

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).

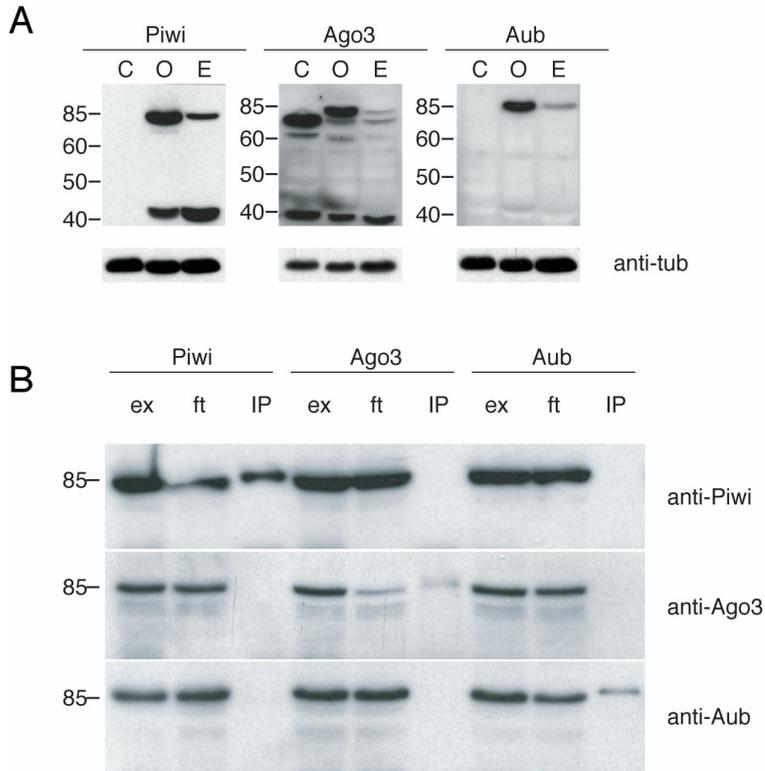


Figure S2. Polyclonal Antibodies Specific for Each *Drosophila* Piwi Family Member

(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 likely represent western-crossreactive proteins. To control for equal loading, the membrane was re-probed 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.

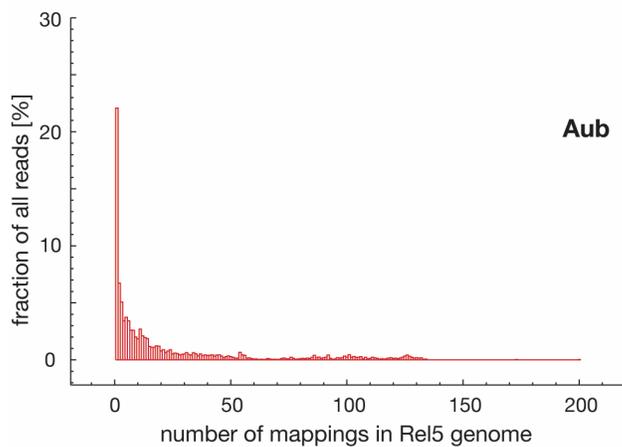
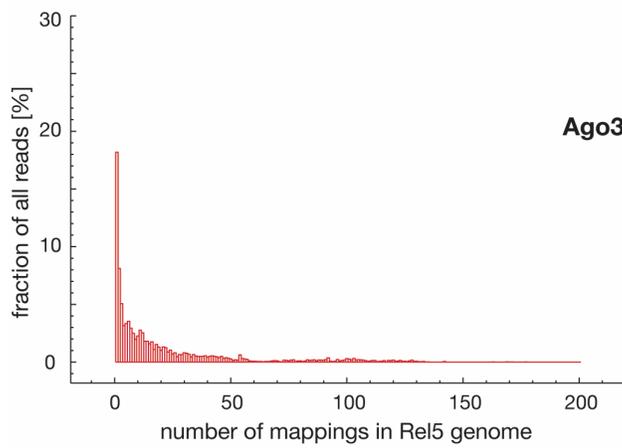
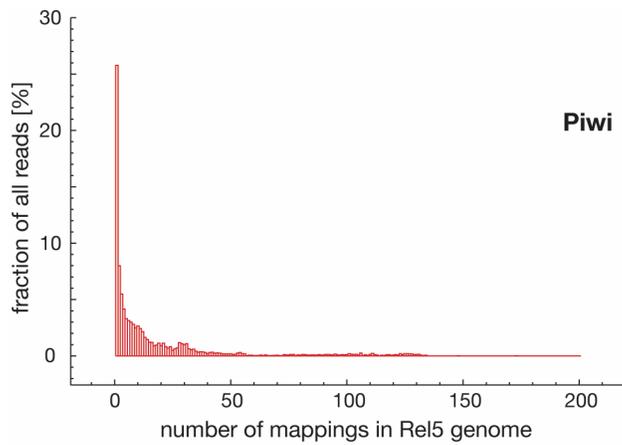


Figure S4. Most *Drosophila* piRNAs Match the Genome in Multiple Locations
 Graphs showing the fraction of piRNAs that map to the Release 5 genomic sequences the number of times indicated on the X-axis. The values are shown for the individual Piwi family proteins Piwi (top), Ago3 (middle) and Aub (bottom).

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 positions to the Release 4.3 assembly and extraction of the corresponding cytological band annotation according to the FlyBase genome browser. For unassembled regions, cytological positions (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	CGCTCGAAGGTTACCAGGTAGGTTCC

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.