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 García Guerreiro

To cite this version:


HAL Id: hal-01596041
https://hal.archives-ouvertes.fr/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 Guerreiro1,*

1Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Spain
2Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Claude Bernard Lyon 1, Villeurbanne, France
3Institut 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.

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...
(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 koepeferae (buzzatii complex, repleta group). Hybridization between these two species can occur in nature (Gomez and Hasson 2003; Piccinini 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. koepeferae. 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. koepeferae 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. koepeferae, 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 (contrary to ovaries) TE deregulation is globally biased toward underexpression, which can be explained by a higher production of piRNAs in hybrid males.
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 (Guilleñ et al. 2015) using BLAT v.35x1 (Kent 2002), with parameters –minIdentity = 80 and –maxIntron = 75000, in order to identify chimeras. 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 (Guilleñ 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.

**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

---

**Figure 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.
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 md5_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).

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 ≤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 and both parental species were considered misregulated.

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/lmodolo/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 ≤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 ≥2-fold differences in their piRNA/endo-siRNA population levels between hybrids and both parental species were considered misregulated.
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 ≤10 aligned reads per sample are also considered to be unexpressed).

As for piRNA and endo-siRNA studies, we considered sequences with ≥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-miSRV target prediction tool with an miSRV score cutoff of <−0.6 (Betel et al. 2010). Finally, we performed a GO enrichment analysis using the GOriilla 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 miSRV 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 S4B, 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.repeatmasker.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 ≥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 dS 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 dS 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 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.
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 back-cross 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, Supplementary Material online). We show that ovaries present significantly higher TE global alignment rate than testes (fig. 2A; comparing all ovarian samples against all testicular samples: Wilcoxon’s $W = 2$, $P = 0.016$). The 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. 2A). 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. 2B). 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. 3A). In both cases, hybrid ovaries present a bias toward TE overexpression compared with parental species (fig. 3A), 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.
(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. 3B), 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. 3B).

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), 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 S4A, 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 S4A, 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.
### Table 1

Overexpressed TE Families in Hybrid Ovaries

<table>
<thead>
<tr>
<th>TE Family</th>
<th>Order</th>
<th>Superfamily</th>
<th>F1 Ovaries</th>
<th>BC1 Ovaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>log2(FC) versus BH Adjusted P Value</td>
<td>log2(FC) versus BH Adjusted P Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dbu</td>
<td>Dko</td>
</tr>
<tr>
<td>Homo6 DNA hAT</td>
<td>2.46</td>
<td>4.32</td>
<td>5.47E-75</td>
<td>7.81E-135</td>
</tr>
<tr>
<td>Homo8 DNA hAT</td>
<td>2.55</td>
<td>6.26</td>
<td>3.35E-40</td>
<td>5.01E-153</td>
</tr>
<tr>
<td>r = B1 DNA hAT</td>
<td>0.68</td>
<td>0.79</td>
<td>1.23E-03</td>
<td>1.44E-04</td>
</tr>
<tr>
<td>rmd-5_family-1117 DNA hAT</td>
<td>0.63</td>
<td>0.37</td>
<td>1.44E-03</td>
<td>7.44E-02</td>
</tr>
<tr>
<td>VEGA_DWβ DNA hAT</td>
<td>1.26</td>
<td>6.53</td>
<td>3.28E-04</td>
<td>2.02E-22</td>
</tr>
<tr>
<td>Rehavkus-2_Nvi MULE-MuDR</td>
<td>0.77</td>
<td>0.46</td>
<td>8.12E-08</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>rmd-5_family-4211 DNA MULE-MuDR</td>
<td>0.37</td>
<td>0.56</td>
<td>7.16E-02</td>
<td>3.61E-03</td>
</tr>
<tr>
<td>DNA8-7_CQ DNA OtherDNA</td>
<td>0.61</td>
<td>0.65</td>
<td>9.85E-06</td>
<td>1.51E-06</td>
</tr>
<tr>
<td>rmd-4_family-786 DNA Transib</td>
<td>0.41</td>
<td>0.67</td>
<td>5.59E-02</td>
<td>9.17E-04</td>
</tr>
<tr>
<td>rmd-5_family-1551 DNA Transib</td>
<td>0.69</td>
<td>0.48</td>
<td>4.49E-04</td>
<td>1.76E-02</td>
</tr>
<tr>
<td>CR1-1_CQ LINE CR1</td>
<td>1.16</td>
<td>0.80</td>
<td>2.25E-04</td>
<td>1.31E-02</td>
</tr>
<tr>
<td>CR1-2_CQ LINE CR1</td>
<td>0.52</td>
<td>0.53</td>
<td>2.94E-02</td>
<td>2.24E-02</td>
</tr>
<tr>
<td>I_DM LINE I</td>
<td>1.28</td>
<td>2.58</td>
<td>1.07E-02</td>
<td>2.61E-07</td>
</tr>
<tr>
<td>rmd-5_family-156 LINE Jockey</td>
<td>1.68</td>
<td>0.96</td>
<td>1.65E-08</td>
<td>1.98E-129</td>
</tr>
<tr>
<td>rnd-3_family-39 LINE Jockey</td>
<td>0.39</td>
<td>0.58</td>
<td>4.60E-03</td>
<td>7.14E-06</td>
</tr>
<tr>
<td>TART_B1b LINE Jockey</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TART LINE Jockey</td>
<td>7.24</td>
<td>3.14</td>
<td>1.13E-58</td>
<td>2.60E-26</td>
</tr>
<tr>
<td>rmd-4_family-338 LINE L2</td>
<td>0.57</td>
<td>0.40</td>
<td>4.36E-04</td>
<td>1.83E-02</td>
</tr>
<tr>
<td>rmd-5_family-2046 LINE L2</td>
<td>0.71</td>
<td>0.65</td>
<td>1.84E-04</td>
<td>6.54E-04</td>
</tr>
<tr>
<td>Bilbo LINE LOA</td>
<td>0.83</td>
<td>1.02</td>
<td>8.33E-13</td>
<td>8.82E-19</td>
</tr>
<tr>
<td>R1_Dps LINE R1</td>
<td>0.56</td>
<td>0.81</td>
<td>3.23E-05</td>
<td>5.52E-10</td>
</tr>
<tr>
<td>rmd-5_family-1630 LINE R1</td>
<td>0.53</td>
<td>0.63</td>
<td>1.03E-04</td>
<td>2.48E-06</td>
</tr>
<tr>
<td>RT2 LINE R1</td>
<td>0.74</td>
<td>0.53</td>
<td>1.21E-08</td>
<td>5.45E-05</td>
</tr>
<tr>
<td>RTAg3 LINE R1</td>
<td>0.93</td>
<td>1.02</td>
<td>3.33E-05</td>
<td>5.48E-06</td>
</tr>
<tr>
<td>RTAg4 LINE R1</td>
<td>0.51</td>
<td>0.60</td>
<td>2.20E-04</td>
<td>6.74E-06</td>
</tr>
<tr>
<td>BELL1-Dmoj LTR BelPao</td>
<td>2.81</td>
<td>4.13</td>
<td>5.42E-24</td>
<td>1.03E-47</td>
</tr>
<tr>
<td>BELL1-LTR LTR BelPao</td>
<td>1.53</td>
<td>1.92</td>
<td>3.80E-03</td>
<td>3.25E-04</td>
</tr>
<tr>
<td>Gypsy-14_Dwil-IIb LTR Gypsy</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gypsy-151_AA-I LTR Gypsy</td>
<td>0.43</td>
<td>0.71</td>
<td>4.33E-03</td>
<td>8.58E-07</td>
</tr>
<tr>
<td>Gypsy16-1_Dps LTR Gypsy</td>
<td>12.76</td>
<td>7.39</td>
<td>2.88E-36</td>
<td>5.41E-150</td>
</tr>
<tr>
<td>Gypsy17-2_Aa-I LTR Gypsy</td>
<td>0.64</td>
<td>0.81</td>
<td>4.66E-02</td>
<td>7.87E-03</td>
</tr>
<tr>
<td>Gypsy18_Dwil-Ib LTR Gypsy</td>
<td>11.10</td>
<td>10.40</td>
<td>1.49E-199</td>
<td>8.22E-174</td>
</tr>
<tr>
<td>Gypsy18_Dwil-LTRa LTR Gypsy</td>
<td>10.35</td>
<td>10.14</td>
<td>2.00E-21</td>
<td>9.12E-52</td>
</tr>
<tr>
<td>Gypsy51-Dya LTR Gypsy</td>
<td>12.40</td>
<td>8.88</td>
<td>0.00E+00</td>
<td>0.00E+00</td>
</tr>
<tr>
<td>Gypsy61-I_AG LTR Gypsy</td>
<td>0.31</td>
<td>1.00</td>
<td>5.90E-07</td>
<td>7.47E-13</td>
</tr>
<tr>
<td>Gypsy51-Dya LTR Gypsy</td>
<td>7.21</td>
<td>3.87</td>
<td>1.15E-91</td>
<td>6.99E-47</td>
</tr>
<tr>
<td>Gypsy6-LTR_Dya LTR Gypsy</td>
<td>0.34</td>
<td>0.25</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gypsy71-Dmoj LTR Gypsy</td>
<td>0.42</td>
<td>0.84</td>
<td>2.23E-03</td>
<td>3.08E-11</td>
</tr>
<tr>
<td>Gypsy81-Dps LTR Gypsy</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>rmd-5_family-2676b LTR Gypsy</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

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

*Overexpressed only in BC1.

*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, 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.
Differences in Parental piRNA Pools Cannot Fully Explain Hybrid TE Expression

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. 5A), 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. 5A): 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. 5B), 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. 5B). 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. 4B–iv) do not present traces of ping-pong
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

---

**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.
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 ≤ 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, table 3). Altogether, these results are consistent with the higher ping-pong fraction reported in this species (fig. 5B). 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 ≤ 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 S3E, 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...
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. 4B-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 S4B, 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 ≥2-fold differences hybrids (supplementary file S4B, 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-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-276b and dmo-miR-iab-4-3p, respectively) (supplementary file S4B, Supplementary Material online); which is in concordance with the underexpression of genes related to developmental growth described in our analyses (supplementary file S4A, 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 of the 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 S4B, Supplementary Material online). Thus, although we cannot completely discard a
Table 3
Summary of Differential Expression Analyses of piRNA Pathway Genes: Comparisons between Parental Species and between Parents and Hybrids

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>D. buzzatii versus D. koepferae</th>
<th>F1 versus Parental Species</th>
<th>BC1 versus Parental Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>D. buzzatii versus D. koepferae</th>
<th>F1 versus Parental Species</th>
<th>BC1 versus Parental Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
<tr>
<td>% id</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
<td>log2(FC)</td>
<td>BH Adjusted P Value</td>
</tr>
</tbody>
</table>

## 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. 7A), and, as in ovaries, most of them are retrotransposons (supplementary file S10A 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. 7B), and the ping-pong fraction is also increased (fig. 7C). Besides, there is a bias toward piRNA overexpression of TE families in hybrids: 130 TE families exhibit more piRNAs...
in hybrids than in \textit{D. buzzatii}, whereas 87 families have lower piRNA levels in hybrids (considering \(\geq 2\)-fold differences, supplementary file S10 and D, Supplementary Material online). Therefore, in the case of males, the bias toward TE under-expression 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 \textit{D. buzzatii}, of which 164 (54.1\%) are more expressed in males than in females (fig. 7A). Furthermore, piRNA production also differs between sexes in \textit{D. buzzatii}: Testes exhibit lower global piRNA amounts (fig. 7B) and lower ping-pong signature levels than ovaries (fig. 7C).

**Discussion**

Interspecific hybridization between \textit{D. buzzatii} and \textit{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
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 cyto-type 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 matern- al 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 deavage 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.
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 >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 Garcia 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. 2011; Romero-Soriano et al. 2015).
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 (Marín 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. 3A). 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. 2B). 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. 7B and Q). 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 cytoplasm (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.
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 Parish for sharing sequencing data of *Drosophila koepeferae* genome; Clément Goubert for his valuable help with dnaPipeTE; Inma Moreno for her technical assistance with RT-PCRs; and Clément Goubert for his valuable help with dnaPipeTE;

**Literature Cited**


