

Transposable Element Misregulation Is Linked to the Divergence between Parental piRNA Pathways in Drosophila Hybrids

Valèria Romero-Soriano, Laurent Modolo, Hélène Lopez-Maestre, Bruno Mugat, Eugénie Pessia, Séverine Chambeyron, Cristina Vieira, Maria Pilar Garcia Guerreiro

▶ **To cite this version:**

Valèria Romero-Soriano, Laurent Modolo, Hélène Lopez-Maestre, Bruno Mugat, Eugénie Pessia, et al.. Transposable Element Misregulation Is Linked to the Divergence between Parental piRNA Pathways in Drosophila Hybrids. Genome Biology and Evolution, 2017, 9 (6), pp.1450 - 1470. 10.1093/gbe/evx091 . hal-01596041

HAL Id: hal-01596041
https://hal.science/hal-01596041

Submitted on 9 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Transposable Element Misregulation Is Linked to the Divergence between Parental piRNA Pathways in *Drosophila* Hybrids

Valèria Romero-Soriano,[1] Laurent Modolo,[2] Hélène Lopez-Maestre,[2] Bruno Mugat,[3] Eugénie Pessia,[2] Séverine Chambeyron,[3] Cristina Vieira,[2] and Maria Pilar Garcia Guerreiro[1,*]

[1]Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Spain

[2]Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Claude Bernard Lyon 1, Villeurbanne, France

[3]Institut de Génétique Humaine, UMR9002, CNRS-Université de Montpellier, France

*Corresponding author: E-mail: mariapilar.garcia.guerreiro@uab.cat.

Accepted: May 9, 2017

Data deposition: Raw sequencing data generated from this article have been deposited at the NCBI Sequence Read Archive (SRA) under accessions SRP096015 and SRP095995.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Abstract

Interspecific hybridization is a genomic stress condition that leads to the activation of transposable elements (TEs) in both animals and plants. In hybrids between *Drosophila buzzatii* and *Drosophila koepferae*, mobilization of at least 28 TEs has been described. However, the molecular mechanisms underlying this TE release remain poorly understood. To give insight on the causes of this TE activation, we performed a TE transcriptomic analysis in ovaries (notorious for playing a major role in TE silencing) of parental species and their F1 and backcrossed (BC) hybrids. We find that 15.2% and 10.6% of the expressed TEs are deregulated in F1 and BC1 ovaries, respectively, with a bias toward overexpression in both cases. Although differences between parental piRNA (Piwi-interacting RNA) populations explain only partially these results, we demonstrate that piRNA pathway proteins have divergent sequences and are differentially expressed between parental species. Thus, a functional divergence of the piRNA pathway between parental species, together with some differences between their piRNA pools, might be at the origin of hybrid instabilities and ultimately cause TE misregulation in ovaries. These analyses were complemented with the study of F1 testes, where TEs tend to be less expressed than in *D. buzzatii*. This can be explained by an increase in piRNA production, which probably acts as a defence mechanism against TE instability in the male germline. Hence, we describe a differential impact of interspecific hybridization in testes and ovaries, which reveals that TE expression and regulation are sex-biased.

Key words: transposable elements, piRNAs, interspecific hybridization, RNA-seq, *Drosophila buzzatii*, *Drosophila koepferae*.

Introduction

Transposable elements (TEs) are mobile DNA fragments that are dispersed throughout the genome of the vast majority of both prokaryotic and eukaryotic organisms. Their capacity to mobilize, together with their repetitive nature, confers them a high mutagenic potential. TE insertions can be responsible for the disruption of genes or regulatory sequences, and can also cause chromosomal rearrangements, representing a threat to their host genome integrity (Hedges and Deininger 2007). To mitigate these deleterious effects, mechanisms of TE control are especially important in the germline, where novel insertions (as well as other mutations) can be transmitted to the progeny (Iwasaki et al. 2015; Czech and Hannon 2016).

Animal genomes have developed a TE silencing system, the piRNA (Piwi-interacting RNA) pathway (Klattenhoff and Theurkauf 2008; Brennecke and Senti 2010), that acts in the germline at both posttranscriptional and transcriptional levels (Rozhkov et al. 2013). piRNA templates form specific genomic clusters, whose transcription produces long piRNA precursors that are cleaved to produce primary piRNAs

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

(Brennecke et al. 2007). The resulting piRNAs can initiate an amplification loop called the ping-pong cycle, giving rise to secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). A third kind of piRNAs are produced by phased cleavage of piRNA cluster transcript remnants that have first been processed during secondary piRNA biogenesis (Han et al. 2015; Mohn et al. 2015). In the soma, another small-RNA mediated silencing system, the endo-siRNA (endogenous small interference RNA) pathway, has been shown to be involved in posttranscriptional silencing of TEs (Ghildiyal et al. 2008).

These strong mechanisms of TE regulation can be relaxed under different stress conditions, leading to unexpected TE mobilization events (García Guerreiro 2012). Hybridization between species causes genomic stress, which can lead to several genome reorganizations that seem to be driven by TEs (Fontdevila 2005; Michalak 2009; García Guerreiro 2014; Romero-Soriano et al. 2016). In the literature, several cases of TE proliferation in interspecific hybrids have been reported for a wide range of species, including plants (Liu and Wendel 2000; Ungerer et al. 2006; Wang et al. 2010) as well as animals (Evgen'ev et al. 1982; O'Neill et al. 1998; Metcalfe et al. 2007). Studies describing an enhanced TE expression in hybrids suggest that this may be caused by a TE silencing breakdown (Kelleher et al. 2012; Carnelossi et al. 2014; Dion-Côté et al. 2014; Renaut et al. 2014; García Guerreiro 2015; Lopez-Maestre et al. 2017). In this work, we propose two possible explanatory hypotheses—not mutually exclusive—to understand this breakdown, as the molecular mechanisms allowing TE releases in hybrids remain unknown.

The first hypothesis, which we call the maternal cytotype failure, recalls the hybrid dysgenesis phenomenon (Picard 1976; Kidwell et al. 1977), where an increase of TE activity is observed. This occurs when *Drosophila* females devoid of piRNAs against a particular TE family are mated with males containing active copies of the cognate TE (Brennecke et al. 2008; Chambeyron et al. 2008), because maternally deposited piRNAs are crucial to initiate an efficient TE silencing response in the progeny (Grentzinger et al. 2012). In the same logic, differences between parental species' piRNA pools could lead to a transcriptional activation of some paternally inherited TEs in interspecific hybrids. Under this hypothesis, only a subset of TE families, specific to the male species, would be deregulated after hybridization.

The second hypothesis claims that a global failure of the piRNA pathway is responsible for the observed TE activation in hybrids. It has been shown that piRNA pathway effector proteins show adaptive evolution marks (Obbard et al. 2009; Simkin et al. 2013) and their expression levels can significantly differ between different populations of the same *Drosophila* species (Fablet et al. 2014). Thus, genetic incompatibilities involving this pathway could arise even between closely related species. The accumulated functional divergence of these proteins would cause widespread transcriptional TE derepression, as suggested in *Drosophila melanogaster–Drosophila simulans* artificial (*Hmr*-rescued) hybrids (Kelleher et al. 2012).

In order to test these hypotheses and provide new insight into the mechanisms underlying TE activation in hybrids, we have performed a whole-genome study of TE expression and regulation using the species *Drosophila buzzatii* and *Drosophila koepferae* (*buzzatii* complex, *repleta* group). Hybridization between these two species can occur in nature (Gomez and Hasson 2003; Piccinali et al. 2004; Franco et al. 2010), giving rise to fertile females that can be backcrossed with parental males (Marín and Fontdevila 1998). Reticulation events thus provide a source of genetic variability that certainly has influenced the evolutionary history of *D. buzzatii* and *D. koepferae*. On the contrary, hybrids from other species' pairs, such as the best-studied *D. melanogaster* and *D. simulans*, cannot be backcrossed (Barbash 2010) and hence are considered evolutionary dead-ends. Hence, our model is particularly suitable and relevant for hybridization and speciation studies. Furthermore, several TE mobilization events have previously been detected in *D. buzzatii–D. koepferae* hybrids by in situ hybridization (Labrador et al. 1999), amplified fragment length polymorphism markers (Vela et al. 2011), and/or transposon display (Vela et al. 2014). At least two of the mobilized elements, the retrotransposons *Osvaldo* and *Helena*, present abnormal patterns of expression in hybrids (García Guerreiro 2015; Romero-Soriano and García Guerreiro 2016), pointing to a failure of TE silencing.

Here we demonstrate that 15.2% of the expressed TE families are deregulated in F1 hybrid ovaries, in most cases overexpressed. This proportion decreases to 10.6% after a generation of backcrossing. However, even if differences between parental piRNA pools can explain the changes in expression of some TE families, they do not account for the whole pattern of deregulation. Accordingly, our analyses of genomic TE content show that parental TE landscapes are very similar, and hence big differences in their piRNA populations are not expected. On the other hand, we demonstrate that the piRNA pathway proteins are particularly divergent between *D. buzzatii* and *D. koepferae*, which seems to lead to dissimilarities in their piRNA production strategies. Interestingly, a high proportion of the TEs overexpressed in hybrids do not have associated piRNA populations in parents (nor in hybrids), pointing out a complex TE deregulation network where a failure of the piRNA pathway together with other TE silencing mechanisms would take place. Finally, we show that the effects of hybridization are sex-biased, as in testes (contrarily to ovaries) TE deregulation is globally biased toward underexpression, which can be explained by a higher production of piRNAs in hybrid males.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Materials and Methods

Drosophila Stocks and Crosses

Interspecific crosses were performed between males of *D. buzzatii* Bu28 strain, an inbred line originated by the union of different populations (LN13, 19, 31, and 33) collected in 1982 in Los Negros (Bolivia); and females of *D. koepferae* Ko2 strain, an inbred line originated from a population collected in 1979 in San Luis (Argentina). Both lines were maintained by brother–sister mating for more than a decade and are now kept by mass culturing.

We performed 45 different interspecific crosses of 10 *D. buzzatii* males with 10 *D. koepferae* virgin females (in order to obtain F1 individuals, fig. 1), then 30 backcrosses of 10 *D. buzzatii* males with 10 hybrid F1 females (which gave rise to BC1 females, fig. 1). All stocks and crosses were reared at 25 °C in a standard *Drosophila* medium supplemented with yeast.

RNA Extraction, Library Preparation, and Sequencing

Flies were dissected in PBT (1× phosphate-buffered saline, 0.2% Tween 20), 5–6 days after their birth. Total RNA was purified from testes ($n = 30$ pairs per sample for *D. buzzatii* and $n = 45$ pairs per sample for F1 hybrids) or ovaries ($n = 20$ pairs per sample) with the Nucleospin RNA purification kit (Macherey-Nagel). RNA quality and concentration were evaluated using Experion Automated Electrophoresis System (Bio-rad), in order to keep only high-quality samples. Two Illumina libraries of 250–500 bp fragments were prepared for each kind of sample (*D. buzzatii*, *D. koepferae*, F1 and BC1 ovaries; and *D. buzzatii* and F1 testes), using 2 µg of purified RNA. Duplicate libraries correspond to biological replicates (ovaries from different crosses and separate RNA extractions). Sequencing was performed using the Illumina mRNA-seq paired-end protocol on a HiSeq2000 platform, at the INRA–UMR AGAP (Montpellier, France). We obtained 53.5–59.1 million paired-end reads for each sample (divided into two replicates) resulting in a total of 332.7 million paired-end reads.

Assembly and Annotation

A de novo reference transcriptome was constructed for each of our target species using Trinity r2013_08_14 (Grabherr et al. 2011) with options *–group_pairs_distance* 500 and *–min_kmer_cov* 2. All contigs were aligned to *D. buzzatii* genome (Guillén et al. 2015) using BLAT v.35x1 (Kent 2002), with parameters *–minIdentity = 80* and *–maxIntron = 75000*, in order to identify chimers. Contigs that aligned partially (≤60%) on up to three genomic locations with a total alignment coverage of ≥80% were considered chimeric and split consequently.

To annotate protein-coding genes, all contigs of both transcriptomes were aligned against the *D. buzzatii*-predicted gene models and the *D. buzzatii* genome (Guillén et al. 2015) using BLAT v.35x1 (same parameters as before). This approach allows us to identify untranslated regions and double-check the genomic position associated with a contig. Only contigs with alignment coverages ≥70% and whose best hit genomic coordinates overlapped in both alignments were annotated. The same approach was applied to the remaining nonannotated contigs with *Drosophila mojavensis*' gene models. The rest of the contigs were clustered using CD-HIT v4.5.4 (Fu et al. 2012) with options *-c 0.8*, -T 0, *-aS 0.8*, *-A 80*, *-p 1*, *-g 1*, *-d 50*; and annotated with the name of the longest sequence of each cluster. Supplementary table S1, Supplementary Material online, depicts a summary of annotation statistics.

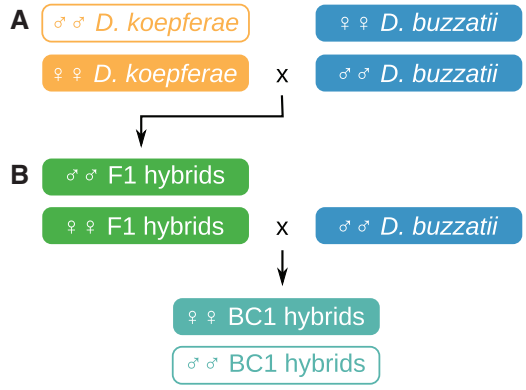


Fig. 1.—Crosses diagram. (*A*) is the first interspecific cross between *D. koepferae* (orange) females and *D. buzzatii* (blue) males, and (*B*) is the backcross between F1 hybrid (green) females and *D. buzzatii* (blue) males that gives rise to BC1 (turquoise). Colors have been assigned according to the *D. buzzatii*/*D. koepferae* genome content: orange for *D. koepferae*, blue for *D. buzzatii*, green for F1 hybrids, and turquoise for BC1 hybrids. Samples marked with a white background rectangle have not been sequenced.

TE Library Construction

Our library is mainly constituted by the list of all TE copies masked in the *D. buzzatii* genome (Rius et al. 2016), as *D. koepferae* has not until now been sequenced. In order to overcome this bias (deeply discussed in supplementary text S1, Supplementary Material online) and increase specificity in further analyses, we sought to have a better representation of *D. koepferae* TE landscape in our TE list. To this aim, we annotated TE transcripts from our de novo assemblies by aligning them to a consensus TE library (the same used to mask the *D. buzzatii* genome) using BLAT v.35x1. Contigs whose alignments covered ≥80% of their sequences with a minimum 80% identity and ≥80 bp long (*three 80 criteria*) were kept as TE transcripts and included in our TE library. To improve our coverage and sensitivity to detect poorly expressed TEs, a third de novo assembly, using all the reads from

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

all sequenced samples (from both parents and hybrids), was performed and annotated as described above.

This resulted in 65,772 final TE copies belonging to 699 TE families, which were assigned to only 658 families after two steps of clustering. Clustering was performed using the *three 80 criteria*; manually through BLAT alignments, and automatically using CD-HIT v4.5.4 (same parameters as in gene annotation). These 658 families were divided into five categories, following Repbase classification (Jurka et al. 2005): LTR and LINE (class I), DNA and RC (class II), and Unknown (unclassified).

In order to assess the biological reliability of our Trinity assemblies, we tried to amplify the most expressed TE-annotated contig of the ten most expressed TE families by RT-PCR. Nine out of these ten reactions (all except that of *Homo5*) gave rise to the expected amplicon, which was recognized by its size (supplementary file S1, Supplementary Material online). Amplification of unspecific sequences was also observed for nine of the reactions (all except that of *rnd5_family-1078*, including that of *Homo5*). These results are not surprising because different copies of the same TE family can share sequence identity while having different lengths (due, for example, to internal deletions).

Small RNA Extraction, Library Preparation, and Sequencing

Small RNA was purified from ovaries ($n = 70$ pairs for all samples) and testes ($n = 96$ pairs for *D. buzzatii* and $n = 333$ pairs for F1 sterile males), following the manual small RNA purifying protocol described by Grentzinger et al. (2013), which significantly reduces endogenous contamination and degradation products abundance. After small RNA isolation, samples were gel-purified and precipitated. A single Illumina library was prepared for each sample and sequencing was performed on an Illumina Hiseq 2500 platform by FASTERIS SA (Switzerland). We obtained a total of 401.1 million reads (21.4–58.7 million reads per sample). Reads of 23–30 nt were kept as piRNAs, reads of 21 nt were considered endo-siRNAs, and reads of 21–23 nt were filtered out as putative microRNAs (miRNAs).

TE Analyses: Read Mapping and Differential Expression

All our sequencing data were treated and analyzed with the TEtools pipeline (Lerat et al. 2017; https://github.com/l-mod olo/TEtools) as described thereafter. First, data were trimmed using UrQt (Modolo and Lerat 2015) in order to remove polyA tails (from RNA-seq reads) and low-quality nucleotides (from both RNA-seq and smallRNA-seq reads). The resulting trimmed reads were aligned to our TE library using Bowtie2 v2.2.4 for RNA-seq (Langmead and Salzberg 2012) and Bowtie1 v1.1.1 for piRNAs and endo-siRNAs (Langmead et al. 2009), with the most sensitive option and keeping a single alignment for reads mapping to multiple positions (*–very-sensitive* for Bowtie2 and *-S* for Bowtie). The read count step was computed per TE family, adding all reads mapped on copies of the same family. Finally, we performed the differential expression analyses between TE families using the R Bioconductor package DESeq2 (Love et al. 2014) on the raw read counts, using the Benjamini–Hochberg multiple test correction (FDR level of 0.1, Benjamini and Hochberg 1995). DESeq2 models the read counts using a negative binomial distribution. We use a likelihood ratio test that compares the full model (which accounts for the differences between samples) with the reduced model (which does not). If the reduced model is rejected in favor of the full model after multiple testing correction (at an FDR level of 0.1), the TE family is considered differentially expressed between samples (Love et al. 2014). Statistical summaries of these analyses are available in supplementary files S2 (RNA-seq) and S3 (smallRNA-seq), Supplementary Material online, including both raw and normalized read count tables. TE families with $\leq$10 aligned reads per sample are considered to be unexpressed in the text. For piRNA and endo-siRNA analyses, no significant differences could be detected at the TE family level due to the lack of replicates, leading us to perform the analyses using fold change (FC) values. In these studies, TE families with $\geq$2-fold differences in their piRNA/endo-siRNA population levels between hybrids and both parental species were considered misregulated.

Gene Analyses: Read Mapping, Differential Expression, and Gene Ontology Enrichment

Gene expression analyses were performed following the same approach used for TEs. RNA-seq reads were aligned against the addition of *D. buzzatii* and *D. koepferae* transcriptomes, and read count was computed per annotated gene (by adding all reads mapped on contigs with the same annotation). We discarded genes with $\leq$10 aligned reads per sample and considered them to be unexpressed in the text.

Trinity's tool TransDecoder (Haas et al. 2013) was employed to predict ORFs within *D. buzzatii* and *D. koepferae* transcriptomes, using Pfam-A database v.29 (Punta et al. 2012). Then, we performed a functional annotation of the resulting proteomes using Gene Ontology (GO) terms (Gene Ontology Consortium 2000) with the eggnog-mapper tool (https://github.com/jhcepas/eggnog-mapper). First, we mapped our sequences to eggNOG orthologous groups from eukaryotic, bacterial, and archaeal databases (Huerta-Cepas et al. 2016) using an *e*-value of 0.001. Then, we transferred the GO terms of the best orthologous group hit for each gene. GO enrichments for deregulated genes in hybrids were analyzed using the Topology-Weighted method built in Ontologizer (Bauer et al. 2008), with a *P*-value threshold of 0.01. GO terms with $\geq$2-fold enrichments shared by F1- and BC1-deregulated genes are listed in the supplementary file S4*A*, Supplementary Material online.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

miRNA Analyses: Read Mapping, Differential Expression, and GO Enrichment

In order to assess the effect of hybridization on miRNA populations, we mapped small RNA-seq reads of 21–23 nt in length to the list of *D. mojavensis* mature miRNAs (Drosophila 12 Genomes Consortium 2007) available in miRBase (Griffiths-Jones et al. 2008; Kozomara and Griffiths-Jones 2011). *Drosophila mojavensis* was chosen as the most closely related species to *D. buzzatii* and *D. koepferae* for which a miRNA complement has been described. We carried out read mapping, read count, and differential expression analyses using the same methodology described for TEs and genes (sequences with $\leq$10 aligned reads per sample are also considered to be unexpressed).

As for piRNA and endo-siRNA studies, we considered sequences with $\geq$2-fold differences in their miRNA levels between hybrids and both parental species to be misregulated. In order to identify the targets of each deregulated miRNA, we first determined its *D. melanogaster* homolog using miRBase (Kozomara and Griffiths-Jones 2014) and then employed the miRanda–mirSVR target prediction tool with an mirSVR score cutoff of ≤ -0.6 (Betel et al. 2010). Finally, we performed a GO enrichment analysis using the GOrilla online tool and the two unranked list of genes running mode (Eden et al. 2009), with an FDR *q*-value threshold of 0.1. The background list consisted in all the genes targeted by miRNA families in *D. melanogaster* given by the microRNA.org resource (Betel et al. 2008) with an mirSVR score ≤ -0.6. A summary of the obtained results (including the FC values of each hybrid-deregulated miRNA and a list of the enriched GO terms of their putative targets) is available in the supplementary file S4*B*, Supplementary Material online.

Parental Species' TE Landscapes

We examined the repeatomes of *D. buzzatii* and *D. koepferae* using dnaPipeTE pipeline (Goubert et al. 2015), which assembles repeats from low coverage genomic NGS data and annotates them with RepeatMasker Open-4.0 (Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2010, http://www.repeat- masker.org, last accessed February 24, 2016) and Tandem repeats finder (Benson 1999). We employed Repbase library version 2014-01-31 (Jurka et al. 2005). For both species, two iterations were performed using a read sample size corresponding to a genome coverage of 0.25× (Guillén et al. 2015), according to genome size estimates in Romero-Soriano et al. (2016). Because mitochondrial DNA is usually assembled, we aligned all dnaPipeTE contigs to BLAST nucleotide collection (McGinnis and Madden 2004) to distinguish nuclear from mitochondrial sequences. We identified reads mapping to mitochondrial contigs using Bowtie2 with default parameters (Langmead and Salzberg 2012) and filtered them out. We then ran dnaPipeTE without mitochondrial reads (same parameters). The assembled mitochondrial sequences of both parental species are available in the supplementary file S5, Supplementary Material online.

Divergence Time between Parental Species

We aligned all sequences $\geq$2,000 bp of the *D. buzzatii* de novo transcriptome against *D. koepferae*'s ones using BLAST (McGinnis and Madden 2004) in order to identify contig pairs between *D. buzzatii* and *D. koepferae*. We kept only the best hit for each query and subject, resulting in a total of 2,656 pairs of contigs, which were translated using EMBOSS getorf (Rice et al. 2000). We used the most likely protein sequences of each contig (i.e., the longest) to perform codon alignments with MUSCLE (Edgar 2004). Finally, the d*S* rate of each pair was calculated using the codeml program in PAML version 4 (Yang 2007). PAML results are available in the supplementary file S6, Supplementary Material online. Finally, we estimated the divergence time between *D. buzzatii* and *D. koepferae* as in Keightley et al. (2014) using the obtained d*S* mode.

Ping-Pong Signature Identification

The ping-pong cycle is mediated by Aubergine and Ago3 proteins, which cleave the piRNA precursor (or TE transcript) preferentially 10 bp after its 5′-end. Thus, sense and antisense reads overlapped by ten nucleotides are produced during secondary piRNA biogenesis (Klattenhoff and Theurkauf 2008). We aligned our piRNA raw reads (23–30 nt, without any trimming step in order to maintain their real size) against the whole TE library using Bowtie1 (*-S* option) and checked for the presence of 10-nt-overlapping sense–antisense read pairs using the *signature.py* pipeline (Antoniewski 2014). The same analysis was carried out separately for each of the TE families of the library.

piRNA Pathway Proteins Ortholog Search

Proteomes of *D. buzzatii* and *D. koepferae* (see Gene Analyses) were aligned against each other using BLAST. Identity percentages of each protein best hit were kept and used to calculate the median identity percentage between *D. buzzatii* and *D. koepferae*.

We identified the orthologs of 30 proteins involved in piRNA biogenesis (Yang and Pillai 2014) in *D. buzzatii* and *D. koepferae* proteomes by reciprocal best blast hit analysis, using their *D. melanogaster* counterparts as seeds (EnsemblMetazoa 27 release, Cunningham et al. 2015), with and *e*-value cutoff of 1e-05. *Drosophila buzzatii* proteins were aligned against their *D. koepferae* ortholog using BLAST, in order to evaluate their identity percentage.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

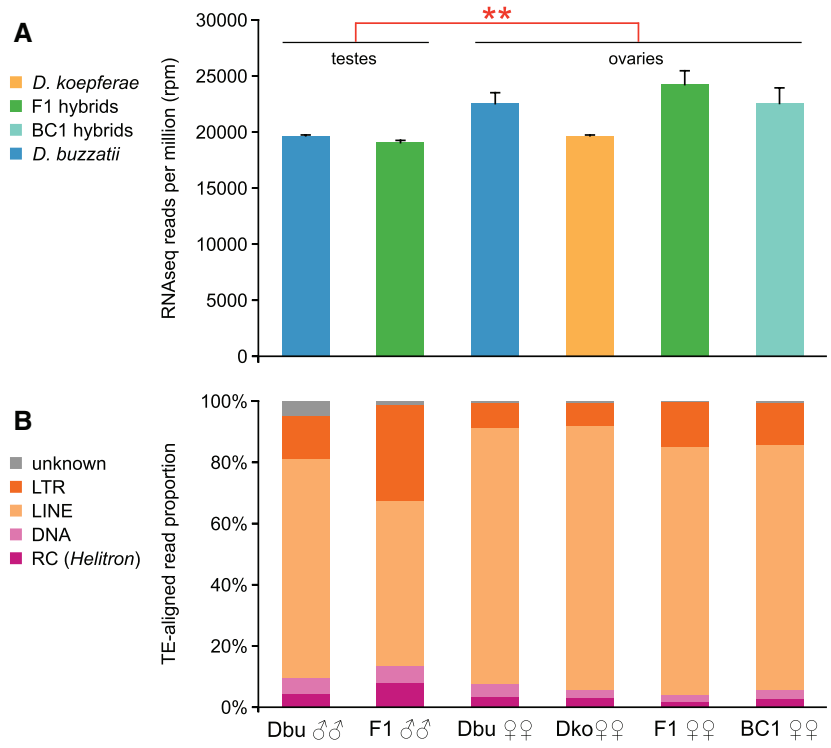


Fig. 2.—Transposable element expression summary. Dbu, *D. buzzatii*; Dko, *D. koepferae*; ♂♂, testes; ♀♀, ovaries. (*A*) Mean proportion of reads aligning to the TE library. Bars represent standard deviation between replicates. **Wilcoxon's $W = 2$, $P = 0.016$. (*B*) TE expression profiles following Repbase classification (Jurka et al. 2005): LTR and LINE (class I), DNA and RC/*Helitron* (class II), Unknown (unclassified). LTR, elements with long terminal repeats; LINE, long interspersed nuclear element; RC, rolling circle elements (or *Helitrons*).

Results

Qualitative Changes in TE Expression after Interspecific Hybridization

We sequenced the ovarian transcriptomes of both parental species and two hybrid generations, the F1 and a first backcross BC1 (fig. 1), and examined their TE expression. We also sequenced and analyzed the testicular transcriptomes of *D. buzzatii* (male parental species) and F1 hybrids. Globally, we detected expression of 414 out of 658 candidate TE families (supplementary file S2*B*, Supplementary Material online). We show that ovaries present significantly higher TE global alignment rate than testes (fig. 2*A*; comparing all ovarian samples against all testicular samples: Wilcoxon's $W = 2$, $P = 0.016$) whereas the global TE alignment rate between hybrids and parental species is not significantly different (comparing all hybrid samples against all parental samples: Wilcoxon's $W = 14$, $P = 0.59$).

At a qualitative level, we observe notable differences between parents and hybrids: LTR proportion is increased in both hybrid testes (from 14.2% to 31.4%) and ovaries (from 7.7–8.3% to 14.4–13.8%), as well as are RC elements (*Helitron*) in F1 testes (from 4.3% to 8.1%, fig. 2*B*). TE expression profiles are very similar between ovaries of *D. buzzatii* and *D. koepferae*, but parental testes (*D. buzzatii*) present a considerably lower LINE proportion (fig. 2*B*). In all cases, TE expression is mainly represented by retrotransposons (LINEs are the most expressed category followed by LTRs). Therefore, even if the global amounts of TE expression remain unchanged after interspecific hybridization, we observe differences at the TE family expression level.

TE Deregulation in Hybrid Ovaries Is Biased toward Overexpression

Compared with *D. buzzatii* and *D. koepferae* separately, F1 ovaries present a similar number of differentially expressed TE families (221 and 234, respectively), whereas in BC1 expression is closer to *D. buzzatii* (149 and 254, fig. 3*A*). In both cases, hybrid ovaries present a bias toward TE overexpression compared with parental species (fig. 3*A*), with 55% of the deregulated families (on average) more expressed in hybrids (supplementary table S2, Supplementary Material online).

When compared with both parental species, 37 TE families are significantly overexpressed in F1 and only 27 in BC1 (most of them are shared between generations, table 1). Among them, 77% are retrotransposons, and *Gypsy* elements exhibit the highest FC values. Surprisingly, we also observe 26 underexpressed families in F1 and 17 in BC1 (table 2). Underexpressed TE families are also mainly retrotransposons

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

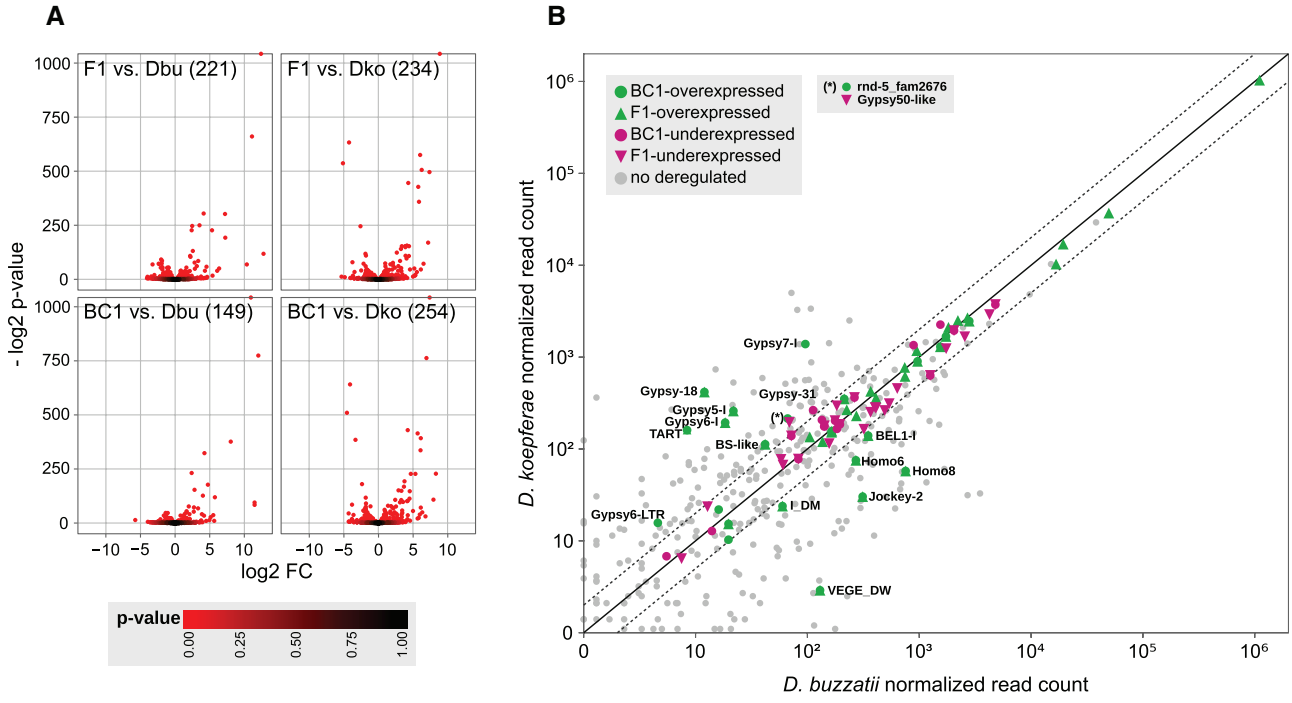


Fig. 3.—TE differential expression analyses in ovaries. (*A*) Differentially expressed TE families in hybrids compared separately with *D. buzzatii* (Dbu) and *D. koepferae* (Dko). The total number of differentially expressed TE families of each comparison is written in parenthesis. FC, fold change (hybrid vs. parent). (*B*) Expression of TE families in *D. koepferae* versus *D. buzzatii*. In color, deregulated TE families in hybrids (compared with both parental species). Dot lines represent 2-fold changes between parental expression and the solid line represents the same amount of expression between Dbu and Dko. Names of those TE families with differences of expression higher than 2-fold between parental species are indicated.

(71%) and their FC values tend to be lower than those of overexpressed families (tables 1 and 2).

Therefore, after a generation of backcrossing, the global amount of TE deregulation decreases from 15.2% to 10.6% of the 414 expressed families. In the same way, we observe that FC values are often lower in BC1 than in F1 (tables 1 and 2). All the deregulated TE families are transcriptionally active in both parental species (fig. 3*B*), but only 21% of them exhibit differences of expression higher than 2-fold between parental species (a total of 16 families; 14 overexpressed and 2 underexpressed, fig. 3*B*).

TE Landscapes and Divergence Influence Deregulation

In *D. simulans–D. melanogaster* artificial hybrids (*Hmr*-rescued), misexpression of TEs in ovaries was found to be widespread compared with that of protein-coding genes (Kelleher et al. 2012). To evaluate the extent of gene deregulation in *D. buzzatii–D. koepferae* hybrids, a de novo transcriptome assembly was produced for each parental species (see Materials and Methods). We annotated 70.9% of the final transcriptome contigs as 11,190 different protein-coding genes (supplementary table S1, Supplementary Material online), of which 11,049 were found to be expressed. Among these, 657 are overexpressed and 821 underexpressed in F1 ovaries (supplementary file S7, Supplementary Material online), reaching a proportion of deregulation of 13.4%. In BC1, it decreases to 12.4%, with 711 overexpressed and 662 underexpressed genes (supplementary file S7, Supplementary Material online).

Our GO terms enrichment analysis shows that F1- and BC1-overexpressed genes have 37 enriched GO terms in common (supplementary file S4*A*, Supplementary Material online), which seem to be mainly related to DNA organization (at the chromosome and chromatin level), cell division (DNA replication, DNA repair, and spindle microtubules), and mRNA processing. In the case of underexpressed genes, a total of 22 enriched GO terms are shared between F1 and BC1 (supplementary file S4*A*, Supplementary Material online), most of them linked to fatty acid metabolism and cell communication, as well as to developmental growth and oogenesis, which may be related to the hybrid loss of fertility.

We show that TE and gene expression are affected at similar levels (~10–15%) in ovaries of *D. buzzatii–D. koepferae* hybrids, although it is noteworthy that their deregulation patterns differ (only TEs are biased toward overexpression). Our results are in contrast to the observed in *D. simulans–D. melanogaster* hybrids, where the extent of deregulation was strikingly higher for TEs (12.1%) than for genes (0.7%, Kelleher et al. 2012). If we estimate the misregulation proportion of *D. buzzatii–D. koepferae* F1 hybrids as in Kelleher et al. (2012), that is, with FDR = 0.05 and filtering differential expression

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Table 1
Overexpressed TE Families in Hybrid Ovaries

TE Family	Order	Superfamily	F1 Ovaries				BC1 Ovaries			
			log2(FC) versus		BH Adjusted *P* Value		log2(FC) versus		BH Adjusted *P* Value	
			Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Homo6	DNA	hAT	2.46	4.32	5.47E-75	7.81E-135	2.38	4.25	2.26E-70	5.04E-130
Homo8	DNA	hAT	2.55	6.26	3.35E-40	5.01E-153	1.97	5.68	8.03E-24	1.77E-125
R = 81	DNA	hAT	0.68	0.79	1.23E-03	1.44E-04	0.62	0.73	5.92E-03	4.50E-04
rnd-5_family-1117	DNA	hAT	0.63	0.37	1.44E-03	7.44E-02	—	—	—	—
VEGE_DW[a]	DNA	hAT	1.26	6.53	3.28E-04	2.02E-22	2.69	7.96	1.64E-16	3.04E-33
Rehavkus-2_Nvi	DNA	MULE-MuDR	0.77	0.46	8.12E-08	2.00E-03	—	—	—	—
rnd-5_family-4211	DNA	MULE-MuDR	0.37	0.56	7.16E-02	3.61E-03	—	—	—	—
DNA8-7_CQ	DNA	OtherDNA	0.61	0.65	9.85E-06	1.51E-06	0.38	0.43	1.49E-02	2.51E-03
rnd-4_family-786	DNA	Transib	0.41	0.67	5.59E-02	9.17E-04	—	—	—	—
rnd-5_family-1551	DNA	Transib	0.69	0.48	4.49E-04	1.76E-02	—	—	—	—
CR1-1_CQ	LINE	CR1	1.16	0.80	2.25E-04	1.31E-02	—	—	—	—
CR1-2_CQ	LINE	CR1	0.52	0.53	2.94E-02	2.24E-02	—	—	—	—
I_DM	LINE	I	1.28	2.58	1.07E-02	2.61E-07	1.27	2.57	1.82E-02	2.27E-07
rnd-5_family-156	LINE	I	1.68	0.96	1.65E-08	1.81E-03	1.36	0.64	1.28E-05	4.89E-02
BS-like	LINE	Jockey	5.33	3.90	5.91E-69	1.82E-45	4.73	3.31	4.52E-54	1.02E-32
Jockey-2_Dya	LINE	Jockey	2.39	5.77	5.28E-69	1.98E-129	0.32	3.70	9.10E-02	2.50E-51
rnd-3_family-39	LINE	Jockey	0.39	0.58	4.60E-03	7.14E-06	—	—	—	—
TART_B1[b]	LINE	Jockey	—	—	—	—	1.46	2.30	3.53E-02	3.45E-04
TART	LINE	Jockey	7.24	3.14	1.13E-58	2.60E-26	5.74	1.64	1.43E-36	1.11E-07
rnd-4_family-338	LINE	L2	0.57	0.40	4.36E-04	1.83E-02	—	—	—	—
rnd-5_family-2046	LINE	L2	0.71	0.65	1.84E-04	6.54E-04	—	—	—	—
Bilbo	LINE	LOA	0.83	1.02	8.33E-13	8.82E-19	0.78	0.97	4.22E-11	4.64E-17
R1_Dps	LINE	R1	0.56	0.81	3.23E-05	5.52E-10	0.53	0.78	1.57E-04	1.91E-09
rnd-5_family-1630	LINE	R1	0.53	0.63	1.03E-04	2.48E-06	0.30	0.40	7.15E-02	4.93E-03
RT2	LINE	R1	0.74	0.53	1.21E-08	5.45E-05	—	—	—	—
RTAg3	LINE	R1	0.93	1.02	3.33E-05	5.48E-06	0.54	0.63	4.22E-02	7.98E-03
RTAg4	LINE	R1	0.51	0.60	2.20E-04	6.74E-06	—	—	—	—
BEL1-I_Dmoj	LTR	BelPao	2.81	4.13	5.42E-24	1.03E-47	1.02	2.34	1.33E-03	1.15E-15
BEL1-LTR	LTR	BelPao	1.53	1.92	3.80E-03	3.25E-04	1.05	1.45	9.10E-02	9.24E-03
Gypsy-14_Dwil-I[b]	LTR	Gypsy	—	—	—	—	3.94	3.91	7.45E-02	4.72E-02
Gypsy-151_AA-I	LTR	Gypsy	0.43	0.71	4.33E-03	8.58E-07	—	—	—	—
Gypsy16-I_Dpse	LTR	Gypsy	12.76	7.39	2.88E-36	5.41E-150	11.47	6.09	2.94E-29	5.80E-102
Gypsy-172_AA-I	LTR	Gypsy	0.64	0.81	4.66E-02	7.87E-03	—	—	—	—
Gypsy-18_Dwil-I[a]	LTR	Gypsy	11.10	6.04	1.49E-199	8.22E-174	12.01	6.95	8.02E-234	2.40E-230
Gypsy-18_Dwil-LTR[a]	LTR	Gypsy	10.35	7.19	2.00E-21	9.12E-52	11.48	8.32	5.49E-26	2.18E-69
Gypsy5-I_Dya	LTR	Gypsy	12.40	8.88	0.00E+00	0.00E+00	10.94	7.41	0.00E+00	0.00E+00
Gypsy61-I_AG	LTR	Gypsy	0.31	1.00	5.90E-02	7.47E-13	—	—	—	—
Gypsy6-I_Dya[a]	LTR	Gypsy	7.21	3.87	1.15E-91	6.99E-47	8.03	4.69	5.22E-114	3.81E-69
Gypsy6-LTR_Dya[b]	LTR	Gypsy	—	—	—	—	4.17	2.48	5.89E-11	5.30E-07
Gypsy7-I_Dmoj[b]	LTR	Gypsy	—	—	—	—	4.23	0.38	5.37E-98	5.37E-02
Gypsy8-I_Dpse	LTR	Gypsy	0.42	0.84	2.23E-03	3.08E-11	—	—	—	—
R = 961[b]	LTR	Gypsy	—	—	—	—	1.71	1.28	6.75E-03	3.08E-02
rnd-5_family-2676[b]	LTR	Gypsy	—	—	—	—	2.72	1.04	1.74E-22	8.93E-05
mean			2.48		6.16E-03		3.34		1.22E-02	

NOTE.—Dbu, *D. buzzatii*; Dko, *D. koepferae*; FC, fold change; BH, Benjamini–Hochberg correction.
[a]Overexpressed only in BC1.
[b]FC increases after BC.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Table 2
Underexpressed TE Families in Hybrid Ovaries

TE Family	Order	Superfamily	F1 Ovaries				BC1 Ovaries			
			log2(FC) versus		BH Adjusted *P* Value		log2(FC) versus		BH Adjusted *P* Value	
			Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Howilli1[a]	DNA	hAT	—	—	—	—	−1.70	−1.59	8.09E-02	7.33E-02
MINOS	DNA	Tc1Mariner	−1.32	−0.53	8.12E-08	6.02E-02	—	—	—	—
rnd-5_family-1477[a]	DNA	Tc1Mariner	—	—	—	—	−0.59	−1.13	1.21E-06	6.24E-24
rnd-5_family-3658[a]	DNA	Tc1Mariner	—	—	—	—	−0.66	−0.97	2.23E-02	8.48E-05
Transib1_DP[b]	DNA	Transib	−0.57	−0.90	8.58E-02	2.44E-03	−0.64	−0.97	6.76E-02	8.76E-04
Transib3_DP	DNA	Transib	−2.01	−2.86	9.45E-02	8.46E-03	—	—	—	—
HELITRON1_DM	RC	Helitron	−3.37	−3.11	1.34E-02	2.37E-02	—	—	—	—
Helitron-1_Dvir	RC	Helitron	−0.81	−0.32	4.66E-08	5.73E-02	—	—	—	—
rnd-3_family-48	RC	Helitron	−0.95	−0.59	1.29E-16	7.62E-07	−0.60	−0.23	6.44E-07	7.37E-02
rnd-4_family-133	RC	Helitron	−1.08	−0.53	1.50E-06	3.50E-02	—	—	—	—
DMCR1A-like	LINE	CR1	−1.21	−0.65	8.95E-11	1.27E-03	—	—	—	—
DPSEMINIME-like	LINE	CR1	−0.76	−0.26	2.38E-08	9.53E-02	—	—	—	—
DMRER1DM-like	LINE	R1	−1.55	−1.08	4.39E-09	1.08E-04	—	—	—	—
BEL-11_Dta-I	LTR	BelPao	−1.91	−1.29	7.37E-18	1.24E-08	—	—	—	—
BEL-20_AA-I[a]	LTR	BelPao	—	—	—	—	−0.67	−0.52	2.23E-02	6.39E-02
BEL-3_Dta-I	LTR	BelPao	−0.70	−0.61	8.23E-03	2.24E-02	−0.57	−0.48	5.13E-02	7.61E-02
BEL-6_Dwil-I	LTR	BelPao	−1.08	−1.47	1.10E-02	2.05E-04	—	—	—	—
BEL-8_Dwil-I	LTR	BelPao	−2.08	−1.10	5.93E-17	3.88E-05	—	—	—	—
Nobel_I[b]	LTR	BelPao	−0.81	−0.73	9.17E-06	6.08E-05	−0.82	−0.74	9.24E-06	3.64E-05
rnd-4_family-529[b]	LTR	BelPao	−0.45	−0.91	9.41E-02	1.06E-04	−0.70	−1.16	8.53E-03	4.98E-07
rnd-5_family-1078	LTR	BelPao	−1.00	−0.44	2.92E-12	3.79E-03	—	—	—	—
rnd-5_family-2670	LTR	BelPao	−2.02	−1.11	2.35E-28	1.50E-08	—	—	—	—
Copia-3-like[a]	LTR	Copia	—	—	—	—	−0.45	−1.04	6.63E-02	8.92E-08
rnd-5_family-4686	LTR	Copia	−0.92	−1.08	1.24E-02	2.22E-03	—	—	—	—
Beagle-like	LTR	Gypsy	−0.59	−1.27	1.58E-02	5.00E-09	—	—	—	—
Gypsy1-I_Dmoj	LTR	Gypsy	−0.85	−1.05	8.73E-04	2.01E-05	−0.53	−0.73	6.52E-02	2.80E-03
Gypsy-22_Dya-I[b]	LTR	Gypsy	−1.74	−1.63	1.23E-04	3.51E-04	−2.13	−2.02	5.53E-06	9.98E-06
Gypsy2-I_DM	LTR	Gypsy	−1.17	−0.65	3.86E-10	1.20E-03	—	—	—	—
Gypsy-31_Dwil-I[a]	LTR	Gypsy	—	—	—	—	−1.11	−2.33	5.27E-02	5.69E-07
Gypsy4-I_Dpse	LTR	Gypsy	−1.90	−0.90	1.40E-26	1.62E-06	−1.37	−0.38	8.49E-15	6.15E-02
Gypsy50-like	LTR	Gypsy	−0.98	−2.47	1.34E-02	4.85E-13	—	—	—	—
QUASIMODO-like[a]	LTR	Gypsy	—	—	—	—	−0.58	−1.20	1.62E-02	1.38E-09
rnd-5_family-1084	LTR	Gypsy	−0.91	−1.85	8.70E-03	1.66E-09	−0.67	−1.61	7.57E-02	2.96E-08
TABOR_DA-LTR[a]	LTR	Gypsy	—	—	—	—	−3.27	−3.46	5.43E-02	2.13E-02
mean			−1.19		1.29E-02		−1.11		2.81E-02	

NOTE.—Dbu, *D. buzzatii*; Dko, *D. koepferae*; FC, fold change; BH, Benjamini–Hochberg correction.
[a]Underexpressed only in BC1.
[b]FC increases after BC.

results by FC ≥ 2 (supplementary table S3, Supplementary Material online), we find that TE deregulation is lower in our hybrids than in *D. melanogaster–D. simulans* ones (5.1% vs. 12.1%) but gene deregulation is strikingly higher (3.1% vs. 0.7%).

The higher alteration of TE expression found in *D. melanogaster–D. simulans* hybrids might be related to the radically different TE contents of these two species: Although mostly recent and active TE copies that account for 15% of the genome are found in *D. melanogaster*, *D. simulans* carries mainly old and deteriorated copies, representing 6.9% of its genome (Modolo et al. 2014). On the contrary, the analyses of our parental species' repeatomes (see Materials and Methods) show that they share similar classes and proportions of recent and active TEs (supplementary fig. S1 and file S8, Supplementary Material online). Hence, the similarity between *D. buzzatii* and *D. koepferae* TE landscapes and abundance concurs with their lower level of TE deregulation.

On the other hand, alteration in gene expression is more widespread in our hybrids than in *D. melanogaster–D.*

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

simulans ones. Our hypothesis is that this is due to the differences in divergence time between these species pairs. We have calculated the most common rate of substitution per synonymous site between our parental species ($dS = 0.139$; supplementary file S6, Supplementary Material online) and estimated their divergence time at 4.96 Ma using Keightley's mutation rate estimate (2014). This result concurs with the few available estimations of divergence between this species pair, which range between 4.02 and 4.63 Ma (Gomez and Hasson 2003; Laayouni et al. 2003; Oliveira et al. 2012). Using the same formula, *D. melanogaster* and *D. simulans* (with $dS = 0.068$, Cutter 2008) would have diverged 2.43 Ma, which is in concordance with the most commonly used estimation (2–3 Ma; Lachaise and Silvain 2004) and confirms that the latter species pair are more closely related.

Altogether, these results suggest that species divergence (rather than differences in TE content) would be the main cause of TE deregulation in *D. buzzatii*–*D. koepferae* hybrids, which would support the piRNA pathway failure hypothesis.

Differences in Parental piRNA Pools Cannot Fully Explain Hybrid TE Expression

Differences in piRNA pools between parental species ovaries can be at the origin of TE silencing impairment (Brennecke et al. 2008; Chambeyron et al. 2008), especially when piRNA levels of a particular TE are lower in the maternal species, *D. koepferae*. To test the maternal cytotype failure hypothesis, we sequenced and analyzed the piRNA populations of the samples presented in figure 1. Globally, antisense regulatory piRNA populations (23–30 nt) were detected for 392 out of 658 candidate TE families, in most cases retrotransposons (supplementary file S3*B*, Supplementary Material online). Differential expression analyses were then performed using FC values (see Materials and Methods).

A total of 196 TE families present differences higher than 2-fold between *D. buzzatii* and *D. koepferae* ovarian antisense piRNA populations (fig. 4*A*). Families having lower levels of piRNAs in the maternal species are not always overexpressed: Among the 98 TE families that exhibit reduced abundance of piRNAs in *D. koepferae*, only eight are overexpressed in hybrids (either in F1 or BC1, fig. 4*B-i*). Reciprocally, families having higher levels of piRNAs in the maternal species are not more commonly underexpressed: Only 12 out of 98 families with higher piRNA abundance in *D. koepferae* are classified as underexpressed (fig. 4*B-iii*). Actually, some deregulated TE families even present the opposite pattern (e.g., *Gypsy6-I* or *Howili1*, fig. 4*A*). Hence, differences between piRNA pools may account only for some specific cases of TE deregulation (e.g., *TART_B1* or *MINOS*, fig. 4*A*).

Interestingly, 12 of the overexpressed families are among those without associated piRNA populations (fig. 4*B-iv*), indicating that other TE regulation mechanisms (if any) could be responsible for their regulation in the ovaries. Accordingly, eight of them present associated endo-siRNA populations (supplementary file S3*E*, Supplementary Material online).

piRNA Production Strategies Differ between Parental Species

In *D. simulans*–*D. melanogaster* hybrids, piRNA production was shown to be deficient (Kelleher et al. 2012), which displaced the size distribution of ovarian piRNAs (23–30 nt) toward miRNAs and siRNAs (18–22 nt). In our case, the small RNA length distribution in hybrids is similar to that of *D. koepferae* (fig. 5*A*), and global levels of piRNAs are similar (or higher) in hybrids than in parental species (supplementary file S3, Supplementary Material online). Thus, *D. buzzatii*–*D. koepferae* hybrids do not present a deficient global piRNA production.

Interestingly, we note that size distribution of small RNA populations differs between our parental species (fig. 5*A*): *D. koepferae* exhibits abundant piRNAs and lower levels of miRNAs and siRNAs, whereas the opposite is observed in *D. buzzatii*. These differential amounts of piRNAs between our parental species might be due to a functional divergence in their piRNA biogenesis pathways. To get greater insight into piRNA production strategies, we have assessed the functionality of the secondary biogenesis pathway in our samples. In the germline, mature piRNAs (either maternal or primary) can initiate an amplification loop called the ping-pong cycle, yielding sense and antisense secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). In this loop, piRNAs are cleaved 10 bp after the 5′-end of their template, a feature that is specific to this pathway and can be used to recognize secondary piRNAs. We have determined the ping-pong signature in our sequenced piRNA populations (Antoniewski 2014) and revealed that *D. buzzatii*'s ping-pong fraction is higher than *D. koepferae*'s (fig. 5*B*), which is in agreement with the idea of divergence in piRNA biogenesis between them.

In F1 and BC1 hybrid ovaries, ping-pong signature levels are intermediate between parental species (F1 is more similar to *D. koepferae* and BC1 to *D. buzzatii*, fig. 5*B*). Contrarily, in *D. simulans*–*D. melanogaster* "artificial" hybrids, a reduced ping-pong fraction was observed (Kelleher et al. 2012). Therefore, our hybrids indeed differ from *D. melanogaster*–*D. simulans* model in that they are not characterized by a widespread decrease of piRNA production: We find a few TE families that present lower levels of piRNAs compared with both parental species (supplementary file S9, Supplementary Material online), but only some coincide with the upregulated ones.

Interestingly, half of the overexpressed TE families (a total of 20, including the 12 without associated piRNA populations described in fig. 4*B-iv*) do not present traces of ping-pong

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

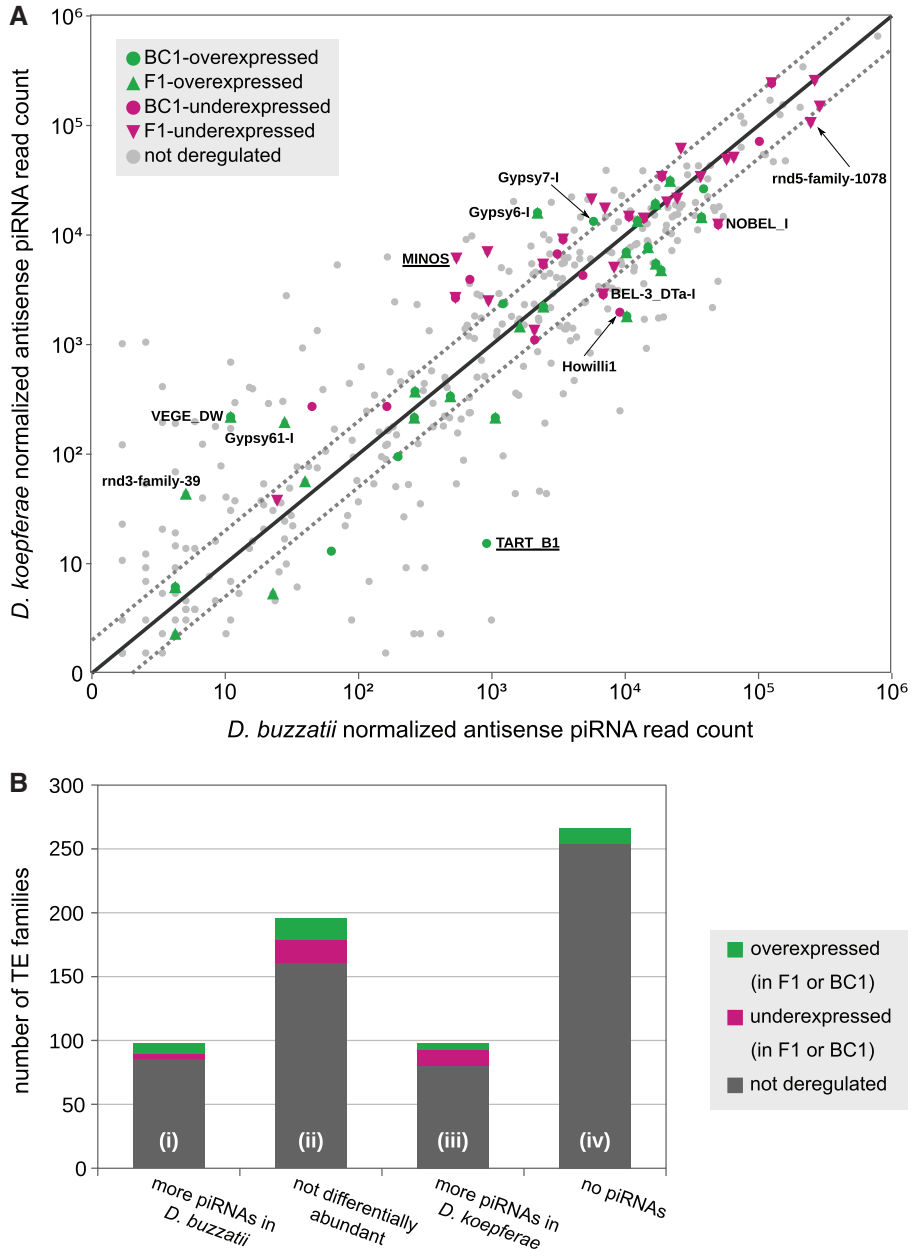


Fig. 4.—Parental piRNA populations and TE deregulation in ovaries. (*A*) Expression of TE-associated piRNA populations in *D. koepferae* (Dko) versus *D. buzzatii* (Dbu). Dot lines represent 2-fold changes between parental piRNA amounts and the solid line represents the same piRNA levels between Dbu and Dko. Underlined TE names are examples of families that may be deregulated due to the maternal cytotype hypothesis (underexpressed with more piRNAs in *D. koepferae*, overexpressed with more piRNAs in *D. buzzatii*). Names of deregulated TE families with unexpected differences in piRNA amounts (underexpressed with more piRNAs in *D. buzzatii*, overexpressed with more piRNAs in *D. koepferae*) are also indicated, with an arrow in some cases. (*B*) Proportion of deregulated TE families of different categories, classified according to differences (of at least 2-fold) between parental piRNA populations: (i) more piRNAs in *D. buzzatii*, (ii) not differentially abundant between parental species, (iii) more piRNAs in *D. koepferae*, (iv) absence of piRNAs in both species.

amplification (supplementary fig. S2, Supplementary Material online). Eleven of them are LINE retrotransposons, of which five belong to the R1 clade, whose members have a high target-specificity for 28S rRNA genes in arthropods (Eickbush et al. 1997; Kojima and Fujiwara 2003). The eight families with associated piRNA populations but without ping-pong signal could possibly be somatic elements, expressed in follicle cells of the ovaries, where secondary piRNA biogenesis does not take place.

piRNA Pathway Proteins Have Rapidly Evolved

Although the piRNA pathway is highly conserved across the metazoan lineage, some of its effector proteins are encoded

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

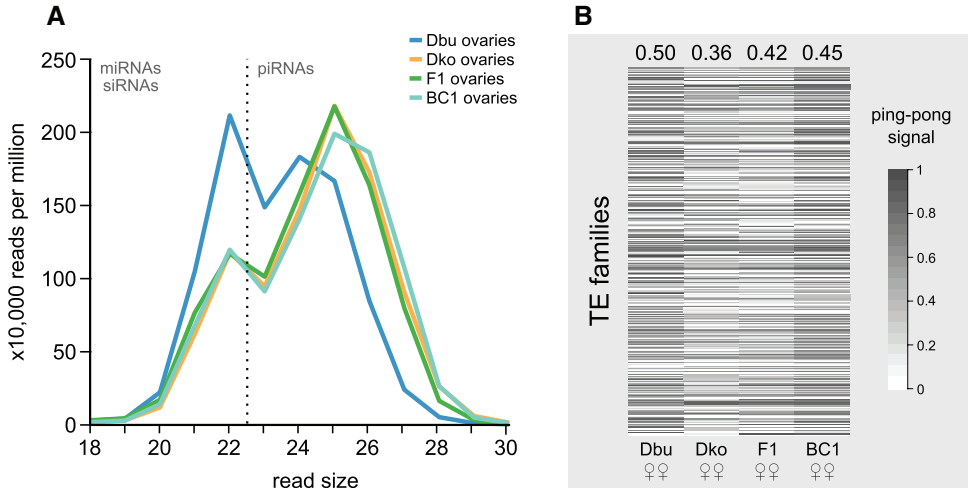


Fig. 5.—Characterization of piRNA populations in parental and hybrid ovaries. Dbu, *D. buzzatii*; Dko, *D. koepferae*; ♀♀, ovaries. (*A*) Read length distribution of ovarian small RNAs. The vertical dot line separates miRNAs and siRNAs (left) from piRNAs (right). (*B*) piRNA ping-pong fraction for each TE family (gray lines) and for the whole piRNA population (upper number). Only families with detectable ping-pong signal (>0) for at least one ovarian sample are represented.

by genes bearing marks of positive selection (Simkin et al. 2013). The accumulated divergence between these proteins has been proposed to account for the TE silencing failure in *Hmr*-rescued interspecific hybrids (Kelleher et al. 2012). To elucidate the global failure hypothesis, we have performed a bioinformatic prediction of *D. buzzatii* and *D. koepferae* coding sequences (henceforth named in silico proteomes) and aligned them against each other. Their identity percentage distribution has been assessed, with a resulting median identity of 97.2% (fig. 6).

We have then identified in *D. buzzatii* and *D. koepferae* in silico proteomes a total of 30 protein-coding genes known to be involved in TE regulation (Yang and Pillai 2014) as reciprocal best BLAST hits of their *D. melanogaster* putative orthologs (their names and symbols are listed in table 3). Alignments of all these genes between our parental species exhibit identity percentages lower than the median—their own median equals 92.5%—with the exception of the helicase *Hel25E*, whose sequence is identical in *D. buzzatii* and *D. koepferae* (fig. 6). Among the ten most divergent proteins (identity $\leq$90%), we find factors involved in both piRNA biogenesis (e.g., zucchini, tejas) and TE silencing (e.g., Panoramix, maelstrom, Hen1, and qin). Thus, protein divergence between our studied species could cause hybrid incompatibilities in both biogenesis and function of piRNAs.

We have also examined the expression of these 30 protein-coding genes and revealed significant differences between our parental species for all of them, with the exception of *Hen1*, *Panoramix* (*Panx*), and *tejas* (*tej*, table 3). The highest FC (log2FC = 5.0) is attributed to *krimper* (*krimp*, more expressed in *D. buzzatii*), known to participate in the ping-pong amplification process (Sato et al. 2015; Webster et al. 2015). Moreover, the two main genes involved in secondary piRNA biogenesis, *Aubergine* (*Aub*) and *Argonaute3* (*Ago3*), are also more expressed in *D. buzzatii* (table 3). Altogether, these results are consistent with the higher ping-pong fraction reported in this species (fig. 5*B*). Therefore, divergence in piRNA production between our parental species can be explained by the accumulated divergence in their piRNA pathway effector proteins as well as by the important differences in their expression levels.

When comparing hybrids with both parental species (table 3), we observe significant underexpression of *Hen1* (involved in primary and secondary piRNA biogenesis) and *Sister of Yb* (*SoYb*, involved in primary piRNA biogenesis) in both F1 and BC1 (Horwich et al. 2007; Saito et al. 2007; Handler et al. 2011). On the other hand, significant overexpression of *Panx* (involved in transcriptional silencing, Yu et al. 2015) also occurs in both hybrid generations. Those three genes are among the most divergent between parental species (identity $\leq$ 90%, fig. 6) and their altered expression could also partially account for TE deregulation.

Role of Endo-siRNAs and miRNAs in TE Deregulation

Aside from piRNAs, other kinds of small RNA pathways can also play a role in animal TE silencing. For instance, endo-siRNAs (21 nt in length) silence TEs posttranscriptionally in the soma (Ghildiyal et al. 2008); and some miRNAs (21–23 nt in length) are required for an effective piRNA-mediated TE silencing in ovarian somatic cells (Mugat et al. 2015).

We analyzed the endo-siRNA populations of our ovarian samples (supplementary file S3*E*, Supplementary Material online), showing that their global TE alignment rates are intermediate in hybrids (14.5% in F1 and 14.7% in BC1) compared with parental species (14.0% in *D. buzzatii* and 19.6% in *D. koepferae*). Although we detected endo-siRNAs associated with 406 out of the 658 studied TE families

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

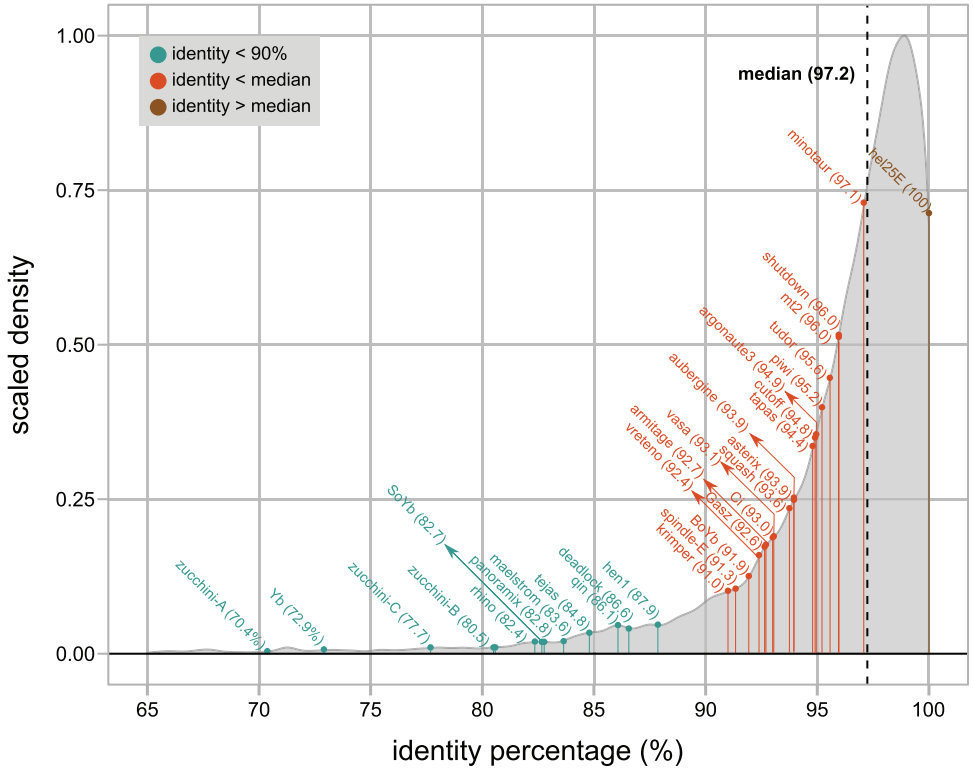


Fig. 6.—Distribution of identity percentages between *D. buzzatii* and *D. koepferae* in silico proteomes. A total of 30 proteins involved in the piRNA pathway were identified as reciprocal best BLAST hits of their *D. melanogaster* orthologs (represented by vertical bars, their identity in parenthesis). For Zucchini, four sequences were recognized as putative paralogs and named zucchini-A, -B, -C, and -D (only zucchini-A, -B, and -C are shown because zucchini-D was only identified in *D. buzzatii*). At least in two other species of the genus *Drosophila, D. melanogaster* and *D. grimshawi*, paralogs of Zucchini have been identified (Drosophila 12 Genomes Consortium 2007).

(supplementary file S3*F*, Supplementary Material online), only four presented differences higher than 2-fold in F1 ovaries compared with parental species (and none in BC1). Interestingly, two of these (*Bilbo* and *TABOR_DA-LTR*) are among the misregulated TE families in the RNA-seq analyses. *Bilbo* is one of the overexpressed TE families (table 1) without associated piRNA populations (fig. 4*B-iv*) and has higher amounts of endo-siRNAs in F1 than in parental species. Reciprocally, *TABOR_DA-LTR* is an underexpressed TE family (table 2) that presents lower amounts of associated endo-siRNAs in F1 than in parents. Hence, endo-siRNAs do not seem to have a crucial role in TE misregulation: A correlation between TE expression and endo-siRNA levels (if any) would be positive.

On the other hand, we evaluated the impact of hybridization on miRNA amounts using *D. mojavensis* miRNA complement as reference (Drosophila 12 Genomes Consortium 2007; Kozomara and Griffiths-Jones 2011). In ovaries, the obtained global alignment rates (supplementary file S4*B*, Supplementary Material online) are more similar in parental species (49.7% in *D. buzzatii* and 50.0% in *D. koepferae*) than in hybrids (48.2% in F1 and 49.2% in BC1). We detected miRNA populations for 63 out of the 71 mature miRNAs described in *D. mojavensis*, of which a total of five presented $\geq$2-fold differences hybrids (supplementary file S4*B*, Supplementary Material online). A single miRNA (*dmo-miR-278*) is less abundant in hybrid ovaries (both F1 and BC1) than in parents, whereas four are more abundant in hybrids (*dmo-miR-iab-4-3p* and *dmo-miR-281-1-5p* in F1; *dmo-miR-10* and *dmo-miR-276b* in BC1). To assess their putative involvement in TE silencing, we performed a GO enrichment analysis on each deregulated miRNA-predicted target set (see Materials and Methods). Several GO terms related to organ development and morphogenesis were found to be enriched among the silencing target genes of two of the upregulated miRNAs (*dmo-miR-iab-4-3p* and *dmo-miR-276b*, supplementary file S4*B*, Supplementary Material online); which is in concordance with the underexpression of genes related to developmental growth described in our analyses (supplementary file S4*A*, Supplementary Material online). Although none of the enriched GO term directly involves TE silencing, we noticed that the piRNA pathway genes *krimp* and *zuc* are putative targets of two upregulated miRNAs (*dmo-miR-276b* and *dmo-miR-iab-4-3p*, respectively). In the same way, the histone methyltransferase genes *Su(z)12* and *Su(var)3-9* seem to be targeted by *dmo-miR-10* and *dmo-miR-iab-4-3p*, respectively (supplementary file S4*B*, Supplementary Material online). Thus, although we cannot completely discard a

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Table 3
Summary of Differential Expression Analyses of piRNA Pathway Genes: Comparisons between Parental Species and between Parents and Hybrids

Gene Name	Gene Symbol	*D. buzzatii* versus *D. koepferae*			F1 versus Parental Species				BC1 versus Parental Species			
		% id	log2(FC)	BH Adjusted *P* Value	log2(FC)		BH Adjusted *P* Value		log2(FC)		BH Adjusted *P* Value	
					Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Argonaute3	*Ago3*	94.90	0.80	3.60E-29*	−0.77	0.02	1.69E-27*	7.68E-01	−0.76	0.04	6.66E-26*	6.41E-01
Armitage	*armi*	92.70	−0.59	1.51E-18*	0.43	−0.16	2.77E-10*	2.24E-02*	0.27	−0.33	1.86E-04*	1.43E-06*
asterix	*arx*	93.89	1.73	4.67E-65*	−0.30	1.43	3.21E-03*	2.45E-44*	−0.02	1.71	8.72E-01	2.43E-63*
aubergine	*aub*	93.92	2.62	3.45E-183*	−0.98	1.64	1.24E-26*	1.56E-72*	−0.46	2.16	1.08E-06*	1.40E-124*
Brother of Yb	*BoYb*	91.93	−0.42	9.63E-09*	0.52	0.10	1.79E-12*	1.83E-01	0.49	0.07	7.25E-11*	3.39E-01
cubitus interruptus	*Ci_tf*	92.97	−1.52	2.73E-18*	0.34	−1.18	6.40E-02*	1.66E-11*	0.24	−1.28	2.55E-01	2.78E-13*
cutoff	*cuff*	94.79	1.85	1.62E-78*	−0.64	1.22	2.77E-10*	2.16E-34*	−0.07	1.78	5.77E-01	1.68E-72*
deadlock	*del*	86.56	−0.88	7.51E-14*	0.32	−0.57	8.98E-03*	2.57E-06*	−0.03	−0.91	8.72E-01	1.82E-14*
GASZ ortholog	*Gasz*	92.64	0.65	1.00E-21*	0.07	0.72	3.05E-01	3.98E-26*	0.37	1.02	1.01E-07*	8.22E-52*
helicase at 25E	*Hel25E*	100	−0.41	1.36E-17*	0.25	−0.16	2.97E-07*	1.29E-03*	0.07	−0.34	2.51E-01	1.40E-12*
Hen1	*Hen1*	87.86	−0.02	9.13E-01	−0.44	−0.46	2.50E-06*	1.87E-06*	−0.50	−0.51	2.48E-07*	7.01E-08*
krimper	*krimp*	91.00	5.04	0.00E+00*	−0.62	4.41	3.02E-32*	0.00E+00*	−0.07	4.97	2.59E-01	0.00E+00*
maelstrom	*mael*	83.64	−1.20	8.48E-66*	0.77	−0.43	1.69E-27*	8.37E-10*	0.39	−0.81	1.13E-07*	6.11E-31*
minotaur	*mino*	97.08	−0.30	1.11E-04*	0.31	0.01	9.79E-05*	9.17E-01	0.03	−0.27	7.79E-01	5.30E-04*
Methyltransferase2	*Mt2*	95.95	0.74	9.90E-18*	−0.07	0.67	3.65E-01	2.95E-14*	−0.06	0.68	5.77E-01	6.58E-15*
Panoramix	*Panx*	95.95	0.01	9.20E-01	0.48	0.50	3.89E-09*	1.81E-09*	0.32	0.33	1.86E-04*	5.27E-05*
piwi	*piwi*	95.21	0.13	4.58E-02*	−0.23	−0.11	2.51E-04*	1.03E-01	−0.20	−0.07	2.49E-03*	2.63E-01
qin	*qin*	86.07	−1.30	9.28E-14*	0.47	−0.83	8.98E-03*	2.85E-06*	0.02	−1.29	9.23E-01	2.94E-13*
rhino	*rhi*	82.35	−1.03	7.85E-27*	0.34	−0.69	6.93E-04*	5.76E-13*	−0.06	−1.09	6.61E-01	1.13E-29*
shutdown	*shu*	95.97	2.26	0.00E+00*	−0.64	1.63	1.09E-53*	4.43E-302*	−0.17	2.10	1.37E-04*	0.00E+00*
Sister of Yb	*SoYb*	82.65	−0.32	4.11E-02*	−1.30	−1.62	1.43E-16*	4.20E-25*	−0.50	−0.82	2.11E-03*	9.76E-08*
spindle E	*spn-E*	91.34	−0.85	3.11E-17*	0.52	−0.33	5.13E-07*	1.29E-03*	0.23	−0.62	3.73E-02*	1.27E-09*
squash	*squ*	93.55	1.34	8.63E-23*	−0.72	0.62	1.10E-07*	9.35E-06*	−0.73	0.61	1.45E-07*	1.09E-05*
tapas	*tapas*	94.42	−0.94	3.03E-19*	0.63	−0.31	3.74E-09*	3.97E-03*	0.17	−0.77	1.67E-01	3.09E-13*
tejas	*tej*	84.79	0.01	9.62E-01	0.15	0.15	1.95E-01	1.83E-01	0.02	0.02	8.90E-01	8.52E-01
tudor	*tud*	95.56	−0.50	7.43E-04*	0.32	−0.19	3.89E-02*	2.26E-01	0.14	−0.37	4.80E-01	1.50E-02*
vasa	*vas*	93.05	0.67	1.41E-43*	−0.16	0.51	1.56E-03*	5.27E-26*	−0.11	0.56	4.90E-02*	3.57E-31*
vret	*vreteno*	92.39	0.68	7.64E-21*	−0.29	0.39	9.79E-05*	1.09E-07*	−0.26	0.42	7.92E-04*	6.71E-09*
Yb	*Yb*	72.89	1.05	4.22E-43*	−0.09	0.96	2.23E-01	1.11E-35*	−0.37	0.68	5.50E-07*	2.91E-18*
zucchini (A)	*zucA*	70.37	−1.55	4.19E-62*	1.21	−0.34	8.74E-38*	3.07E-04*	0.87	−0.67	5.21E-20*	4.03E-13*
zucchini (B)	*zucB*	80.50	−2.17	2.02E-04*	1.02	−1.15	1.10E-01	2.24E-02*	0.71	−1.45	3.57E-01	4.31E-03*
zucchini (C)	*zucC*	77.68	1.16	8.18E-53*	−0.28	0.88	1.65E-04*	2.05E-30*	−0.22	0.95	5.11E-03*	4.67E-35*
zucchini (D)	*zucD*	–	−0.43	6.87E-01	0.04	−0.39	9.62E-01	7.01E-01	0.48	0.05	6.61E-01	9.55E-01

NOTE.—Dbu, *D. buzzatii*; Dko, *D. koepferae*; FC, fold change; BH, Benjamini–Hochberg correction.
*Significant *P* value.

putative involvement of the miRNA pathway in our hybrids' TE expression (via the regulation of the piRNA pathway and of histone modification machinery), the fact that neither *krimp* nor *zuc* is deregulated in our hybrids does not support this hypothesis.

Interspecific Hybridization Has Sex-Biased Effects on TE Deregulation

An Enhanced piRNA Production May Cause TE Underexpression in Hybrid Testes

F1 testes present 256 differentially expressed TE families compared with *D. buzzatii* (more than any hybrid-parent comparison in ovaries, fig. 7*A*), and, as in ovaries, most of them are retrotransposons (supplementary file S10*A* and *B*, Supplementary Material online). Although we cannot compare hybrids with both parental species, we observe that TE underexpression in hybrid testes prevails over their overexpression (supplementary table S2, Supplementary Material online), showing that TE deregulation exhibits sex-biased patterns.

Regarding piRNA populations, the global piRNA production seems to be enhanced in F1 hybrids compared with *D. buzzatii* (fig. 7*B*), and the ping-pong fraction is also increased (fig. 7*C*). Besides, there is a bias toward piRNA overexpression of TE families in hybrids: 130 TE families exhibit more piRNAs

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

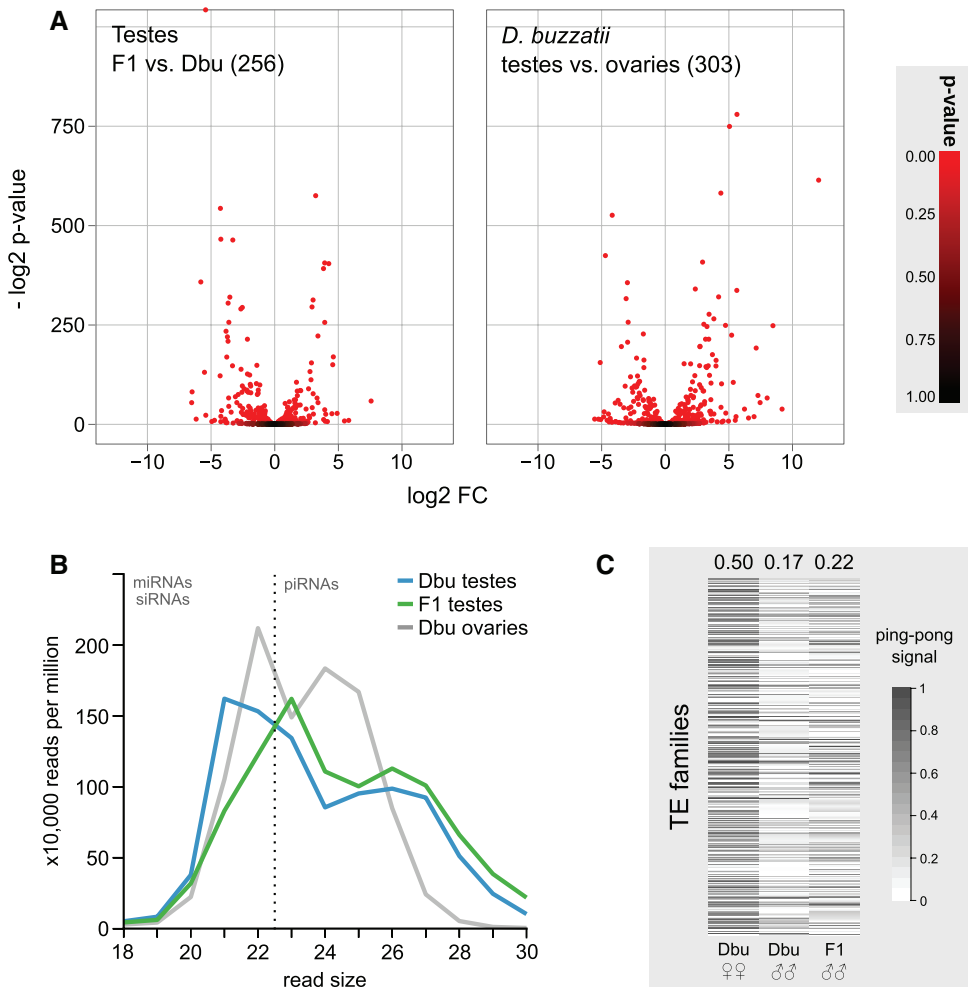


Fig. 7.—Differential expression analyses in testes. Dbu, *D. buzzatii;* ♂♂, testes; ♀♀, ovaries. (*A*) Differentially expressed TE families between F1 testes and Dbu (left) and between sexes of *D. buzzatii* (right). The total number of significant differences of each comparison is written in parenthesis. FC, fold change. (*B*) Read length distribution of *D. buzzatii* (testes and ovaries) and F1 testes small RNAs. The vertical dot line separates miRNAs and siRNAs (left) from piRNAs (right). (*C*) piRNA ping-pong fraction for each TE family (gray lines) and for the whole piRNA population (upper number). Only families with detectable ping-pong signal (>0) for at least one sample are represented.

in hybrids than in *D. buzzatii*, whereas 87 families have lower piRNA levels in hybrids (considering ≥2-fold differences, supplementary file S10*C* and *D*, Supplementary Material online). Therefore, in the case of males, the bias toward TE underexpression seems to be explained by a higher production of piRNAs.

TE Expression and piRNA Production Are Sex-Biased

The described sex-biased TE deregulation patterns are consistent with the remarkable differences in TE expression observed between testes and ovaries. Our results show that opposite sex samples always present more differences than samples of the same sex (supplementary table S2, Supplementary Material online). In particular, testes tend to present higher TE expression than ovaries (supplementary table S2, Supplementary Material online): For instance, 303 TE families present differential expression between ovaries and testes of *D. buzzatii*, of which 164 (54.1%) are more expressed in males than in females (fig. 7*A*). Furthermore, piRNA production also differs between sexes in *D. buzzatii*: Testes exhibit lower global piRNA amounts (fig. 7*B*) and lower ping-pong signature levels than ovaries (fig. 7*C*).

Discussion

Interspecific hybridization between *D. buzzatii* and *D. koepferae* causes different changes in TE expression: Some TE families are more expressed in hybrids, whereas others are more expressed in parental species. TE overexpression in hybrids might be caused not only by a failure of TE regulation mechanisms but also by an increase in TE copy number. In our study, these two events cannot be distinguished, but they are considered to be linked to each other because

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

transcription precedes transposition events (especially of retrotransposons). On the other hand, TE families that are underexpressed in hybrids might present more efficient repression mechanisms or simply a lower copy number in hybrids.

In ovaries, hybrid TE overexpression prevails over underexpression (tables 1 and 2 and supplementary table S2, Supplementary Material online). This concurs with several studies focused on a single or few TEs, where higher transcription levels in hybrids than in parents were observed (Kawakami et al. 2011; Carnelossi et al. 2014; García Guerreiro 2015). At a whole-genome level, a few surveys also report cases of TE families underexpressed in hybrids, but these results are generally out of the main attention focus and consequently poorly discussed. For instance, in lake whitefish hybrids, approximately 38% of differentially expressed TEs are underexpressed (Dion-Côté et al. 2014). Another well-studied case is that of hybrid sunflowers, where F1 hybrids present lower expression of the majority of TEs compared with parental species (Renaut et al. 2014). The presence of both overexpressed and underexpressed TEs suggests that hybrid TE deregulation is more complex than previously expected and may depend on the TE family.

Functional Divergence between Parental piRNA Pathways Can Lead to Hybrid Incompatibilities

We demonstrate that TE families with differences higher than 2-fold in their piRNA amounts between *D. buzzatii* and *D. koepferae* are not more commonly deregulated than families with similar levels (fig. 4). This shows that the maternal cytotype failure hypothesis cannot completely account for the observed pattern of TE deregulation, which is consistent with the similarity of TE landscapes between our parental species (supplementary fig. S1, Supplementary Material online). Thus, this explanation might be valid only for some particular TE families (fig. 4).

On the other hand, sequence divergence between maternal piRNAs and paternal TE transcripts (and the reciprocal) could also lead to a decrease of silencing efficacy in hybrids. A genome-wide comparison of sequences within a TE family between parental species cannot be performed because sequenced TEs in *D. koepferae* are scarce and its genome has not been sequenced yet (see supplementary text S1, Supplementary Material online, for a discussion on this putative bias). However, the presence of underexpressed TEs in hybrids, together with the knowledge that some TE families (such as *Helena*) are highly conserved between our parental species (Romero-Soriano and García Guerreiro 2016), seems to rule out this explanation.

Therefore, our results point rather to the piRNA pathway global failure hypothesis, which states that accumulated divergence of piRNA pathway effector proteins is responsible for hybrid TE deregulation. In this way, we show that proteins involved in piRNA biogenesis and function are more divergent than expected between *D. buzzatii* and *D. koepferae* (fig. 6). Consistent with this observation, previous studies in other *Drosophila* species have demonstrated that some of these proteins are encoded by rapidly evolving genes with marks of adaptive selection (Obbard et al. 2009; Simkin et al. 2013). Furthermore, we find that almost all piRNA pathway genes present significant differences in expression between *D. buzzatii* and *D. koepferae* (table 3). Such level of variability was also observed between different populations of a same species, *D. simulans* (Fablet et al. 2014).

Drosophila koepferae seems to produce higher amounts of piRNAs compared with *D. buzzatii*, which exhibits higher levels of ping-pong signature (fig. 5). Those differences in global piRNA production strategies between parental species could be linked to the divergence and variability in expression between piRNA pathway genes. Indeed, the two main effectors of ping-pong amplification, *Aub* and *Ago3*, are more expressed in *D. buzzatii* than in *D. koepferae* ($\log2FC = 2.62$ and 0.80, table 3), which is consistent with the higher ping-pong fraction detected in this species. Furthermore, an excess of *Aub* expression relative to *Piwi* could lead to a decrease of piRNA production due to a less efficient phased piRNA biogenesis. After the cleavage of a piRNA cluster transcript by Ago3 in the ping-pong cycle, the remnants of this transcript are loaded into Aub and processed to form the 3′-end of an antisense Aub-bound piRNA (Czech and Hannon 2016). The excised fragment of the piRNA cluster transcript is then loaded into Piwi (and to a lesser extent, into Aub) and cut by Zucchini (Zuc) every 27–29 nucleotides, producing phased antisense piRNAs that allow sequence diversification (Han et al. 2015; Mohn et al. 2015). We can hypothesize that an excess of *Aub* expression leads to a more frequent loading of this protein for phased piRNA production; impairing the efficiency of phasing in *D. buzzatii*. This would lead to lower levels of piRNAs in *D. buzzatii*, which would mostly be produced by ping-pong amplification.

Contrary to *Aub*, *qin* is more expressed in *D. koepferae* than in *D. buzzatii* ($\log2FC = -1.30$, table 3), which can be at the origin of the observed lower amounts of antisense piRNAs in *D. buzzatii* (supplementary file S3, Supplementary Material online). Qin is known to enforce heterotypic ping-pong between Aub and Ago3 by preventing futile homotypic Aub:Aub cycles, which mainly produce sense piRNAs (Zhang et al. 2011). A recent study has demonstrated that homotypic Aub:Aub ping-pong also generates lower Piwi-bound antisense-phased piRNAs, because qin ensures the correct loading of Piwi with antisense sequences (Wang et al. 2015). Therefore, a lower expression of *qin* (coupled with an excess of Aub) could lead to a less efficient production of antisense piRNAs (both secondary and phased) in *D. buzzatii* compared with *D. koepferae*. However, we must note that the remarkably higher expression levels of *krimper* in *D. buzzatii* ($\log2FC = 5.0$, table 3) may diminish these effects, because krimper contributes to heterotypic ping-pong cycle formation

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

by sequestering unloaded Ago3 proteins to prevent illegitimate access of other RNA sequences into them (Sato et al. 2015; Webster et al. 2015).

Drosophila buzzatii and D. koepferae seem to present a functional divergence of the piRNA pathway, which could likely be at the origin of TE misregulation in hybrids. However, contrarily to the observed in *D. melanogaster–D. simulans* artificial hybrids (Kelleher et al. 2012), our hybrids do not exhibit deficient piRNA production. Indeed, global piRNA amounts in hybrids are higher than in *D. buzzatii* and resemble those observed in *D. koepferae* (fig. 5*B* and supplementary file S3, Supplementary Material online); and hybrid secondary piRNA biogenesis presents intermediate levels between parental species (fig. 5*A*). Thus, incompatibilities in our hybrids may entail piRNA-mediated silencing effectors rather than proteins involved in piRNA biogenesis, even though both kinds of proteins are among those with the lowest identity percentages (fig. 6).

Misexpression of *SoYb*, *Hen1*, and *Panoramix* Can Influence Hybrid TE Expression

Two of the piRNA pathway genes, *SoYb* and *Hen1*, are underexpressed in hybrids (table 3). Hen1 is known to methylate piRNAs at their 3′-ends in both follicle and germ cells (Horwich et al. 2007; Saito et al. 2007), but the impact of its mutation on TE expression may depend on the TE family. For instance, overexpression of *HeT-A* retrotransposon was observed in *Hen1* mutants due to a higher instability of piRNAs (Horwich et al. 2007), but other mutants exhibited an unchanged expression of retrotransposons (Saito et al. 2007). *SoYb* seems to be involved in primary piRNA biogenesis and has a partially redundant function with its paralog *BoYb* (Handler et al. 2011). Thus, even a complete gene loss of *SoYb* could be compensated by *BoYb* and would not lead to a widespread TE overexpression. Curiously, *BoYb* was underexpressed in *D. simulans–D. melanogaster* artificial hybrids (Kelleher et al. 2012). Although downregulation of *Hen1* and *SoYb* cannot explain the whole pattern of TE deregulation, we cannot dismiss it as a possible contributor to TE overexpression in some cases.

On the other hand, overexpression of *Panoramix*, known to be essential for TE transcriptional silencing (Czech et al. 2013; Handler et al. 2013; Sienski et al. 2015; Yu et al. 2015), may compensate silencing deficiencies (especially at a posttranscriptional level) and be at the origin of TE underexpression.

TE Deregulation May Involve Other Mechanisms

We have shown that TE deregulation in hybrid ovaries may be related to the piRNA pathway in terms of 1) incompatibilities due to its divergence between parental species, 2) misregulation of some genes involved in TE silencing, and 3) differences between parental piRNA pools (for a few TE families). However, changes in this pathway may not explain the whole set of alterations of TE expression observed in hybrids.

For instance, the endo-siRNA pathway is known to silence TEs in somatic and germinal tissues, with a partially redundant function with the piRNA pathway in gonads (Saito and Siomi 2010). *Drosophila buzzatii–D. koepferae* hybrids do not present lower global levels of TE-related endo-siRNAs than parental species (supplementary file S3, Supplementary Material online), and the few $\geq$2-fold changes in TE-specific endo-siRNA populations seem to be positively correlated to changes in TE expression. Therefore, there is no evidence pointing out a high impact of endo-siRNAs in hybrid TE deregulation, although we cannot discard a mild role in somatic TE silencing. Unfortunately, our data do not allow the distinction between somatic and germinal elements (and related bibliography in our species model is virtually nonexistent), but the presence of the usually somatic *gypsy* elements among deregulated families (tables 1 and 2) could indicate that some of them are indeed expressed in follicle somatic cells.

On the other hand, histone methylation marks linked with permissive or repressive chromatin states have frequently been associated with TE sequences and their surroundings (Klenov et al. 2007; Yasuhara and Wakimoto 2008; Riddle et al. 2011; Yin et al. 2011). This has been shown to be tightly connected with the piRNA pathway: For instance, expression of piRNA clusters depends (directly or indirectly) on methylation marks (Rangan et al. 2011; Goriaux et al. 2014; Mohn et al. 2014; Molla-Herman et al. 2015), and piRNA-mediated transcriptional silencing triggers the deposition of repressive H3K9me3 marks. Other mechanisms—such as endo-siRNAs and miRNAs—are also able to recruit this silencing machinery leading to heterochromatin formation (Holoch and Moazed 2015). In *D. buzzatii–D. koepferae* hybrids, two of the upregulated miRNAs target two histone methyltransferase genes (*Su(z)12* and *Su(var)3-9*). We could hypothesize that abnormal silencing of these genes might cause a failure in the deposition of histone modifications, resulting in abnormal TE expression.

Finally, two other TE defence mechanisms have been proposed to be activated in wild wheat hybrids: Deletion and methylation (Senerchia et al. 2015). Even though DNA methylation is not common in *Drosophila*, internal or complete deletions of TE copies have been suggested to act as a prevention mechanism against TE genome invasions (Petrov and Hartl 1998; Lerat et al. 2011; Romero-Soriano and García Guerreiro 2016). In that case, suppression of active insertions could reduce the RNA amounts of some TE families, contributing to their underexpression. Furthermore, recombination between copies is known to control R1 elements that are specifically inserted in 28S rRNA genes in *Drosophila* (Eickbush and Eickbush 2014).

The pattern of TE deregulation observed in *D. buzzatii–D. koepferae* hybrids seems to be the result of several interacting phenomena involving different regulation pathways, as has been observed in plants during stress episodes (Slotkin et al.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

2009; Ito et al. 2011; Marí-Ordóñez et al. 2013; Creasey et al. 2014). For instance, when a de novo invasion of an active retrotransposon takes place (Marí-Ordóñez et al. 2013), the action of the DCL4/Ago1 small RNA pathway (21-nt siRNAs) necessarily precedes the achievement of efficient TE silencing by another small RNA pathway (DCL3/Ago4, 24-nt siRNAs).

TE Deregulation across Generations of Hybridization

Interspecific gene flow between *D. buzzatii* and *D. koepferae* is a natural source of genetic diversity that can only be maintained through introgression of a parental genome in F1 females, as F1 males are all sterile (Marin et al. 1993). Therefore, the study of backcrossed hybrids delves into the understanding of the real impact of hybridization in nature. We show that differences in ovarian TE expression between hybrids and parents are concordant with the expected *D. buzzatii*/*D. koepferae* genome fraction at each generation: F1 is equally distant from both parental species, whereas BC1 drifts apart from *D. koepferae* (fig. 3*A*). Furthermore, the total amount of deregulated TE families is lower in BC1 (10.6% of the expressed TEs) than in F1 (15.2%): A generation of backcrossing seems to be sufficient to restore the regulatory mechanisms of some families, but not of the totality. A similar result was reported in inbred lines of *Oryza sativa* introgressed with genetic material from the wild species *Zizania latifolia*, where *copia* and *gypsy* retrotransposons were activated and then rapidly repressed within a few selfed generations (Liu and Wendel 2000). F1 and BC1 ovaries exhibit the lowest number of differentially expressed TEs within one-to-one sample comparisons (supplementary table S2, Supplementary Material online) and present similar TE expression profiles (fig. 2*B*). This points to the hypothesis that more generations would be necessary to restore TE expression to the parental levels. Indeed, if TE activation in hybrids is caused by the failure of different epigenetic mechanisms (Michalak 2009), these are expected to be mitigated after several backcrosses thanks to the dominance of one of the parental genomes. In agreement to this hypothesis, we showed in a recent study that TE activation causes a genome expansion in *D. buzzatii–D. koepferae* hybrid females, but the *C*-value decreases after the first backcross (Romero-Soriano et al. 2016).

Tendency to TE Repression in Hybrid Testes Demonstrates That TE Regulation Is Sex-Biased

We show that TE expression presents different patterns between ovaries and testes, both at the quantitative and qualitative levels (fig. 2). Other studies have reported tissue-specific expression of transposons between male and female gonads. For instance, in *D. simulans* and *D. melanogaster*, transcripts of *412* are only found in testes (Borie et al. 2002), *I-like* elements are more expressed in testes than in ovaries of *D. mojavensis* and *Drosophila arizonae* (Carnelossi et al. 2014), as well as are *Osvaldo* and *Helena* in *D. buzzatii* and *D. koepferae* (García Guerreiro 2015; Romero-Soriano and García Guerreiro 2016). All these studies show higher transcript abundances in male gonads, which is consistent with the bias we observe toward testes overexpression compared with ovaries (supplementary table S2, Supplementary Material online).

These findings point out a differential TE regulation between male and female gonads, which was previously suggested by studies in *Drosophila* testes demonstrating that male piRNA biogenesis is not always performed by the same mechanisms as in ovaries (Nagao et al. 2010; Siomi et al. 2010). Concordantly, we observe that testes have lower piRNA amounts and a less efficient ping-pong cycle than ovaries (fig. 7). It has indeed been shown that piRNAs in testes are involved not only in TE repression but also in gene silencing, particularly of *Stellate* and *vasa* (Nishida et al. 2007).

Our results on TE deregulation in hybrids fully support the idea of sex-specificity in TE silencing. Contrarily to ovaries, hybrid testes exhibit a bias toward TE underexpression compared with *D. buzzatii* (supplementary table S2, Supplementary Material online). Accordingly, the retrotransposon *Helena* was shown to exhibit lower transcript abundances in F1 testes than in *D. buzzatii* and *D. koepferae* (Romero-Soriano and García Guerreiro 2016), as was the case for most TE families in a transcriptomic study in F1 sunflower hybrids (Renaut et al. 2014). Although two other studies in *Drosophila* hybrids, focused on individual TEs, displayed the opposite effect (Carnelossi et al. 2014; García Guerreiro 2015), we consider that disparity between specific studies fits in our global results.

TE underexpression prevalence in our hybrid testes can be explained by an increase of piRNA production and ping-pong signal in F1 testes (fig. 7*B* and *C*). Thus, activation of piRNA biogenesis, especially through the ping-pong cycle, seems to be responsible for TE repression in testes. Consistent with this tight repression of TE activity in males, the genome size increase observed in *D. buzzatii–D. koepferae* hybrids occurs only in females, whereas the hybridization impact on male genome size is undetectable (Romero-Soriano et al. 2016).

Conclusions

We suggest that TE deregulation in ovaries of *D. buzzatii–D. koepferae* hybrids might be the result of several interacting phenomena: A partial failure of the piRNA pathway due to a functional divergence between parental species, misexpression of some piRNA pathway genes, and differences in the amounts of TE-specific piRNAs between maternal cytoplasms (for some TE families). Furthermore, we cannot discard that other TE repression mechanisms might partially account for the observed set of deregulations. For instance, endo-siRNAs might be controlling somatic elements, deletions could play a role in TE underexpression, and alteration of histone posttranslational modifications may alter the chromatin state

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

pattern of the hybrid genome and cause either overexpression or underexpression. The study of these mechanisms would be an interesting focus for future investigations, as it could shed light on other causes of hybrid TE deregulation.

On the other hand, comparisons between ovaries and testes show that TE regulation is sex-biased. Surprisingly, piRNA biogenesis is enhanced in hybrid testes, which underlines that hybridization is a genomic stress that can activate response pathways to counteract TE deregulation. Further work in testes needs to be performed to elucidate the observed differences in TE silencing, which could be crucial to understand the molecular basis of hybrid breakdown and sterility.

Supplementary Material

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

Acknowledgments

The authors wish to thank Nuria Rius and Alfredo Ruiz for providing advanced access to the *Drosophila buzzatii* genome and its TE list; Xavier Grau-Bové for his useful help with orthology methods; Esteban Hasson and Diego Nicolás de Panis for sharing sequencing data of *Drosophila koepferae* genome; Clément Goubert for his valuable help with dnaPipeTE; Inma Moreno for her technical assistance with RT-PCRs; and INRA—UMR AGAP (France) and Fasteris SA (Switzerland) for RNA sequencing services. This work was supported by research grant CGL2013-42432-P from the Ministerio de Economía y Competitividad (Spain) and grant 2014 SGR 1346 from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GGBE). VRS was supported by a PIF PhD fellowship from the Universitat Autûnoma de Barcelona (Spain).

Literature Cited

Antoniewski C. 2014. Computing siRNA and piRNA overlap signature. In: Werner A, editor. Animal endo-SiRNAs: methods and protocols. New York: Springer. p. 135–146.

Barbash DA. 2010. Ninety years of *Drosophila melanogaster* hybrids. Genetics 186:1–8.

Bauer S, Grossmann S, Vingron M, Robinson PN. 2008. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics 24:1650–1651.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57:289–300.

Benson G. 1999. Tandem Repeats Finder: a program to analyse DNA sequences. Nucleic Acids Res. 27:573–578.

Betel D, Koppal A, Agius P, Sander C, Leslie C. 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. 11:R90.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. Nucleic Acids Res. 36:149–153.

Borie N, Maisonhaute C, Sarrazin S, Loevenbruck C, Biémont C. 2002. Tissue-specificity of 412 retrotransposon expression in *Drosophila simulans* and *D. melanogaster*. Heredity (Edinb) 89:247–252.

Brennecke J, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128:1089–1103.

Brennecke J, et al. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science (80-.) 322:1387–1392.

Brennecke J, Senti K-A. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. 26:499–509.

Carnelossi EAG, et al. 2014. Specific activation of an I-like element in *Drosophila* interspecific hybrids. Genome Biol Evol. 6:1806–1817.

Chambeyron S, et al. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. Proc Natl Acad Sci USA. 105:14964–14969.

Creasey KM, et al. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. Nature 508:411–415.

Cunningham F, et al. 2015. Ensembl 2015. Nucleic Acids Res. 43:D662–D669.

Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. Mol Biol Evol. 25:778–786.

Czech B, Hannon GJ. 2016. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. Trends Biochem Sci. 41:324–337.

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. Mol Cell. 50:749–761.

Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. Mol Biol Evol. 31:1188–1199.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203–218.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:48.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eickbush TH, Burke WD, Eickbush DG, Lathe WC. 1997. Evolution of R1 and R2 in the rDNA units of the genus *Drosophila*. Genetica 100:49–61.

Eickbush TH, Eickbush DG. 2014. Integration, regulation, and long-term stability of R2 retrotransposons. Microbiol Spectr. 3:MDNA3-0011-2014.

Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin YV. 1982. Transposition of mobile genetic elements in interspecific hybrids of *Drosophila*. Chromosoma 85:375–386.

Fablet M, Akkouche A, Braman V, Vieira C. 2014. Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. Gene 537:149–153.

Fontdevila A. 2005. Hybrid genome evolution by transposition. Cytogenet Genome Res. 110:49–55.

Franco FF, Silva-Bernardi ECC, Sene FM, Hasson ER, Manfrin MH. 2010. Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the *Drosophila buzzatii* cluster. J Zool Syst Evol Res. 48:322–331.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

García Guerreiro M. 2012. What makes transposable elements move in the *Drosophila* genome? Heredity (Edinb) 108:461–468.

García Guerreiro MP. 2014. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in *Drosophila*. Mob Genet Elements. 4:e34394. Available from: http://www.pubmedcentral.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

nih.gov/articlerender.fcgi?artid=4132227&tool=pmcentrez&rendertype=abstract
García Guerreiro MP. 2015. Changes of Osvaldo expression patterns in germline of male hybrids between the species *Drosophila buzzatii* and *Drosophila koepferae*. Mol Genet Genomics. 290:1471–1483.
Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25:25–29.
Ghildiyal M, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. Science (80-.) 320:1077–1081.
Gomez GA, Hasson E. 2003. Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii*. Mol Biol Evol. 20:410–423.
Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. 2014. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep. 15:411–418.
Goubert C, et al. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Genome Biol Evol. 7:1192–1205.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29:644–652.
Grentzinger T, et al. 2012. piRNA-mediated transgenerational inheritance of an acquired trait piRNA-mediated transgenerational inheritance of an acquired trait. Genome Res. 22:1877–1888.
Grentzinger T, et al. 2013. A user-friendly chromatographic method to purify small regulatory RNAs. Methods 67:91–101.
Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36:154–158.
Guillén Y, et al. 2015. Genomics of ecological adaptation in cactophilic *Drosophila*. Genome Biol Evol. 7:349–366.
Gunawardane LS, et al. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. Science (80-.) 315:1587–1590.
Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8:1494–1512.
Han BW, Wang W, Li C, Weng Z, Zamore PD. 2015. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science (80-.) 348:817–821.
Handler D, et al. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 30:3977–3993.
Handler D, et al. 2013. The genetic makeup of the *Drosophila* piRNA pathway. Mol Cell. 50:762–777.
Hedges DJ, Deininger PL. 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res. 616:46–59.
Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 16:71–84.
Horwich MD, et al. 2007. The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 17:1265–1272.
Huerta-Cepas J, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44:D286–D293.
Ito H, et al. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472:115–119.
Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-interacting RNA: its biogenesis and functions. Annu Rev Biochem. 84:405–433.
Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462–467.
Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol Evol. 3:156–167.
Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics 196:313–320.
Kelleher ES, Edelman NB, Barbash DA. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 10:e1001428.
Kent WJ. 2002. BLAT—the BLAST-like alignment tool. Genome Res. 12:656–664.
Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86:813–833.
Klattenhoff C, Theurkauf W. 2008. Biogenesis and germline functions of piRNAs. Development 135:3–9.
Klenov MS, et al. 2007. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucleic Acids Res. 35:5430–5438.
Kojima KK, Fujiwara H. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol. 20:351–361.
Kozomara A, Griffiths-Jones S. 2011. MiRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39:152–157.
Kozomara A, Griffiths-Jones S. 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42:68–73.
Laayouni H, Hasson E, Santos M, Fontdevila A. 2003. The evolutionary history of *Drosophila buzzatii*. XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. Mol Biol Evol. 20:931–944.
Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol. 16:931–937.
Lachaise D, Silvain J-F. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. Genetica 120:17–39.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
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:R25.
Lerat E, Burlet N, Biémont C, Vieira C. 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473:100–109.
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:e17.
Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880.
Lopez-Maestre H, et al. 2017. Identification of misexpressed genetic elements in hybrids between *Drosophila*-related species. Sci Rep. 7:1–13.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.
Marí-Ordóñez A, et al. 2013. Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet. 45:1029–1039.
Marín I, Fontdevila A. 1998. Stable *Drosophila buzzatii-Drosophila koepferae* hybrids. J Hered. 89:336–339.
Marin I, Ruiz A, Pla C, Fontdevila A. 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. Evolution (N. Y) 47:1616–1624.
McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32:W20–W25.
Metcalfe CJ, et al. 2007. Genomic instability within centromeres of interspecific marsupial hybrids. Genetics 177:2507–2517.

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021

Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity (Edinb) 102:45–50.
Modolo L, Lerat E. 2015. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. BMC Bioinformatics 16:137.
Modolo L, Picard F, Lerat E. 2014. A new genome-wide method to track horizontally transferred sequences: application to *Drosophila*. Genome Biol Evol. 6:416–432.
Mohn F, Handler D, Brennecke J. 2015. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science (80-.) 348:812–817.
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 157:1364–1379.
Molla-Herman A, Vallés AM, Ganem-Elbaz C, Antoniewski C, Huynh J-R. 2015. tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription. EMBO J. 34:3009–3027.
Mugat B, et al. 2015. MicroRNA-dependent transcriptional silencing of transposable elements in *Drosophila* follicle cells. PLoS Genet. 11:e1005194.
Nagao A, Mituyama T, Huang H, Chen D, Siomi MC. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila testis*. RNA 16:2503–2515.
Nishida KM, Saito K, Mori T. 2007. Gene silencing mechanisms mediated by Aubergine–piRNA complexes in *Drosophila* male gonad. RNA 13:1911–1922.
Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci. 364:99–115.
Oliveira DCSG, et al. 2012. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. Mol Phylogenet Evol. 64:533–544.
O'Neill RJW, O'Neill MJ, Marshall Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–73.
Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol. 15:293–302.
Picard G. 1976. Non mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor. Genetics 83:107–123.
Piccinali R, Aguadé M, Hasson E. 2004. Comparative molecular population genetics of the Xdh locus in the cactophilic sibling species *Drosophila buzzatii* and *D. koepferae*. Mol Biol Evol. 21:141–152.
Punta M, et al. 2012. The Pfam protein families databases. Nucleic Acids Res. 40:D290–D301.
Rangan P, et al. 2011. PiRNA production requires heterochromatin formation in *Drosophila*. Curr Biol. 21:1373–1379.
Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philos Trans R Soc Lond B Biol Sci. 369:20130345.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.
Riddle NC, et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. Genome Res. 21:147–163.
Rius N, et al. 2016. Exploration of the *Drosophila buzzatii* transposable element content suggests underestimation of repeats in *Drosophila* genomes. BMC Genomics 17:344.
Romero-Soriano V, et al. 2016. *Drosophila* females undergo genome expansion after interspecific hybridization. Genome Biol Evol. 8:556–561.
Romero-Soriano V, García Guerreiro MP. 2016. Expression of the retrotransposon Helena reveals a complex pattern of TE deregulation in *Drosophila* hybrids. PLoS One 11:e0147903.
Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing *Drosophila* transposons. Genes Dev. 27:400–412.
Saito K, et al. 2007. Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. Genes Dev. 21:1603–1608.
Saito K, Siomi MC. 2010. Small RNA-mediated quiescence of transposable elements in animals. Dev Cell. 19:687–697.
Sato K, et al. 2015. Krimper enforces an antisense bias on piRNA pools by binding AGO3 in the *Drosophila* germline. Mol Cell. 59:553–563.
Senerchia N, Parisod C, Parisod C. 2015. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proc R Soc B Biol Sci. 282:20142874.
Sienski G, et al. 2015. Silencio/CG9754 connects the Piwi–piRNA complex to the cellular heterochromatin machinery. Genes Dev. 29:1–14.
Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD. 2013. Recurrent and recent selective sweeps in the piRNA pathway. Evolution (N. Y.) 67:1081–1090.
Siomi MC, Miyoshi T, Siomi H. 2010. piRNA-mediated silencing in *Drosophila* germlines. Semin Cell Dev Biol. 21:754–759.
Slotkin RK, et al. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136:461–472.
Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol. 16:R872–R873.
Vela D, Fontdevila A, Vieira C, García Guerreiro MP. 2014. A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. PLoS One 9:e88992.
Vela D, García Guerreiro MP, Fontdevila A. 2011. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques 50:247–250.
Wang N, et al. 2010. Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. BMC Plant Biol. 10:190.
Wang W, et al. 2015. Slicing and binding by Ago3 or Aub trigger Piwi-bound piRNA production by distinct mechanisms. Mol Cell. 59:819–830.
Webster A, et al. 2015. Aub and Ago3 are recruited to nuage through two mechanisms to form a ping-pong complex assembled by Krimper. Mol Cell. 59:564–575.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Pillai RS. 2014. Fly piRNA biogenesis: tap dancing with Tej. BMC Biol. 12:77.
Yasuhara JC, Wakimoto BT. 2008. Molecular landscape of modified histones in *Drosophila* heterochromatic genes and euchromatin-heterochromatin transition zones. PLoS Genet. 4:0159–0172.
Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2011. A High-Resolution Whole-Genome map of key chromatin modifications in the adult *Drosophila melanogaster*. PLoS Genet. 7:e1002380.
Yu Y, et al. 2015. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science (80-.) 350:339–342.
Zhang Z, et al. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell. 44:572–584.

Associate editor: Sarah Schaack

Downloaded from https://academic.oup.com/gbe/article/9/6/1450/3813252 by BIU Montpellier user on 09 June 2021