

# *pi*RNA and Transposon Dynamics in *Drosophila*: A Female Story

Bastien Saint-Leandre, Pierre Capy, Aurelie Hua-Van <sup>†</sup> and Jonathan Filée<sup>†</sup>

Laboratoire Evolution, Génomes, Comportement, Ecologie CNRS, Université Paris-Sud, IRD, Université Paris-Saclay, Gif-sur-Yvette, France

Corresponding author: E-mail: basai@sas.upenn.edu.

<sup>†</sup>These authors contributed equally to this work.

Accepted: 6 May 2020

The germlines of metazoans contain transposable elements (TEs) causing genetic instability and affecting fitness. To protect the germline from TE activity, gonads of metazoans produce TE-derived *PIWI-interacting* RNAs (*pi*RNAs) that silence TE expression. In *Drosophila*, our understanding of *pi*RNA biogenesis is mainly based on studies of the *Drosophila melanogaster* female germline. However, it is not known whether *pi*RNA functions are also important in the male germline or whether and how *pi*RNAs are affected by the global genomic context. To address these questions, we compared genome sequences, transcriptomes, and small RNA libraries extracted from entire testes and ovaries of two sister species: *D. melanogaster* and *Drosophila simulans*. We found that most TE-derived *pi*RNAs were produced in ovaries and that *pi*RNA pathway genes were strongly overexpressed in ovaries compared with testes, indicating that the silencing of TEs by the *pi*RNA pathway mainly took place in the female germline. To study the relationship between host *pi*RNAs and TE landscape, we analyzed TE genomic features and how they correlate with *pi*RNA production in the two species. In *D. melanogaster*, we found that TE-derived *pi*RNAs target recently active TEs. In contrast, although *Drosophila simulans* TEs do not display any features of recent activity, the host still intensively produced silencing *pi*RNAs targeting old TE relics. Together, our results show that the *pi*RNA silencing response mainly takes place in *Drosophila* ovaries and indicate that the host *pi*RNA response is implemented following a burst of TE activity and could persist long after the extinction of active TE families.

**Key words:** transposable elements, germline, *pi*RNA, *Drosophila melanogaster*.

## Introduction

In sexually reproducing organisms, germline cells transmit genetic information from generation to generation. The maintenance of genome integrity in these cells is crucial in ensuring the progeny an optimal fitness. Transposable elements (TEs) are selfish genetic elements that have the ability to insert at any genomic location, thus constituting an important source of genetic variability and instability within the germline. In rare cases, the host can take advantage of beneficial TE insertions to establish new genetic functions (Jangam et al. 2017). However, evolutionary trajectories of TEs also rely on negative selective pressures acting against deleterious insertions (Petrov et al. 2003; Le Rouzic and Decelie 2005; Dolgin and Charlesworth 2008). Indeed, the germline deploys important genetic and epigenetic resources to silence TEs and limit their harmful consequences on host genomes.

Conserved across metazoans, the *PIWI-interacting* RNA (*pi*RNA) pathway is a germline-specific mechanism that plays a predominant role in restricting TE propagation (Lau et al.

2006; Houwing et al. 2007; Kawaoka et al. 2009; Robine et al. 2009). This small RNA-based mechanism involves members of the PIWI family proteins that can bind *pi*RNAs (23–29 nt) and act as transcriptional and posttranscriptional silencers of the TE expression (Brennecke et al. 2007; Czech et al. 2018). Over the two last decades, considerable efforts have been carried out to understand the molecular basis of the *pi*RNA pathway. In *Drosophila melanogaster*, it has been shown that a discrete number of genomic loci, called the *pi*RNA clusters, are dedicated to the production of *pi*RNAs. From these *loci*, which represent <3% of the total genome, hundreds of thousands of different *pi*RNAs are produced and most of them derive from TEs themselves (up to 90% for some clusters, Brennecke et al. 2007). RNA precursors are produced from *pi*RNA clusters and are processed into *pi*RNAs serving as guides to target TE *mRNAs*. Proteins of the PIWI family load *pi*RNAs to mediate both the recognition of complementary TE-derived transcripts and their slicing into small RNAs. Depending on the nature of the *pi*RNA clusters

and the PIWI-interacting proteins, populations of *pi*RNAs can eventually feed a secondary amplification process called the *ping-pong* amplification loop. This process leads to massive production of *pi*RNAs against a specific subset of active TE families (Brennecke et al. 2007; Gunawardane et al. 2007; Mohn et al. 2014).

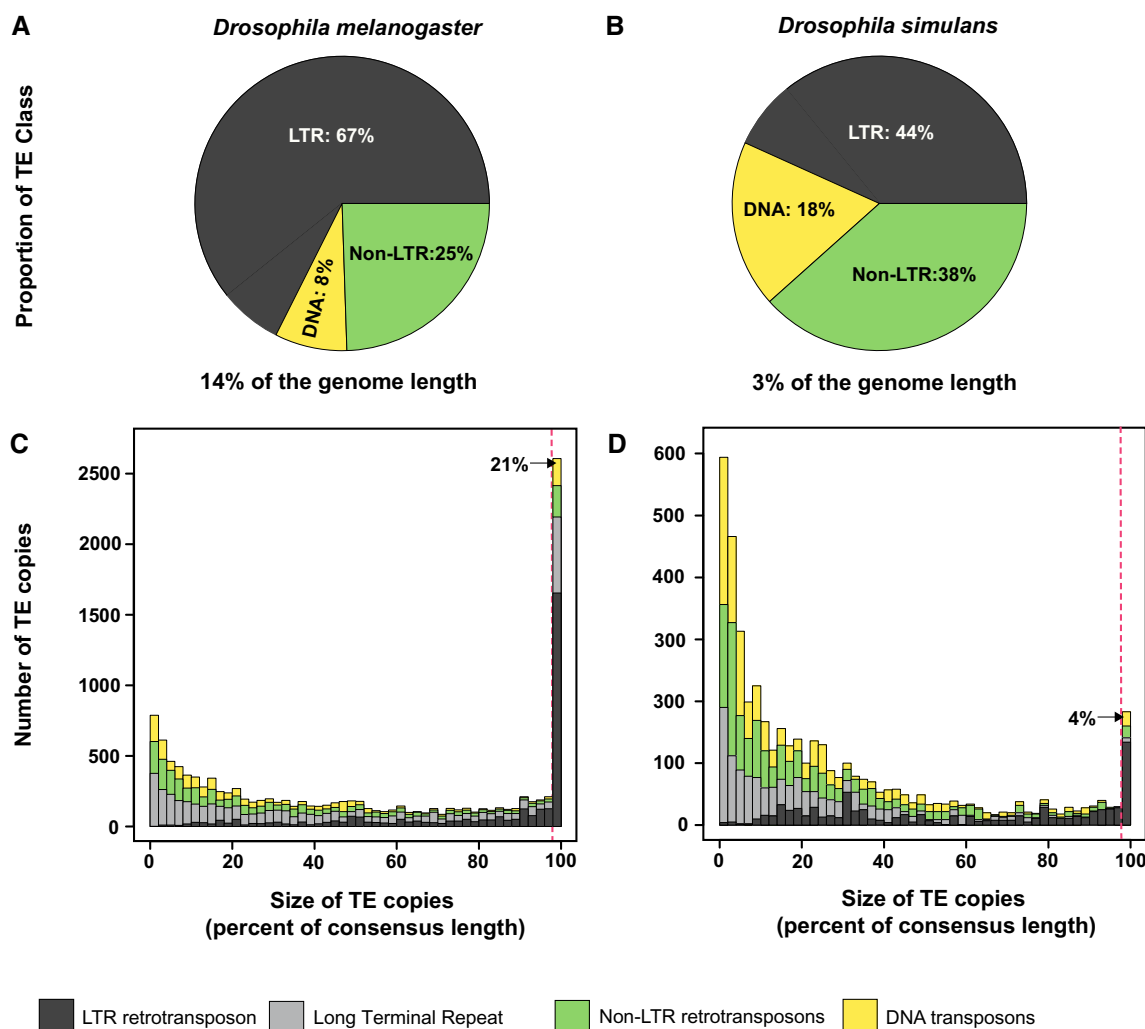
Although models of *pi*RNA biogenesis have been extensively studied from *D. melanogaster* ovaries, they remain poorly studied in the male germline. However, notable differences have been already observed between the *pi*RNA pathway functions in male and female germline. For instance, the PIWI family proteins *Argonaute3* (*Ago3*) and *Aubergine* (*Aub*), known to be essential for the *ping-pong* amplification cycle, display contrasting patterns of expression and cellular localization between the two germlines. On one hand, although *Ago3* protein expression is observed at almost all developmental stages in female germline cells, its expression in testes is restricted to very early stages of spermatogenesis (up to the first four mitotic divisions, Nagao et al. 2010). On the other hand, *Aub* proteins are predominantly associated with TE-derived *pi*RNAs in ovaries, whereas in testes <7% of the *pi*RNAs associated with *Aub* are derived from TEs (Nagao et al. 2010). Indeed, in testes, *Aub* are mainly associated with *pi*RNAs derived from specific Y and X chromosome repeats, but not with transposons (Nishida et al. 2007). Moreover, population analyses in *Drosophila simulans* revealed a general transcriptional bias of both *ago3* and *aub* gene expressions in ovaries compared with testes (Saint-Leandre et al. 2017).

Regardless of lack of understanding in testes, *pi*RNA pathway functions are assumed to serve as the main genome-defense mechanism against new invading TE families. Indeed, the *pi*RNA pathway is regularly compared to an immune system, due to its ability to promptly identify and stop the proliferation of new invading TEs. A number of studies relative to the *P* DNA transposon and the *I* non-long terminal repeat (LTR) retrotransposon have demonstrated that the acquisition of new TE lineages in natural populations is followed by the *de novo* production of their corresponding *pi*RNAs (Brennecke et al. 2008; Chambeyron et al. 2008; Khurana et al. 2011; Grentzinger et al. 2012). Comparative studies of *D. melanogaster* and its sister species *D. simulans* have shown that the expression of TE-derived *pi*RNAs between populations displays low levels of variation (Akkouche et al. 2013; Song et al. 2014). However, recent populational studies suggest that an increase of *pi*RNA gene expression levels could facilitate TE silencing in *D. simulans* (Lerat et al. 2017; Saint-Leandre et al. 2017). At the genome level, most of the *pi*RNA production likely depends on the presence of TE families that reach a high copy number in the genome, particularly those accumulating within *pi*RNA clusters (Kelleher and Barbash 2013). These observations raise interesting questions regarding the exact relationship between TE activity and their regulation by the *pi*RNA pathway in these two sibling species.

Notably, can TE history recapitulate the evolution of lineage-specific *pi*RNA repertoires?

*Drosophila melanogaster* and *D. simulans* are two closely related species that diverged 3–5 Ma (Hey and Kliman 1993; Kelleher and Barbash 2013). Genomic comparison between the two sibling species has revealed that the TE content and landscape are dramatically different (Vieira et al. 1999; Lerat et al. 2011; Kofler, Hill, et al. 2015; Kofler, Nolte, et al. 2015). A range of evidence suggests that the *D. melanogaster* genome has undergone recent transpositional bursts of many TE families (Bowen and McDonald 2001; Bergman and Bensasson 2007; Kofler, Hill, et al. 2015; Kofler, Nolte, et al. 2015). Furthermore, the sequence similarity of the TEs from divergent lineages suggests that the *D. melanogaster* genome has been repeatedly invaded by novel TE families (Sanchez-Gracia et al. 2005; Bartolome et al. 2009; Gilbert et al. 2010). Consistent with the recent activation of many TE families, the genome of *D. melanogaster* contains a large number of full-length copies (Lerat et al. 2011). By contrast, the *D. simulans* genome displays a large number of old and degraded copies indicating that TEs have lost most of their activity (Lerat et al. 2011). Thus, these two sister species represent good models to study the impact of TE evolution and the genome-defense response mediated by the *pi*RNA regulatory machinery.

In the present work, we first reannotated TEs of *D. melanogaster* and *D. simulans* genomes and confirmed that both species display very different TE histories in terms of amplification time and extent. We compared transcriptional levels of TEs in both male and female gonads from several populations of *D. simulans* and *D. melanogaster*. Both species display severe patterns of sex-biased TE transcription in gonads. Comparison of *pi*RNA deep-sequencing libraries showed that ovaries intensely produce *pi*RNAs derived from TEs, whereas TE-specific *pi*RNAs were barely present in testes. We found that variation in the TE content (TE age and structure) between the two species has strongly impacted the populations of *pi*RNAs expressed in the ovaries. Furthermore, we noticed variation in the PIWI pathway (*ping-pong* efficiency and expression of PIWI effectors) that could also reflect different TE invasion histories. Indeed, *D. melanogaster* ovarian *pi*RNAs preferentially match to TE families overexpressed in testes showing signatures of relatively recent transposition bursts. Although *D. simulans* present signatures of lower TE activity and *pi*RNAs derive from old (fragmented and inactivated) copies, the *ping-pong* silencing has been efficiently maintained. We propose an evolutionary dynamics model that includes 1) after a new TE invasion, the setup of a progressive implementation of the *pi*RNA machinery in order to moderate and ultimately control TE expansion; 2) after efficient silencing, a long-term persistence of the *pi*RNA production against extinct TE lineages that may help protecting the host from future reinvasions. The differences observed in *D. melanogaster* and *D. simulans* suggest that they may be at different steps of this process.



**Fig. 1.**—*Drosophila melanogaster* and *Drosophila simulans* share divergent TE histories. Pie charts show the proportion of the different types of TEs in the *D. melanogaster* *iso1* (dmel r6) (A) and *D. simulans* *w501* (dsim r2) (B) genomes. TEs represent 14% of the *D. melanogaster* genome length (based on nonredundant annotations of [supplementary table S2, Supplementary Material](#) online) and 3% of *D. simulans* genome length (based on nonredundant annotations of [supplementary table S2, Supplementary Material](#) online). Histograms show the size distribution of TE insertions in *D. melanogaster* (C) and *D. simulans* (D). The y axis displays the number of copies found in the reference genome, according to their size. Size of TE insertions was normalized by the length of their respective consensus sequence (x axis). Red-dotted vertical lines delimit the full-length elements ( $\geq 98\%$  of consensus size) with percentages given. The same color code is used for TE class. We distinguished internal portion of LTR retrotransposons (black) from their LTR (dark gray).

## Results

### *Drosophila melanogaster* TEs Are Younger and More Abundant to Those Found in *D. simulans*

Using a library of *Drosophila* consensus elements derived from Repbase, we first analyzed the relative TE proportions in *D. melanogaster* and *D. simulans*. All main types of TEs are found in both genomes in slightly different proportions (fig. 1A and B). In *D. melanogaster* and *D. simulans*, LTR retrotransposons constitute the main class of elements (respectively 67% and 44% of the total TE fraction), followed by non-LTR retroelements (25% and 38%, respectively) and DNA transposons (8% and 18%, respectively).

Although both genomes display relative similarities in terms of TE diversity, the total fraction of all repeated sequences (including TEs, satellites, and simple repeats) is quite different between the two species. It represents 15% of the *D. melanogaster* genome (14% of TEs and 1% of other repeated sequences) and 4% of the *D. simulans* genome (3% of TEs and 1% of other repeated sequences). Moreover, genome size differences corroborate levels of TE degradation between the sibling species (fig. 1C and D). Out of 12,803 TE insertions in *D. melanogaster*, 21% of copies are full length, whereas only 4% of TEs are full length in *D. simulans* (out of 4,583 insertions). These observations are

consistent with other comparative studies showing that *D. melanogaster* display abundant full-length TE copies rather than highly degraded copies as in its sibling species *D. simulans* (Lerat et al. 2011).

The genome assembly quality could partially account for this important difference in TE load. The heterochromatic regions of the *D. melanogaster* genome are more contiguous and chromosome arms usually span several additional megabases compared with the *D. simulans* genome. Although the TE excess of the *D. melanogaster* genome (25 Mb) seems a sufficient factor to explain the genome size difference observed with *D. simulans* (175 and 150 Mb, respectively), we compared genomic TE contents on alignable portion of both genomes (i.e., removing deep heterochromatin regions of the *D. melanogaster* assembly). This analysis shows the same qualitative differences between the two species (supplementary fig. S1, Supplementary Material online), notably, a higher genomic TE fraction, higher copy number, and a 5-fold excess of full-length insertions in *D. melanogaster* compared with *D. simulans*.

In summary, we found that the *D. melanogaster* genome TE load is considerably higher compared with that of *D. simulans*, and the copies are less degraded. As suggested before (Lerat et al. 2011), this difference could be explained by different evolutionary histories: a more recent or a continuous TE invasion in *D. melanogaster*, and an older one in *D. simulans*. Such differences in terms of TE histories might have serious consequences on TE transcription levels in the germline.

### TEs Display Strong Sex-Biased Patterns of Expression

To understand the relationship between the TE load and their activity in the germline, we sequenced testis and ovary transcriptomes for two populations of *D. melanogaster* (Gotheron and Zimbabwe) and two populations of *D. simulans* (Fukuoka and Nairobi). We mapped these population transcriptomes on the Repbase *Drosophila* TE data set and first computed intra-specific variation of TE transcription. Between populations of the same species, we detected only a limited number of differentially expressed families (fig. 2A and B). By contrast, comparing TEs between species revealed that 62% of the TE families present in both species ( $n = 232$ ) were differentially expressed (fig. 2C). A principal component analysis (PCA) on all data revealed two main axes explaining 61% of the total variance between transcriptomes (supplementary fig. S2A, Supplementary Material online). *Drosophila melanogaster* and *D. simulans* are clearly distinguished on the first PCA axis (38%), whereas the second PCA axis (23%) split transcriptomes according to the gonad type (testes vs. ovaries). Differences between sex are confirmed by PCAs performed on each species individually (supplementary fig. S2B and C, Supplementary Material online). In both cases, the first PCA axis (>55%) is strongly correlated to TE transcription changes

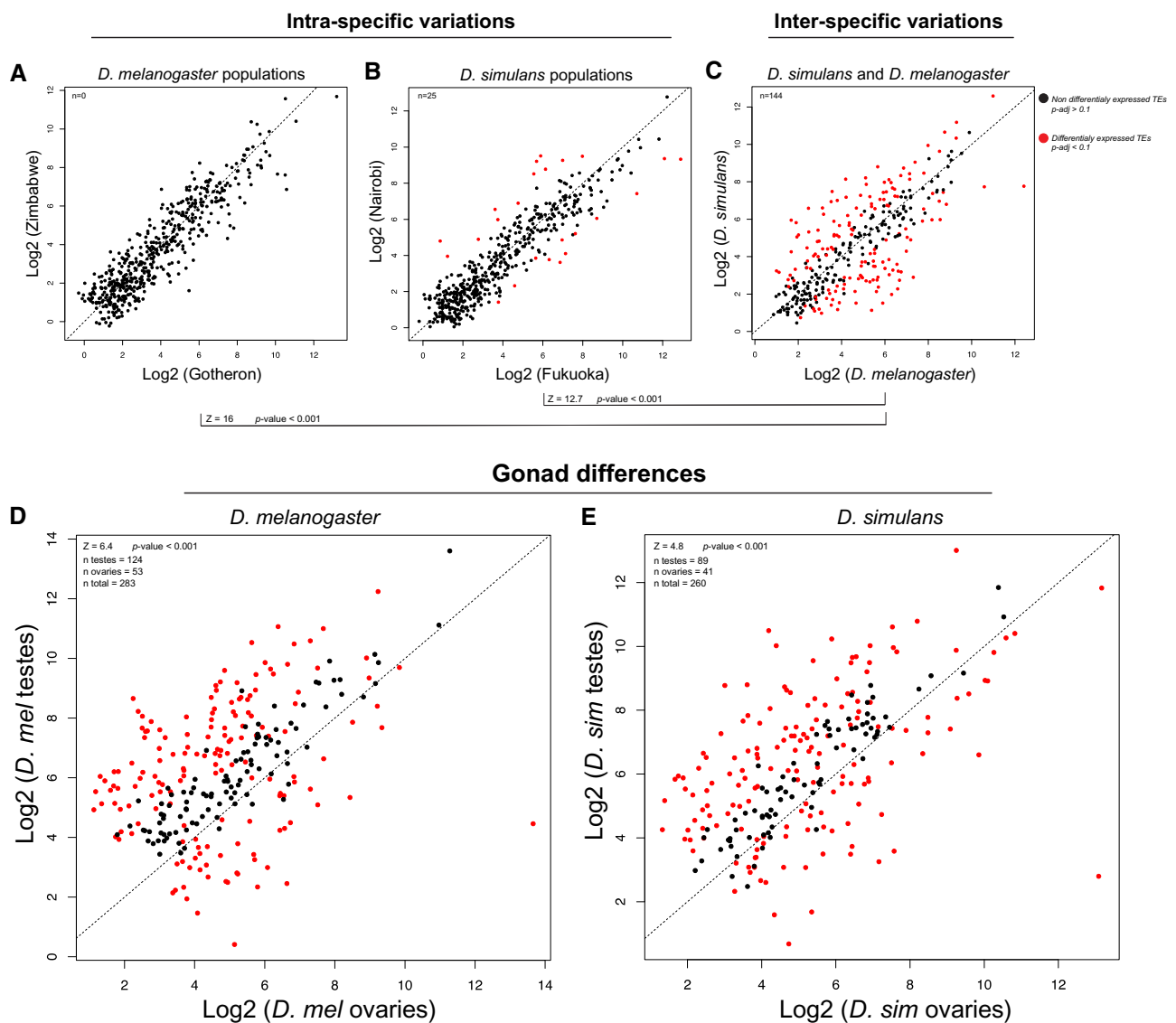
between the two germinal lines, whereas differences according to the population captured <20% of the variance on the second axis. Indeed, sex-biased TEs (TEs differentially expressed between male and female germlines) represent 63% of the TE consensus in *D. melanogaster* and 50% in *D. simulans* (fig. 2D and E). In both species, this sex-biased pattern of expression is mainly due to a global higher expression of TEs in testes. TEs more expressed in testes represent 70% of the sex-biased TEs in *D. melanogaster* and 68% in *D. simulans*. Finally, we observed in both species (supplementary fig. S3, Supplementary Material online) that sex-biased TEs are similarly distributed among the main super-families of TE (i.e., LTR, non-LTR, or DNA transposons).

In summary, two major factors seem to influence TE expression: on one hand, interspecific variation (*D. melanogaster* vs. *D. simulans*) that is much higher than within species, and in the other hand gonad-specific variation (testes vs. ovaries). We observed a global higher expression of most TE families in testes compared with ovaries for both populations of *D. melanogaster* and *D. simulans*. We also observed some variations between populations within each species, as previously reported (Lerat et al. 2017). Yet, intra-specific variations remain much lower than those observed both between the two species and between sex (supplementary table S1, Supplementary Material online).

### *pi*RNA-Mediated Silencing of TE Is Predominant in Ovaries but Weak in Testes

Could the global TE overexpression in testes underlie major sexual differences of the *pi*RNA regulatory pathway between gonads? To evaluate this hypothesis, we first compared levels of TE-derived *pi*RNAs across ovaries of different laboratory strains and populations for which small RNA sequences were publicly available (see supplementary table S1, Supplementary Material online). We observed only slight variation (supplementary fig. S4A–C, Supplementary Material online). Indeed, <4% of TEs show a *pi*RNA expression change higher than 2-fold. This result agrees with previous independent studies showing that pools of ovarian *pi*RNAs were stable across strains and populations (Akkouche et al. 2013; Song et al. 2014).

For sex comparisons, we used data set of laboratory strains, produced in this study or publicly available (M19, a strain derived from  $w^{1118}$  for *D. melanogaster* and  $w^{501}$  for *D. simulans*). Between male and female gonads, not only *pi*RNAs but also other small RNA species greatly differ (supplementary fig. S5, Supplementary Material online). Indeed, *pi*RNAs and miRNAs in testes represent a small fraction of the total small RNA pool compared with ovaries (supplementary table S1, Supplementary Material online). These ostensible differences may reflect the very distinct biological functions carried out in male and female germlines and direct comparisons may be challenging. To avoid eventual bias linked to

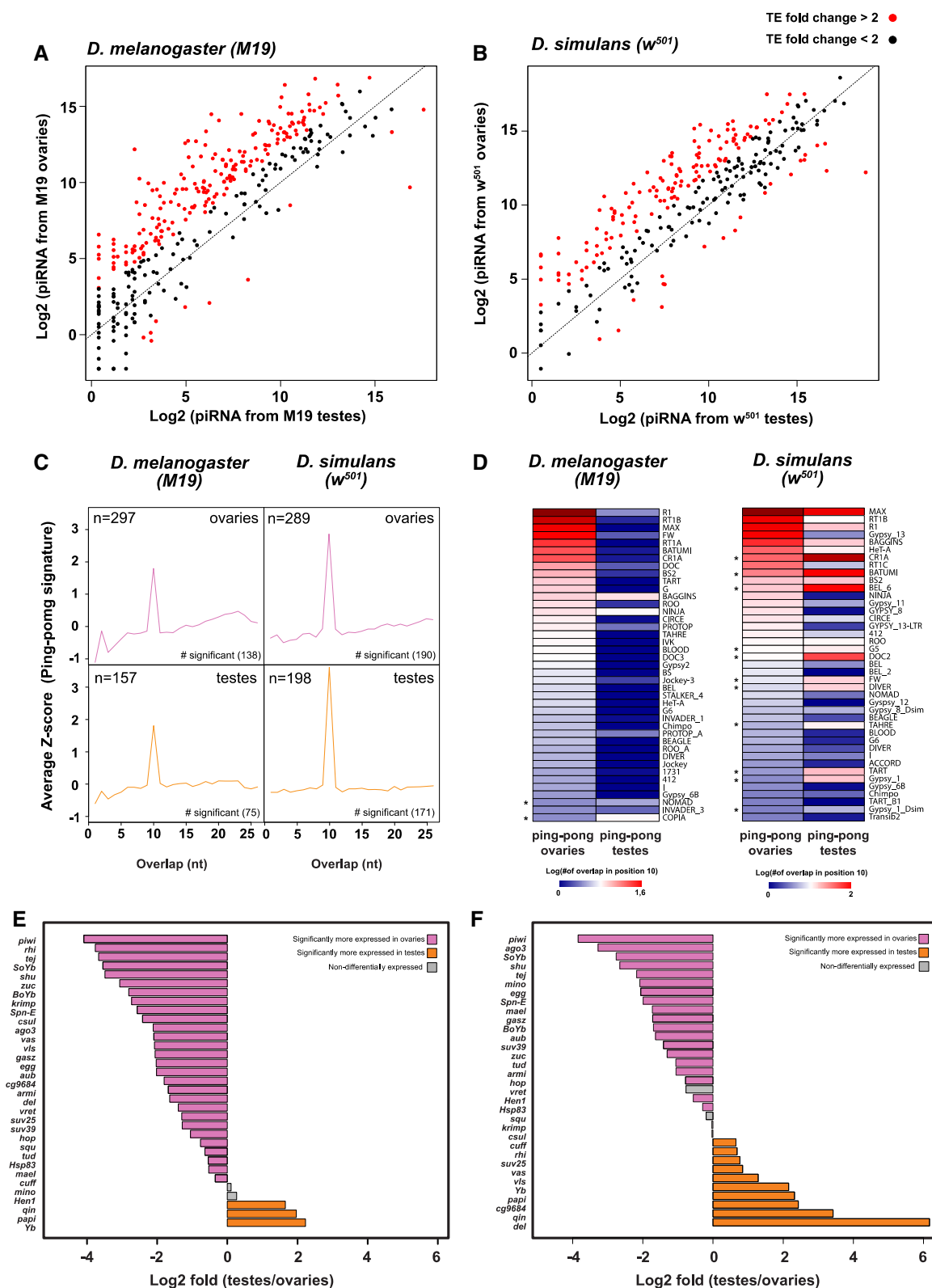


**Fig. 2.**—*Drosophila melanogaster* and *Drosophila simulans* TEs display strong sex-biased patterns of expression. Scatterplots (A–E) showing the transposon expression between species and populations of *D. melanogaster* and *D. simulans*. RNAseq was performed on ovaries and testes and each point shows normalized (DESeq2) values for a transposon family according to conditions. Diagonals represent  $x = y$ . Points in red show TE families significantly differentially expressed (DESeq2,  $P_{adj} < 0.1$ , FDR = 0.1). (A) Relative TE expression between Zimbabwe (y axis) and Gothenron (France) populations (x axis) of *D. melanogaster*. (B) Relative TE expression between Nairobi (y axis) and Fukuoka populations (x axis) of *D. simulans*. (C) Relative TE expression between pooled populations of *D. simulans* (y axis) and *D. melanogaster* (x axis). (D) Relative TE expression between testes (y axis) and ovaries (x axis) of *D. melanogaster* populations. (E) Relative TE expression between testes (y axis) and ovaries (x axis) of *D. simulans* populations.  $P$  values for differences were obtained by Z-statistics.

products of mRNA degradation, we normalized *pi*RNAs by the total number of miRNAs. In both species, after normalization, the observed *pi*RNA drop in testes still persists when compared with ovarian *pi*RNAs (supplementary fig. S5, Supplementary Material online). In *D. melanogaster* ovaries, 52% of TE families display a >2-fold *pi*RNA enrichment compared with testes (fig. 3A). This pattern is conserved in *D. simulans* where 47% of TEs show such a biased expression

in ovaries (fig. 3B). Indeed, significant changes correspond almost exclusively to a higher production in ovaries. These observations were conserved as well, when we mapped *pi*RNAs with other mismatch thresholds (supplementary fig. S6A and B, Supplementary Material online).

The *ping-pong* process is a secondary amplification of *pi*RNAs generated from the slicing of mRNA precursors. Slicing of precursors generates secondary sense *pi*RNAs that



**Fig. 3.—***piRNA*-mediated silencing of TE is predominant in ovaries, weak in testes. (A, B) Scatterplots showing the *piRNA* expression (number of *piRNA* normalized by total number of miRNA) per TE family between ovaries and testes of *Drosophila melanogaster* (A) and *Drosophila simulans* (B). Diagonals represent  $x = y$ . Points in red show TEs displaying a fold change expression >2. (A) Relative *piRNA* expression from ovaries of the *D. melanogaster*  $w^{1118}$ -derived strain M19 ovaries(y axis) and M19 testes (x axis). (B) Relative *piRNA* expression from ovaries of the *D. simulans* strain  $w^{501}$  (y axis) and  $w^{501}$  testes (x

typically overlap by 10 nt with the complementary antisense guiding piRNAs (Brennecke et al. 2007). An excess of 10-nt overlap observed between sense and antisense piRNAs is then the signature of a ping-pong mechanism. We computed the length of overlap in our data set (Antoniewski 2014) and detected global significant ping-pong signatures in both gonads of both species. However, comparing the overlap signals for a set of TE consensus shared by the two sibling species, we noticed that ping-pong signals were predominantly higher in ovaries than in testes (fig. 3D). TE overlap signatures in *D. simulans* testes are stronger compared with *D. melanogaster* testes (fig. 3C), suggesting that the ping-pong mechanism is more efficient in *D. simulans* testes. Consistent with this observation, TE families with higher ping-pong signal in testes than in ovaries (fig. 3D) are more important in *D. simulans* (11 out of 40) compared with *D. melanogaster* (2 out of 40).

Using our RNAseq data, we further compared male and female gonad patterns of expression (fig. 3E and F) for a set of genes known to be essential for piRNA silencing (Handler et al. 2013). In *D. melanogaster*, 85% of the piRNA regulatory genes are more expressed in ovaries. In *D. simulans*, a similar trend is observed with 58% of piRNA genes showing enriched expression in ovaries. The number piRNA genes with higher expression in testes is greater in *D. simulans* (30%) compared with *D. melanogaster* (9%). Interestingly, this pattern is consistent with a more efficient ping-pong amplification loop in *D. simulans* testes.

Altogether, these data suggest that the TE silencing via piRNAs presents a female-biased activity: 1) expression of genes involved in the piRNA pathway is higher in the female germline, 2) ovaries produce remarkably larger amounts of piRNAs derived from TEs than testes do, and 3) TE families have generally higher ping-pong signatures in ovaries than in testes. This pattern could be a major contributor of the TE overexpression pattern in testes evidenced in this study. Nevertheless, based on these observations, it remains difficult to propose a comprehensive view of the relationship between the germline piRNA repertoire, the TE sex-specific patterns of expression, and the TE dynamics within and between genomes. To this end, we performed qualitative and quantitative analyses of piRNA variation to understand the genomic features of TEs preferentially targeted.

### TE Histories Have Shaped the Dynamics of piRNA Biogenesis

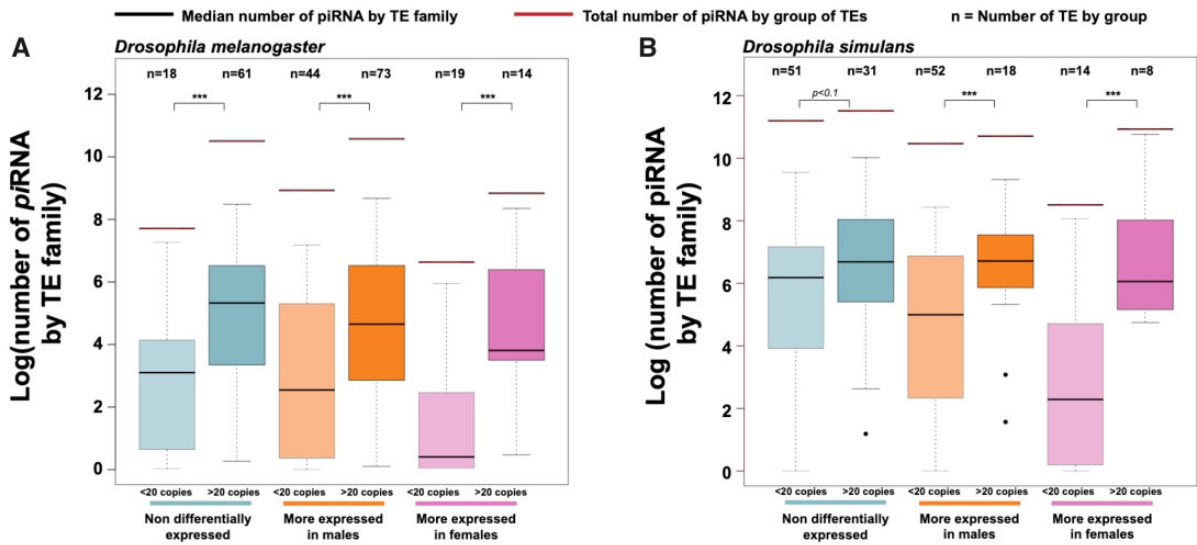
The piRNA-mediated silencing of TEs might have been primarily co-opted to slow down transposition rates of the most active families. However, an efficient silencing could eventually suppress the activity of a targeted lineage. In this context, an intense piRNA response could reflect either a recent transposition burst or an abundant TE family that stop transposing and start to go extinct. Here, we analyzed the relationship between amplification levels of TE families and the strength of their specific piRNA responses (fig. 4A and B) using ovarian piRNA data sets presented in the previous section (fig. 3A and B). We first used copy number (more or <20 copies in the reference genome) as an arbitrary criterion to distinguish between highly and poorly amplified TE families and grouped TEs according to their expression status (more expressed in testes, in ovaries, or nondifferentially expressed). We detected a clear relationship between piRNA levels and TE abundance in the genome. In both species, the average number of piRNA per TE was significantly higher for the most amplified TE families (>20 copies) compared with less abundant families (<20 copies), whatever their status concerning differential expression between gonads. In addition, we observed the same pattern for piRNA production in testes (supplementary fig. S7, Supplementary Material online), suggesting that piRNAs of both germlines are mainly responding to the most abundant TE families. These results suggest that piRNAs primarily respond to the most abundant TE lineages in the genome although copy age and degradation are strikingly different between the two focal species.

To provide additional supports to this observation, we performed a more complete analysis of the relationship between the genomic characteristics of each TE family and the strength of the corresponding piRNA defense response (fig. 4C and D). PIWI proteins loaded with piRNAs target TE mRNAs and degrade TE transcript through their slicing activity. During the ping-pong cycle, PIWI proteins directly use TE mRNAs as substrate to generate novel piRNAs indicating that TE mRNA levels may have critical impacts on piRNA biogenesis output. Consistent with this phenomenon, both species display a strong positive relationship between mRNA and piRNA levels ( $r=0.49^{***}$  for *D. melanogaster* and  $r=0.44^{***}$  for *D. simulans*). Levels of TE transcription are mainly explained

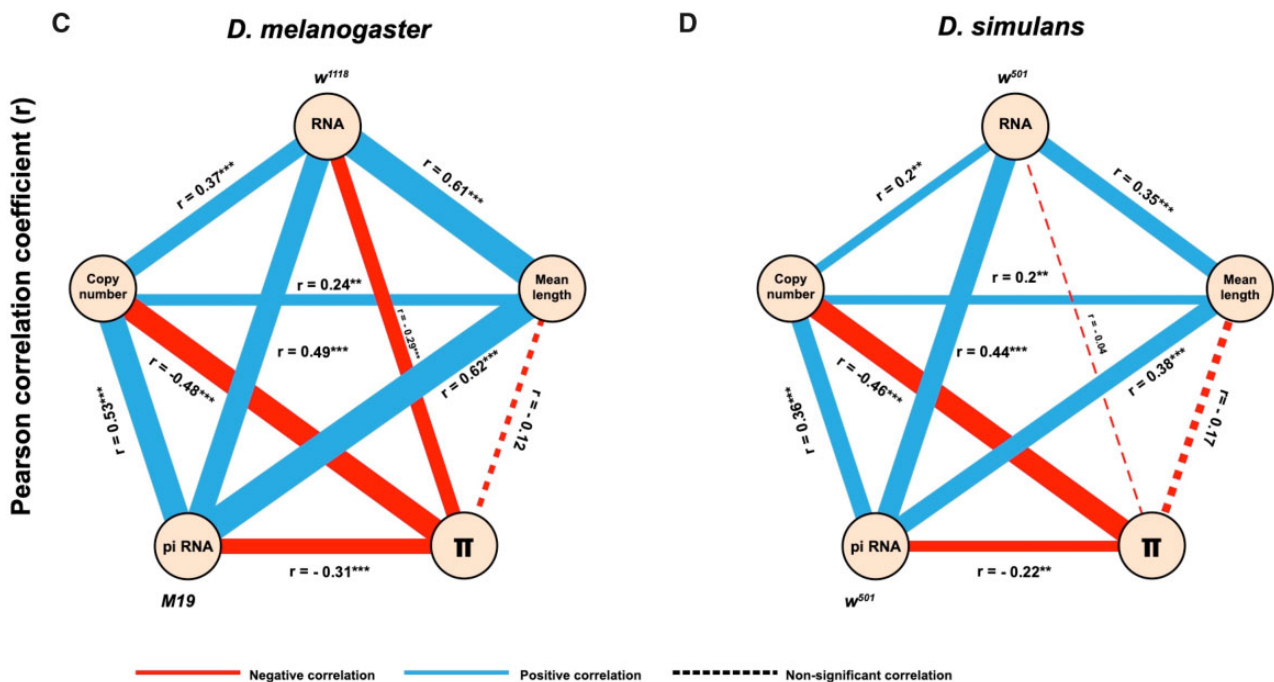
**Fig. 3.—Continued**

axis). (C, D) Plots showing probability of overlapping piRNAs and the length of the overlap according to their starting position on 3' piRNAs in *D. melanogaster* (left) and *D. simulans* (right). Pink lines show average z-score in ovaries, whereas orange lines show average z-score in testes. Total number of TE families presenting overlapping piRNAs and number of TE families presenting significant overlap (z-score > 1.96;  $P < 0.05$ ) are indicated. (D) Heatmaps comparing ping-pong signatures in testes and ovaries for a set of TEs shared between *D. melanogaster* (left) and *D. simulans* (right). The ping-pong signature is expressed as the number of overlapping pairs (first 10 nt) normalized by number of piRNAs. Black asterisks highlight TEs with stronger ping-pong signatures in testes. Bar graphs (E, F) shows relative expressions of piRNA pathway genes between ovaries and testes of *D. melanogaster* (E) and *D. simulans* (F). Pink bars show genes more expressed in ovaries. Orange bars show genes more expressed in testes. Genes that are not significantly differentially expressed are shown in gray.

### Relationship between TE piRNA transcription levels, expression status and degradation levels



### Relationship between TE copy number, mRNA and piRNA transcription levels



**Fig. 4.**—Relationship between TE piRNA transcription levels and genomic features of TEs. (A, B) Boxplots show the distribution of the normalized number of ovarian piRNA per TE family in *Drosophila melanogaster* (A) and *Drosophila simulans* (B). Clear colors represent TE families present in <20 copies and dark colors represent TE families displaying more than 20 copies in the genome. Blue shows nondifferentially expressed TE families, pink is for TE families more expressed in ovaries, and orange for TE families more expressed in testes. Stars indicate *P* value of Wilcoxon statistics. (C, D) Pearson' correlation coefficients between piRNA transcription levels and TE genomic traits (i.e., mRNA transcription levels, number of copies, median length of the TE families, and the nucleotide diversity index  $\pi$ ) in *D. melanogaster* (C) and *D. simulans* (D). Dashed lines represent nonsignificant correlations and solid lines significant correlations. Thickness of the lines is proportional to their respective Pearson' coefficients *r*. Blue lines stand for positive correlations, whereas red for negative correlations. The significant *P* values are indicated for each solid line.



by an increase in TE copy number ( $r=0.37^{***}$  for *D. melanogaster* and  $r=0.2^{**}$  for *D. simulans*) and TE length ( $r=0.61^{***}$  for *D. melanogaster* and  $r=0.35^{***}$  for *D. simulans*). The nucleotide diversity index ( $\pi$ ) is related to the conservation level between copies of a given TE family (a low  $\pi$  corresponds to highly similar copies indicating a recent expansion). We observed a relationship between transcription levels and  $\pi$  in *D. melanogaster* ( $r = -0.29^{***}$ ) but not in *D. simulans* ( $r = -0.04$  for *D. simulans*).

However, in both species, the *piRNA* production is significantly associated with these four variables that altogether describe the intensity of TE family's activity. Levels of *piRNAs* dramatically increase for TE families displaying high copy number, high length, and low  $\pi$  ( $r=0.53^{***}$ ,  $0.62^{***}$ , and  $-0.31^{***}$ , respectively, for *D. melanogaster* and  $r=0.36^{***}$ ,  $0.38^{***}$ ,  $-0.22^{**}$ , respectively, for *D. simulans*), suggesting that production of *piRNAs* preferentially target expanding, or expanded but not too old, TE families. These trends are conserved regardless of the data set used (RNAseq, small RNAseq, or genomes, see [supplementary fig. S7](#) and [table S3, Supplementary Material](#) online). However, relationships between *piRNA* and recent activity markers are globally weaker in *D. simulans* (old TE invasions) compared with *D. melanogaster* (more recent TE invasions), which support a preferential link between *piRNA* production and recent TE expansion ([fig. 4C and D](#)).

In summary, differences between the sibling species appear to be the result of different tempo and activities of TE invasion: a recent invasion in *D. melanogaster* where TEs spread actively and an ancient invasion in *D. simulans* where TEs slowly go extinct. These results suggest that the host *piRNA*-mediated defense was activated first to slow down the invasion of the most active TE lineages and later to maintain a long-term protection against former successful TEs. If this assertion is true, we should observe 1) an accumulation of the most active TE families within *piRNA* clusters and 2) their persistence within *piRNA* clusters when TE families get old and lose their activity.

### Recent Bursts of Transposition Enhance the Ovarian Specificity of *piRNA* Clusters

To investigate the dynamics of TEs within *piRNA*-producing loci and the consequences on sex-biased expression, we compared the *piRNA* clusters density of TEs according to their expression pattern (overexpressed in testes, overexpressed in ovaries, and nondifferentially expressed). In both species, we localized *piRNA* clusters and compared their genomic distribution in testes and ovaries ([fig. 5A–C](#)). The cluster 42AB, known as a *piRNA* “master locus” in *D. melanogaster* (Brennecke et al. 2007), is transcriptionally active in both testes and ovaries ([fig. 5A](#)). Along this genome, however, we could also identify other clusters that are expressed only in ovaries (e.g., pericentric *piRNA* cluster of the chromosome 2R,

[fig. 5B](#)), and clusters that are transcriptionally active only in testes ([fig. 5C](#)). We performed a global screen of the genome for *piRNA*-producing loci in a 1-kb window and revealed that most of the *piRNA* clusters are active in ovaries ([fig. 5D and H](#)). In *D. melanogaster*, 4% of *piRNA* loci are testes specific, 5% were found expressed in both germline, and 91% were only expressed in ovaries. In *D. simulans*, female-specific *piRNA* clusters represent as well 91% of all *piRNA* clusters. These data clearly reinforce our previous observations that the female germline is the main tissue involved in TE silencing and also explain the global tendency to produce less *piRNA* in testes.

We further characterized TE density along *piRNA* clusters according to their pattern of expression ([fig. 5E–G and I–K](#)). In *D. melanogaster*, we observed that TEs more expressed in testes display the highest density within female-specific *piRNA* clusters. These TEs constitute almost 40% of ovary-specific *piRNA* clusters. Nondifferentially expressed TEs are predominant (~20%) in *piRNA* clusters active in both germlines, closely followed by TEs with enriched expression in testes (~15%). In contrast, TEs with higher expression in ovaries exhibit an extremely low density across all *piRNA* clusters. Interestingly, in *D. simulans*, we noticed a lower TE density within ovary-specific *piRNA* clusters ([fig. 5J](#)) compared with nonspecific ones ([fig. 5I](#)) and also compared with *D. melanogaster* ([fig. 5F](#)). This pattern in TE density in *D. simulans* could be a direct consequence of the TE degradation process, suggesting that ovary-specific *piRNA* clusters are progressively purged from TEs once their invasion has been successfully tackled.

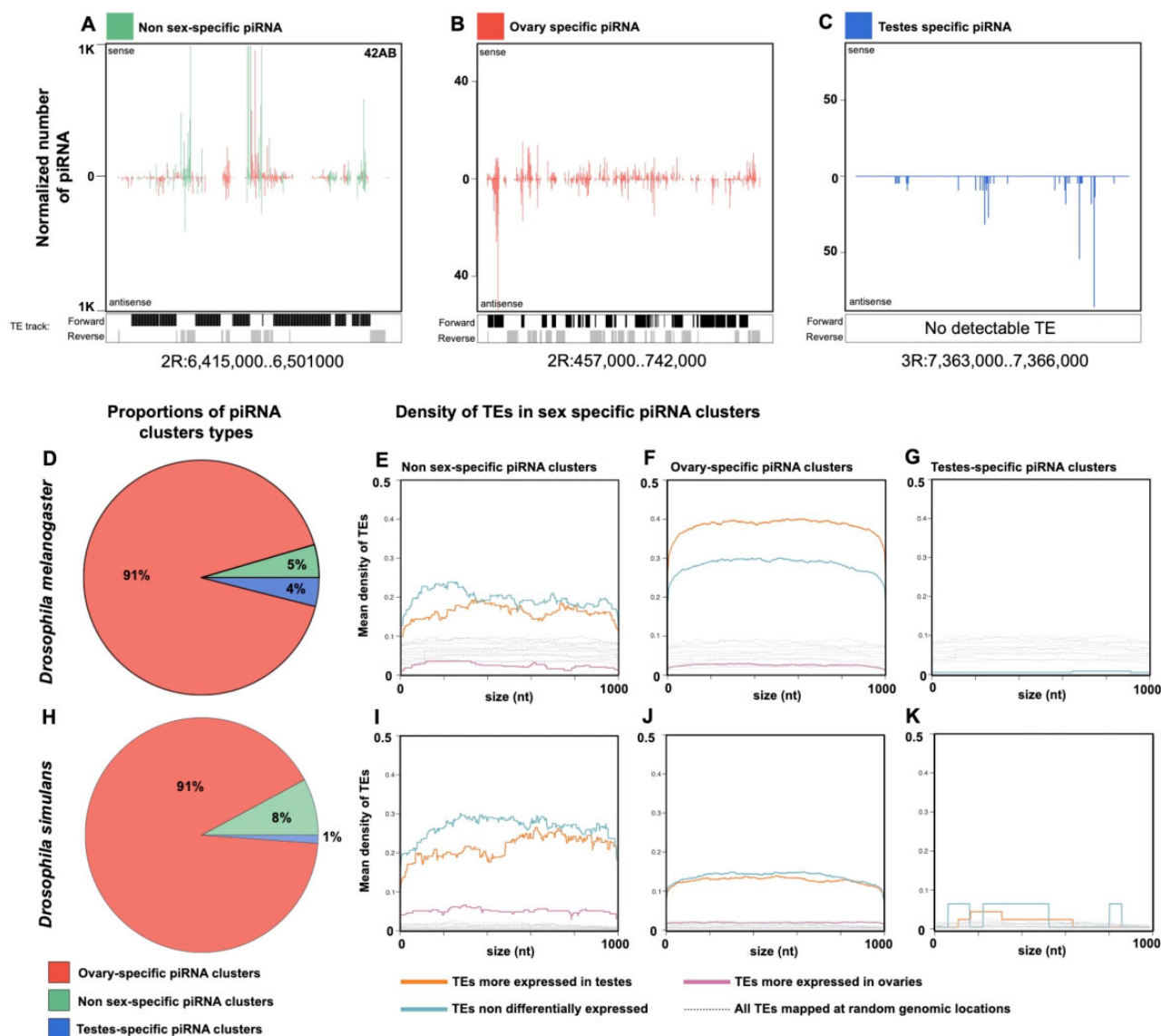
Therefore, it seems that when a species experiences an intense TE expansion (as for *D. melanogaster*), an accumulation of TE fragments occurs in genomic regions dedicated to the ovarian-specific production of silencing *piRNAs* leading to a contrast of TEs expression in the two germlines. Then, when the TE expansion is under control (as for *D. simulans*), a progressive TE loss in these exclusively female *piRNA* clusters occurs, which re-equilibrates the pattern of TE expression between testes and ovaries.

## Discussion

### Divergent TE Evolutionary History between *D. melanogaster* and *D. simulans*

Our results show that during 5 Myr of divergence, *D. melanogaster* and *D. simulans* genomes have accumulated very different TE content. This is consistent with several previous studies comparing the two sibling species (Vieira et al. 1999, 2012; Lerat et al. 2011). Recently, a large-scale analysis from natural populations of both species has revealed that most of the new TE insertions are due to the ongoing expansion of 58 TE families (Kofler, Nolte, et al. 2015). In addition, the distribution among populations of these recently invading

## Density of piRNA mapping at unique genomic locations (1kb window)



**Fig. 5.**—Sex-biased evolution of piRNA clusters is shaped by the tempo of TE activity. The genomic maps (A–C) show the number of uniquely mapping piRNAs along the regions of chromosome 2R (A, B) and 3R (C) of *Drosophila melanogaster* species (x axis). Colored bars show the number of piRNAs according to their expression pattern: green bars display piRNAs expressed in both testes and ovaries, red bars piRNAs exclusively expressed in ovaries, and blue bars piRNAs exclusively expressed in testes. All small RNAs display a size comprised between 23 and 30 nucleotides. Composition in TEs is indicated underneath: black shows TEs inserted in forward orientation and gray TEs inserted in reverse orientation relative to the genome. (D, H) Pie charts showing the relative proportion of piRNA clusters across the whole genome. (E–G, I–K) The average TE density (y axis) along the 1-kb window of piRNA clusters screened all along the genome (x axis). The TE density was analyzed according to the sex expression pattern of TE families. Pink lines represent TEs more expressed in females, blue lines stand for TEs nondifferentially expressed, and orange lines for TEs more expressed in testes. The piRNA clusters were analyzed according to their germline expression. (E, I) Stand for piRNA clusters expressed in both testes and ovaries, (F, J) for piRNA clusters exclusively expressed in ovaries, and (G, K) for piRNA clusters exclusively expressed in testes.

TEs described high levels of heterogeneity, consistent with some non-annotated TE families in the reference genomes. In our study, we only consider TE families that have been present in the reference genome, excluding low frequency TE lineages that are not yet established. For instance, the

ongoing P element invasion of *D. simulans* (Kofler, Hill, et al. 2015) is not present in the reference genome. However, our estimations of the TE diversity in the reference genome (fig. 1) and in other assemblies (supplementary table S3, Supplementary Material online) are similar to those

determined by pool sequencing analyses (Kofler, Nolte, et al. 2015). Besides the global TE load, the most striking difference observed here is the strong overrepresentation of deleted copies in *D. simulans* compared with *D. melanogaster*. This suggests that the main differences between both species are related to the tempo of TE activity: *D. melanogaster* could be characterized by recent TE invasion or transposition bursts of several TE families, whereas *D. simulans* TE content consists mainly in fragmented and inactive elements probably due to ancient invasions.

Transcriptomic data of gonads show that differences in TE expression between populations are quite limited compared with those detected between male and female germlines. We have therefore concentrated our analyses on TE families differentially expressed in these tissues. Interestingly, we found that levels of TE transcription were related to TE copy number and *piRNA* levels of expression in the germline. Because the *piRNA* pathway is crucial in modulating TE activity in the germline, we further analyzed the type of relationships between features of TE activity and the subsequent host-mediated silencing response.

### TE Activities and the *piRNA* Genome Response

We analyzed TE-derived *piRNA* profiles in order to clarify the relationship between TE activity and *piRNA* regulation. We observed that the majority of the TE-derived *piRNAs* matched to TE families that are highly transcribed. The positive correlation between *piRNA* and TE mRNA levels likely result from the PIWI protein slicing activity that use TE mRNAs as substrates to generate novel *piRNAs* during the *ping-pong* amplification cycle. At the genome level, the response against TE invasions is predominantly achieved by multiple insertions within *piRNA* clusters involved in the secondary *piRNAs* biogenesis. These observations suggest that a TE family inserted at high density within *piRNA* clusters, and still producing abundant mRNA transcripts would represent an ideal *piRNA* target.

Consistent with Kelleher and Barbash (2013) model, we observed that high levels of mRNA and *piRNA* expression are features associated with TE families displaying higher copy number, suggesting that the host TE silencing response was shaped by successful TE amplification. In addition, we observed a strong relationship between *piRNA* levels and features of recent TE activity (long length and low diversity between copies) indicating that *piRNAs* preferentially target relatively recent waves of TE expansion. However, these correlations persist in *D. simulans* where TEs are more degraded and less active compared with *D. melanogaster*. This last observation suggests that among a pool of ancient TEs, the relatively youngest families will still be preferentially targeted by *piRNAs*. Altogether, our data favor a model in which *piRNA* production is acquired during TE expansion, as soon as copies are accumulated and fixed in *piRNA* clusters. Then, the

maintenance of an active *piRNA* production relies on the absolute mRNA levels of a given TE family, its rate of degradation in the genome, and ultimately, on its rate of elimination from *piRNA* clusters.

Indeed, *piRNA* clusters are composed of repeated sequences derived from TEs and their fragmented derivatives (Brennecke et al. 2007). Their genomic locations are conserved across *Drosophila* species suggesting that natural selection has favored the maintenance of TE silencing regions producing *piRNAs* (Brennecke et al. 2007; Malone and Hannon 2009; Castaneda et al. 2011). It has been shown that the pool of TEs within a *piRNA* cluster can be easily updated by new TE insertions (Malone and Hannon 2009; Khurana et al. 2011), indicating that the TE composition of the *piRNA* clusters might be directly dependent on the species pool of successfully invading TEs. In addition, models of TE dynamics predict that TE can take advantage of the *piRNA* silencing machinery to reach fixation within *piRNA* clusters (Lu and Clark 2010; Kofler 2019).

*piRNA* clusters are located in highly heterochromatic regions (Brennecke et al. 2007). Then, a new TE insertion inside these regions may confer numerous advantages in a selective context. Such an insertion is not deleterious to the host and may ultimately give the host the ability to silence other TEs due to similarities between TEs. In contrast, euchromatic copies are often associated with deleterious effects and thus frequently removed by purifying selective forces (Gonzalez et al. 2008; Lu and Clark 2010). This scenario is consistent with the positive correlation observed between copy number, *piRNA* abundance observed in the present work, and the predominance of heterochromatic TE insertions in both species (Junakovic et al. 1998; Bartolome et al. 2002; Kaminker et al. 2002). In this respect, the comparison between *D. melanogaster* and *D. simulans* is of particular interest because they display different tempos of TE activity. Compared with *D. melanogaster*, *D. simulans* presents a dramatic TE loss characterized by a low copy number and a lower TE size (e.g., fig. 1). However, despite lower TE density within *piRNA* clusters in *D. simulans*, these TE fragments are sufficient enough to maintain a TE *piRNA* production (figs. 3 and 5). This may be due to the efficiency of the *D. simulans* PIWI pathway. In any case, it seems that once established, the *piRNA* silencing persists until the complete decay of the ancient families. In the late steps of invasion, although full-length active elements keep on declining or become extinct, copies are still able to persist as small TE relics embedded within the *piRNA* clusters and act against transcription of euchromatic ones.

### The TE *piRNA* Regulatory Machinery Is a Female-Specific System

Only very few works have paid attention to TE silencing by the *piRNA* pathway in testes. First, it was shown that most *piRNAs*

derived from *Stellate* in *D. melanogaster* (supplementary fig. S5 and table S1, Supplementary Material online, and Nishida et al. [2007]). A biochemical approach evidenced that most of the *pi*RNAs derived from TEs are loaded by *Ago3* but not by *Aub*, and that *Ago3* expression is restricted to the first four cellular divisions in *D. melanogaster* (Nagao et al. 2010). These results are consistent with our observations showing that in *D. melanogaster* testes, *pi*RNAs constitute a weak fraction of our small RNA libraries and generally display weaker *ping-pong* signatures (fig. 3D). The TE-derived *pi*RNAs observed in testes are likely to be analogous to those described by Nagao et al. (2010) and are thus probably restricted to the *Ago3*-loaded *pi*RNA at the extreme part of the testes. Indeed, TE-derived *pi*RNA populations collapse when testicular germ cells differentiate into spermatocytes (Qu  nerch'du et al. 2016). Here, we compared small RNA libraries of testis developmentally arrested mutants in early mitotic division (Qu  nerch'du et al. 2016) to our libraries from entire wild type testes. Only 10% of TEs display significant differences between these two conditions (supplementary fig. S3D, Supplementary Material online), suggesting that most of the *pi*RNA production in testes is limited to the very first stages of cell differentiation. Moreover, we mapped testes *pi*RNA clusters in both *D. melanogaster* and *D. simulans* and found that *pi*RNA clusters exclusively active in testes are rare and have a very low TE density. Indeed, testes *pi*RNA clusters containing TEs are the ones also active in females. Altogether, these results suggest that the role of *pi*RNA-mediated TE silencing in testes is relatively limited.

We suggest that this large difference directly contributes to the overall bias of TE expression in testes. In *D. melanogaster*, 119 TE families are overexpressed in testes, whereas they are mainly silenced in ovaries. This trend is more balanced in *D. simulans* for which only 70 TE families were testes biased.

Levels of *pi*RNAs and *ping-pong* signatures are higher in *D. simulans* testes than in *D. melanogaster*, indicating that a more efficient *pi*RNA production in testes could reduce this bias.

### Sex-Biased Evolution of *pi*RNA Clusters

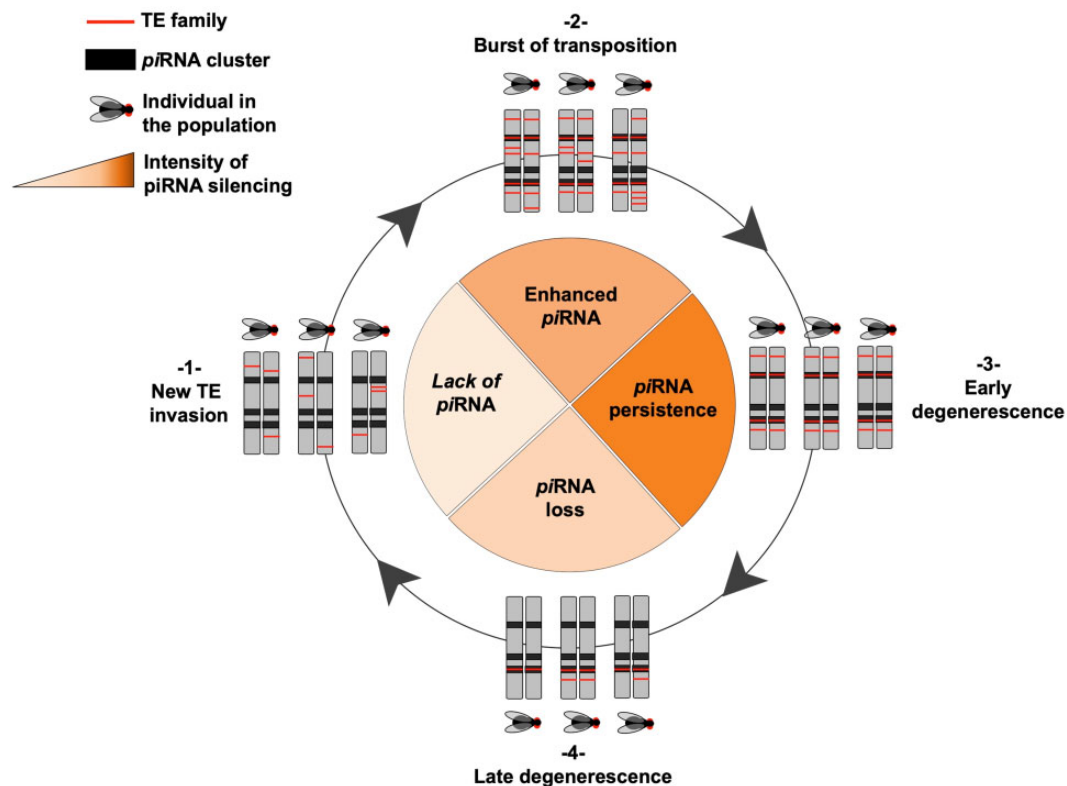
More than 90% of *pi*RNA clusters are exclusively expressed in ovaries. In *D. melanogaster*, TEs are recent active lineages, whereas mostly degraded in *D. simulans*. Our results show that TEs cover ~60% of female-specific *pi*RNA clusters in *D. melanogaster* and ~20% in *D. simulans*. These observations indicate that female-specific *pi*RNA clusters become saturated in TEs during current invasions and are progressively purged once invasions have stopped. As opposed to female-specific clusters, TE density of nonsex-specific clusters still remains high in *D. simulans*. These results suggest that selection is favoring TE accumulation within female-specific *pi*RNA clusters during pervasive expansion and that selection

maintains TEs within *pi*RNA clusters expressed in both sexes once TE activity has been controlled by the host.

The evolutionary arms race between host and TEs also has direct consequences on the evolution rate of *pi*RNA effector proteins (Kidwell and Lisch 2001; Aravin et al. 2007; Siomi et al. 2008; Blumenstiel 2011; Lee and Langley 2012). Independent works support that *pi*RNA proteins belong to the faster evolving component of coding sequences in *Drosophila* genomes and are further subject to recurrent adaptive mutations (Heger and Ponting 2007; Pane et al. 2007; Berry et al. 2009; Obbard et al. 2009; Kolaczowski et al. 2011; Lee and Langley 2012). In this study, *pi*RNA genes display different patterns of expression between the two sibling species, suggesting that host proteins have adapted to species-specific constraints. Indeed, we observed stronger *ping-pong* signatures in *D. simulans* compared with *D. melanogaster*. In terms of evolutionary strategies, a more efficient *pi*RNA amplification may suppress TE activity more promptly and in the end make the host less permissive to new TE invasions (Lerat et al. 2017; Saint-Leandre et al. 2017). In this context, shaping the efficiency of the *pi*RNA machinery could reflect an adaptation reminiscent of former pervasive transposition. Alternatively, it could reflect an adaptation to compensate the lack of TE material required to generate novel *pi*RNAs when genome TE content is too low.

Despite these species specificities, we found that the majority of essential *pi*RNA genes (Handler et al. 2013) display patterns of overexpression in ovaries. It is usually expected that genes under sex-specific selection display a sex-biased expression (Ellegren and Parsch 2007). Under this assumption, the *pi*RNA regulatory genes are under a strong female selection in both species. However, the female-biased expression of *pi*RNA pathway genes is more balanced in *D. simulans* indicating that female-specific selective pressures were relaxed because TE propagation has been stopped.

Altogether, our results indicate that increased TE activity may enhance female-biased investments in mobilizing genomic defense resources. However, an increase of TE activity in the male germline may constitute an important source of genetic variability and rearrangements that can feed the emergence of new genetic conflicts. In a wide range of species, male germline was described as a crucial tissue driving the evolution of genomes. The male-driven hypothesis was built on the observation that mutation rates in male gametes is always higher compared with female gametes (Hurst and Ellegren 1998; Connallon and Knowles 2005; Connallon and Clark 2010; Parsch and Ellegren 2013). Another concept, named the "out of testes" hypothesis, is based on the observation that the vast majority of newly emerging genes start to be expressed in a testes-specific manner (Paulding et al. 2003; She et al. 2004; Levine et al. 2006; Ponce and Hartl 2006; Heinen et al. 2009; Kaessmann 2010; Light et al. 2014). In this context, TEs were shown to stand as important contributors of the "male-biased" evolution of genomes (Bennetzen



**Fig. 6.**—Evolutionary dynamics of a new emerging TE family under control of the piRNA silencing pathway. Step 1 represents a young emerging TE (red line) in a population of diploid genomes (pair of gray bars). The activity of the new element is associated with a strong insertion polymorphism, and thus present at diverse genomic locations that are not fixed in the population. Step 2 corresponds to the establishment of the family within the population of genomes. Following the TE burst of amplification, some TE insertions are now found to be fixed in many genomic loci and some of them into piRNA-producing regions (black line). This is the stage found in *Drosophila melanogaster*. These insertions appear selectively advantageous for the host as they are able to limit the expansion of the TE family. Step 3 corresponds to the long-term establishment of the TE family: Due to the implementation of the piRNA machinery, most of the TE insertions are now fixed and not able to transpose. The piRNA response will persist until step 4 as observed in *Drosophila simulans*. Step 4 corresponds to the very long-term dynamics in which most of the TE insertions are found completely degenerated and fragmented. TEs are progressively removed from the genome and the loss of the fragmented copies inserted in the piRNA cluster lead to the progressive loss of the genomic piRNA response. Once completely lost, the cycle is over and a new reinvasion can occur.

2000; Toll-Riera et al. 2009; Wilson Sayres and Makova 2011; Wissler et al. 2013). In the future, it would be challenging to test to what extant TEs in testes can constitute a force that facilitates genetic rearrangements and thus, to what extant TEs in testes enlarge the selective spectrum required to promote diversifying selection and adaptive innovations.

## Conclusion

In this work, we found that the tempo and the dynamics of TE invasion are clearly different between two closely related species of *Drosophila*: *D. simulans* have experienced ancient waves of TE invasion, whereas *D. melanogaster* still undergo recent TE bursts. We proposed that trajectories of TE invasion have strongly affected the host defense machinery involved in TE silencing through the production specific pools of piRNAs. Moreover, we found that the “postinvasion” piRNA-mediated response is dramatically enhanced in ovaries compared

with testes. Therefore, we proposed a dynamic model describing how the piRNA silencing machinery takes place through the female germline (fig. 6).

In the early stage of the invasion, new TE insertions are characterized by a high insertion polymorphism and fixations at a given locus are rare. As a consequence, these recently expanding TE families are rarely targeted by piRNAs. Alongside the TE amplification process, the targeted piRNA response becomes progressively active as some insertions become fixed into piRNA clusters. This step corresponds to what we observe in *D. melanogaster*. Thereafter, when piRNA post-transcriptional silencing is stably established, the mobilization of active TEs considerably slows down. In the long term, as observed in *D. simulans*, although active copies will progressively degenerate and finally disappear (except few relic copies in piRNA clusters), the persistence of the piRNA response will act as a long-term genomic memory to protect the genomes from future invasions.

## Materials and Methods

### *Drosophila* Stocks

*Drosophila* natural populations were collected from geographically distinct areas. We used *D. melanogaster* populations from Zimbabwe (Harare) and Gotheron (France) and *D. simulans* populations from Nairobi (Kenya) and Fukuoka (Japan) for transcriptome analysis. Populations were maintained at 25 °C from the date of capture as well as laboratory strains. We also used the laboratory strains M19 and  $w^{1118}$  of *D. melanogaster*, the strain  $w^{501}$  and natural population from Makindu (Kenya) and Chicharo (Portugal) of *D. simulans* for small RNA sequencing analysis. The list of populations and strains is given in [supplementary table S1, Supplementary Material](#) online.

### mRNA Library Preparation and Small RNA Library Preparation

We extracted total RNA from 30 pairs of ovaries and testes in 2–4-day-old adults, according to the manufacturer's instructions (Macherey-Nagel). PolyA mRNAs were extracted using the "FastTrack MAG Micro mRNA isolation kit" (Life Technologies), fragmented with RNA fragmentation reagents (Ambion), and treated with antarctic phosphatase (NEB) and polynucleotide kinase (NEB), according to the recommendations. We prepared strand-orientated libraries with the "Truseq Small RNA sample prep Kit" (Illumina). The final gel purification step has been replaced by a polymerase chain reaction cleanup with AMPureXP beads (Beckman-Coulter). We then proceeded to mRNA libraries illumina sequencing.

Small RNAs were extracted from total RNA of 50 pairs of ovaries and 100 pairs of testes, dissected from 2- to 4-day-old adults, using a TRIzol extraction according to the manufacturer's procedure (TRIzol reagent, Invitrogen). We size fractionated small RNAs from 1 µg total RNA on a TBE-urea 15% acrylamide gel. We treated the resulting RNAs with the Illumina "Truseq Small RNA sample prep Kit" according to the manufacturer's recommendations and send small RNA libraries to deep sequencing.

### Sequencing

Libraries were prepared and sequenced by the IMAGIF sequencing platform (Gif-sur-Yvette—France) on an Illumina HiSeq 1000 instrument, with a TruSeq SR Cluster Kit v3-cBot-HS (Illumina) and a TruSeq SBS v3-HS—50 cycles Kit (Illumina), using a single read 50-bp recipe. Libraries were pooled in equimolar proportions and diluted libraries to a final concentration of 12 pM, according to Illumina recommendations. The data were demultiplexed using the distribution of CASAVA software (CASAVA-1.8.2) (Mortazavi et al. 2008). The quality of the data was checked with the software FastQC 0.10.1 (available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

### Read Mapping and Differential Expression Analysis

We filtered single-end reads from each library with UrQt software (Modolo and Lerat 2015); we retained only high quality reads (phred score > 33) for the analyses. We remove Illumina adapters using scythe software (<https://github.com/najoshi/>) and kept reads with a minimal length of 15 nucleotides. Quality-control data are presented in [supplementary table S1, Supplementary Material](#) online. We performed RNAseq mappings using the STAR software (Dobin et al. 2013) on the *Drosophila* TEs and reference genomes. We identified mRNA transcripts from TEs by mapping reads against a custom TE library derived from Repbase (Jurka et al. 2005). This database contains consensus sequences of all known TE in the *Drosophila* genomes. We kept reads mapping to a single consensus sequence (i.e., one TE family) and we generated count tables of TE transcripts with HTSeq. We performed differential expression analysis from these count matrices using the R bioconductor package DESeq2 (Love et al. 2014). The package DESeq2 implements a generalized linear model in which counts for each gene, in each sample, are modeled using a negative binomial distribution. We used one factor generalized linear model formulas depending on the tested conditions (i.e., species, population, or gonad). We selected differentially expressed TEs according to their adjusted *P* value (corrected *P* value < 0.1 and 10% false discovery rate). We compared DESeq2 results with another normalization method: TE expression was normalized by a pool of 100 housekeeping genes with stable expression between ovaries and testes samples (not shown). DESeq2 results were more stringent; we thus kept these in the main text. We used the same procedure to find differentially expressed genes mapping reads to the *D. melanogaster* and *D. simulans* reference genomes. Differential expression analysis of piRNA pathway genes was performed on count matrices containing all orthologous genes shared by the two sibling species using DESeq2. Between versus within specific variations of TE transcription levels were compared using PCAs (implemented in DESeq2) on TE count matrices.

### Small RNA Mapping and Analysis

We removed barcodes and adapters from small RNA libraries of testes and ovaries using the Cutadapt tool and reads that are between 5 and 45 bp after stripping were kept. For each sample, we characterized small RNA species ([supplementary table S1, Supplementary Material](#) online). Then, we cleaned small RNA libraries from contaminant mRNA species (tRNA, rRNA, and genic mRNA in sense orientation). We identified TE-derived small RNA mapping small RNAs libraries (19–30 nt) to our set of sex-biased TEs and nonsex-biased TEs by using bowtie (Langmead et al. 2009), allowing up to three mismatch and multiple matches to one position (-v [3] -M 1 -best -strata -p 12). The same analysis was performed with 0 mismatch ([supplementary fig. S6, Supplementary Material](#)

online). All reads mapping to a unique TE consensus were pooled, and reads mapping to more than one TE consensus were discarded. To account for differences in sequencing depth between libraries and levels of sample contamination, the number of piRNA per TE family was normalized by the total number of miRNAs, a piRNA comigrant RNA species (supplementary table S3, Supplementary Material online). The ping-pong signature is the probability that a randomly sampled piRNA from a given TE family has an antisense binding RNA overlapping on the first 10 bp. It was calculated using the tool described in Antoniewski (2014). In order to estimate piRNA variation between populations and strains, we downloaded several sets of small RNA and compared them with our sequenced libraries (supplementary figs. S4 and S7, Supplementary Material online). To this end, we used two small RNA libraries presented supplementary table S1, Supplementary Material online.

### Comparison of Genome TE Content and Genome Annotation of TEs

We used the *Drosophila* Repbase data set (2,289 TE consensus) to identify TE insertions on recent releases of *D. melanogaster* and *D. simulans* reference genomes (dmel-r6.17 and dsim-r2.02) from which we removed contigs <15 kb (size of the longest TE in the data set).

From the Repbase TE list, we first discarded consensus sequences for which no RNAseq reads could map (449 sequences for *D. melanogaster* and 400 sequences for *D. simulans*) as queries for BlastN searches with default parameters (BlastN, e-value 10) on full genomes of both species. We removed all BLAST hits sharing <80% identity with TE consensus and merged successive hits belonging to the same TE family, when overlapping or when the lengths of the hits plus the gap distance in between were inferior to the size the TE consensus. When overlaps concerned different TE families, we kept the TE family with the highest identity to the consensus. In rare cases, some TE families displayed both overlapping and nonoverlapping regions. Insertions of this type were treated as two independent insertions. The final annotation files are summarized in supplementary table S2, Supplementary Material online.

For each TE family, we aligned copies to their consensus using the Geneious mapper (high sensitivity; Geneious 10.2.3) and calculated the nucleotide diversity ( $\pi$ ) using a custom script and the expression:

$$\pi = 2 \cdot \sum_{i=1}^n \sum_{j=1}^{i-1} x_i x_j \pi_{ij}$$

where  $x_i$  and  $x_j$  are the respective frequencies of the  $i$ th and  $j$ th sequences,  $\pi_{ij}$  is the number of nucleotide differences per nucleotide site between the  $i$ th and the  $j$ th sequences, and  $n$  is the number of sequences per TE family. All consensus

genomic features with both normalized mRNA and piRNA expression features are summarized supplementary table S3, Supplementary Material online, respectively, for *D. melanogaster* and *D. simulans*. Supplementary table S3, Supplementary Material online, also sum-up Pearson' correlation statistics performed to highlight relationships between piRNAs variations, TE genomic, and transcriptomic features.

### piRNA Cluster Analysis

We extracted piRNAs (23–30 nt) from our cleaned small RNA libraries and mapped these to both species reference genomes (dmel-r6.17 and dsim-r2.02). We mapped piRNAs using bowtie, allowing up to one mismatch. We used a 1-kb window to identify all regions with densities greater than five piRNA/kb. Only piRNAs that uniquely mapped to the cluster were retained. Presence or absence of TEs was analyzed along each 1-kb window containing uniquely mapping piRNAs. We obtained the total TE density along piRNA cluster by averaging TE' presence/absence by nucleotide position along 1-kb piRNA cluster windows. We performed this analysis for each class TE according to their differential expression pattern. We summarized piRNA cluster mapping for testes and ovaries in supplementary table S4, Supplementary Material online.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

This work was supported by the APEGE program of the CNRS in environmental genomics. We thank Christophe Antoniewski who provided computational resources and tools to perform piRNA analyses. We also thank Cristina Vieira' lab that helped for the preparation of the samples used for testes and ovaries RNA sequencing. We also thank Malcolm Eden for the English review of the text.

### Author Contributions

The experiments were conceived and designed by J.F. and B.S.-L. The experiments were performed by B.S.-L. and J.F. Data analysis and presentation were performed by J.F., B.S.-L., and A.H.-V. The paper was written by B.S.-L., P.C., J.F., and A.H.-V.

### Literature Cited

- Akkouche A, et al. 2013. Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* 14(5):458–464.  
 Antoniewski C. 2014. Computing siRNA and piRNA overlap signatures. *Methods Mol Biol.* 1173:135–146.

- Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318(5851):761–764.
- Bartolome C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol.* 10(2):R22.
- Bartolome C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol.* 19(6):926–937.
- Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol.* 42(1):251–269.
- Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 104(27):11340–11345.
- Berry B, Deddouche S, Kirschner D, Imler JL, Antoniewski C. 2009. Viral suppressors of RNA silencing hinder exogenous and endogenous small RNA pathways in *Drosophila*. *PLoS One* 4(6):e5866.
- Blumenstiel JP. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet.* 27(1):23–31.
- Bowen NJ, McDonald JF. 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* 11(9):1527–1540.
- Brennecke J, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128(6):1089–1103.
- Brennecke J, et al. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322(5906):1387–1392.
- Castaneda J, Genzor P, Bortvin A. 2011. piRNAs, transposon silencing, and germline genome integrity. *Mutat Res.* 714(1–2):95–104.
- Chambeyron S, et al. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci U S A.* 105(39):14964–14969.
- Connallon T, Clark AG. 2010. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution (N Y).* 64(12):3417–3442.
- Connallon T, Knowles LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* 21(9):495–499.
- Czech B, et al. 2018. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet.* 52(1):131–157.
- Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.
- Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* 178(4):2169–2177.
- Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8(9):689–698.
- Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464(7293):1347–1350.
- Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol.* 6(10):e251.
- Grentzinger T, et al. 2012. piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res.* 22(10):1877–1888.
- Gunawardane LS, et al. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315(5818):1587–1590.
- Handler D, et al. 2013. The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell* 50(5):762–777.
- Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res.* 17(12):1837–1849.
- Heinen TJ, Staubach F, Haming D, Tautz D. 2009. Emergence of a new gene from an intergenic region. *Curr Biol.* 19(18):1527–1531.
- Hey J, Kliman RM. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol Biol Evol.* 10:804–822.
- Houwing S, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129(1):69–82.
- Hurst LD, Ellegren H. 1998. Sex biases in the mutation rate. *Trends Genet.* 14(11):446–452.
- Jangam D, Feschotte C, Betran E. 2017. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* 33(11):817–831.
- Junakovac N, Terrinoni A, Di Franco C, Vieira C, Loevenbruck C. 1998. Accumulation of transposable elements in the heterochromatin and on the Y chromosome of *Drosophila simulans* and *Drosophila melanogaster*. *J Mol Evol.* 46(6):661–668.
- Jurka J, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110(1–4):462–467.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20(10):1313–1326.
- Kaminker JS, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3(12):research0084.1.
- Kawaoka S, et al. 2009. The *Bombyx* ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* 15(7):1258–1264.
- Kelleher ES, Barbash DA. 2013. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. *Mol Biol Evol.* 30(8):1816–1829.
- Khurana JS, et al. 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* 147(7):1551–1563.
- Kidwell MG, Lisch DR. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution (N Y).* 55(1):1–24.
- Kofler R. 2019. Dynamics of transposable element invasions with piRNA clusters. *Mol Biol Evol.* 36(7):1457–1472.
- Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. 2015. The recent invasion of natural *Drosophila simulans* populations by the P-element. *Proc Natl Acad Sci U S A.* 112(21):6659–6663.
- Kofler R, Nolte V, Schlotterer C. 2015. Tempo and mode of transposable element activity in *Drosophila*. *PLoS Genet.* 11(7):e1005406.
- Kolaczowski B, Hupaló DN, Kern AD. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol Biol Evol.* 28(2):1033–1042.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):R25.
- Lau NC, et al. 2006. Characterization of the piRNA complex from rat testes. *Science* 313(5785):363–367.
- Le Rouzic A, Deceliere G. 2005. Models of the population genetics of transposable elements. *Genet Res.* 85(3):171–181.
- Lee YC, Langley CH. 2012. Long-term and short-term evolutionary impacts of transposable elements on *Drosophila*. *Genetics* 192(4):1411–1432.
- Lerat E, Burret N, Biemont C, Vieira C. 2011. Comparative analysis of transposable elements in the *melanogaster* subgroup sequenced genomes. *Gene* 473(2):100–109.
- Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. 2017. TETools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* 45(4):e17.
- Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A.* 103(26):9935–9939.



- Light S, Basile W, Elofsson A. 2014. Orphans and new gene origination, a structural and evolutionary perspective. *Curr Opin Struct Biol.* 26:73–83.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550.
- Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res.* 20(2):212–227.
- Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* 136(4):656–668.
- Modolo L, Lerat E. 2015. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC Bioinformatics* 16(1):137.
- Mohn F, Sienski G, Handler D, Brennecke J. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157(6):1364–1379.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 5(7):621–628.
- Nagao A, et al. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. *RNA* 16(12):2503–2515.
- Nishida KM, et al. 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA* 13(11):1911–1922.
- Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5(10):e1000698.
- Pane A, Wehr K, Schupbach T. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell* 12(6):851–862.
- Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14(2):83–87.
- Paulding CA, Ruvolo M, Haber DA. 2003. The Tre2 (USP6) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci U S A.* 100(5):2507–2511.
- Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol.* 20(6):880–892.
- Ponce R, Hartl DL. 2006. The evolution of the novel Sdic gene cluster in *Drosophila melanogaster*. *Gene* 376(2):174–183.
- Quénerch' du E, Anand A, Kai T. 2016. The piRNA pathway is developmentally regulated during spermatogenesis in *Drosophila*. *RNA* 22(7):1044–1054.
- Robine N, et al. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol.* 19(24):2066–2076.
- Saint-Leandre B, Clavereau I, Hua-Van A, Capy P. 2017. Transcriptional polymorphism of piRNA regulatory genes underlies the mariner activity in *Drosophila simulans* testes. *Mol Ecol.* 26(14):3715–3731.
- Sanchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet.* 21(4):200–203.
- She X, et al. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature* 430(7002):857–864.
- Siomi MC, Saito K, Siomi H. 2008. How selfish retrotransposons are silenced in *Drosophila* germline and somatic cells. *FEBS Lett.* 582(17):2473–2478.
- Song J, et al. 2014. Variation in piRNA and transposable element content in strains of *Drosophila melanogaster*. *Genome Biol Evol.* 6(10):2786–2798.
- Toll-Riera M, et al. 2009. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol.* 26(3):603–612.
- Vieira C, Lepetit D, Dumont S, Biemont C. 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol.* 16(9):1251–1255.
- Vieira C, et al. 2012. A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *J Environ Radioact.* 113:83–86.
- Wilson Sayres MA, Makova KD. 2011. Genome analyses substantiate male mutation bias in many species. *BioEssays* 33(12):938–945.
- Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E. 2013. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol Evol.* 5(2):439–455.

Associate editor: Josefa Gonzalez