3. Model for the evolution of unistrand piRNA clusters. An initiator element's motif recruits transcriptional machinery. Following TE insertion nearby, the motif is co-opted to produce piRNA-producing transcripts. Successful TE suppression leads to additional insertions and cluster expansion. Blue rectangles, TEs; black arrows, transcription start sites; blue arrows, transcription direction; wavy lines and red marks, piRNA processing factors. Abbreviations: piRNA, PIWI-interacting RNA; TE, transposable element.

Evolution of the germline TE suppression system

While ovarian somatic piRNA clusters such as *flamenco* are required for TE suppression and organismal fertility, the case with germline piRNA clusters is more complicated. In laboratory strains of *D. melanogaster*, a few large piRNA clusters produce the majority of piRNA. This led to the development of the concept of piRNA clusters – that there would be large, dedicated, and conserved 'TE graveyards' that are primarily responsible for suppressing TEs [2]. The existence of these clusters is also important for the trap model, because a TE must be able to transpose into the large cluster, or trap, and then get suppressed. However, several recent studies have found that germline piRNA clusters are very evolutionarily labile. Between genotypes of the same species of *Drosophila*, 70–80% of cluster regions are shared, meaning that the piRNA-producing regions exist but not necessarily that they produce piRNA [41]. If the existence of the region and shared activity are considered, a recent study found that only around 40% of cluster regions that produce piRNA in one genotype also produced significant amounts of piRNA in other genotypes [42]. For example, *42AB* is the canonical piRNA cluster upon which most work to understand piRNA silencing is targeted. In 2/8 studied genotypes in *D. melanogaster*, *42AB* did not contribute substantially to the pool of sequenced piRNA [42]. Between closely related species such as *D. melanogaster* and *D. simulans*, only around 8% of cluster regions were shared, indicating a lack of overall evolutionary conservation [42]. This phenomenon was also observed in mice, where polymorphic TE insertions were acting as novel piRNA clusters which differed between strains [43].

Germline piRNA cluster regions are also generally thought to be redundant, that is, piRNA-producing TE insertions exist in more than one cluster and the function of any single cluster is not

unique. While this has not been definitely shown, it was supported by a recent study where three of the major piRNA clusters in *D. melanogaster* were deleted without triggering TE activity or altering fertility [34]. For some families of TEs these clusters contained the only piRNA cluster insertions and produced 75% of piRNA targeted to the TE. This actually suggests two somewhat contradictory things. First, it suggests that piRNA cluster regions are not entirely functionally redundant, as other clusters did not contain insertions of the same TE. Second, it suggests that even non-redundant clusters are not entirely necessary for TE silencing, which brings in to question the piRNA cluster model of TE suppression. Is it possible that TEs are actually being silenced by dispersed TE insertions that independently nucleate piRNA? Or are these TEs being silenced by a mechanism other than piRNA? The idea that any (or some) TE insertions could ultimately serve as a piRNA-producing region was supported by several studies which found that numerous insertions spread across the genome contribute to piRNA production [44,45].

Srivastav (2024) suggested a birth death process in which *de novo* TE insertions nucleate piRNA production and grow over time, potentially becoming a large piRNA cluster. A TE insertion could trigger the production of piRNA, which in turn recruits additional TE insertions. This is supported by the finding that large piRNA clusters are evolutionarily older than smaller piRNA clusters. There could be considerable evolutionary turnover of these smaller clusters, with some eventually growing to be larger clusters such as *42AB*. As new invasions occur, the large piRNA cluster may not contain insertions of the newly invading TEs and it will lose its role in TE suppression and be lost [42]. Germline clusters do not require the co-option of an initiator sequence, but it is unclear what specifies a locus for piRNA production in this pathway.

Overall, the evolution of piRNA clusters is dominated by large insertions from new TE invasions and small deletions of older inactive TEs [41,42]. These insertions occur so rapidly that homologous clusters often do not share insertions of the same TEs [41]. Regardless of whether piRNA clusters conform to either of these models – that they are discrete loci with a dedicated function or that most TE copies can serve functionally as a piRNA cluster– individual piRNA-producing loci are dispensable and evolutionarily labile.

Unistrand to double-stranded evolutionary pipeline?

In both *D. simulans* and *D. mauritiana*, *flamenco* has been colonized by abundant *R1* elements, which are also a significant source of double-stranded piRNA production from *flamenco* [38]. *R1* elements are not *gypsy* class elements and they are exclusively vertically transmitted [46,47]. We propose a model of unistrand cluster evolution in which the locus is initially limited to *gypsy*-class *envelope* containing transposons, but over time other classes of transposons colonize the locus (Figure 4). This would stimulate the production of double-stranded piRNA from the *flamenco* locus. Once double-stranded piRNA is produced from the *flamenco* locus it is less efficient in the unistrand pathway and over time a different unistrand locus will evolve, while the original locus is abandoned to the double-stranded piRNA pathway (Figure 4). For example, in *D. melanogaster*, *flamenco* does not contain insertions that belong to TEs active in the germline. However, in *D. simulans*, large islands of repeated *R1* insertions have colonized *flamenco* and double-stranded piRNA is produced from these insertions [38]. Likewise, in *D. mauritiana*, *flamenco* contains *R1* insertions and produces double-stranded piRNA, while *flamenco* in *D. sechellia* does not contain full-length *R1* insertions and is exclusively unistrand. This would explain the frequent conversion of *flamenco* into double-stranded clusters, the existence of intermediate clusters, and the presence of non *gypsy*-class TEs at the locus. This process is distinct from the *de novo* birth of germline piRNA clusters, which is likely the most common mechanism for the evolution of new piRNA clusters. The transition of unistrand ovarian somatic piRNA clusters to germline clusters is hypothetical and would be an interesting area of future research.

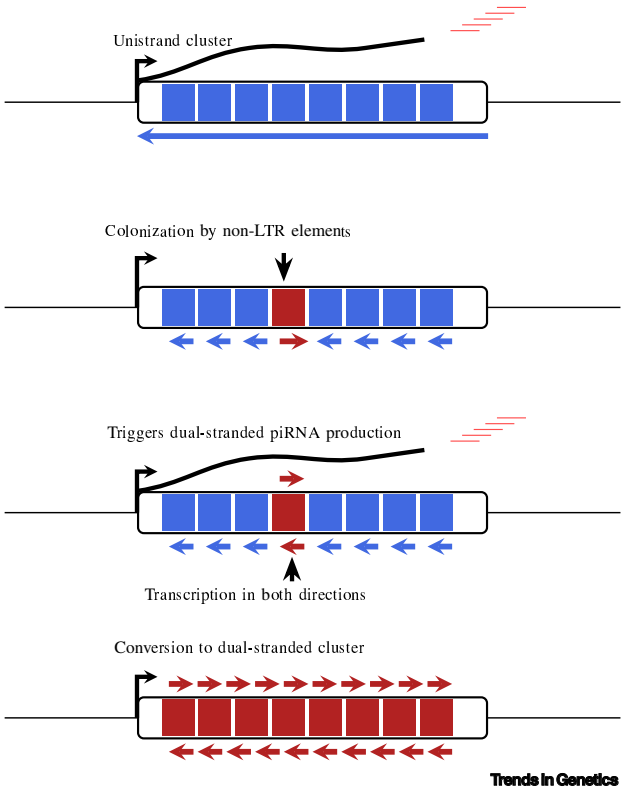

Figure 4. Model for the conversion of unistrand to dual-strand piRNA clusters. A unistrand cluster initially suppresses gypsy-class elements (blue). Non-LTR element colonization (red) triggers bidirectional transcription, leading to dual-strand piRNA production. Over time, the cluster transitions to a dual-strand configuration, while somatic suppression likely shifts to a new unistrand cluster. Blue/red rectangles, gypsy/non-LTR elements; arrows, transcription direction; wavy lines and red marks, piRNA machinery. Abbreviations: LTR, long terminal repeat; piRNA, PIWI-interacting RNA; TE, transposable element.

Evolution of piRNA pathway genes

While the piRNA pathway as a whole has an ancient origin and is conserved amongst most animals, the proteins in the pathway evolve rapidly and are often gained and lost. One recent study in particular highlights this process, where it was found that in the *obscura* group and *Drosophila eugracilis* there are no *Yb* homologs [40]. In most species, *Yb* is required for the formation of *Yb* bodies at the nuclear periphery which recruit piRNA biogenesis factors [48–50]. In particular, *Yb* is required for the preferential production of piRNA from *flamenco*. Despite this, species of the *obscura* group and *D. eugracilis* produce antisense piRNA in the ovarian somatic cells [40]. Indeed, in *Drosophila pseudoobscura* and *D. eugracilis* piRNA biogenesis bodies exist despite the absence of *Yb*, suggesting another protein may have evolved to fill this role. In addition, *D. eugracilis* lacks *Ago3*. *Ago3* is typically necessary for the processing of piRNAs in the ping-pong pathway, and indeed there is no evidence of a ping-pong signature in the piRNAs of this species. Primary piRNA biogenesis occurs as expected, but the vast majority of the piRNA is antisense rather than a mix of both strands, as observed in *D. melanogaster*. Rhino-licensed transcription is dual-stranded, so it is possible that *D. eugracilis* uses promoters outside of clusters for single-stranded transcription, much as occurs within *flamenco* [40]. However, recent models posited that the initiation of the ping-pong cycle was the necessary step for transposon silencing, thus it is unclear if: (i) that is not the case, or (ii) this species has evolved an alternative to this pathway.

The genes in the piRNA pathway also evolve rapidly despite conservation of the overall pathway. For example, *D. melanogaster* and *D. simulans* are sister species that can mate and produce viable (but sterile) offspring. The piRNA pathway genes Rhino and Deadlock are present in

both species, and in both species they co-localize and co-precipitate. However, if you knock out Rhino in *D. melanogaster* and express *D. simulans* Rhino, it no longer co-precipitates with Deadlock [51]. This appears to be due to species-specific compensatory mutations in the interaction surfaces of the two proteins [51].

Concluding remarks

Since the discovery of piRNA simultaneously in dipterans, mice, rats, and zebrafish there has been enormous progress in understanding how TEs are silenced [2,52–54]. However, as time goes on many of the described mechanisms of TE silencing appear to be unique to particular species or even genotypes. For example, it was thought that a few large piRNA clusters were primarily responsible for TE silencing and that the ping-pong cycle was a necessary part of TE silencing. Yet even within genotypes of *D. melanogaster* there is variation in the importance of piRNA clusters: they can be deleted with no consequences, and outside of *D. melanogaster* the entire ping-pong cycle can be lost. Understanding this type of variation in the piRNA system is essential to deciphering how it works (see Outstanding questions). For example, the trap model is based on the idea that a TE will insert into a piRNA cluster like *42AB* and then be silenced. As we move outside of these laboratory genotypes of *D. melanogaster* and find that *42AB* produces nominal amounts of piRNA and that small insertions of TEs appear to be able to nucleate their own piRNA, this brings the trap model into question. This type of evolutionary perspective will be essential to unraveling the mysteries of piRNA-based TE defense from dipterans to humans. As hundreds of new genomes become available every year based on highly accurate long read sequencing, our field is primed for exciting discoveries in this arena in the years to come.

Outstanding questions

How are germline piRNA clusters defined from other heterochromatic regions?

How is the piRNA system initially deployed against novel genomic invaders?

What additional silencing mechanisms are in play to suppress transposition? For example, is splicing an important mechanism of TE suppression?

What is the necessary step to initiate transposon silencing - insertion into a cluster, initiation of the ping-pong cycle, or some other signal?

Acknowledgments

S.S. would like to thank R. Kofler and J. Blumenstiel for useful discussion about the topics discussed within this review paper. S. S. would also like to thank the other members of the laboratory – G. Sarfo Boateng, P. Narayanana, F. Belt, and K. Gabel for useful input. Lastly, S.S. would like to acknowledge P. Senn, without whom this manuscript would not be possible.

S.P. expresses heartfelt gratitude to Siddharth, Linus, Srishti, Penelope, Baibhav, and Kuhu for their unwavering support and cherished presence. This work was supported by the National Science Foundation Established Program to Stimulate Competitive Research (NSF-EPSCoR-1826834 and NSF-EPSCoR-2032756). This work was also supported by the National Institutes of Health MIRA R35GM155272-01.

Declaration of interests

The authors declare no competing interests.

References

1. Schrader, L. and Schmitz, J. (2019) The impact of transposable elements in adaptive evolution. *Mol. Ecol.* 28, 1537–1549
2. Brennecke, J. *et al.* (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103
3. Gunawardane, L.S. *et al.* (2007) A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* 315, 1587–1590
4. Le Thomas, A. *et al.* (2013) Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399
5. Sienski, G. *et al.* (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980
6. Kalmykova, A.I. *et al.* (2005) Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* 33, 2052–2059
7. Peters, L. and Meister, G. (2007) Argonaute proteins: mediators of RNA silencing. *Mol. Cell* 26, 611–623
8. Darricarrère, N. *et al.* (2013) Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1297–1302
9. Wang, L. *et al.* (2023) P-element invasion fuels molecular adaptation in laboratory populations of *Drosophila melanogaster*. *Evolution* 77, 980–994
10. Huang, X. *et al.* (2017) piRNA biogenesis in *Drosophila melanogaster*. *Trends Genet.* 33, 882–894
11. Mohn, F. *et al.* (2014) The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157, 1364–1379
12. Rozhkov, N.V. *et al.* (2013) Evolution and dynamics of small RNA response to a retroelement invasion in Drosophila. *Mol. Biol. Evol.* 30, 397–408
13. Andersen, P.R. *et al.* (2017) A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549, 54–59

14. Klattenhoff, C. *et al.* (2009) The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138, 1137–1149
15. Chen, J. *et al.* (2015) Temperature related reaction norms of gene expression: regulatory architecture and functional implications. *Mol. Biol. Evol.*, msv120
16. Ipsaro, J.J. *et al.* (2012) The structural biochemistry of zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491, 279–283
17. Nishimasu, H. *et al.* (2012) Structure and function of zucchini endoribonuclease in piRNA biogenesis. *Nature* 491, 284–287
18. Goriaux, C. *et al.* (2014) History of the discovery of a master locus producing piRNAs: The flamenco/COM locus in *Drosophila melanogaster*. *Front. Genet.* 5, 257
19. Aguiar, E.R.G.R. *et al.* (2020) A single unidirectional piRNA cluster similar to the flamenco locus is the major source of eve-derived transcription and small RNAs in aedes aegypti mosquitoes. *RNA* 26, 581–594
20. 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
21. Senti, K.A. *et al.* (2015) piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev.* 29, 1747–1762
22. Wang, W. *et al.* (2015) Slicing and binding by ago3 or aub trigger piwi-bound piRNA production by distinct mechanisms. *Mol. Cell* 59, 819–830
23. Sato, K. *et al.* (2015) Krimper enforces an antisense bias on piRNA pools by binding ago3 in the drosophila germline. *Mol. Cell* 59, 553–563
24. Zhang, Z. *et al.* (2011) Heterotypic piRNA ping-pong requires qin, a protein with both e3 ligase and tudor domains. *Mol. Cell* 44, 572–584
25. Czech, B. *et al.* (2018) piRNA-guided genome defense: From biogenesis to silencing. *Annu. Rev. Genet.* 52, 131–157
26. Bergman, C.M. *et al.* (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol.* 7, R112
27. Malone, C.D. *et al.* (2009) Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535
28. Ozata, D.M. *et al.* (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20, 89–108
29. Yamanaka, S. *et al.* (2014) piRNA clusters and open chromatin structure. *Mob. DNA* 5, 22
30. Zanni, V. *et al.* (2013) Distribution, evolution, and diversity of retrotransposons at the *flamenco* locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci.* 110, 19842–19847
31. Muerdter, F. *et al.* (2012) Production of artificial piRNAs in flies and mice. *RNA* 18, 42–52
32. Josse, T. *et al.* (2007) Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet.* 3, 1633–1643
33. Luo, Y. *et al.* (2023) Maternally inherited sirnas initiate piRNA cluster formation. *Mol. Cell* 83, 3835–3851
34. Gebert, D. *et al.* (2021) Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol. Cell* 81, 3965–3978.e5
35. Kofler, R. (2019) Dynamics of transposable element invasions with piRNA clusters. *Mol. Biol. Evol.* 36, 1457–1472
36. Selvaraju, D. *et al.* (2022) P-element invasions in *Drosophila erecta* shed light on the establishment of host control over a transposable element. *bioRxiv*
37. Van Lopik, J. *et al.* (2023) Unistrand piRNA clusters are an evolutionarily conserved mechanism to suppress endogenous retroviruses across the Drosophila genus. *Nat. Commun.* 14, 7337
38. Signor, S. *et al.* (2023) Rapid evolutionary diversification of the flamenco locus across simulans clade drosophila species. *PLoS Genet.* 19, e1010914
39. Senti, K.A. *et al.* (2023) Functional adaptations of endogenous retroviruses to the *Drosophila* host underlie their evolutionary diversification. *bioRxiv*, 2023–08
40. Chary, S. and Hayashi, R. (2023) The absence of core piRNA biogenesis factors does not impact efficient transposon silencing in *Drosophila*. *PLoS Biol.* 21, e3002099
41. Wierzbicki, F. *et al.* (2023) Evolutionary dynamics of piRNA clusters in *Drosophila*. *Mol. Ecol.* 32, 1306–1322
42. Srivastav, S. *et al.* (2023) Rapid evolution of piRNA clusters in the *Drosophila melanogaster* ovary. *bioRxiv*
43. Casas, E. *et al.* (2022) Genetic polymorphisms lead to major, locus-specific, variation in piRNA production in mouse. *bioRxiv*, 2022–10
44. Luo, S. *et al.* (2020) The evolutionary arms race between transposable elements and pirnas in *Drosophila melanogaster*. *BMC Evol. Biol.* 20, 14
45. Moon, S. *et al.* (2018) A robust transposon-endogenizing response from germline stem cells. *Dev. Cell* 47, 660–671.e3
46. Eickbush, D.G. and Eickbush, T.H. (1995) Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics* 139, 671–684
47. Eickbush, T.H. and Malik, H.S. (2002) *Origins and evolution of retrotransposons*, volume 93. ASM Press
48. Ishizu, H. *et al.* (2019) Distinct and collaborative functions of yb and armitage in transposon-targeting piRNA biogenesis. *Cell Rep.* 27, 1822–1835
49. Olivieri, D. *et al.* (2010) An *in vivo* RNAi assay identifies major genetic and cellular requirements for primary pirna biogenesis in *Drosophila*. *EMBO J.* 29, 3301–3317
50. Pandey, R.R. *et al.* (2017) Recruitment of armitage and yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *PLoS Genet.* 13, e1006956
51. Parhad, S.S. *et al.* (2017) Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. *Dev. Cell* 43, 60–70.e5
52. Girard, A. *et al.* (2006) A germline-specific class of small RNAs binds mammalian piwi proteins. *Nature* 442, 199–202
53. Houwing, S. *et al.* (2007) A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* 129, 69–82
54. Lau, N.C. *et al.* (2006) Characterization of the piRNA complex from rat testes. *Science* 313, 363–367
55. Aravin, A.A. *et al.* (2008) A piRNA pathway primed by individual transposons is linked to de novo dna methylation in mice. *Mol. Cell* 31, 785–799
56. Kuramochi-Miyagawa, S. *et al.* (2008) DNA methylation of retrotransposon genes is regulated by piwi family members mili and miwi2 in murine fetal testes. *Genes Dev.* 22, 908–917
57. Watanabe, T. *et al.* (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse rasgrf1 locus. *Science* 332, 848–852
58. Aravin, A. *et al.* (2006) A novel class of small RNAs bind to mili protein in mouse testes. *Nature* 442, 203–207
59. Li, X.Z. *et al.* (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell* 50, 67–81
60. Deng, W. and Lin, H. (2002) Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2, 819–830
61. Cosby, R.L. *et al.* (2019) Host–transposon interactions: Conflict, cooperation, and cooption. *Genes Dev.* 33, 1098–1116
62. Brito, T.F. *et al.* (2024) Embryonic piRNAs target horizontally transferred vertebrate transposons in assassin bugs. *Front Cell. Dev. Biol.* 12, 1481881
63. Ninova, M. *et al.* (2017) Abundant expression of somatic transposon-derived piRNAs throughout tribolium castaneum embryogenesis. *Genome Biol.* 18, 184
64. George, P. *et al.* (2015) Increased production of piRNAs from euchromatic clusters and genes in anopheles gambiae compared with drosophila melanogaster. *Epigenetics Chromatin* 8, 1–21
65. Konstantinidou, P. *et al.* (2024) A comparati
66. Llonga, N. *et al.* (2018) Diversity of piRNA expression patterns during the ontogeny of the german cockroach. *J. Exp. Zool. B Mol. Dev. Evol.* 330, 288–295
67. Aravin, A.A. *et al.* (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* 11, 1017–1027

68. Wang, W. *et al.* (2017) Contrasting sex-and caste-dependent piRNA profiles in the transposon depleted haplodiploid honeybee apis mellifera. *Genome Biol. Evol.* 9, 1341–1356
69. Juliano, C.E. *et al.* (2014) Piwi proteins and piwi-interacting RNAs function in hydra somatic stem cells. *Proc. Natl. Acad. Sci.* 111, 337–342
70. Gainetdinov, I. *et al.* (2018) A single mechanism of biogenesis, initiated and directed by piwi proteins, explains piRNA production in most animals. *Mol. Cell* 71, 775–790
71. Sarkies, P. *et al.* (2015) Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biol.* 13, e1002061
72. Goh, W.S.S. *et al.* (2015) piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.* 29, 1032–1044
73. Weick, E.M. and Miska, E.A. (2014) piRNAs: from biogenesis to function. *Development* 141, 3458–3471