Cell

Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*

Julius Brennecke,[1] Alexei A. Aravin,[1,3] Alexander Stark,[2,3] Monica Dus,[1] Manolis Kellis,[2] Ravi Sachidanandam,[1] and Gregory J. Hannon[1,*]

[1] Cold Spring Harbor Laboratory, Watson School of Biological Sciences and Howard Hughes Medical Institute, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

[2] Broad Institute, MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

[3] These authors contributed equally to this work.

*Correspondence: hannon@cshl.edu

DOI 10.1016/j.cell.2007.01.043

SUMMARY

Drosophila Piwi-family proteins have been implicated in transposon control. Here, we examine piwi-interacting RNAs (piRNAs) associated with each *Drosophila* Piwi protein and find that Piwi and Aubergine bind RNAs that are predominantly antisense to transposons, whereas Ago3 complexes contain predominantly sense piRNAs. As in mammals, the majority of *Drosophila* piRNAs are derived from discrete genomic loci. These loci comprise mainly defective transposon sequences, and some have previously been identified as master regulators of transposon activity. Our data suggest that heterochromatic piRNA loci interact with potentially active, euchromatic transposons to form an adaptive system for transposon control. Complementary relationships between sense and antisense piRNA populations suggest an amplification loop wherein each piRNA-directed cleavage event generates the 5′ end of a new piRNA. Thus, sense piRNAs, formed following cleavage of transposon mRNAs may enhance production of antisense piRNAs, complementary to active elements, by directing cleavage of transcripts from master control loci.

INTRODUCTION

Mobile genetic elements, or their remnants, populate the genomes of nearly every living organism. Potential negative effects of mobile elements on the fitness of their hosts necessitate the development of strategies for transposon control. This is critical in the germline, where transposon activity can create a substantial mutational burden that would accumulate with each passing generation.

Hybrid dysgenesis exemplifies the deleterious effects of colonization of a host by an uncontrolled mobile element. The progeny of intercrosses between certain *Drosophila* strains reproducibly show high germline mutation rates with elevated frequencies of chromosomal abnormalities and partial or complete sterility (Bucheton, 1990; Castro and Carareto, 2004; Kidwell et al., 1977). Studies of the molecular basis of this phenomenon linked the phenotype to transposon mobilization (Pelisson, 1981; Rubin et al., 1982).

Hybrid dysgenesis occurs when a transposon, carried by a male that has established control over that element, is introduced into a naïve female that does not carry the element. The transposon becomes active in the progeny of the naïve female, causing a variety of abnormalities in reproductive tissues that ultimately result in sterility (Engels and Preston, 1979). Since the dysgenic phenotype is often not completely penetrant, a fraction of the progeny from affected females may survive to adulthood. Such animals can develop resistance to the mobilized element, although in many cases, several generations are required for resistance to become fully established (Pelisson and Bregliano, 1987). Immunity to transposons can only be passed through the female germline, indicating that there are both cytoplasmic and genetic components to inherited resistance (Bregliano et al., 1980).

Studies of hybrid dysgenesis have served a critical role in revealing mechanisms of transposon control. In general, two seemingly contradictory models have emerged. The first model correlates resistance with an increasing copy number of the mobile element. A second model suggests that discrete genomic loci encode transposon resistance.

The first model is supported by studies of the I element. Crossing a male carrying full-length copies of the I element to a naïve female leads to I mobilization and hybrid dysgenesis (Bregliano et al., 1980; Bucheton et al., 1984). The number of I copies builds during subsequent crosses of surviving female progeny until it reaches an average of 10–15 per genome (Pelisson and Bregliano, 1987). At this point, I mobility is suppressed, as the initially naïve strain

gains control over this element. Thus, a gradual increase in I element copy number over multiple generations was implicated in the development of transposon resistance.

The second model, which attributes transposon resistance to specific genetic loci, is illustrated by studies of *gypsy* transposon control (Bucheton, 1995). Genetic mapping of *gypsy* resistance determinants led to a discrete locus in the pericentric β-heterochromatin of the X chromosome that was named *flamenco* (Pelisson et al., 1994). Females carrying a permissive *flamenco* allele (one that allows *gypsy* activity) showed a dysgenic phenotype when crossed to males carrying functional *gypsy* elements. Permissive *flamenco* alleles exist in natural *Drosophila* populations but can also be produced by insertional mutagenesis of animals carrying a restrictive *flamenco* allele (Robert et al., 2001). Despite extensive deletion mapping over the *flamenco* locus, no transposon repressor from *flamenco* has been identified. For P elements, a repressor of transposition has been identified as a 66 kDa version of the P element transposase. Expression of the repressor was proposed to correlate with increasing P element copy number, leading to a self-imposed limitation on P element mobility (Misra and Rio, 1990). However, studies of resistance determinants indicated that control over P elements could also be established by insertion of P elements into specific genomic loci, arguing for an alternative, copy number-independent control pathway (Biemont et al., 1990). Studies of inbred lines or of wild isolates with natural P element resistance indicated that P insertions near the telomere of X (cytological position 1A) were sufficient to confer resistance if maternally inherited (Biemont et al., 1990; Ronsseray et al., 1991). Additionally, several groups isolated insertions of incomplete P elements in this same cytological location that acted as dominant transposition suppressors (Marin et al., 2000; Stuart et al., 2002). Importantly, these defective P elements lacked sequences encoding the repressor fragment of transposase.

Both models of transposon resistance, those determined by specific genomic loci and those caused by copy number-dependent responses might be linked to small RNA-based regulatory pathways. Copy number-dependent silencing of mobile elements is reminiscent of copy number-dependent transgene silencing in plants (cosuppression) (Smyth, 1997) and *Drosophila* (Pal-Bhadra et al., 1997). In both cases, silencing occurs through an RNAi-like response where high-copy transgenes provoke the generation of small RNAs, presumably through a double-stranded RNA intermediate (Hamilton and Baulcombe, 1999; Pal-Bhadra et al., 2002). Moreover, mutations in RNAi pathway genes impact transposon mobility in flies (Kalmykova et al., 2005; Sarot et al., 2004; Savitsky et al., 2006) and *C.elegans* (Ketting et al., 1999; Tabara et al., 1999). Finally, small RNAs (rasiRNAs) corresponding to transposons and repeats have been isolated from flies and zebrafish (Aravin et al., 2001, 2003; Chen et al., 2005).

At the core of the RNAi machinery are the Argonaute proteins, which directly bind to small RNAs and use these as guides for the identification and cleavage of their targets (Liu et al., 2004). In animals, Argonautes can be divided into two clades (Carmell et al., 2002). One contains the Argonautes, which act with microRNAs and siRNAs to mediate gene silencing. The second contains the Piwi proteins. Genetic studies have implicated Piwi proteins in germline integrity (Cox et al., 1998; Harris and Macdonald, 2001). For example, *piwi* mutations cause sterility and loss of germline stem cells (Cox et al., 1998; Lin and Spradling, 1997). *aubergine* is a spindle-class gene that is required in the germline for the production of functional oocytes (Harris and Macdonald, 2001). The third *Drosophila* Piwi gene, *Ago3*, has yet to be studied. Mutation of Piwi-family genes also affects mobile elements. For example, *piwi* mutations mobilize *gypsy* (Sarot et al., 2004), and *aubergine* mutations impact *TART* (Savitsky et al., 2006) and P elements (Reiss et al., 2004). Finally, both Piwi and Aubergine bind rasiRNAs (Saito et al., 2006; Vagin et al., 2006) targeting a number of mobile and repetitive elements. These complexes are enriched for antisense small RNAs, as might be expected if they were actively involved in silencing transposons by recognition of their RNA products.

Recently, a new class of small RNAs, the piRNAs, was identified through association with Piwi proteins in mammalian testes (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006). These 26–30 nt RNAs are produced from discrete loci, generally spanning 50–100 kb. Interestingly, mammalian piRNAs are relatively depleted of transposon sequences. Despite apparent differences in the content of Piwi-associated RNA populations in mammals and *Drosophila*, Piwi-family proteins share essential roles in gametogenesis, with all three murine family members, Miwi2 (M.A. Carmell et al., submitted), Mili (Kuramochi-Miyagawa et al., 2004), and Miwi (Deng and Lin, 2002), being required for male fertility.

In order to probe mechanisms of transposon control in *Drosophila* and to understand the relationship between Piwi protein function in flies and mammals, we undertook a detailed analysis of small RNAs associated with Piwi proteins in the *Drosophila* female germline. Our studies indicate that *Drosophila* Piwi-family members function in a transposon surveillance pathway that not only preserves a genetic memory of transposon exposure but also has the potential to adapt its response upon contact with active transposons.

RESULTS

Piwi-Family Members Have Distinct Expression Patterns

In *Drosophila*, Piwi-clade consists of three members: Piwi, Aubergine (Aub), and Ago3. As a prerequisite to further studies of this family, we experimentally determined the sequence of *Ago3* and raised antisera specific to each *Drosophila* Piwi-family protein (Figures S1 and S2).

Previous studies have used myc-tagged Piwi and green fluorescent protein (GFP)-tagged Aub transgenes to investigate their spatial and temporal expression patterns

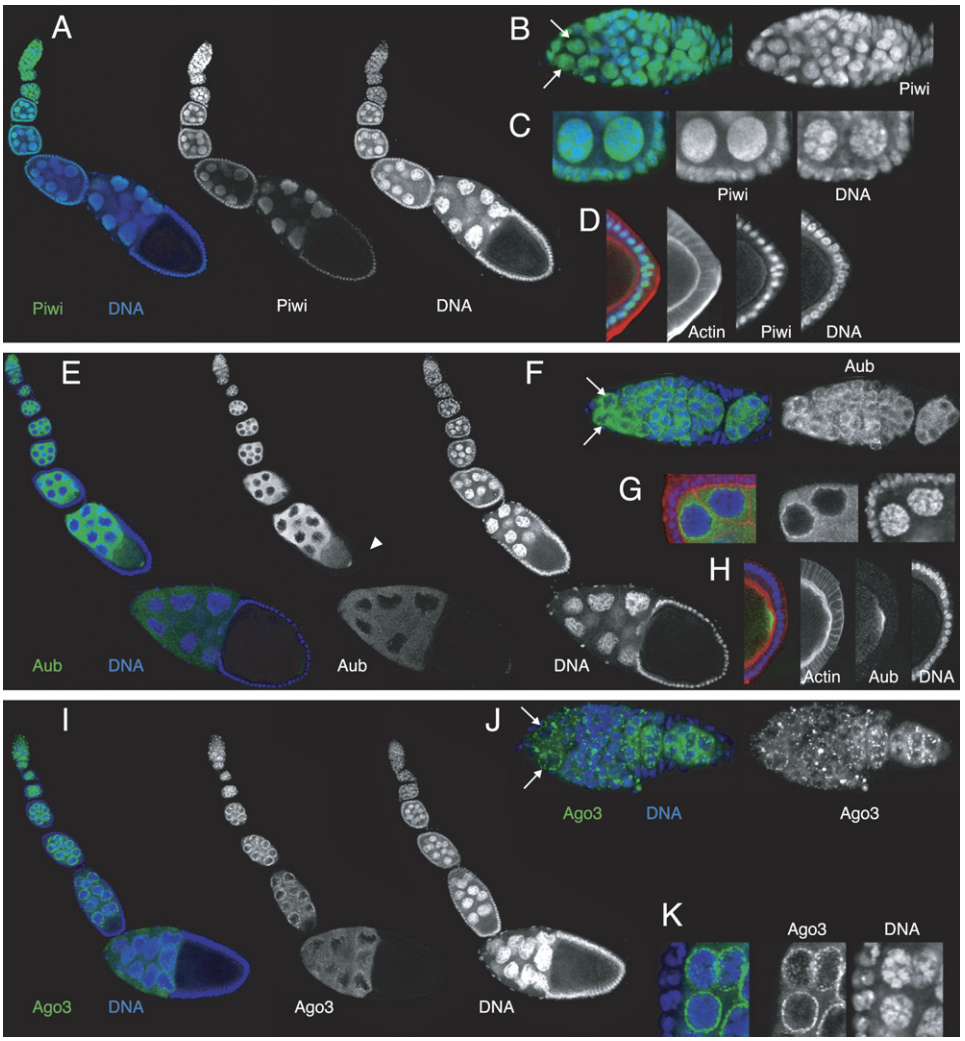


Figure 1. Expression and Localization of Piwi-Family Members in Ovarioles

In all panels, signal from the indicated Piwi-family member is in green; DNA is in blue and actin in red.

(A) Overview of Piwi localization in the ovariole.

(B) Detailed view of the germarium containing the two stem cells (arrows).

(C) Nuclear localization of Piwi in nurse cells and surrounding follicle cells.

(D) Indicates a weak accumulation of maternally deposited Piwi at the posterior pole of stage 10 oocytes.

(E) Overview of Aubergine localization in the ovariole with enrichment of Aub at the posterior oocyte pole in late egg chambers (arrow head).

(F) Detailed view of Aub localization in the germarium containing the two stem cells (arrows).

(G) Enrichment of Aub in the cytoplasm and perinuclear nuage of germline cells; staining is absent from surrounding somatic follicle cells.

(H) Accumulation of Aub at the posterior pole of a stage 10 oocyte.

(I) Overview of Ago3 localization in the ovariole.

(J) Detailed view of Ago3 staining in the germarium showing strong enrichment around the stem cell nuclei (arrows) and in discrete foci.

(K) Detailed view of Ago3 localization to nuage in nurse cells.

during oogenesis (Cox et al., 2000; Harris and Macdonald, 2001). We used our specific antibodies to examine the expression patterns of the endogenous proteins and to extend analyses to the third family member, Ago3.

As previously reported (Cox et al., 2000), Piwi is predominantly nuclear and is present not only in germline cells but also in the somatic cells of the ovary (Figures 1A–1D). Strong Piwi staining is seen in the cap cells that surround the germline stem cells and in the follicle cells that envelop the developing egg chamber. In later stage egg chambers, Piwi is detectable in the cytoplasm of the developing oocyte with a slight enrichment at the posterior where the germline primordia of the embryo will form.

Aubergine is expressed at very low or undetectable levels outside the germline (Figures 1E–1H). Aub is primarily cytoplasmic. As reported previously for GFP-Aub, we detect endogenous protein in germline stem cells, developing cystoblasts, and the nurse cells of developing egg

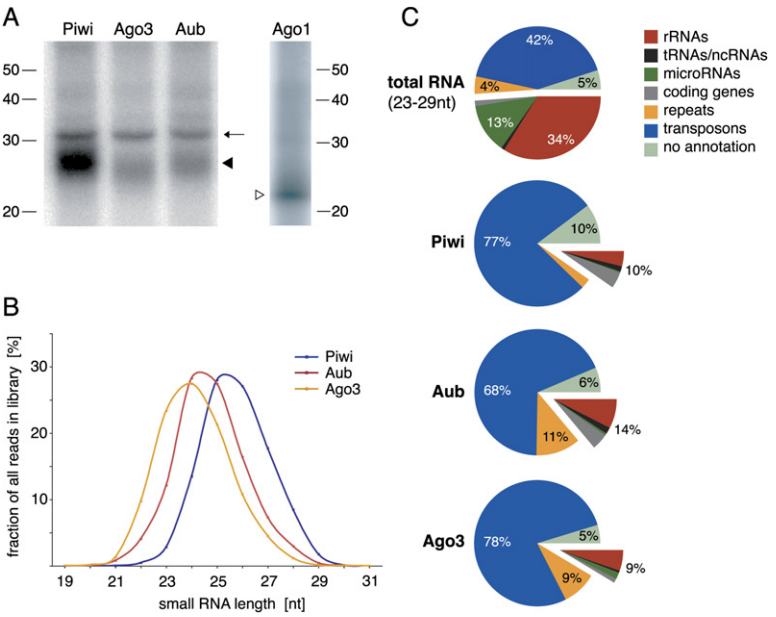


Figure 2. Characteristics of *Drosophila* piRNAs

(A) Radioactively labeled RNA isolated under identical conditions from specific Piwi-family RNPs and Ago1 was analyzed on a denaturing polyacrylamide gel. The positions of RNA size markers, electrophoresed in parallel, are shown to the left. Indicated are piRNAs (solid arrowhead), miRNAs (open arrow head), and 2S rRNA (arrow), which is also present in purifications using control antibodies.

(B) Size distributions of sequenced piRNAs specifically bound by the three Piwi-family members.

(C) Pie chart summarizing the annotation of piRNA populations in total RNA and those bound by Piwi, Aub, and Ago3.

chambers. Aubergine is enriched in nuage, a perinuclear, electron dense structure, displaying a localization pattern very similar to Vasa. Like Vasa, Aubergine is deposited into the developing oocyte from early stage 10 onward and is localized to the pole plasm.

Ago3 protein is also predominantly cytoplasmic (Figure 1I–1K). It is present in the germline but not detectable in somatic cells of the egg chamber, although we do find Ago3 in the somatic cap cells of the germarium. Ago3 shows a more striking accumulation in nuage than does Aub, and it is also found in prominent but discrete foci of unknown character in the germarium. Despite its localization to nuage, Ago3 does not accumulate at the posterior pole of the developing oocyte, and Ago3 is not detected in the pole plasm of early embryos.

Piwi-Family Members Bind Distinct Populations of Small RNAs

To investigate the small RNA populations bound by each *Drosophila* Piwi protein, we purified RNP complexes from ovary lysates. All three proteins associate with small RNAs ranging in length from 23 to 29 nt (Figure 2A). We prepared cDNA libraries from each complex and from 23–29 nt RNAs purified from total ovary RNA. 454 sequencing yielded 60,691 reads (17,709 for Piwi; 23,376 for Ago3; 14,872 for Aub; and 4,734 for ovary total RNA) that match perfectly to Release 5 of the *Drosophila* genome or to nonassembled *Drosophila* sequences from GenBank. These were used for subsequent analysis. It should be noted that we isolated small RNAs from the Oregon R strain, rather than the sequenced strain (y^1; cn^1 bw^1 sp^1). A detailed discussion of potential differences between these genomes and any impacts on the data that we present can be found in the Supplemental Data.

Based both upon gel mobility (Figure 2A) and size distributions (Figure 2B), each Piwi protein bound a specific class of piwi-interacting RNA (piRNA). Piwi-associated RNAs were the largest (25.7 nt), followed by Aub (24.7) and Ago3 bound (24.1) RNAs. As with mammalian piRNAs, Piwi and Aub bound RNAs have a strong preference for a 5′ terminal uridine (83% and 72%, respectively), a trend that is essentially absent from the Ago3 bound population (37% terminal U). *Drosophila* piRNA populations are quite complex, with most RNAs being cloned only once (87% for Piwi, 81% for Aub, and 73% for Ago3).

Despite their size differences, the small RNAs obtained from each complex were remarkably similar in the types of genomic elements to which they correspond. Overall, roughly three quarters of all sequences from each complex could be assigned to annotated transposons or transposon remnants, with nearly all known transposons being represented (Figures 2C and S3). An additional 2%–11% of small RNAs were derived from regions of local repeat structure, such as the subtelomeric TAS repeats or pericentromeric satellite repeats. Thus, nearly 80% of *Drosophila* piRNAs can also be characterized as rasiRNAs. An additional group of sequences (10% for Piwi, 6% for Aub, and 5% for Ago3) map to unannotated regions of the genome, most often to heterochromatic, transposon-rich loci.

Drosophila piRNAs Are Derived from Discrete Genomic Loci

In *Drosophila*, intact and potentially active transposons populate the euchromatin as well as pericentromeric and telomeric heterochromatin. There are also numerous transposon remnants that have mutated sufficiently to negate their potential for transposition. These are strongly enriched in the β-heterochromatin that borders

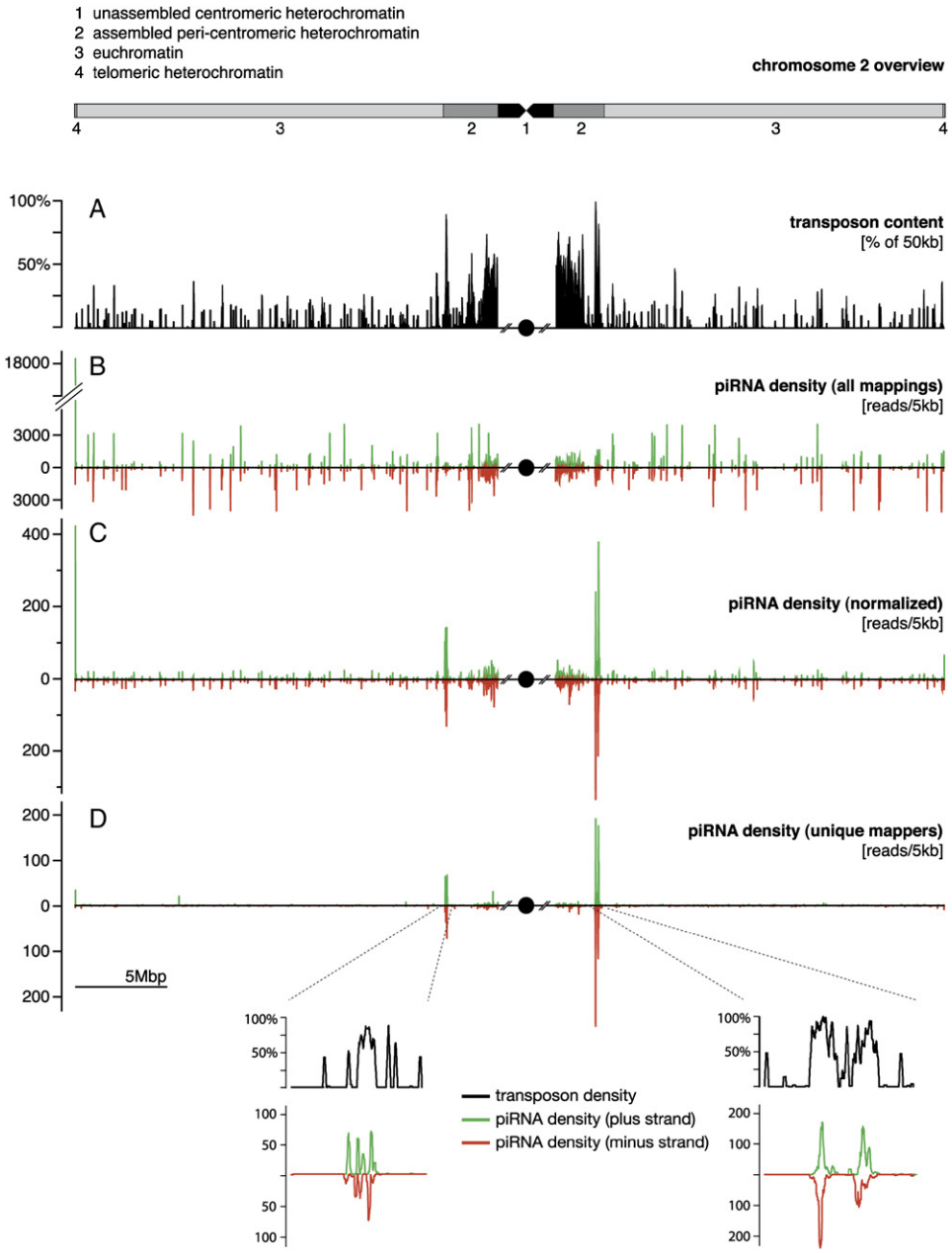


Figure 3. *Drosophila* piRNAs Map to Discrete Genomic Loci

For reference, a sketch of chromosome 2 with the major chromatin domains is provided.

(A) Density of annotated transposons.

(B–D) Density of cloned piRNAs (green: sense; red: antisense) along chromosome 2; the y axis indicates relative units, and the centromere is shown as a circle. In (B) each genomic position corresponding to a cloned piRNA is given equal weight; in (C) each piRNA-genomic match is normalized for the number of times it maps to the genome, resulting in proportionally less signal for piRNAs that map many times. In (D) only those piRNAs that map uniquely to the genome are plotted. The most prominent piRNA clusters (enlarged below and shown together with transposon density in black) are typically found at the border of the pericentromeric heterochromatin.

centromeres and are generally absent from euchromatic chromosome arms (Hoskins et al., 2002).

A depiction of the chromosomal locations matched by *Drosophila* piRNAs closely resembles a plot of transposon density (Figure 3). However, most piRNAs match multiple chromosomal sites. Therefore, to address the genomic origin of piRNAs, it was necessary to restrict our analysis to the ~20% that match the genome at a unique position (Figure S4). A density plot of this small RNA subset shows a striking clustering of piRNAs at discrete loci (e.g.,

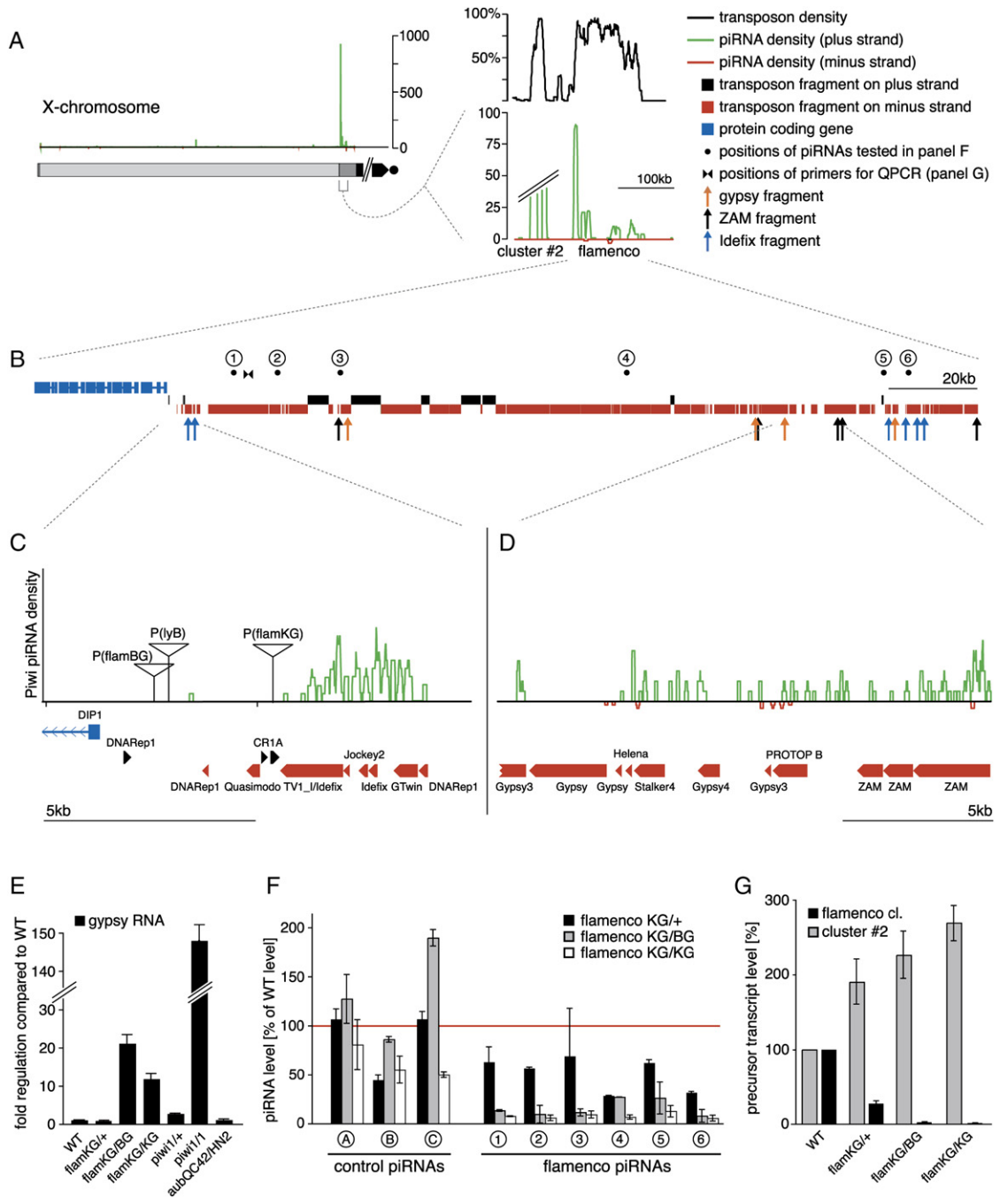


Figure 4. *flamenco* Is a piRNA Cluster

(A) piRNA density on the X chromosome showing two large clusters (numbers as in Table S1) in the pericentromeric heterochromatin (chromatin domains are shaded as in Figure 3). Densities of uniquely mapping sense (green) and antisense (red) piRNAs in two clusters is shown enlarged alongside with corresponding transposon density (black).

(B) Shown is a more detailed map of the *flamenco* locus, indicating protein-coding genes (blue), sense (black), and antisense (red) transposon fragments. The *flamenco* cluster ends ~180 kb proximal to *DIP1* in a gap of unknown size in the genome assembly. Arrows highlight the retroelements *gypsy* (orange), *Idefix* (blue), and *ZAM* (black), which are known to be regulated by the locus.

(C) Shown are the first 10 kb of the *flamenco* locus with the flanking *DIP1* gene (blue), annotated transposon fragments, the P element insertions that result in inactive *flamenco* alleles (triangles), and the density of Piwi-associated piRNAs that map to this region (note that over 99% of the uniquely mapping piRNAs in the *flamenco* cluster derive from the sense strand and that more than 95% associate with Piwi).

(D) Fifteen kilobases of the *flamenco* locus showing Piwi bound piRNAs and a detailed view of transposon fragments from the region, including *gypsy* and *ZAM*.

Figure 3D). A similar plot can be obtained for all piRNAs if each piRNA-genomic match is divided by the number of genomic hits for that sequence (normalized) (Figure 3C). We identified 142 genomic locations as sites of abundant piRNA generation (Tables S1 and S2). These clusters produce 81% of all piRNAs that match the genome at a single site. Although these sites comprise only 3.5% of the assembled genome, more than 92% of all sequenced piRNAs could potentially be derived from these loci.

Only seven piRNA clusters occur in potentially euchromatic regions, with the remainder being present in pericentromeric and telomeric heterochromatin. Telomeric clusters most often consist of satellite sequences and correspond to the subtelomeric terminal associated sequence (TAS) repeats (Karpen and Spradling, 1992). These flank the telomeric arrays of *HetA* and *TART* transposons, for which we also find corresponding piRNAs. Telomeric clusters, supported by the presence of uniquely mapping piRNAs, are found on most chromosome arms (X, 2R, 2L, and 3R). Clusters in the pericentromeric β-heterochromatin display a high content of annotated transoposons (typically from 70% to 90%), with the majority being partial or defective nested elements.

The size of *Drosophila* piRNA clusters varies substantially with the largest being a 240 kB locus in the pericentromeric heterochomatin of chromosome 2R (cytological position 42AB). This cluster produces 20.8% of all uniquely mapping piRNA sequences and could potentially give rise to 30.1% of all the piRNAs, which we identified (Table S1). Overall, the largest 15 clusters account for 57% of the unique piRNAs and up to 70% of the total.

piRNA Clusters Are Master Regulators of Transposon Activity

Numerous genetic studies have pointed to discrete genomic loci that suppress the activity of specific transposons. The best understood is the recessive *flamenco/COM* locus (Prud'homme et al., 1995). *flamenco* was originally identified as a locus controlling the activity of the retroviral *gypsy* element (Pelisson et al., 1994). This locus has subsequently been shown to regulate two additional retroelements, *Idefix* and *ZAM* (Desset et al., 2003).

Through the use of numerous deficiencies, *flamenco* was mapped proximally to the *DIP-1* gene and proposed to span a region of at least 130 kb. This locus corresponds precisely to a piRNA cluster (cluster 8; Table S1, Figures 4A and 4B). Eighty-seven percent of the sequence in the locus consists of nested transposable elements spanning a total length of 179 kb. The locus includes numerous fragments of all three transposable elements that were shown to be controlled by *flamenco/COM* (*gypsy*, *Idefix*, and *ZAM*; Figurse 4B–4D) in addition to many other transposon families.

A second piRNA cluster that has been genetically linked to transposon control corresponds to the subtelomeric TAS repeat on the X chromosome (X-TAS) (Table S1; cluster 4). Numerous studies indicate that insertions of one or two P elements into X-TAS are sufficient to suppress P-M hybrid dysgenesis (Marin et al., 2000; Ronsseray et al., 1991; Stuart et al., 2002). Transposon silencing by these insertions has been linked to the Piwi family, as it is relieved by mutations in *aubergine* (Reiss et al., 2004). The precise insertion sites of three suppressive P elements in X-TAS have been mapped, and they correspond to areas, which give rise to multiple small RNAs (not shown). In accord with a *trans*-acting model for silencing, lacZ-containing P elements inserted into X-TAS can suppress euchromatic lacZ transgenes in the female germline (Roche and Rio, 1998; Ronsseray et al., 1998).

A Functional Pathway Links *flamenco*-Derived piRNAs to *gypsy* Suppression

To probe the relevance of the piRNA cluster mapped to *flamenco*, we made use of mutations that negate the ability of this locus to silence *gypsy*. The only molecularly defined *flamenco* allele corresponds to a P element insertion ~2 kb proximal to *DIP1* and 550 bp upstream of the first piRNA uniquely mapped to this cluster (P(lyB); Figure 4C) (Robert et al., 2001). We obtained two additional lines that harbor P element insertions near P(lyB) and tested their effects on *gypsy* expression. *gypsy* RNA levels increased by ~20-fold in strains carrying homozygous or trans-heterozygous insertions, indicating that P(KG00476) (flamKG) and P(BG02658) (flamBG) (Figure 4C) are indeed *flamenco* mutant alleles (Figure 4E).

We next examined the levels of mature piRNAs from *flamenco* in wild-type animals and *flamenco* mutants. Using quantitative RT-PCR, we found substantial reductions in piRNAs that uniquely map to the *flamenco* piRNA cluster in mutant animals (2, 3, and 5; Figures 4B and 4F). In contrast, piRNAs definitively derived from other clusters were unaffected. We also probed levels of piRNAs that did not map uniquely to *flamenco* (1,4, and 6; Figures 4B and 4F). These were also substantially reduced in *flamenco* mutants, indicating that they arise mainly from this cluster despite not being uniquely assignable based upon sequence information alone.

The *flamenco* piRNA cluster preferentially loads the Piwi protein, with 94% of its uniquely mapping RNAs being Piwi partners. An examination of *gypsy* RNA levels reveals a 150-fold increase in *piwi* mutants (Figure 4E) (Sarot et al., 2004). In contrast, *aubergine* mutations show no

(E) Quantitative RT-PCR analysis on *gypsy* transcript levels in RNA isolated from wild-type, *flamenco* mutant, *piwi* mutant, and *aub* mutant ovaries. Shown are average levels (n = 3), and error bars indicate standard deviation (SD).

(F) Quantitative RT-PCR analysis on several individual piRNAs from small RNA libraries prepared from wild-type or *flamenco* mutant ovaries. Positions of tested piRNAs in *flamenco* are indicated in (B). Error bars indicate SD.

(G) Quantitative RT-PCR analysis on precursor transcripts from two different piRNA clusters in ovaries from wild-type and *flamenco* mutant flies. Position of primers used for the *flamenco* primary transcript are indicated in (B). Shown are average levels (n = 3), and error bars indicate SD.

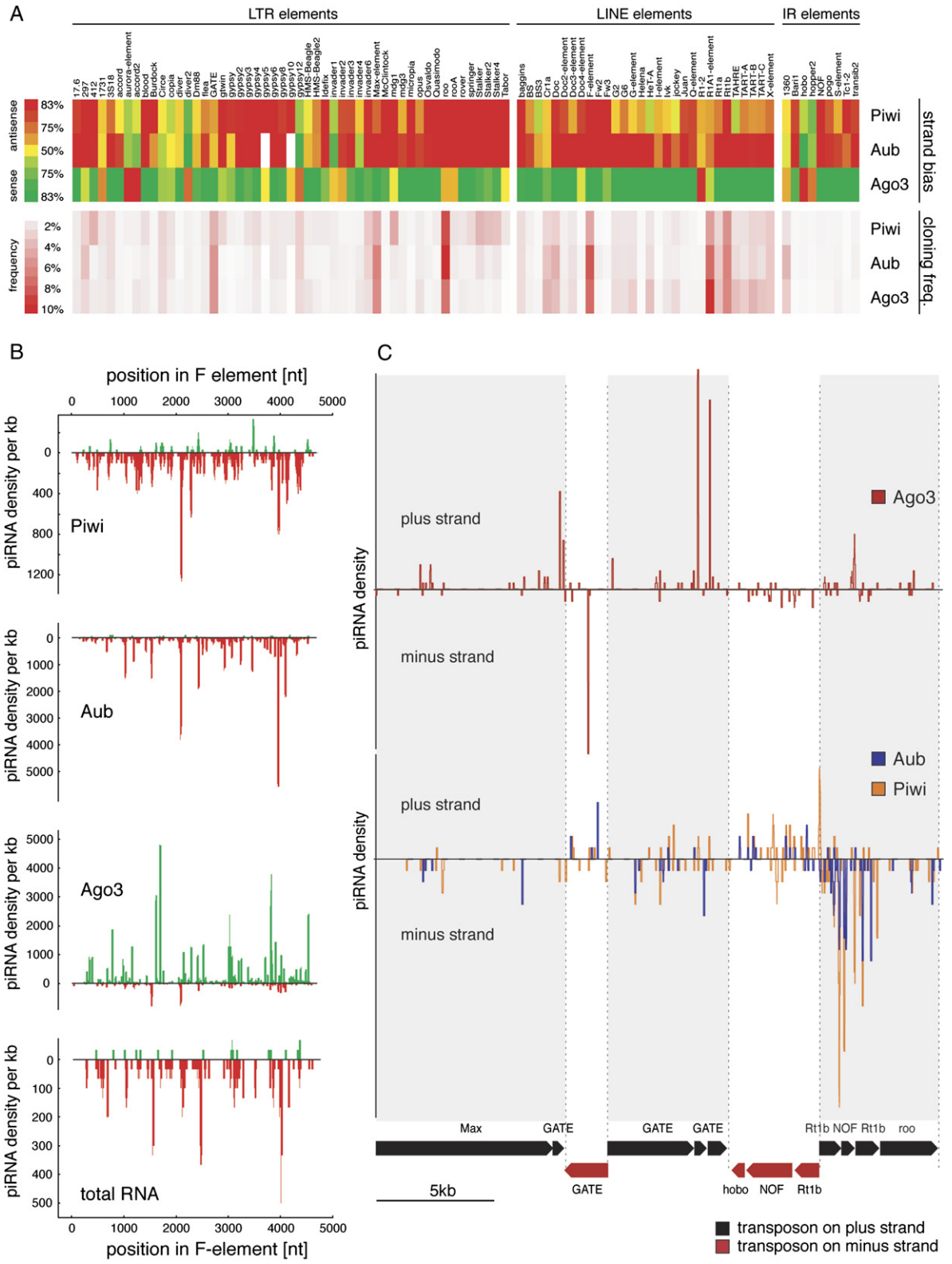


Figure 5. Piwi Complexes Show Profound Strand Biases

(A) Shown in the upper panel is a heat map indicating the strand bias of cloned piRNAs with respect to canonical transposon sequences (indicated at the top). Transposons are grouped into long terminal repeat (LTR) elements, long interspersed nuclear elements (LINE) elements, and inverted repeat

elevation of *gypsy* levels, consistent with a minority of *flamenco* piRNAs entering this complex and with a lack of Aub expression in follicle cells where *gypsy* is expressed. The greater effect of *piwi* compared to *flamenco* mutations is consistent with *flamenco* locus contributing a substantial fraction of but not all *gypsy* repressive piRNAs. In this regard, the *flamenco* cluster has the potential to produce 79% of all piRNAs that target *ZAM*, 30% of those matching *Idefix*, and 33% of piRNAs complementary to *gypsy*.

The P element insertions that we analyzed strongly reduced the abundance of piRNAs generated from sequences up to 168 kb away (Figures 4B and 4F). Considering sequences that map uniquely, *flamenco* produces piRNAs with a marked strand asymmetry that correlates with a strongly biased orientation of transposon fragments in the locus. These observations can be accommodated by a model in which piRNAs are produced from long, unidirectional, precursor transcripts that traverse *flamenco*. Indeed transcripts containing multiple transposons, and several kb of the *flamenco* locus can be easily detected by RT-PCR, and these are lost in *flamenco* mutants (Figure 4G and not shown).

Argonaute3 Binds Sense-Strand piRNAs

Drosophila rasiRNAs show a strong bias for sequences that are antisense to transposable elements (Vagin et al., 2006). We asked whether this observation held for our sequenced piRNAs by examining the strand biases of those derived from Piwi, Aub, and Ago3 complexes. We aligned all piRNA sequences to a comprehensive set of consensus sequences for *D. melanogaster* transposons (canonical set v9.41, Flybase). Piwi and Aub preferentially incorporate piRNAs matching the antisense strand of transposons (76% and 83%, respectively). In contrast, Ago3 complexes contain piRNAs that are strongly biased for sense transposon strands (75%). As piRNAs derived from total RNA retain an antisense bias, Ago3 complexes must be less abundant overall.

The pattern of asymmetry among the three RNPs was preserved when we evaluated each transposable element separately (Figure 5A). As an example, a plot of piRNAs along the F element reveals numerous antisense piRNAs that are loaded into Piwi and Aub and numerous sense piRNAs that enter Ago3 (Figure 5B). There are a few notable exceptions where asymmetry remains marked but is reversed for Piwi/Aub and Ago3 complexes (see for example *accord2*, *gypsy12*, *diver2*, and *hopper2*; Figure 5A).

Unlike *flamenco*, transposons within most piRNA clusters lack an orientation bias. For example, the largest piRNA cluster, at 42AB, contains a high density of randomly oriented, nested transposons and produces uniquely mapping piRNAs from both strands (Figure 5C). Even within this cluster, the strand asymmetry of Piwi complexes is preserved. An interesting illustration is two adjacent *GATE* fragments that are in opposite orientations. Uniquely mapping RNAs in the Ago3 complex correspond to the sense strand of each copy, while Aub, and to some extent Piwi, show the opposite trend (Figure 5C).

Mechanisms of piRNA Biogenesis

To investigate mechanisms of piRNA biogenesis, we examined the relationship between sense and antisense piRNAs. A processing mechanism resembling siRNAs or miRNAs would predict the detection of sense-antisense piRNA pairs that reflect the 2 nt 3′ overhangs produced by RNase III enzymes. According to this scenario, complementary sense and antisense piRNAs should have 5′ ends separated by 23 nt (assuming an average piRNA size of 25 nt) and correspondingly show 23 nt of complementary sequence.

We plotted the distance between each piRNA 5′ end and the 5′ end of its neighbors on the opposite strand. Instead of the expected peak at 23 nucleotides, we found that 5′ ends of complementary piRNAs are most frequently separated by exactly 10 nucleotides (Figure 6A). On the whole, 20% of all piRNAs have a partner whose 5′ end can be mapped ten nucleotides away on the complementary strand. We found the strongest complementary relationships between piRNAs in Ago3 and Aub complexes (Figure 6B). Even though our sequencing efforts are not saturating, more than 48% of small RNAs in the Ago3 library had complementary partners in the Aub library. Statistically significant, although less pronounced, interactions are indicated between Piwi and Ago3, and self complementarity is seen in Aub and Ago3 comparisons.

The 10 nt overlap between sense and antisense piRNAs provoked the hypothesis that the Piwi proteins have a role in piRNA biogenesis. In such a model, an antisense piRNA, derived from a piRNA locus and complexed with Aub or Piwi, would recognize and cleave a sense transposon transcript. This cleavage event would occur opposite nucleotides 10 and 11 of the antisense piRNA, generating a 5′ end 10 nt distant, and on the opposite strand, from the end of the original piRNA (see Figures 6A and 7). The cleaved product would be loaded into a second Piwi family protein, likely Ago3 based upon observed strand biases, ultimately becoming new piRNA after processing at the 3′ end by an unknown mechanism (see Figure 7). piRNAs generated by Piwi-mediated cleavage events are designated as secondary piRNAs.

elements (IR). Color intensities indicate the degree of the strand bias (green: sense; red: antisense; yellow: unbiased). In the lower panel, the cloning frequency for individual transposons in all three complexes is indicated as a heat map. The key for relative cloning frequency is shown at the left.
(B) Shown is the density (sense in green, antisense in red) of all cloned piRNAs assigned to the canonical F element sequence with up to three mismatches. Frequencies in each Piwi-family RNP and in total ovarian RNA are shown individually, as indicated. The relative nucleotide position within the consensus sequence is indicated along the x axes.
(C) Shown is a fragment of the largest piRNA cluster (position 42AB) with only those piRNAs that map uniquely to this region (Ago3 piRNAs in red, Piwi piRNAs in orange, and Aub piRNAs in blue). Top strand piRNAs are shown above the x axes, while bottom strand piRNAs are shown below. Annotated transposon fragments are indicated at the bottom of the panel. Shaded areas mark the boundaries where transposon orientation reverses.

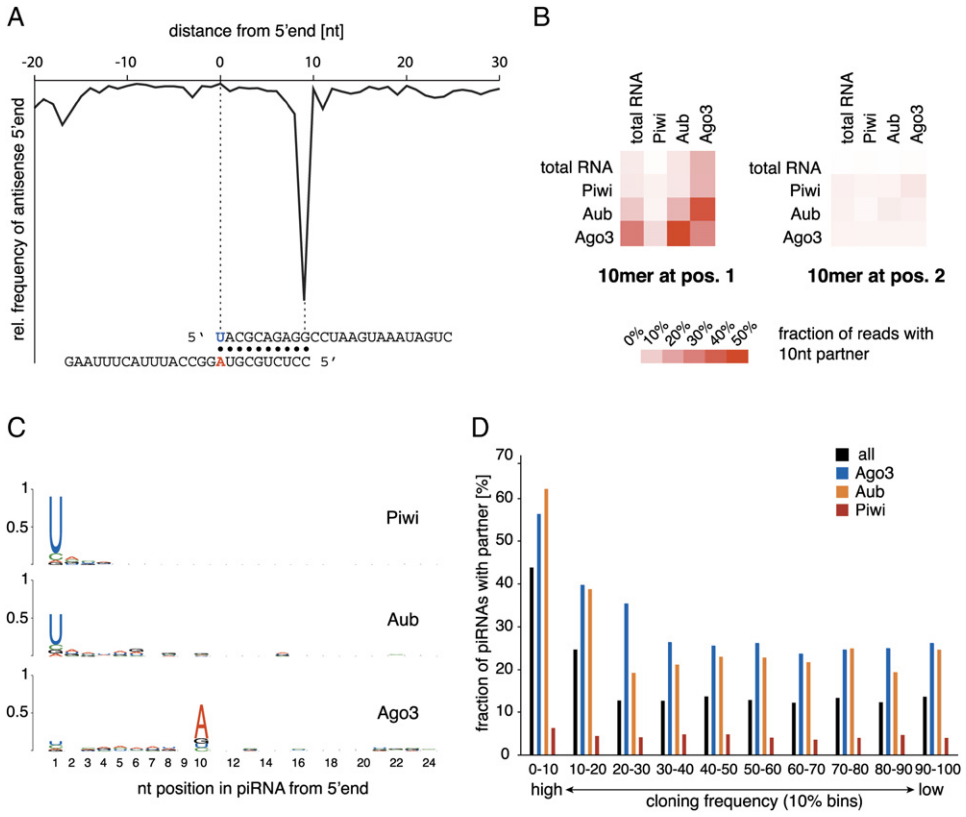


Figure 6. A Slicer-Mediated Amplification Loop for piRNAs

(A) Shown is a frequency map of the separation of piRNAs mapping to opposite genomic strands. The spike at position 9 (the graph starts at 0) indicates the position of maximal probability of finding the 5′ end of a complementary piRNA. One example of this 10 nt offset is shown below.

(B) Shown are heat maps that indicate the degree to which complementary 5′ 10-mers are found in pairwise library comparisons, with a key to the intensity of the signal shown below. The panel to the right represents a control analysis performed with the 10-mer from position 2–11.

(C) The relative nucleotide bias of each position in all piRNAs obtained from Piwi, Aub, and Ago3 complexes, as indicated.

(D) Ten bins were constructed for each Piwi complex (as indicated) and for all sequences combined (all) by sorting piRNAs according to their cloning frequency (e.g., the bin labeled 0–10 contains the 10% of sequences that were most frequently cloned). The fraction of piRNAs within each bin that has a complementary partner was graphed on the y axis.

This model is consistent with the known biochemical properties of Piwi and Argonaute proteins (Lau et al., 2006; Liu et al., 2004; Saito et al., 2006). Moreover, it explains the observed lack of a U bias at the 5′ end of sense-strand piRNAs in Ago3 complexes. If 5′ U-biased antisense piRNAs produce sense piRNAs following cleavage, these should have an unbiased 5′ end. However, sense piRNAs should show an enrichment for A, the complement of the 5′ U, at position 10 (Figure 7). Strikingly, 73% of all Ago3 bound RNAs conform to this prediction, suggesting that the majority of these arise from piRNA-directed cleavage events (Figure 6C).

DISCUSSION

An Amplification Loop that Reinforces Transposon Silencing

In *C. elegans*, effective RNAi depends upon an amplification mechanism (Sijen et al., 2001). Small RNAs from the primary dsRNA trigger are largely dedicated to promoting the use of complementary targets as templates for RNA-dependent RNA polymerases (RdRPs) in the generation of secondary siRNAs. In *Drosophila*, no RdRPs have been identified. However, the ability of Piwi-mediated cleavage to prompt the production of new piRNAs could create an amplification cycle that serves the same purpose as the RdRP-driven secondary siRNA generation systems in worms (Figure 7).

The cycle is initiated by generating primary piRNAs, sampled from the piRNA clusters that we have identified. As these are composed mainly of defective transposon copies, they serve as a genetic memory of transposons to which the population has been exposed. piRNAs that are antisense to expressed, dispersed transposons would identify and cleave their targets, resulting in the genesis of a new, sense piRNA in an Ago3 complex. The Ago3 bound sense piRNA would then seek a target, probably a precursor transcript from a master control locus that contains

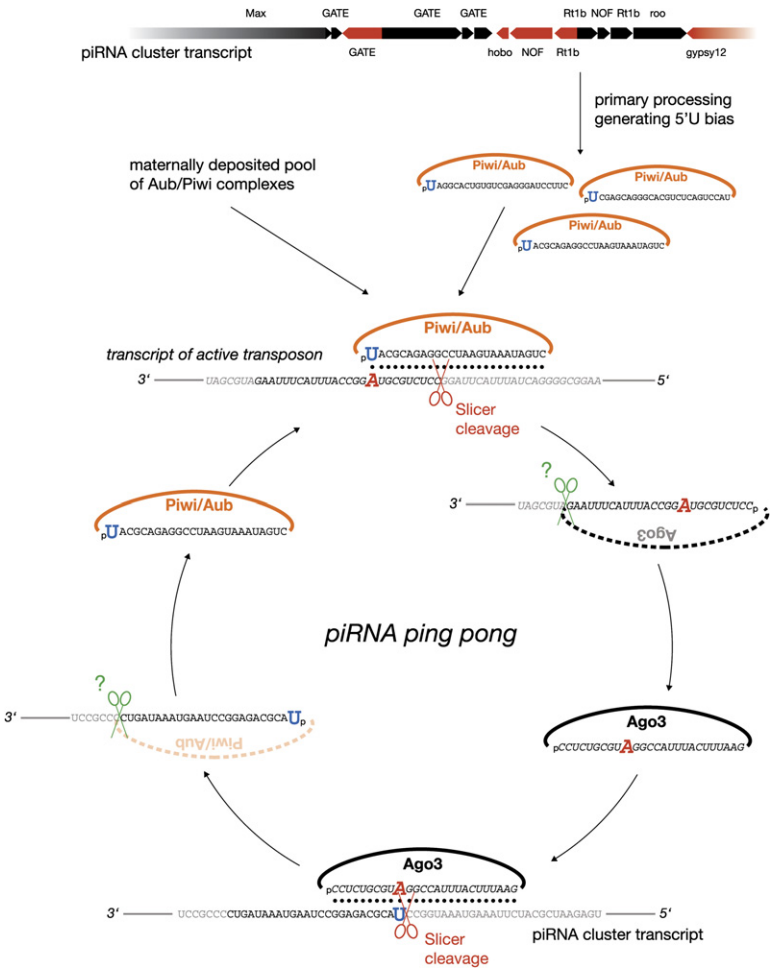


Figure 7. The piRNA Ping-Pong Model

Illustrated is the amplification loop consisting of Piwi/Aub complexes, Ago3 complexes, piRNA cluster transcripts, and transcripts of active transposons. Nucleotide cleavage events are shown as scissors. Potential sources of primary piRNAs are piRNA cluster transcripts and maternally inherited piRNA complexes.

antisense transposon sequences. Ago3-directed cleavage would then generate additional antisense piRNAs capable both of actively silencing their target element and reinforcing the cycle through the creation of additional sense piRNAs.

The existence of such an amplification cycle essentially permits sense and antisense piRNAs act in concert to increase production of silencing-competent RNAs in response to the activity of individual transposons. Since Argonautes act catalytically, a significant amplification of the response could be achieved by even a relatively low level of sense piRNAs in Ago3 complexes. This model predicts that piRNAs participating in this process, namely those with complementary partners, should be more abundant than piRNAs without detectable partners. In accord with this hypothesis, the most frequently cloned Aub and Ago3-associated piRNAs show an increased probability of having partners within the data set (Figure 6D). Interestingly, Piwi-associated RNAs do not follow this pattern. Since the amplification cycle consumes target transposon transcripts as part of its mechanism, post-transcriptional gene silencing may be sufficient to explain transposon repression. However, we cannot rule out the possibility that transcriptional silencing may also be triggered by Piwi-family RNPs.

The amplification cycle may not be absolutely essential for silencing of all elements, as loci such as *flamenco* may operate in somatic follicle cells where the absence of Aub and Ago3 forces it to act in a stoichiometric fashion. In this regard, *flamenco* is unusual in that the vast majority of transposon fragments within this locus exist in a common orientation, which can lead to the production of antisense primary piRNAs given a long, unidirectional, precursor transcript.

Origins of Strand Bias in piRNA Populations

In contrast to *flamenco*, most piRNA loci appear to be both bidirectionally transcribed and contain transposon sequences in random orientation. Nevertheless, the marked asymmetry of Piwi/Aub and Ago3 complexes is conserved in piRNAs that uniquely map to clusters (e.g., Figure 5C). Among piRNAs that match transposons, 77% and 79% of unambiguously cluster-derived Piwi- and Aub-associated RNAs are antisense, while 73% of those in Ago3 are sense. These observations strongly suggest that piRNA clusters themselves participate in

an amplification cycle in a manner that informs strand choice. According to our model, the remarkable strand asymmetry in piRNA populations hinges on informative interactions between master control loci and active transposons, which by their nature produce sense RNAs. Our observations identify Ago3 as the principal recipient of piRNAs derived from transposon mRNAs. Thus, as long as there is an input to the system from active transposon transcripts via Ago3 and a preferential relationship between Ago3 and Aub for generating secondary piRNAs in their reciprocal complexes, a strand bias can be maintained even if primary and secondary piRNAs can both be derived from master control loci.

Biogenesis of Primary piRNAs

The amplification cycle must be initiated by primary piRNAs. Presently, the biogenesis pathway that generates primary piRNAs from piRNA clusters remains obscure. Our data suggest that the piRNA precursor is a long, single-stranded transcript that is cleaved, preferentially at U residues. We detect transcripts from piRNA loci by RT-PCR that encompass multiple transposon fragments (not shown) and find numerous small RNAs that cross junctions between adjacent transposons. In the case of *flamenco*, P element insertions near the proximal end of the locus have a polar effect both on these long RNA transcripts and on *flamenco* piRNAs.

Equally mysterious is the generation of piRNA 3′ ends. Mature piRNAs could arise by two cleavage events and subsequent loading into Piwi complexes. Alternatively, piRNAs could be created following 5′ end formation and incorporation of a long RNA into Piwi by resection of their 3′ ends. The latter model is attractive, since it could provide an explanation for observed size differences between RNAs bound to individual Piwi proteins, as piRNA size would simply reflect the footprint of each Piwi protein.

Although de novo biogenesis mechanisms must exist, maternally inherited piRNA complexes could also serve to initiate the amplification cycle. All three Piwi proteins are loaded into the developing oocyte (Figure S2) (Harris and Macdonald, 2001; Megosh et al., 2006), and Piwi and Aub are concentrated in the pole plasm, which will give rise to the germline of the next generation. Inherited piRNAs could enhance resistance to transposons that are an ongoing challenge to the organism, augmenting zygotic production of primary piRNAs. Indeed, maternally loaded rasiRNAs were detected in early embryos (Aravin et al., 2003), and their presence was correlated with suppression of hybrid dysgenesis in *D. virilis* (Blumenstiel and Hartl, 2005).

A Model for Transposon Silencing in *Drosophila*

Our data point to a comprehensive strategy for transposon repression in *Drosophila* that incorporates both a long-term genetic memory and an acute response to the presence of potentially active elements in the genome. The model that emerges from our studies shows many parallels to adaptive immune systems. The piRNA loci themselves encode a diversity of small RNA fragments that have the potential to recognize invading parasitic genetic elements. Throughout the evolution of *Drosophila* species, a record of transposon exposure may have been preserved by selection for transposition events into these master control loci, as this is one key mechanism through which control over a specific element can be achieved. Evidence has already emerged that X-TAS can act as a transposition hotspot for P elements (Karpen and Spradling, 1992), raising the possibility the piRNA clusters in general may attract transposons. Once an element enters a piRNA locus, it can act, in *trans*, to silencing remaining elements in the genome, either directly through primary piRNAs or by engaging in the amplification model described above. A comparison of *D. melanogaster* piRNAs to transposons present in related *Drosophilids* shows a lack of complementarity when comparisons are made at high stringency. However, when even a few mismatches are permitted, it is clear that piRNA loci might have some limited potential to protect against horizontal transmission of these heterologous elements (Figure S3). The existence of a feed-forward amplification loop can be compared to clonal expansion of immune cells with the appropriate specificity following antigen stimulation, leading to a robust and adaptable response.

EXPERIMENTAL PROCEDURES

Antibodies

Rabbit polyclonal antisera directed against the N-terminal 14–16 AA of Piwi, Aub, and Ago3 (see Figure S1) were raised and affinity purified as described (Denli et al., 2004). Primary antibody dilutions were 1:2000 for western analysis and 1:500 for immunohistochemistry. Actin staining was with Rhodamine-coupled Phalloidin (1:100). Ovaries were dissected on ice (phosphate-buffered saline) and fixed for 20 min in 4% Formaldehyde/PBS/0.1%Triton X-100.

Immunoprecipitation of Piwi RNP Complexes

Ovary extract was prepared in Lysis buffer (20 mM HEPES-NaOH at pH 7.0, 150 mM NaCl, 2.5 mM MgCl2, 250 mM Sucrose, 0.05% NP40, 0.5% Triton X-100, 1x Roche-Complete). Cleared extracts were incubated with primary antibodies (1:50) for 4 hr at 4°C. Antibody complexes were isolated and analyzed for RNA content as described (Aravin et al., 2006).

Small RNA Cloning and Sequencing

Small RNA cloning was performed as described in Pfeffer et al. (2005), and a detailed protocol is available upon request. Sequencing of small RNA libraries was at 454 Life Sciences.

Quantitative RT-PCR

See Supplemental Data for primer sequences and for the amplification of primary cluster transcripts, *gypsy* transcript, and piRNAs.

Bioinformatic Analysis of piRNAs

Sequence extraction and genomic mapping was as described in Girard et al. (2006). We used the Release 5 assembly of the *Drosophila melanogaster* genome and GenBank to identify all piRNAs matching 100% to genomic sequences. The only GenBank entry (L03284) that recovered hits not present in Release 5 corresponds to the tip of the X chromosome, which differs significantly between the sequenced strain and Oregon R, the strain used for our analysis (Abad et al., 2004). Annotation of small RNAs was done using the following

databases: Repbase (http://www.girinst.org/) on the Release 5 assembly, Transposable element canonical sequences (http://www.fruitfly.org/p_disrupt/TE.html), Flybase annotations for coding and noncoding genes (extracted from http://genome.ucsc.edu), and microRNA annotations from Rfam (http://microrna.sanger.ac.uk/sequences). Density analysis of transposons along Release 5 chromosome arms was by counting all nucleotides in a 50 kb window (10 kb offset) annotated as transposons.

piRNA Cluster Analysis

All piRNAs except the 10% of reads corresponding to miRNAs, rRNAs, tRNAs, other ncRNAs, and the sense strand of annotated genes were mapped to Release 5 and the telomeric X-TAS repeat L03284. Nucleotides corresponding to the 5′ end of each piRNA were weighted according to N/M with N = cloning frequency and M = number of genomic mappings. We used a 5 kb sliding window to identify all regions with densities greater than 1 piRNA/kb. Windows within 20 kb of each other were collapsed into clusters. Clusters with at least 5 piRNAs that uniquely matched to the cluster were retained.

Analysis of piRNAs Mapping to Transposable Elements

All piRNAs were matched to the canonical sequences of *Drosophila* transposable elements with high (0 mismatches), medium (3 mm), or low (5 mm) stringency, and the strand relative to the transposon sense strand was determined. We calculated the ratio of all piRNAs per library that match exclusively to the plus or minus strand and excluded those that matched to both (e.g., in inverted repeat [IR] elements). The relative density of piRNAs for each transposon is the number of piRNAs matching a specific element as fraction of all transposon-matching piRNAs.

10 nt Offset Analysis and Nucleotide Bias

The relative relationship of each piRNA to its nearest neighbor was calculated as described in the text, except that all genomic matches were weighted according to the normalized model to avoid skewing by piRNAs that match the genome multiple times. For all subsequent analyses, partners were defined as piRNAs, whose 5′ 10 nt are reverse complements. For Figure 6D we sorted all piRNAs in each library according to their cloning frequency and determined the fraction of piRNAs with partners in bins, each of which contained 10% of all sequences. Position-dependent nt biases for each library were determined by their log-odds score relative to library-specific nt compositions.

Supplemental Data

Supplemental Data include four figures, two table, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://www.cell.com/cgi/content/full/128/6/1089/DC1/.

ACKNOWLEDGMENTS

We thank members of the Hannon laboratory for helpful discussions and support. Ahmet Denli (CSHL) provided antibodies, and Stuart Shuman (MSKCC) provided the expression vector for Rnl2. We thank Andrew Olson and Ted Roeder for bioinformatic support and Stephen Hearn (CSHL) for assistance with confocal microscopy. J.B. and A.S. are supported by fellowships from the Ernst Schering Foundation. A.A. is supported by a fellowship from the CSHL Association. M.D. is an Engelhorn Scholar of the Watson School of Biological Sciences. G.J.H. is an investigator of the Howard Hughes Medical Institute. This work was supported by grants from the N.I.H. (G.J.H.).

Received: October 23, 2006
Revised: December 8, 2006
Accepted: January 19, 2007
Published online: March 8, 2007

REFERENCES

Abad, J.P., De Pablos, B., Osoegawa, K., De Jong, P.J., Martin-Gallardo, A., and Villasante, A. (2004). Genomic analysis of Drosophila melanogaster telomeres: full-length copies of HeT-A and TART elements at telomeres. Mol. Biol. Evol. *21*, 1613–1619.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr. Biol. *11*, 1017–1027.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev. Cell *5*, 337–350.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203–207.

Biemont, C., Ronsseray, S., Anxolabehere, D., Izaabel, H., and Gautier, C. (1990). Localization of P elements, copy number regulation, and cytotype determination in Drosophila melanogaster. Genet. Res. *56*, 3–14.

Blumenstiel, J.P., and Hartl, D.L. (2005). Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. Proc. Natl. Acad. Sci. USA *102*, 15965–15970.

Bregliano, J.C., Picard, G., Bucheton, A., Pelisson, A., Lavige, J.M., and L'Heritier, P. (1980). Hybrid dysgenesis in Drosophila melanogaster. Science *207*, 606–611.

Bucheton, A. (1990). I transposable elements and I-R hybrid dysgenesis in Drosophila. Trends Genet. *6*, 16–21.

Bucheton, A. (1995). The relationship between the flamenco gene and gypsy in Drosophila: how to tame a retrovirus. Trends Genet. *11*, 349–353.

Bucheton, A., Paro, R., Sang, H.M., Pelisson, A., and Finnegan, D.J. (1984). The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell *38*, 153–163.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Castro, J.P., and Carareto, C.M. (2004). Drosophila melanogaster P transposable elements: mechanisms of transposition and regulation. Genetica *121*, 107–118.

Chen, P.Y., Manninga, H., Slanchev, K., Chien, M., Russo, J.J., Ju, J., Sheridan, R., John, B., Marks, D.S., Gaidatzis, D., et al. (2005). The developmental miRNA profiles of zebrafish as determined by small RNA cloning. Genes Dev. *19*, 1288–1293.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev. *12*, 3715–3727.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503–514.

Deng, W., and Lin, H. (2002). miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Dev. Cell *2*, 819–830.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231–235.

Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. Genetics *164*, 501–509.

Engels, W.R., and Preston, C.R. (1979). Hybrid dysgenesis in Drosophila melanogaster: the biology of female and male sterility. Genetics *92*, 161–174.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Grivna, S.T., Pyhtila, B., and Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc. Natl. Acad. Sci. USA *103*, 13415–13420.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Harris, A.N., and Macdonald, P.M. (2001). Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823–2832.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., et al. (2002). Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol. *3*, RESEARCH0085.

Kalmykova, A.I., Klenov, M.S., and Gvozdev, V.A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic Acids Res. *33*, 2052–2059.

Karpen, G., and Spradling, A. (1992). Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. Genetics *132*, 737–753.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. (1977). Hybrid dysgenesis in drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics *86*, 813–833.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839–849.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363–367.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463–2476.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437–1441.

Marin, L., Lehmann, M., Nouaud, D., Izaabel, H., Anxolabehere, D., and Ronsseray, S. (2000). P-Element repression in Drosophila melanogaster by a naturally occurring defective telomeric P copy. Genetics *155*, 1841–1854.

Megosh, H.B., Cox, D.N., Campbell, C., and Lin, H. (2006). The Role of PIWI and the miRNA Machinery in Drosophila Germline Determination. Curr. Biol. *16*, 1884–1894.

Misra, S., and Rio, D.C. (1990). Cytotype control of Drosophila P element transposition: the 66 kd protein is a repressor of transposase activity. Cell *62*, 269–284.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (1997). Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell *90*, 479–490.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol. Cell *9*, 315–327.

Pelisson, A. (1981). The I–R system of hybrid dysgenesis in Drosophila melanogaster: are I factor insertions responsible for the mutator effect of the I–R interaction? Mol. Gen. Genet. *183*, 123–129.

Pelisson, A., and Bregliano, J.C. (1987). Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in Drosophila melanogaster. Mol. Gen. Genet. *207*, 306–313.

Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J. *13*, 4401–4411.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., et al. (2005). Identification of microRNAs of the herpesvirus family. Nat. Methods *2*, 269–276.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697–711.

Reiss, D., Josse, T., Anxolabehere, D., and Ronsseray, S. (2004). aubergine mutations in Drosophila melanogaster impair P cytotype determination by telomeric P elements inserted in heterochromatin. Mol. Genet. Genomics *272*, 336–343.

Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pelisson, A. (2001). Characterization of the flamenco region of the Drosophila melanogaster genome. Genetics *158*, 701–713.

Roche, S.E., and Rio, D.C. (1998). Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the Drosophila Polycomb group gene, Enhancer of zeste. Genetics *149*, 1839–1855.

Ronsseray, S., Lehmann, M., and Anxolabehere, D. (1991). The maternally inherited regulation of P elements in Drosophila melanogaster can be elicited by two P copies at cytological site 1A on the X chromosome. Genetics *129*, 501–512.

Ronsseray, S., Marin, L., Lehmann, M., and Anxolabehere, D. (1998). Repression of hybrid dysgenesis in Drosophila melanogaster by combinations of telomeric P-element reporters and naturally occurring P elements. Genetics *149*, 1857–1866.

Rubin, G.M., Kidwell, M.G., and Bingham, P.M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell *29*, 987–994.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. *20*, 2214–2222.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics *166*, 1313–1321.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. Genes Dev. *20*, 345–354.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465–476.

Smyth, D.R. (1997). Gene silencing: cosuppression at a distance. Curr. Biol. *7*, R793–R795.

Stuart, J.R., Haley, K.J., Swedzinski, D., Lockner, S., Kocian, P.E., Merriman, P.J., and Simmons, M.J. (2002). Telomeric P elements associated with cytotype regulation of the P transposon family in Drosophila melanogaster. Genetics *162*, 1641–1654.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123–132.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320–324.

Accession Numbers

The GenBank accession number for Ago3 in this paper is EF211827. The GEO accession numbers for all piRNAs from this study are GSE6734, GSM154618, GSM154620, GSM154621, and GSM154622.

Cell, Volume *128*

Supplemental Data

Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*

Julius Brennecke, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J. Hannon

Figure S1. ClustalW Alignment of the Three Drosophila Piwi Family Proteins
The Ago3 sequence represents the largest open reading frame in the putative full length cDNA clone RE57814 that we obtained from BDGP (Genbank accession # : EF211827). The N-terminal peptides used for polyclonal antibody production are highlighted in red. PAZ and PIWI domains are boxed in grey and green, respectively. The position of the catalytic DDH residues essential for slicer mediated cleavage are indicated by arrowheads and highlighted in blue. Note, that although Piwi contains a DDK motif, Slicer activity has been demonstrated for this protein (Saito et al., 2006).

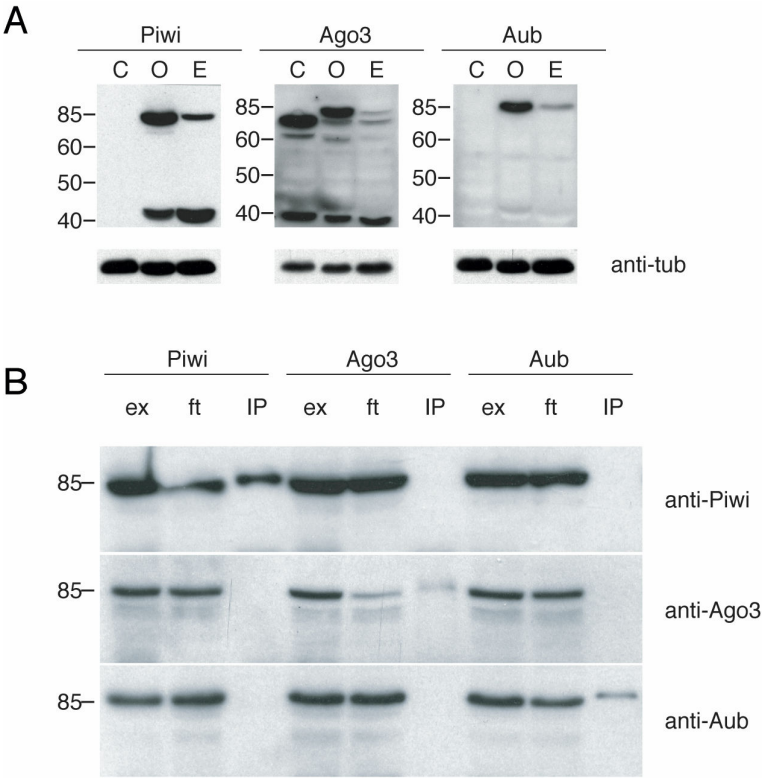


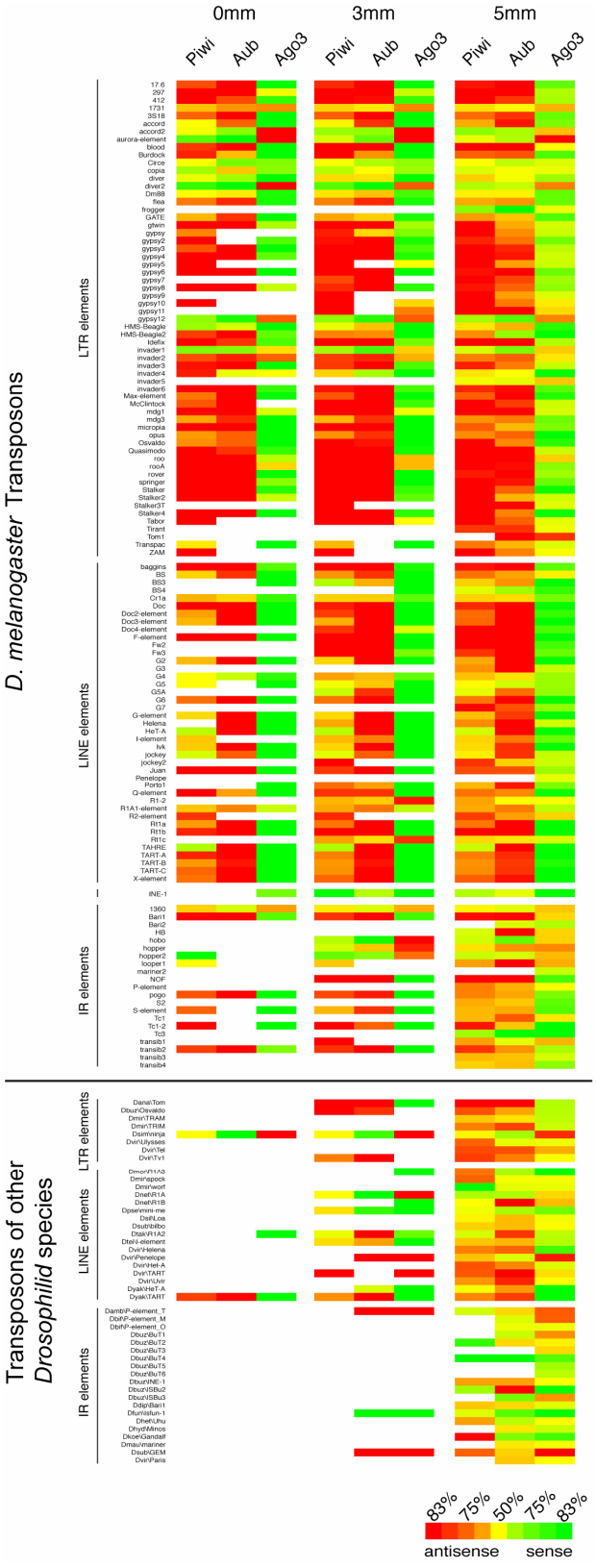
Figure S2. Polyclonal Antibodies Specific for Each Drosophila Piwi Family Member
(A) Western blotting was performed on total protein lysates from female carcasses (C), ovaries (O) and 0-2h embryos (E) using antibodies raised against the N-termini of Piwi, Ago3 and Aub, as indicated in Fig. S1. Besides their specific signal at ~85-90 kDa, the Piwi and Ago3 antibodies recognize additional bands, none of which was enriched in upon immunoprecipitation (not shown) and therefore likley represent western-crossreactive proteins. To control for equal loading, the membrane was reprobed with mouse monoclonal anti-tubulin (tub) from SIGMA.
(B) Western blot analysis was performed on immunoprecipitations prepared with Piwi, Ago3 and Aub specific antibodies from ovary extract. Immunoprecipitates (IP), as well as the total extract (ex) and
supernatant from the immunoprecipitate (ft) were blotted individually with each of the three Piwi family antibodies, as indicated.
In both panels, the positions of protein size markers in kDa, electrophoresed in parallel, are indicated to the left of each panel. We conclude that all three antibodies are specific to their respective proteins and allow specific immuno-precipitation of them. To minimize the potential that some of the staining in ovaries shown in Figure 1 results from crossreactive antigens, we verified that the staining is absent in piwi and aub mutant ovaries, respectively (not shown). No mutant is presently available for Ago3, raising the formal possibility that some component of the staining pattern we observe could be contributed by a cross-reactive species.

Figure S3. Strand Asymmetry of piRNAs mapping to All LTR/LINE/IR Transposons from Drosophila melanogaster and from Related Drosophilid Species

Analysis was performed and data displayed exactly as described in Figure 5A. Included here are all additional transposons, for which we cloned less than 50 piRNAs in total.

Shown are also transposons from related Drosophilid species (all sequences extracted from www.fruitfly.org/p_disrupt/TE.html).

Heat maps were constructed for matches to the consensus sequences at different stringencies (0 mismatches, 3 mismatches and 5 mismatches). Note, that for non-melanogaster transposons, we cloned several piRNAs that match at lower stringencies. Interestingly, in these cases, the pronounced asymmetry between Piwi/Aub and Ago3 is mostly absent.



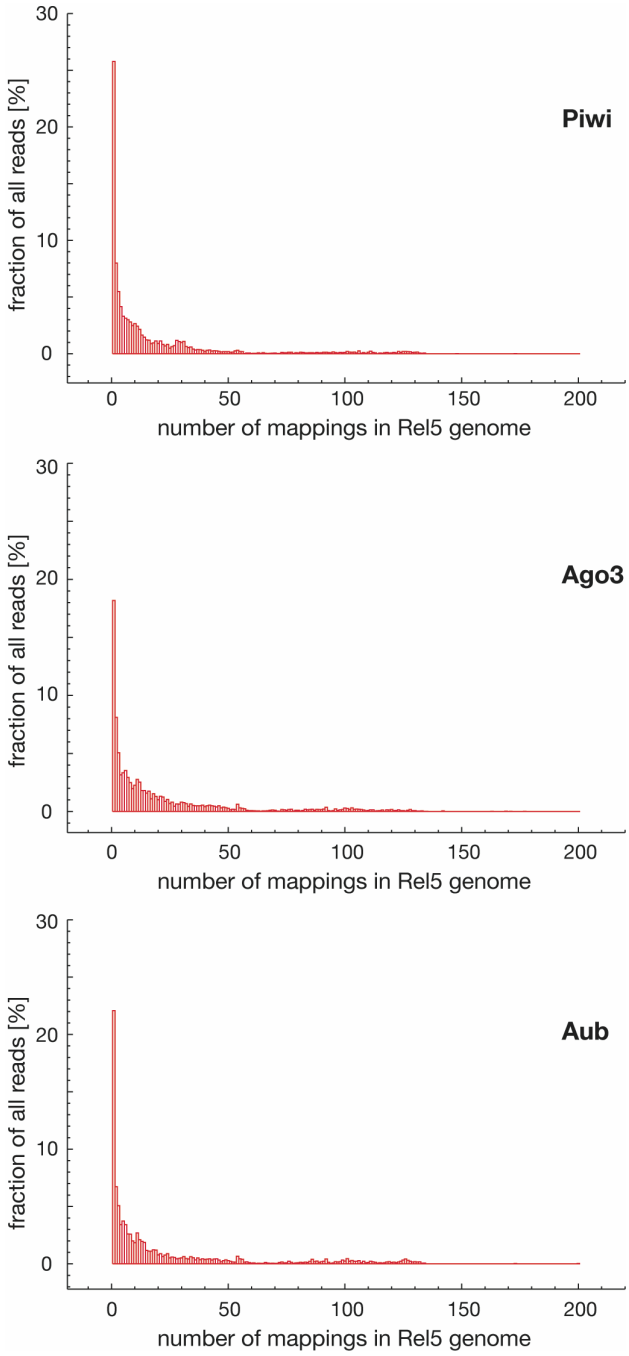


Figure S4. Most Drosophila piRNAs Match the Genome in Multiple Locations
Graphs showing the fraction of piRNAs that map to the Release 5 genomic sequences the number of times indicated on the X-axis. The values are shown for the individual Piwi family proteins Piwi (top), Ago3 (middle) and Aub (bottom).

cluster ID	chromosome	chromosomal band	start	stop	length [kb]	Transposon content (+/- in %)	cluster unique piRNAs	number of potential piRNAs derived from this cluster and fraction of all piRNAs	piRNA strand distribution (+/- strand in %)
1	2R	42AB	2144349	2386719	242	38/32	1686	15102/30.1%	49/51
2	X	20A	21392175	21431907	40	0/78	986	8621/17.2%	100/0
3	4	102E	1258473	1348320	90	6/83	684	2519/5.0%	23/77
4	X-TAS	1A	-	-	7	0/3	484	1306/2.6%	44/56
5	2L	38C	20148259	20227581	79	23/64	482	1851/3.7%	54/46
6	3L	80E-F	23273964	23314199	40	29/37	228	1455/2.9%	64/36
7	U	-	4015849	4029971	14	67/0	176	317/0.6%	62/38
8	X	20A-B	21505666	21684449	179	12/75	170	6649/13.2%	99/1
9	X	20B	21759393	21844063	85	23/55	155	2187/4.4%	63/37
10	U	-	5766708	5772171	5	100/0	133	281/0.6%	54/46
11	3R	100F	27895169	27905030	10	11/4	107	932/1.9%	0/100
12	3LHet	-	1402377	1557939	156	28/39	102	4789/9.5%	51/49
13	3LHet	-	2011004	2180268	169	35/37	86	7062/14.1%	31/69
14	U	-	7542733	7545114	2	100/0	84	149/0.3%	59/41
15	3LHet	-	238123	332969	95	27/46	71	4266/8.5%	43/57

Table S1. Top piRNA Clusters in the Drosophila melanogaster Genome
piRNA clusters are collapsed overlapping windows, which have a normalized piRNA density of at least 1 piRNA/kb and that are supported by at least 5 piRNAs mapping exclusively to the cluster. piRNA clusters were ranked according to the number of cluster-unique piRNAs (column 8). The genomic positions are according to the Release5 assembly (BDGP). X-TAS refers to Genbank entry L03284 and represents X-telomeric TAS repeats present in the OregonR strain but absent in the Celera sequence strain. (Het) refers to unassembled portions of pericentromeric heterochromatin, while (U) refers to heterochromatic contigs that have not been assigned to a chromosome. Positions of piRNA clusters on the polytene chromosome map (column 3) were determined by mapping genomic poitions to the Release 4.3 assembly and extraction of the corresponding cytological band annotation according to the FlyBase genome browser. For unassembled regions, cytological poitions (column 3) could not be determined. Clusters shaded in grey map to telomers, those in orange map to pericentromeric and centromeric heterochromatin. To determine the piRNA strand distribution, only piRNAs which map the genome uniquely were considered.

Additional Supplemental Information

1. Drosophila Strain Differences and Mapping of piRNAs to Heterochromatic Regions

The strain used throughout this study was OregonR, a laboratory wild-type *Drosophila melanogaster* strain. The genomic sequence of *Drosophila* was determined using the isogenized *y; cn bw sp* strain (Adams et al., 2000). As most piRNAs map to transposons and heterochromatic regions of the genome, strain differences potentially impact various aspects of the bioinformatics analysis presented in this study. Nevertheless, 75% of all piRNA sequences match the annotated genome 100% and an additional 14% can be aligned with up to 3 mismatches. With a calculated 454 sequencing error rate of roughly one error in 10 piRNAs (based on sequences matching known microRNAs), we conclude that the strain differences do not prevent a meaningful analysis of the data set. We also note that the Release5 assembly (http://www.fruitfly.org/sequence/release5genomic.shtml), which contains large assembled heterochromatic regions accounts for the origin of most uniquely mapping piRNAs.

For our analysis we exclusively used piRNAs matching the Release5 genome assembly 100%. Excluded from our analysis was the “Uextra” file from Release5, which includes short, un-assembled shotgun reads with low sequence quality and often unverified origin. Less than 10% of the piRNAs that matched the Release5 genome assembly uniquely had additional mappings in Uextra file, supporting the claim that these sequences can be used to unambiguously identify the genomic origins of piRNAs.

The only Genbank sequence, that was evaluated in addition to the Release5 assembly was a ~10kb entry (L03284), which corresponds to the telomeric TAS repeat of the X-chromosome (Karpen and Spradling, 1992). (Abad et al., 2004) have shown that the sequenced strain lacks X-TAS repeats while other strains such as OregonR contain them. We therefore felt justified in including this in our analysis. We find, that ~500 piRNAs match uniquely to this entry and that up to 2.6% of all piRNAs potentially derive from this site.

Large portions of the Release5 assembly comprise uninterrupted contigs. However, in heterochromatic regions, particularly in the file termed “arm_U”, which contains exclusively heterochromatic sequences of unknown chromosomal origin, contigs are often only a few kb long and are interrupted by stretches of 100Ns. These denote the boundaries of definitively assembled sequences. For the identification of piRNA clusters, we did not bridge windows with high piRNA content, if they did not unambiguously arise from the same contig.

2. Supplemental Experimental Procedures

2.1 *Drosophila* Fly Strains Used in This Study

Oregon R was used as a wild type strain; the *piwi*[1] allele is described in (Cox et al., 1998); *aub*[HN] *cn*[1] *bw*[1]/*CyO* and *aub*[QC42] *cn*[1] *bw*[1]/*CyO*, l(2)DTS513[1]) are described in (Wilson et al., 1996); stocks carrying P-element insertions into the *flamenco* region were generated by the BDGP Gene Disruption Project and obtained from Bloomington Stock Center; KG refers to y[1] P{y[+mDint2] w[BR.E.BR]=SUPor-

P}KG00476/FM4 (stock 0016453) BG refers to w[1118] P{w[+mGT]=GT1}BG02658 (stock 0013912).

2.2 Quantification of Individual piRNAs, piRNA Cluster Transcripts and gypsy RNA

cDNA libraries were prepared from 24-29 nt small RNA fractions isolated from 50 µg of total ovarian RNA by purification from 15% denaturing poly acrylamide gels. Libraries from wild-type flies (OreR), *flamenco* heterozygotes (KG/+) and two *flamenco* allelic combinations (BG/KG and KG/KG) were prepared in parallel. Two synthetic RNA oligonucleotides (24 and 28 nt) of known sequence were added to each sample. cDNA libraries were prepared by sequential ligation of 5' and 3' linker followed by reverse transcription as described in (Pfeffer et al., 2005). cDNA pools were amplified with primers that match linker sequences for 20 cycles and resulting pools of PCR products were used for quantitative PCR on individual piRNAs. Individual piRNAs were amplified with specific primers that match the piRNA sequence in the sense or antisense orientation and a universal primer matching the 5' or 3' linker. Each piRNA was quantified using two different primer pairs (5' linker primer/piRNA antisense and 3'linker primer/ piRNA sense) and reactions with each primer set were repeated in duplicate. Therefore, the values obtained represent the average from four reactions that use two different primer pairs. Synthetic oligonucleotides spiked into the samples before cDNA library preparation were quantified and used for normalization of results. For each piRNA we calculated its abundance in ovaries of flamenco heterozygotes and two allelic combinations relative to abundance in wild-type ovaries.

Quantitative RTPCR on piRNA cluster precursor transcripts and gypsy RNA was done according to standard procedures using total ovarian RNA from genotypes indicated in Fig 4.

Primer Sequences

Primer name	**Sequence**
Spiked RNA oligonucleotides	
24 nt	CGUACGGUUUAAACUUCGAAAUGU
28 nt	UAAAAGACGAGUGAGAACUAACAAGGAG
Primers used for qPCR on individual piRNAs	
Universal primer to 5' linker	CGCCATCAGATCGTAGGCACCTGA
Universal primer to 3' linker	CCGCTCAGATTGATGGTGCCTACAG
piRNA A sense	CACTGTACGCAGAGGCCTAAGTAAATAGTC
piRNA A antisense	CACTGGACTATTTACTTAGGCCTCTGCGTA
piRNA B sense	CACTGTGACTGACTCGTGTAGTGTGCACT
piRNA B antisense	CACTGAGTGCACACTACACGAGTCAGTCA
piRNA C sense	CACTGCCCGCCTTATTGAGGTCCCAC
piRNA C antisense	CACTGGGAGCGTGGGACCTCAATAAG
piRNA 1 sense	CACTGTCAACTGCAATGTCTTCAAATGGT

piRNA 1 antisense	CACTGGACCATTTGAAGACATTGCAGTTG
piRNA 2 sense	CACTGTCCACGGTTAGCTGCCTCTCTG
piRNA 2 antisense	CACTGACAGCAGAGAGGCAGCTAACCGT
piRNA 3 sense	CACTGTCAACTAGTATTTCTGGGCTGCCA
piRNA 3 antisense	CACTGATGGCAGCCCAGAAATACTAGTTG
piRNA 4 sense	CACTGTCCACAGTATCGGTTATGCCCTTG
piRNA 4 antisense	CACTGGACAAGGGCATAACCGATACTGTG
piRNA 5 sense	CACTGTACACTGAGCCGTTGATGACTG
piRNA 5 antisense	CACTGAACGCAGTCATCAACGGCTCAG
piRNA 6 sense	CACTGTAAACTTACAGATGCTTCCTGGGT
piRNA 6 antisense	CACTGTACCCAGGAAGCATCTGTAAGTTTA
28 nt oligo antisense	CGCTGCTCCTTGTTAGTTCTCACTCG
24 nt oligo antisense	CGCTGACATTTCGAAGTTTAAACCGTAC

Primers used for qPCR of gypsy transcript

Gypsy forward	CTTCACGTTCTGCGAGCGGTCT
Gypsy reverse	CGCTCGAAGGTTACCAGGTAGGTTC

Primers used for qPCR of precursor transcript from flamenco locus

Flamenco forward	CAGATTACCATTTGGCTATGAGGATCAGAC
Flamenco reverse	TGGTGAAATACCAAAGTCTTGGGTCAAC

References

Abad, J. P., De Pablos, B., Osoegawa, K., De Jong, P. J., Martin-Gallardo, A., and Villasante, A. (2004). Genomic analysis of Drosophila melanogaster telomeres: full-length copies of HeT-A and TART elements at telomeres. Mol Biol Evol *21*, 1613-1619.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F.*, et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

Karpen, G., and Spradling, A. (1992). Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. Genetics *132*, 737-753.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M.*, et al.* (2005). Identification of microRNAs of the herpesvirus family. Nat Methods *2*, 269-276.

Wilson, J. E., Connell, J. E., and Macdonald, P. M. (1996). aubergine enhances oskar translation in the Drosophila ovary. Development *122*, 1631-1639.