SUPPLEMENTARY DATA for

piRNA silencing contributes to interspecies hybrid sterility and reproductive isolation in *Drosophila melanogaster*

Alexei A. Kotov, Vladimir E.Adashev, Baira K. Godneeva, Maria Ninova, Aleksei S. Shatskikh, Sergei S. Bazylev, Alexei A. Aravin* and Ludmila V. Olenina*

*Corresponding authors:

A.A.A.: aaa@caltech.edu;

L.V.O.: olenina_ludmila@mail.ru

This file includes:

Supplementary Figures S2, S4 to S6 with Figure legends

Supplementary Tables S1-S5 and S7.

Supplementary Figures

Figure S2

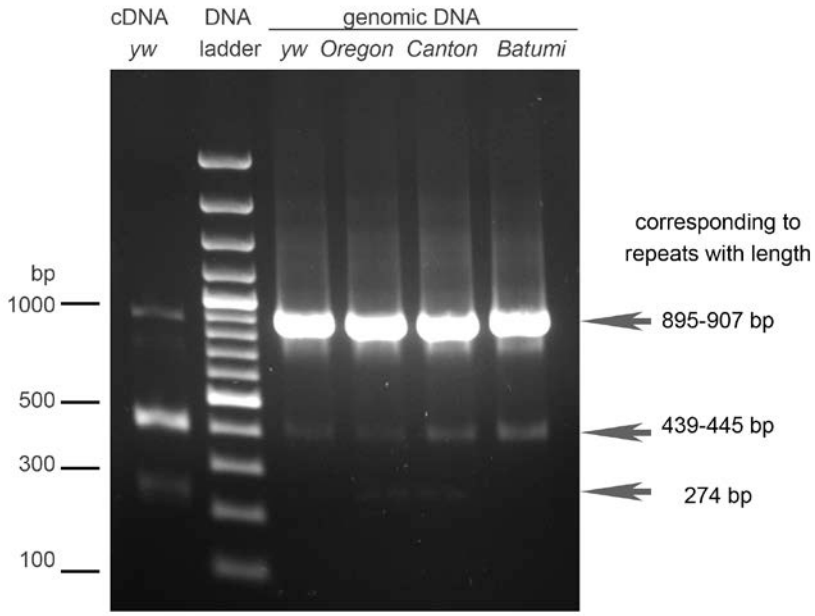


Figure S2. Comparative PCR analysis of *AT-chX* repeats in different strains *Drosophila melanogaster*.

Flies of *yw*, *Oregon R, Canton S* and *Batumi L* strains were used for isolation of genomic DNA. Genomes of all tested fly strains contain *AT-chX* repeats with the lengths of 895-907 bp and 439-445 bp (upper grey arrows). Fragments corresponding to repeats with length 274 bp were amplified from *Oregon R* and *Canton S* genomic DNA preparations in these PCR conditions (lower grey arrow). Repeat #3 presented in the reference genome was not detected, probably because it contains the largest internal insertions of *gr*, *DM297 1* and *TIRANT* sequences that require significantly longer amplification time. Left lane: RT-PCR analysis of transcription of *AT-chX* repeats using cDNA isolated from the testes of *yw* males. Positions of the primers are indicated by black arrowheads on the Figure 1D.

Figure S4

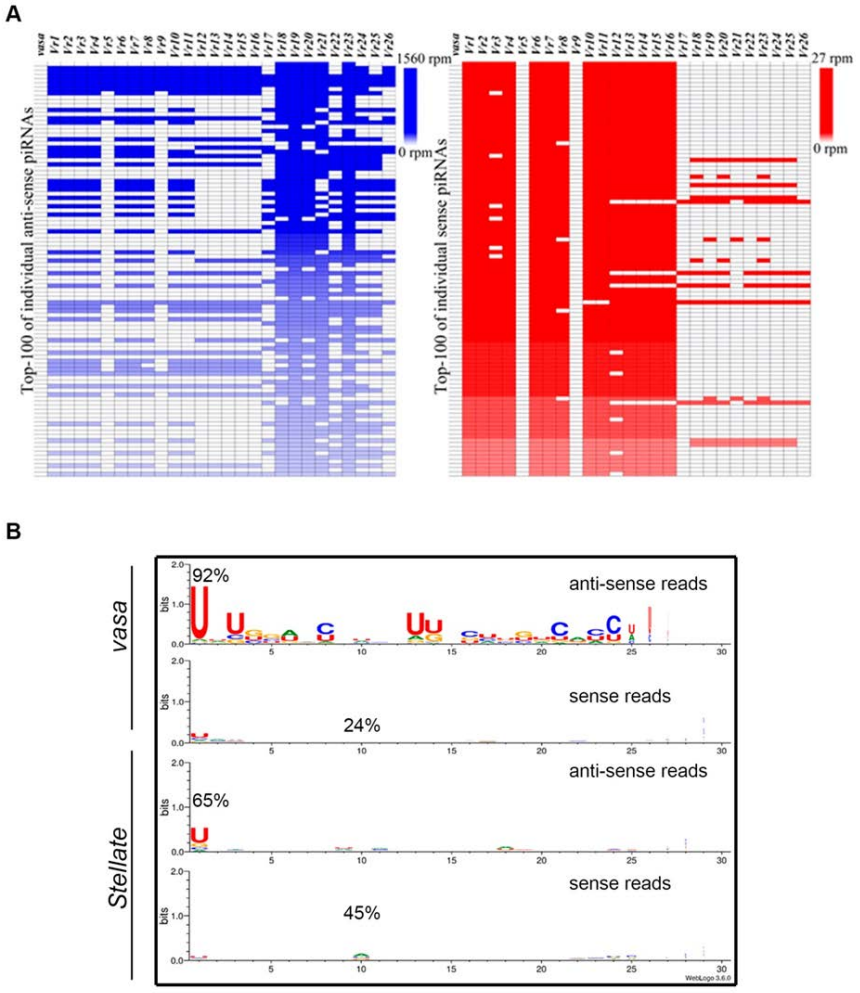
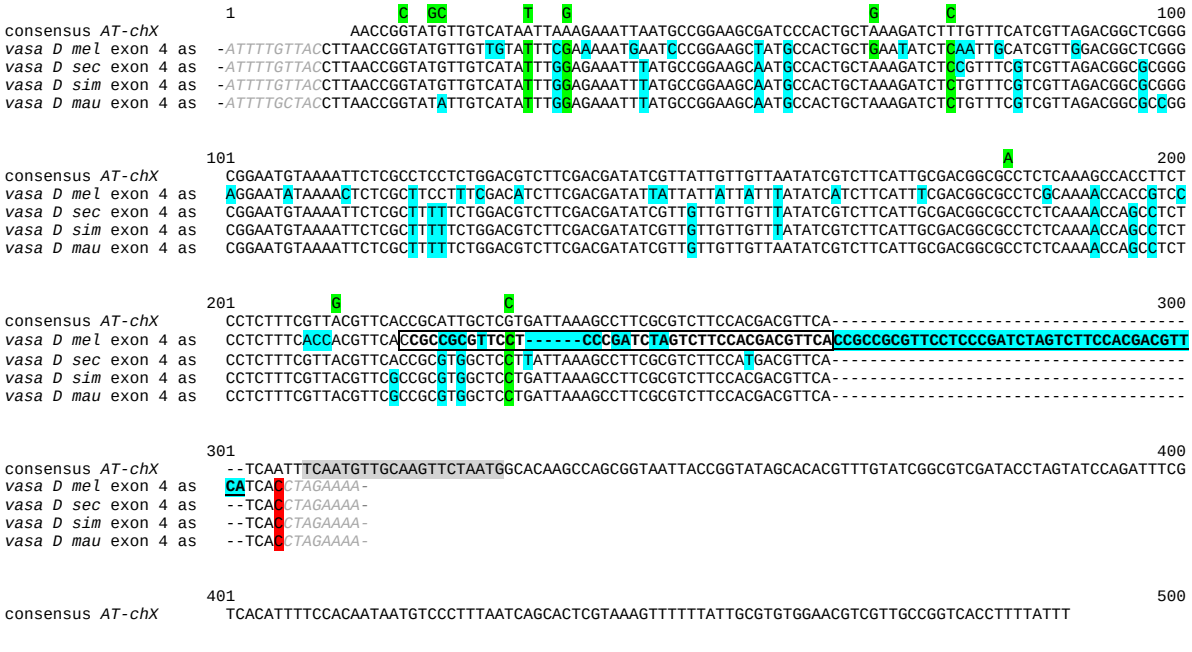


Figure S4. Analysis of piRNAs mapped to *AT-chX* repeats, *vasa* and *Stellate.*
(**A**) Analysis of piRNA (23-29 nt) distribution from *yw* testis library of small RNAs across all *AT-chX* repeats permitting 0 mm. The analysis allowed to reveal features of two main groups of repeats. Top-100 of the most abundant individual antisense (blue, left part) and sense (red, right part) piRNAs are represented here as heap maps. A fraction of antisense piRNAs is mapped to all repeats (12.5%), while predominant cohort of antisense *AT-chX* piRNAs (46.9%) is mapped only to repeats of the second group (##18-26); only 0.5% of piRNAs is mapped to the first group members. At the same time, a bulk of sense piRNAs (86.6%) is perfectly mapped to the repeats of the first group (##1-4, 6-8, 10-16). It srongly supports the hypothesis that the expression of sense and antisense piRNA precursors occurs from different groups of repeats in the germline of *D.melanogaster*. The most abundant *AT-chX* piRNAs are not mapped to *D.melanogaster vasa* without mismatches (leftward columns of both maps). (**B**) Properties of piRNAs mapped to *vasa* and *Stellate* with 0-3 mm. Ping-pong signature determination for antisense and sense piRNAs mapped to *vasa* (upper) and *Stellate* transcripts (bottom). U1 (for antisense reads) and A10 (for sense reads) biases are shown in percent.

Figure S5

A



B

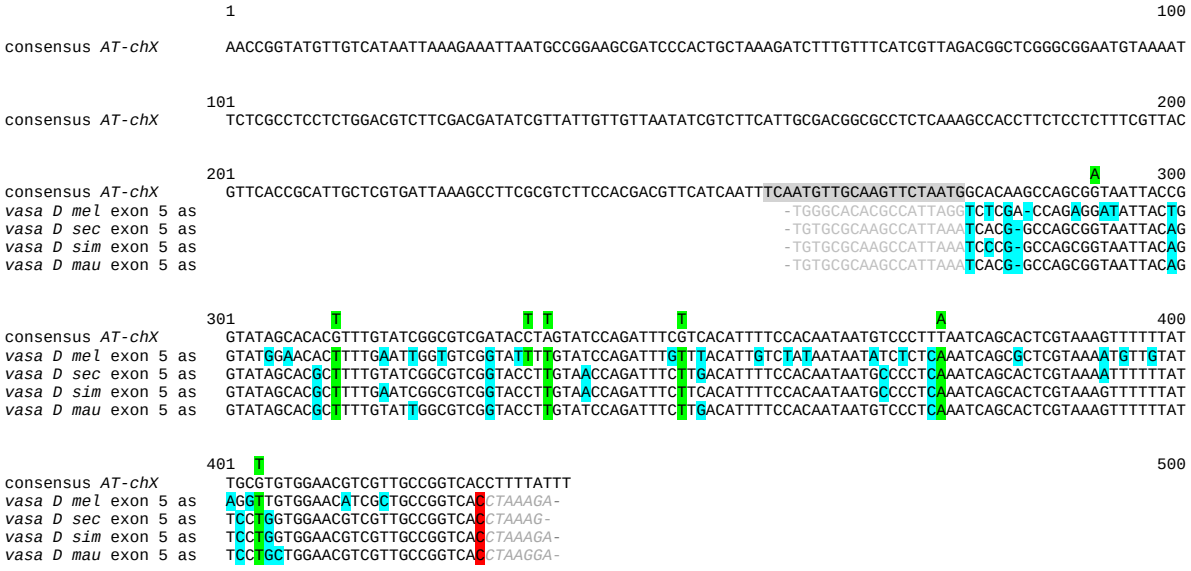


Figure S5. Multiple alignment of the *AT-chX* consensus with *vasa* fourth and fifth exons (in antisense orientation) of *D. melanogaster* and sibling species *D. simulans*, *D. sechellia* and *D. mauritiana*.

Alignment was performed using Clustal W algorithm with subsequent manual adjustment. (**A**) Alignment of the *AT-chX* consensus with *vasa* fourth exons (in antisense orientation relative to the coding *vasa* sequence). The first nucleotide positions of *vasa* fourth exons are marked by red. Fragments of flanking intron sequences are shown by gray italic letters. Nucleotide substitutions in the *vasa* sequences which distinguish it from the *Vr-ex4* part of consensus are marked by cyan or green color. Green color indicates nucleotides that are not included in the consensus, but are often

found in *AT-chX* repeats (they are also indicated above the corresponding positions of the consensus). The simple repeat spacer that connects the *Vr-ex4* and *Vr-ex5* parts is shown in light grey bar. The start of fourth exon of *D. melanogaster vasa* bears twice repeated 39 nt sequences (both in bold letters, one repeat shown in the box, and the other one is underlined). This region does not perfectly match to the *Vr-ex4* part of consensus. (**B**) Alignment of the *AT-chX* consensus with *vasa* fifth exons. Designations are as in (A). Downstream simple repeat included in light grey bar. The *Vr-ex5* part of the consensus maintains homology with starting fragments of *vasa* fifth exons. The first nucleotides of *vasa* fifth exons are marked by red.

Figure S6

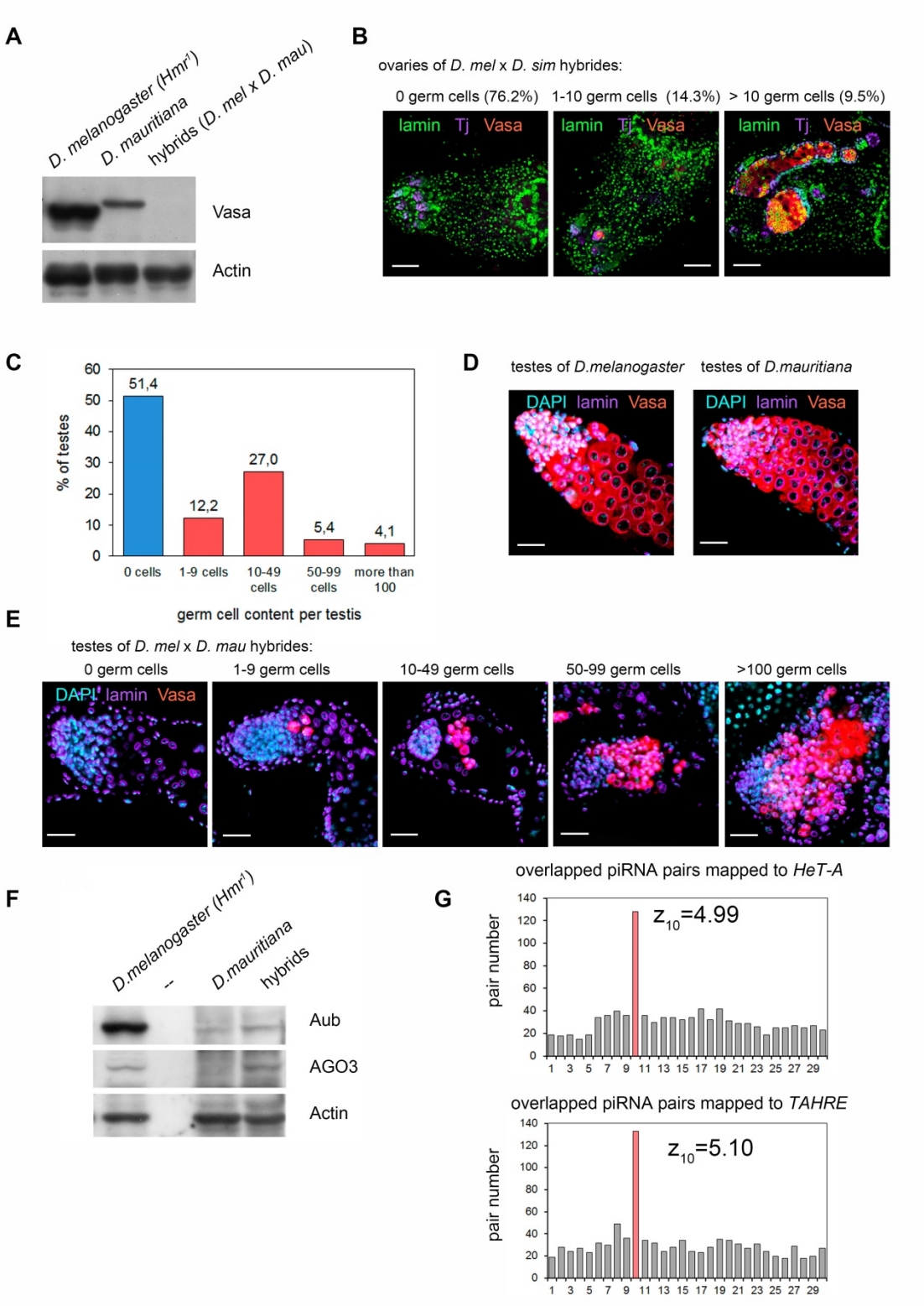


Figure S6. Analysis of germline content in the ovaries and testes of interspecies hybrids.

(**A**) Expression of *D. melanogaster vasa* (marker of germ cells) was detected in the ovaries of parent species and hybrid *Hmr*^{1} *D. melanogaster* × *D. mauritiana* females by RT-qPCR. Error bars represent standard errors of mean. The level of *vasa* expression in hybrid

ovaries is close to the detection limit. (**B**) Confocal images of *Hmr*^{1} *D. melanogaster* × *D. simulans* hybrid ovaries. Ovaries were stained with anti-Vasa (red), anti-lamin (green) and anti-Tj (violet) antibodies. Scale bars are 50 μm. More than 75% hybrid ovaries do not contain germ cells. (**C**) Bar diagram presents the distribution of the hybrid testes with different numbers of germ cells. (**D**) Confocal images of apical ends of *Hmr*^{1} *D. melanogaster* and *D. mauritiana* testes. Testes were stained with anti-Vasa (red) and anti-lamin (violet) antibodies, chromatin was stained by DAPI. Scale bars are 30 μm. (**E**) Confocal slices of apical ends of hybrid testes with different germ cell content. Testes were stained as (B). Scale bars are 30 μm. (**F**) Expression of Aub and AGO3 was detected in the testes of parental species and of hybrid males by Western blotting. Antibodies to Actin were used for loading control. (**G**) Analysis of overlap between 5'-ends of sense and anti-sense piRNAs mapped to germline-specific transposons *HeT-A* and *TAHRE* in the hybrid testis library with corresponding z_{10} scores. The bias for 10 nt overlap (z_{10} score value) indicates piRNA pairs generated through the ping-pong mechanism.

Supplementary tables

Table S1

Library ID	Description	Number of reads mapped to the genome of *D. melanogaster* or/and *D. mauririana*	Reads length
tbl_kotov_dm6_1_mappers - P192_3_mult	Total small RNA library (19-29 nt) from testes of adult *yw* flies (0-3 days). Expression profiling by high throughput sequencing. Illumina HiSeq 4000. Barcode: TTAGGC	4 344 413	1 x 50 bp mode
Unique name: BG2_Index9 Long name: BG2_Index9_HmrI_merged	Total small RNA library (19-29 nt) from testes of adult *D. melanogaster hmr*^{1} flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: CTGATC	3 359 760	1 x 50 bp mode
Unique name: BG3_Index44 Long name: BG3_Index44_D_maur_merged	Total small RNA library (19-29 nt) from testes of adult *D. mauritiana* flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: ATTATA	3 872 451	1 x 50 bp mode
Unique name: BG1_Index11 Long name: BG1_Index11_HmrIxD.maur_hybrid_merged	Total small RNA library (19-29 nt) from testes of hybrids of *D. melanogaster hmr*^{1} and *D. mauritiana* flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: GTAGCC	4 735 699	1 x 50 bp mode
Unique name: BG14 Long name: BG14_HMR1_repl1_testis	Ribo-depleted RNA-seq library from testes of adult *D. melanogaster hmr*^{1} flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: GGAACT	13 038 725	1 x 50 bp mode
Unique name: BG15 Long name: BG15_D.maur_repl1_testis	Ribo-depleted RNA-seq library from testes of adult *D. mauritiana* flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: TGACAT	13 504 376	1 x 50 bp mode
Unique name: BG13 Long name: BG13_HMR1xD.maur_repl1_testis	Ribo-depleted RNA-seq library from testes of hybrids of *D.melanogaster hmr*^{1} and *D. mauritiana* flies (0-5 days). Expression profiling by high throughput sequencing. Illumina HiSeq2500. Barcode: TTGACT	14 533 279	1 x 50 bp mode

Table S1. Short descriptions of RNA libraries which were prepared and analyzed in this study.

Table S2

Number of repeat	Genomic coordinates	Length, bp	*Vr-ex4*, bp	*Vr-ex5*, bp	Properties of repeat
1	chrX:21,818,956-21,819,862	907	247	159	insertions of *gr* and *TIRANT*
2	chrX:21,889,652-21,890,558	907	247	159	insertions of *gr* and *TIRANT*
3	chrX:21,964,841-21,972,749	7909	247	159	insertions of *gr*, *DM297 1* and *TIRANT*
4	chrX:21,996,645-21,997,551	907	247	159	insertions of *gr* and *TIRANT*
5	chrX:22,021,080-22,021,186	107	95	-	first 107 nt of *vasa* fourth exon homologous region
6	chrX:22,024,364-22,025,270	907	247	159	insertions of *gr* and *TIRANT*
7	chrX:22,027,873-22,028,779	907	247	159	insertions of *gr* and *TIRANT*
8	chrX:22,031,382-22,032,288	907	247	159	insertions of *gr* and *TIRANT*
9	chrX:22,057,508-22,057,614	107	95	-	first 107 nt of *vasa* fourth exon homologous region
10	chrX:22,060,790-22,061,696	907	247	159	insertions of *gr* and *TIRANT*
11	chrX:22,064,263-22,065,169	907	247	159	insertions of *gr* and *TIRANT*
12	chrX:22,095,312-22,096,206	895	247	159	insertions of *gr* and *TIRANT*
13	chrX:22,098,809-22,099,715	907	247	159	insertions of *gr* and *TIRANT*
14	chrX:22,102,318-22,103,224	907	247	159	insertions of *gr* and *TIRANT*
15	chrX:22,105,827-22,106,733	907	247	159	insertions of *gr* and *TIRANT*
16	chrX:22,109,334-22,110,240	907	247	159	insertions of *gr* and *TIRANT*
17	chrX:22,137,556-22,137,994	439	233	158	flanked by *Stalker2* at the left, greatly disrupted
18	chrX:22,275,723-22,276,167	445	251	159	without internal transposon insertions
19	chrX:22,276,843-22,277,282	440	251	154	without internal transposon insertions
20	chrX:22,278,970-22,279,414	445	251	159	without internal transposon insertions
21	chrX:22,280,435-22,280,874	440	251	154	without internal transposon insertions
22	chrX:22,315,516-22,315,789	274	251	23	without internal transposon insertions; contains homologous region of *vasa* fourth exon and 23 nt sequence of fifth exon
23	chrX:22,355,554-22,355,998	445	251	159	without internal transposon insertions

24	chrX:22,356,675-22,356,948	274	251	23	without internal transposon insertions; contains homologous region of *vasa* fourth exon and 23 nt sequence of fifth exon
25	chrX:22,364,143-22,364,416	274	251	23	without internal transposon insertions; contains homologous region of *vasa* fourth exon and 23 nt sequence of fifth exon
26	chrX:22,371,776-22,371,931	156	156	-	without internal transposon insertions; first 156 nt of *vasa* fourth exon homologous region; flanked by *R1 DM* at the right

Table S2. The list of *AT-chX* repeats with their genomic coordinates. The length of each repeat without flanked *baggins1* fragments is indicated. The lengths of regions that are homologous to vasa exons, *Vr-ex4* and *Vr-ex5*, are indicated.

Table S3

vasa exons from species	*Vr-ex4* part		*Vr-ex5* part		Whole *AT-chX*	
	Identical positions	% identity	Identical positions	% identity	Identical positions	% identity
D. melanogaster	196	76.56	121	76.10	317	76.39
D. simulans	235	91.80	141	88.68	376	90.60
D. sechellia	233	91.02	141	88.68	374	90.12
D. mauritiana	235	91.80	142	89.31	377	90.84
D. yakuba	209	81.64	111	69.81	320	77.11
D. erecta	215	83.98	109	68.55	324	78.07
D. virilis	143	55.86	105	66.04	248	59.76
D. ananassae	175	68.36	92	57.86	267	64.34

Table S3. Determination of sequence identity between *AT-chX* consensus and *vasa* genes of different *Drosophila* species. Light grey color is marked three species with most homology.

Table S4

Overlap, nt	Pair number	z-score	Overlap probability
1	467	-0.88	0.00
2	539	-0.72	0.00
3	662	-0.44	0.01
4	594	-0.59	0.01
5	669	-0.42	0.01
6	1050	0.45	0.01
7	1330	1.09	0.02
8	1303	1.03	0.01
9	1707	1.95	0.05
10	2148	2.96	0.59
11	1279	0.97	0.02
12	1046	0.44	0.02
13	1020	0.38	0.02
14	1264	0.94	0.01
15	901	0.11	0.01
16	1522	1.53	0.02
17	1123	0.62	0.01
18	786	-0.15	0.01
19	765	-0.20	0.01
20	710	-0.33	0.01
21	478	-0.86	0.01
22	451	-0.92	0.01
23	495	-0.82	0.02
24	632	-0.50	0.01
25	505	-0.80	0.04
26	525	-0.75	0.01
27	387	-1.06	0.01
28	414	-1.00	0.01
29	410	-1.01	0.01
30	402	-1.03	0.00

Table S4. Analysis of relative distributions of antisense and sense *AT-chX* piRNA pairs whose 5'-ends are overlapped by indicated nucleotide number. Z-scores and overlap probabilities were determined for each overlap sequence using signature.py script (31).

Table S5

	piRNA orientation	0 mm	1 mm	2 mm	3 mm	Total
AT-chX	sense	410	120	36	1	567
	antisense	11974	11822	3121	25	26942
Su(Ste)	sense	1530	1284	862	11	3687
	antisense	29556	13050	2531	28	45166
vasa	sense	175	37	19	0	231
	antisense	3	5	209	12	229
Stellate	sense	881	835	741	8	2465
	antisense	2179	5196	4374	71	11820

Table S5. Abundance of piRNAs (reads per million, rpm) in the testes of *D. melanogaster* that are mapped to the consensus sequences of *AT-chX* and *Su(Ste)* as well as to the *vasa* and *Stellate* sequences with specified number of mismatches (0-3).

Table S7.

Small RNA library	piRNAs mapped to *AT-chX* consensus		piRNAs mapped to *D. melanogaster vasa*		piRNAs mapped to *D. mauritiana vasa*	
	antisense, rpm	sense, rpm	antisense, rpm	sense, rpm	antisense, rpm	sense, rpm
Testes of *D. mel* × *D. mau* hybrids	294.99	13.5	0	2.75	58.7	4.86
Testes of *D. melanogaster Hmr*[1]	6465.34	217.28	0.89	78.87	1728.10	12.20
Testes of *D. mauritiana*	14.46	0.52	0	0.52	3.62	0.52

Table S7. Amounts of *AT-chX* and *vasa* piRNAs in the small RNA libraries of testes of parent species and interspecies hybrids. The numbers (rpm) of piRNAs that are mapped to *AT-chX* consensus and *vasa* sequences of *D. melanogaster* and *D. mauritiana* with no mismatches are shown.