BMC Biology

RESEARCH

Open Access

Male sterility in *Drosophila* hybrids revealed by a multi-generational transcriptomic analysis of genes and transposable elements in testes

Alejandra Bodelón[1†], Vicente Chillida[1†], Daniel Siqueira de Oliveira[2,3], Cristina Vieira[2] and Maria Pilar García Guerreiro[1*]

Abstract

Background Hybridization can trigger phenotypic and transcriptomic changes, leading to reproductive isolation and genome evolution. *Drosophila* is a key model for understanding the genetic basis of speciation, yet the complete process leading to hybrid sterility remains unknown. We investigated the regulation of genes and transposable elements (TEs) from the onset of sterility to fertility recovery through backcrossing.

Results We studied the testicular transcriptomes of *Drosophila buzzatii*, *D. koepferae*, sterile F1 hybrids, from several unidirectional crosses, and subsequent backcross generations. F1 hybrids presented testicular atrophy, absence of sperm, increased body size, and gene misregulation. Successive backcrosses with *D. buzzatii* restored gene expression and fertility in the third generation, but a set of 13 genes remained deregulated, potentially contributing to persistent gene flow barriers. TE expression exhibited a tendency towards underexpression, particularly for *D. buzzatii*-derived elements, with partial derepression in subsequent backcrosses. Importantly, 32% of differentially expressed genes (DEG) between parental species harbored lineage-specific TE insertions in flanking or intragenic regions, suggesting a putative role of these elements in shaping regulatory differences between these species. Regulatory divergence analyses identified DEG with *cis* and *trans* regulatory changes as prevalent in parental species, whereas compensatory regulation was enriched in misexpressed genes in hybrids.

Conclusions Hybrid incompatibilities may result from extensive deregulation of genes related to development, metabolism, and reproduction, possibly combined with the activity of specific TE copies. These findings demonstrate the complexity of regulatory incompatibilities in hybrids and the contribution of persistent deregulated genes and TE dynamics to postzygotic isolation.

Keywords *Drosophila*, Sterility, Transposable elements, Interspecific hybridization, Gene misregulation

†Alejandra Bodelón and Vicente Chillida contributed equally to this work.

*Correspondence:
Maria Pilar García Guerreiro
mariapilar.garcia.guerreiro@uab.es
Full list of author information is available at the end of the article

Background

Hybridization of closely related species is a powerful evolutionary force that brings new phenotypes, disrupts the established genetic networks, and reinforces the processes of reproductive isolation [1, 2]. Postzygotic isolation often manifests through hybrid sterility or inviability, typically affecting the heterogametic sex in most animals, especially in Drosophila species, as described by Haldane's rule [3, 4]. The *Drosophila* genus has long served as a model for studying the genetic basis of speciation, particularly through the lens of hybrid sterility [3, 5, 6]. In the past 20 years, numerous studies have investigated this phenomenon through a molecular perspective, primarily based on transcriptomics [4, 7, 8], even leading to the identification of a list of putative genes involved in hybrid incompatibilities [9]. Indeed, many recent studies reported gene [4, 7, 8, 10] and transposable elements (TEs) misregulation in hybrids [8, 11, 12]. These studies also suggested a putative role of TEs in hybrid anomalies [13, 14]. Understanding the changes of expression in hybrids compared to parental species is essential to unravel their potential role in speciation processes [15, 16]. The present study investigates the phenotypic and transcriptomic consequences in males resulting from the hybridization between *Drosophila buzzatii* and *Drosophila koepferae*, two species closely related but having independent evolutionary trajectories. These cactophilic species represent a unique model for studying hybrid incompatibility because they combine a well-characterized evolutionary history with an ecological specialization to arid environments. These species have been shown to exhibit strong postzygotic isolation, sterility, and instability induced by TEs [17–19], making them highly relevant to investigate how regulatory incompatibilities arise in natural systems. Crosses between these species give rise to sterile males and fertile females that can be backcrossed (BC) with males of either parental species. However, because crosses with *D. buzzatii* males are more feasible [20], backcrossing was performed using *D. buzzatii* males [6, 17], gradually restoring male fertility in the offspring from the third BC generation onwards (Fig. 1). Previous work, focused on hybrid females, suggested that interspecific hybridization between these species could alter TEs and piwi pathway gene expression of F1 hybrids [10, 21]. In addition, a female genome expansion has also been observed in the first generation of backcrossing [19], as previously observed in plants [22, 23]. By examining multiple generations of hybrids and backcrosses, the present research provides a wide-ranging overview of how hybridization can disrupt developmental, metabolic, and reproductive processes in males, and how these disruptions are resolved or persist across generations. Earlier research on these species revealed that sterility resulted from a polygenic genetic architecture, driven by the combined effects of multiple genes, where only two chromosomal regions had major phenotypic impacts [17]. These results reinforced the evidence accumulated during several decades claiming that the number of sterility genes is large, and often with small individual effects [24, 25]. Molecular information is currently insufficient to elucidate the underpinnings of phenotypic disruptions in males. Hybrid gene expression has primarily focused on F1 hybrids, with limited attention to later generations of introgression and without monitoring changes across multiple successive backcross generations [2, 26, 27]. In fact, our own previous work provided a more limited dataset that included testes from F1 hybrid males and one parental species [21]. A recent review [28] explicitly highlights this gap, noting that the dynamics of gene expression beyond the F1 stage remain largely unexplored. Our study addresses this limitation by systematically analyzing transcriptomic changes across F1 and three successive backcross generations and by providing a unique longitudinal perspective on regulatory restoration during introgression. Our analysis of gene expression across generations identifies a subset of deregulated genes that persist through multiple backcrosses. While most of the observed misregulation is resolved by the successive backcrosses, a core group of 13 genes remained deregulated in BC3, which are mostly involved in developmental processes and metabolism. These persistent expression changes may represent entrenched incompatibilities that are difficult to purge through backcrossing, potentially contributing to long-term barriers to gene flow. By integrating phenotypic, transcriptomic, and genetic analyses across multiple generations, the study advances our understanding of the mechanisms driving reproductive isolation, TE misregulation, and the evolutionary fate of hybrid genomes.

Results

Increased size and sperm absence in F1 hybrids

We first focused on the phenotypic characteristics of the parental species *Drosophila buzzatii* and *Drosophila koepferae*, and the first generation F1 hybrids. Different phenotypic characteristics were measured: colour, size, testicular shape, and sperm production. To date, parental species were phenotypically undistinguishable except by the aedeagal size and shape, where the aedeagus (male intromittent organ) of *D. buzzatii* is smaller than that of *D. koepferae* [29]. This finding aligns with prior work linking aedeagal morphology to reproductive success in *Drosophila* hybrids [30]. Nevertheless, we described for the first time a new distinguishing feature: *D. buzzatii* exhibits a slight orange pigmentation on the thorax which is absent in *D. koepferae* (Additional file 1: Fig. S1).

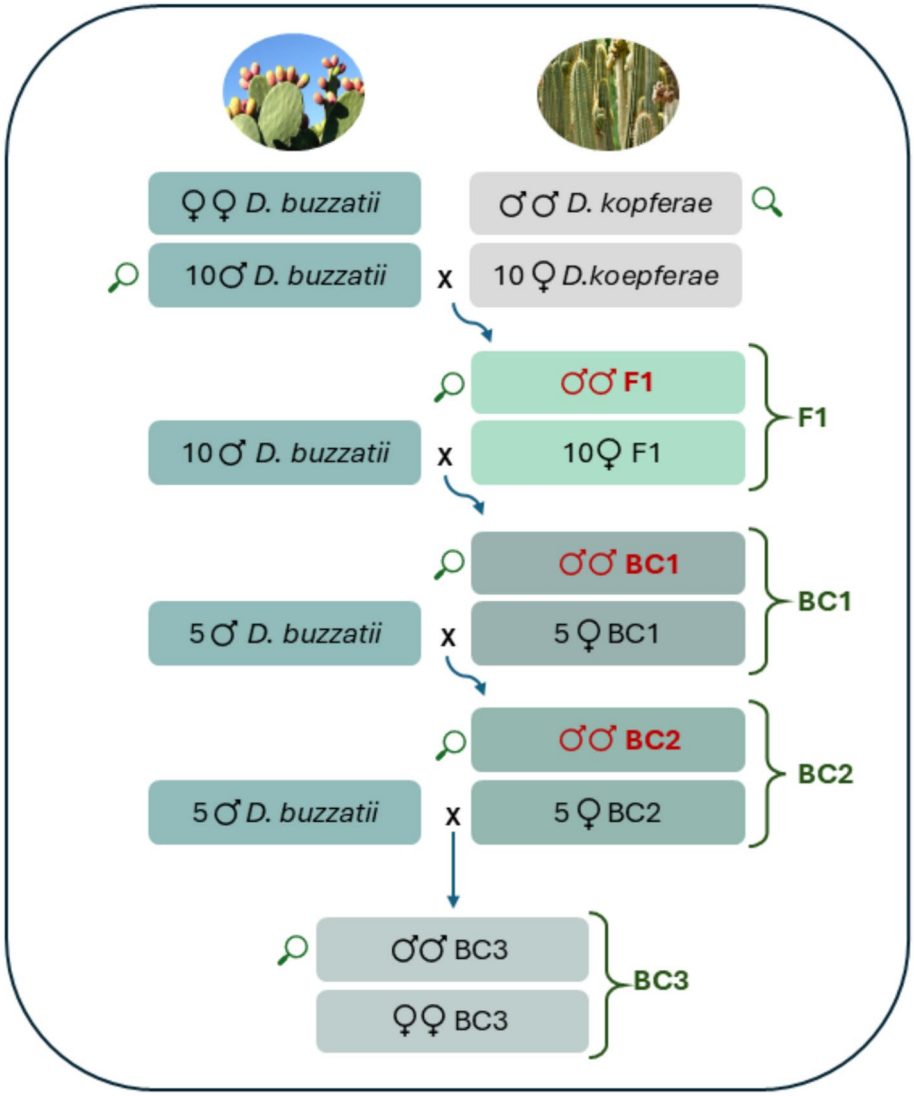

Fig. 1 Diagram of the interspecific crosses. The sequenced and analyzed samples are indicated by the magnifying glass. Next to each parental species is the cactus typical of its habitat. Rectangle colours represent the different genomic backgrounds of the individuals involved in each cross. Males in red are sterile

This new differential colour will allow us to easily distinguish both species, avoiding using the *aedeagus* shape, a method that requires applying pressure to the fly's abdomen, often causing its death. Regarding hybrids, they maintain *D. buzzatii* thorax colour along all generations, making them clearly distinguishable from *D. koepferae*.

Body size in parental species and every generation of hybrids was also measured (Fig. 2A) and compared (Additional file 2: Tables S1 and S2). *D. koepferae* males were statistically bigger (p-adj $= 5.66 \times 10^{-4}$) than *D. buzzatii* ones (Additional file 2: Tables S1 and S2). F1 hybrids exhibited the biggest size (Fig. 2B) with a mean of 2.78 mm and a progressive size decrease from the F1 generation onwards (Additional file 2: Table S1). When hybrid generations were compared with each other and with their parents, we observed that all pairwise comparisons were significant, except BC1 *vs* BC2 (Additional file 2: Table S2).

F1 hybrid males, in addition to being larger, exhibited testes atrophy compared to both parental species (Additional file 1: Fig. S2A). By dissecting and puncturing the testis and their seminal vesicles, we observed that F1 and BC1 male hybrids did not present any spermatozoa (Additional file 1: Fig. S2B). Most of the BC2 hybrids had testes atrophy and sperm absence, but some of them had more developed testes with cysts, although no mature

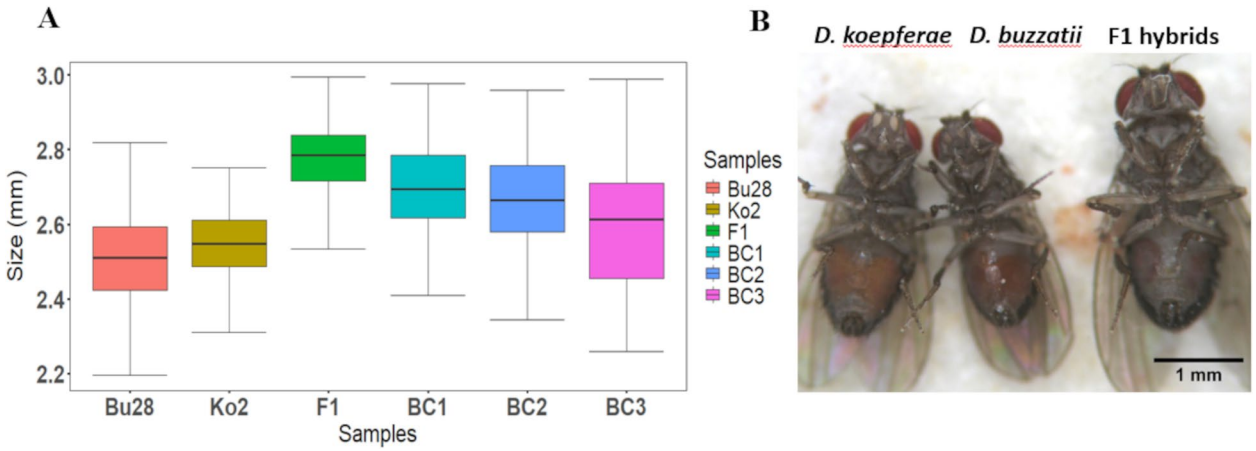

Fig. 2 Male body size of parental species and hybrids. **A** boxplot representing the measures of male body size for parental species and every hybrid generation. **B** picture of males from *D. koepferae*, *D. buzzatii*, and F1 hybrids

sperm was found in the seminal vesicle. Finally, BC3 hybrids presented a *D. buzzatii*-like gonad phenotype with motile spermatozoa (Additional file 1: Fig. S2C).

High gene deregulation in F_1 hybrid testes vs parental species

We first compared the testicular transcriptome of the parental species *D. buzzatii* and *D. koepferae*, and we found that 21.74% of the genes were differentially expressed (Additional file 2: Table S3), with an enrichment of GO terms involved in general processes (metabolic, cellular and multicellular organismal processes, localization and biological regulation), but also in some interesting functions, such as response to stimulus, immune system and developmental processes, reproduction and biological processes involved in interspecies interaction between organisms (Additional file 2: orange in Table S4). We then studied the impact of hybridization on the testicular gene expression by comparing the transcriptome of F1 hybrids and three backcross generations to each parental species. We detected approximately twice as many DEG in F1 hybrid testes *vs* parental species (~41% of the genes, Additional file 2: Table S5) than when comparing the two parental species (~22%, Additional file 2: Table S3). However, the number of DEG was markedly reduced in the backcrosses, with DEG accounting for only 23–27% *vs D. koepferae* and 0.74–5.7% *vs D. buzzatii* (Additional file 2: Table S5 and Two proportion Z-test, all $p < 0.001$, Additional file 2: Table S6). Among these DEG in F1 hybrids *vs* parental species, there is a notable enrichment of GO terms involved in developmental processes and reproduction (Additional file 2: Table S4). In fact, we observed some sperm-related GO terms enriched in F1 hybrids *vs* both parental species, such as flagellated sperm motility, sperm individualization, cilium-dependent cell motility, and axonemal dynein complex assembly (names in bold in Additional file 2: Table S4). Gene expression in F1 hybrids was also detected to be more similar to the maternal species *D. koepferae* than to the paternal species *D. buzzatii* (Two proportion Z-test, $p = 0.001$, Additional file 2: Table S6). As expected, the percentage of DEG in each backcross *vs D. buzzatii* decreased by generation until stabilization (BC1, 5.71%; BC2, 2.20%; BC3, 0.74%, Additional file 2: Table S5). Nonetheless, the percentage of DEG in each backcross *vs D. koepferae* approached the differences between parental species (BC1, 27.60%; BC2, 23.63%; and BC3: 24.26%, Additional file 2: Table S5).

We then classified the DEG in hybrids *vs* parental species in three categories: *D. buzzatii* or *D. koepferae*-like expression, additive expression, and deregulated genes, as described in our previous work [8] (Fig. 3A). We found a higher number of DEG in F1 hybrids *vs* both parental species (deregulated genes) than in all backcrosses (Fig. 3B) with a tendency towards underexpression. However, deregulated genes had a tendency towards overexpression in the backcrosses (Fig. 3B and Additional file 2: Table S7). Moreover, deregulated genes were enriched in some interesting GO terms involved in reproduction and developmental processes that are worth mentioning, such as sperm individualization and flagellated sperm mobility in F1 hybrids, in embryonic morphogenesis in BC1, anatomical structure development in BC2 (examples in bold, Additional file 2: Table S8). Interestingly, most deregulated genes in F1 hybrids were also involved in general processes that could cause a systematic collapse of the individual, such as GO terms involved in protein processing and

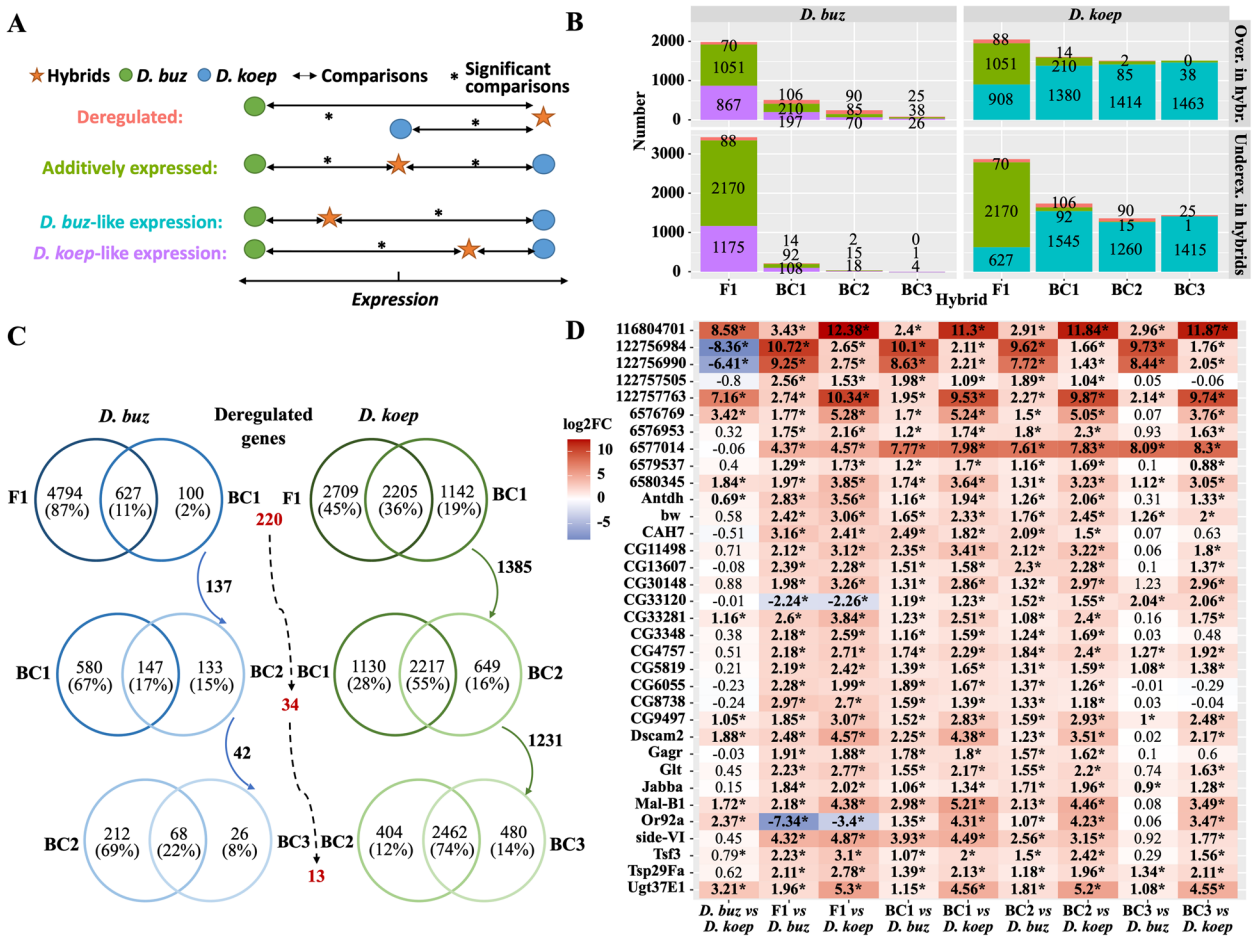

Fig. 3 Differential expression analyses of genes. **A** Expression categories in hybrids *vs* parental species. **B** Number of differentially expressed genes (over and underexpressed) in hybrids and backcrosses *vs* parental species from the total of annotated genes. Colours indicate gene expression categories. **C** Venn diagrams showing the number of differentially expressed genes shared by generation in F1 hybrids and backcrosses *vs D. buzzatii* (blue) and *D. koepferae* (green). Differentially expressed genes maintained in each generation are shown in bold black. Deregulated genes in hybrids *vs* both parental species maintained in each generation are shown in bold red. **D** Heatmap of the 34 differentially expressed genes maintained as deregulated in F1 hybrids, BC1, and BC2. Names of the orthologs in *D. melanogaster* are shown when available (bottom 24 genes). For the rest, the annotation identification number is shown (top 10 genes)

transport, mitochondrial functions, metabolic processes, or cell organization (Additional file 2: Table S8). In fact, the most affected KEGG pathway in F1 hybrids (Additional file 2: Table S9) was the metabolic pathway (306 DEG), mainly those genes involved in energetic metabolism as oxidative phosphorylation or TCA (tricarboxylic acid cycle), where a clear downregulation was observed throughout the cycle (Additional file 1: Fig. S3A). The motor protein pathway was also affected by the deregulation of the motor protein structural genes, which were mostly underexpressed and affected mainly the axonemal dynein, kinesin, and myosin (Additional file 1: Fig. S3B).

Due to the observed increase in body size in hybrids, we checked two interconnected pathways related to this phenotype in *Drosophila melanogaster*: mTOR and insulin signaling, where several genes were deregulated in our F1 hybrids. These pathways include genes involved in cell growth and proliferation processes (*Ras*, *Lkb1*) and some genes related to male sterility, such as *gskt* or *eIF4E* genes (Additional file 1: Fig. S4). When we extended the analysis to genes in F1, related to an increase in body size reported in FlyBase, we detected 26 DEG *vs* parental species. Among these genes, 69% exhibited overexpression (Additional file 1: Fig. S5), such as *sit*, *RPs6*, and *Ras85D*, whereas 23% were underexpressed *vs* both parental species (e.g., *Tor, Hr4, Dis3L2*).

Differential expression of specific genes is maintained across hybrid generations

We then studied the DEG in hybrids *vs* parental species that were shared by generations and maintained in each cross (Fig. 3C). We found lower percentages of genes that shared differential expression in different generations in comparison to *D. buzzatii* than to *D. koepferae* and also of genes that maintained differential expression until the last backcross (42 *vs D. buzzatti* and 1231 *vs D. koepferae*, bold numbers in Fig. 3C). Moreover, the percentage of DEG shared between backcrosses increased across generations (Fig. 3C) *vs D. koepferae* and decreased *vs D. buzzatii*. This can be explained by the introgression of the *D. buzzatii* genome in the hybrid genome through the different generations of backcrossing. This idea is supported by the detection of enriched GO terms between backcrosses involved in functions already detected when the parental species were compared, such as proteolysis or peptidoglycan catabolic and mannose metabolic processes (bold names in Additional file 2: Table S10).

Regarding deregulated genes, we observed that 220 genes were maintained as deregulated *vs* both parental species from the first generation to the BC1 (red numbers, Fig. 3C) and were involved in general processes that could affect individual homeostasis, such as protein autoprocessing, embryonic morphogenesis, stabilization of membrane potential, or intein-mediated protein splicing (bold names in Additional file 2: Table S11). In addition, 34 of these genes were still deregulated in BC2 (red numbers, Fig. 3C), and this number dropped to 13 in BC3. The study of these 34 deregulated genes (Fig. 3D) showed that they were mostly overexpressed and commonly expressed in primal developmental embryo/larval stages. They are also involved in the formation of different structures, such as the digestive or nervous system (Additional file 2: Table S12). In addition, we detected deregulation of a gene (CG11498) involved in mitochondrial protein catabolic process expressed in spermatozoa and a gene whose expression increases in response to the *Gypsy* retrovirus (*Gagr*). The expression of most of these genes, including the latter, was recovered in BC3 (Fig. 3D and yellow genes in Additional file 2: Table S12). They are associated with different functions such as development (Jabba, *CG3348*) and metabolism (*Tsf3, Mal-B1, CG9497 and Antdh*), among others. Only 13 genes were maintained deregulated along all generations (Fig. 3D), and most of them are involved in developmental processes (*Ugt37E1* and *Tsp29Fa*) and metabolism (*CG4757* and *CG33120)*.

Finally, we studied the distribution of DEGs between autosomes and sex chromosomes (Additional file 2: Table S13). We observed a different number of DEGs in most F1 hybrids and backcrosses, compared to the parental species, with a trend to be overrepresented in the sex chromosomes. Comparisons between the sexual chromosomes and autosomes showed significant differences in the number of DEGs in F1 hybrids *vs D. koepferae* species, but not *vs D. buzzatii*. Regarding backcrosses, the three BCs were significantly different between autosomes and sex chromosomes *vs D. buzzatii,* but not in *D. koepferae* (Additional file 2: Table S13).

Deregulation of genes involved in reproduction and sterility in hybrids

To gain insight into the effect of hybridization on reproduction and sterility, we analyzed the genes involved in the following GO terms: spermatogenesis, male gamete generation, male sterility, male infertility, sexual reproduction, reproduction process, and reproduction. From the 1497 annotated genes involved in these functions, we observed that ~45% already showed differential expression between parental species (Additional file 2: Table S14). Focusing on the hybrids, we detected that around 68% and 65% of these genes were differentially expressed in F1 hybrids *vs D. buzzatii* or *D. koepferae*, respectively (Additional file 2: Table S14). Among them, only 26% exhibited differential expression with respect to both parental species (383 deregulated genes, Additional file 2: Table S15). In BC1, a high proportion of these genes were still differentially expressed *vs D. buzzatii* or *D. koepferae* (35.40% and 57.38% respectively, Additional file 2: Table S14), but only ~2% were deregulated (Additional file 2: Table S15). In the following backcrosses, the percentage decreased when compared to *D. buzzatii* and stabilized when compared to *D. koepferae* (Additional file 2: Table S14), and most of these genes were not deregulated (maximum of 0.53%, Additional file 2: Table S15). Regarding spermatogenesis genes (total 297) reported in FlyBase (FlyBase 2024_05 Dmel Release 6.60), 105 were underexpressed and 11 were overexpressed in F1 hybrids *vs* both parental species (Additional file 1: Fig. S6A). In addition, we analyzed sperm motility genes, finding 21 deregulated genes underexpressed in F1 hybrids (Additional file 1: Fig. S6B).

When we focused on the deregulated genes maintained by generation, we observed that only 20 genes involved in reproduction and sterility were deregulated in F1 and BC1 *vs* both parental species (Additional file 2: Table S15), which corresponded to 9.1% of the deregulated genes maintained in the first two generations. The in-depth study of these genes (Fig. 4A) showed that some of them (*mil*, *Fas3*, *Eip74EF*, *fan*, and *rpk)* were involved in spermatogenesis or spermatozoa mobility and were mostly overexpressed in the hybrids, except *fan* and *mil* (bold red genes in Additional file 1: Fig. S6A) and their specific functions are described in Additional file 2:

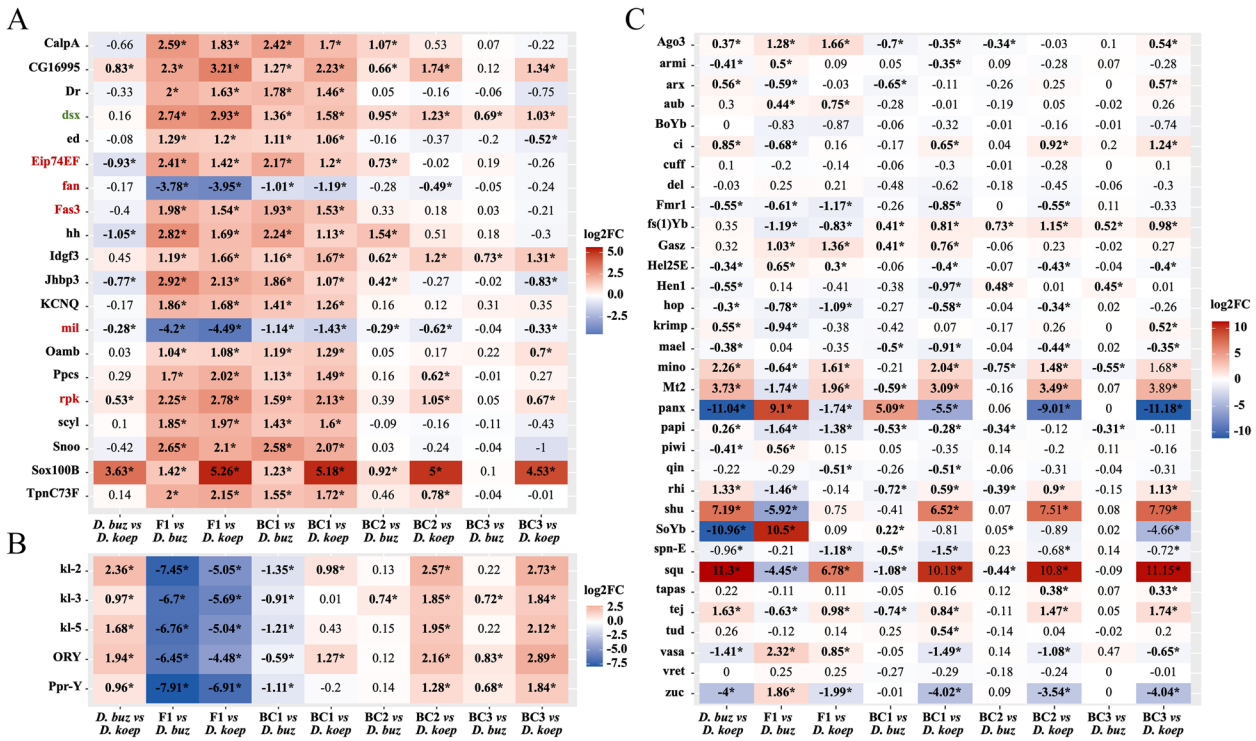

Fig. 4 Expression changes on testes of specific gene groups in hybrids and backcrosses vs parental species. **A** Deregulated genes in F1 hybrids and BC1 involved in reproduction and sterility. Genes in red letters are involved in spermatogenesis. The gene in green letters is related to a sterile phenotype. **B** Chromosome Y genes. **C** piRNA pathway genes. Significant values ($p < 0.05$) are indicated with an asterisk and in bold. Colors indicate the values of the differences in gene expression ($\log_2$FC)

Table S16. Only 2 and 5 genes involved in reproduction and sterility were still deregulated ($\log_2$FC > |1|, Additional file 2: Table S15) in BC2 compared to *D. buzzatii* (*CalpA* and *hh*) and *D. koepferae* (*CG16995*, *dsx*, *Idgf3*, *rpk*, *Sox 100B*), respectively (Fig. 4A).

The genes from the Y chromosome play an important role in *Drosophila* male fertility [31]. Thus, we analyzed their expression between parental species and observed that all of them were differentially expressed (Fig. 4B). In F1 hybrids, we detected that all genes were underexpressed *vs* both parental species. However, some of these genes recovered their expression in BC1 by showing additive expression. Gene expression in BC2 tended to be like *D. buzzatii*, except for *kl-3*, which was overexpressed in most backcrosses *vs* both parental species (Fig. 4B). It is important to note that some Y genes in *D. melanogaster* migrated to autosomes in *D. buzzatii* and *D. koepferae* species, and they show the same underexpression tendency as those that remained on the Y chromosome in *D. melanogaster* (Additional file 1: Fig. S7). Nevertheless, they recovered their expression in BC1, one generation earlier than those of the Y chromosome, except for *PRY*, which recovers in BC2.

We finally examined the expression of the piRNA pathway genes for their role in germline TE silencing [32] and interspecific male sterility [33]. As observed in Fig. 4C, we found differential expression for most of these genes in testes when parental species were compared. Regarding F1 hybrids, we detected a few deregulated genes *vs* both parental species: overexpression of *Ago3*, *aubergine*, *Gasz*, *Hel25E*, and *vasa*, and underexpression *of Fmr1*, *fs(1)Yb*, *hop*, and *papi*. In the backcrosses, we observed that, in general, gene expression progressively shifted toward the values observed in *D. buzzatii* or was additive between parental species, and hence parental differences were maintained. However, some deregulation was still detected in BC1: overexpression of *Gasz vs* both parental species, and underexpression of *Ago3*, *mael*, *papi*, and *spn-E*. Finally, we detected an overexpression of *fs(1)Yb vs* both parental species in all backcrosses.

Transcriptional variation associated with cis- and trans-regulatory changes

Differences in gene expression between species are often attributed to regulatory divergence, which can occur in *cis*, through mutations in regulatory DNA sequences; or in *trans*, via changes in the abundance or activity of

regulatory factors such as transcription factors or non-coding RNAs. To assess the extent and nature of regulatory divergence between *D. buzzatii* and *D. koepferae*, we performed allele-specific expression (ASE) analyses in testes, building on our previous framework [8] and categorizing genes in the classes described in McManus et al. [34]. These categories differentiate between *cis*-only, *trans*-only, *cis* + *trans*, *cis* × *trans*, compensatory, and conserved regulation. As shown in Fig. 5A, most genes (32.87%) did not show regulatory divergence between parental species and were categorized as conserved. Regarding genes with regulatory divergence, we detected more genes showing *cis*-regulatory divergence (10.18%) than *trans* (8.64%) (*Z*-test, $p = 3.19 \cdot 10^{-4}$, Additional file 2: Table S17). We then analyzed how the regulatory divergence influenced gene expression between parental species, and we found more DEG with both *cis*- and *trans*-regulatory divergence (Fig. 5B, Fisher's exact test, $p < 0.01$, Additional file 2: Table S18) than expected. Therefore, the regulatory divergence between parental species seems to account for the differences in expression, slightly stronger in the case of *trans*-regulatory divergence. Finally, we studied how the regulatory divergence between *D. buzzatii* and *D. koepferae* influenced gene expression in hybrids by grouping DEG in hybrids *vs* parental species by their expression category (Fig. 3A). We found a significantly higher number of genes with compensatory regulation (*cis*- and *trans*-regulatory differences that compensate each other) in F1 deregulated genes than in other expression categories (Fig. 5C, Fisher's exact test, all $p < 0.001$, Additional file 2: Table S19). Regarding the backcrosses, we observed that compensatory regulation still seemed to be common in deregulated genes.

Reduced TE activity in F1 hybrids and backcrosses

To analyze the effect of hybridization on TE expression, we first annotated the TE content of *D. buzzatii* and *D.*

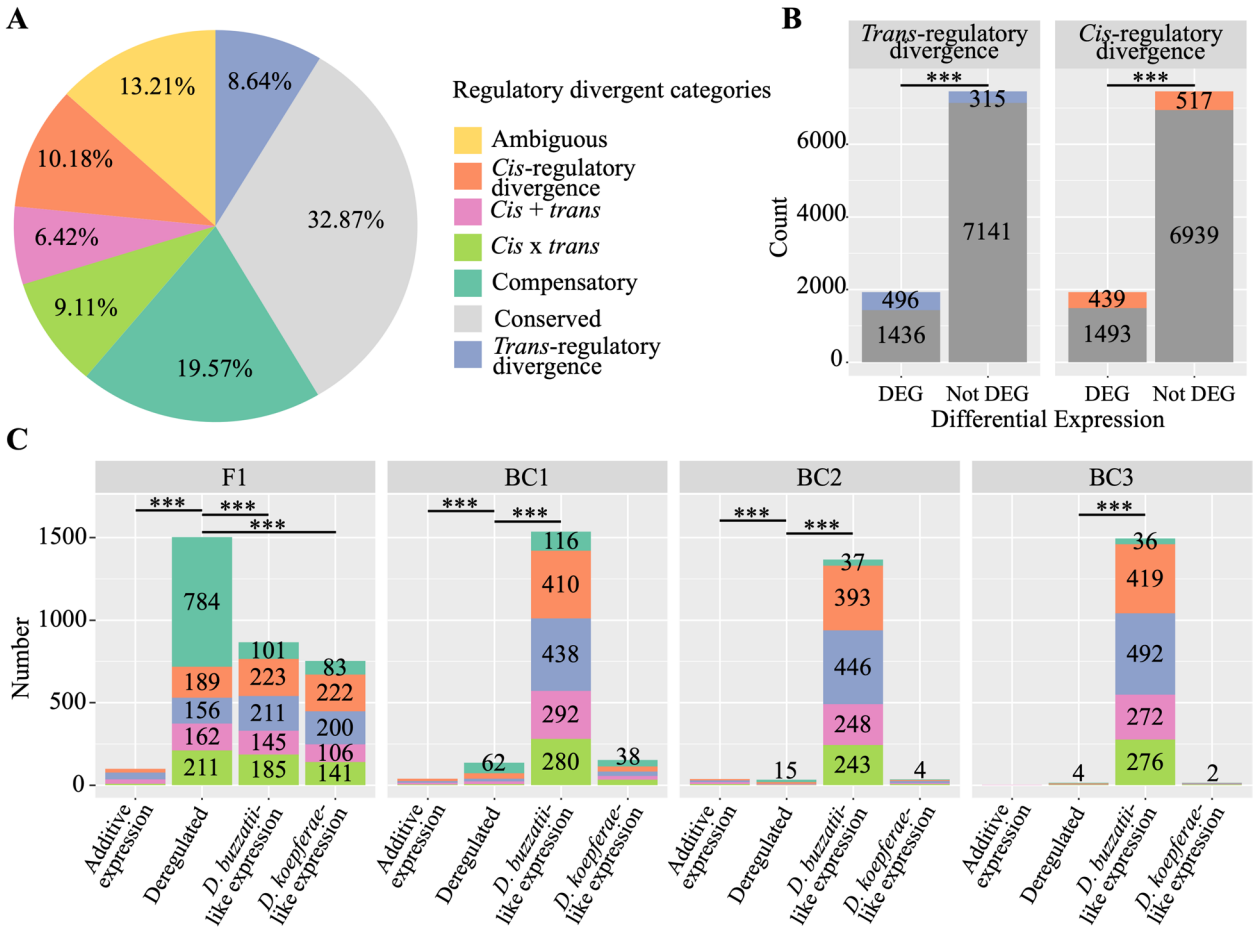

Fig. 5 Regulatory changes affecting gene expression between parental species. **A** pie chart showing the percentage of total genes categorized by regulatory divergence between parental species. **B** number of differentially expressed genes between parental species with *trans*- or *cis*-regulatory divergence. **C** number of differentially expressed genes in hybrids and backcrosses *vs* parental species, grouped by each regulatory category and per differential expression category. Fisher's exact test results are shown in the top of **B** and **C**. ***: $p < 0.001$

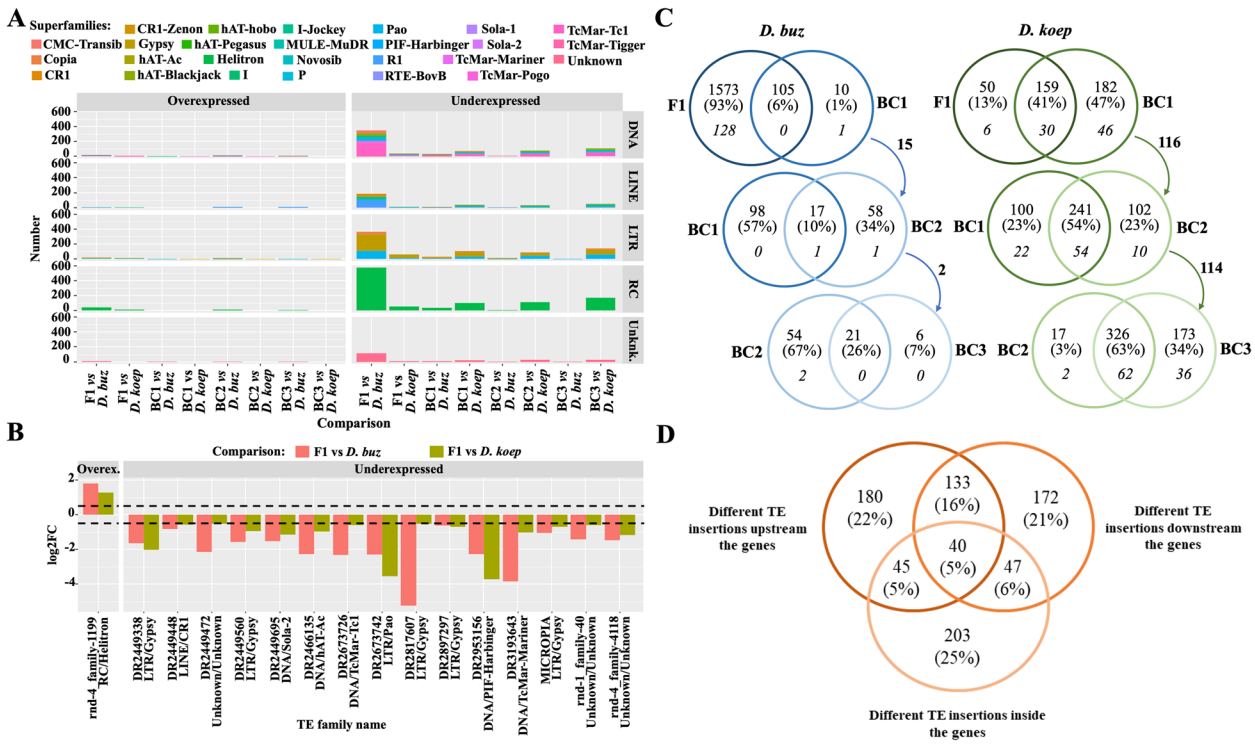

Fig. 6 Differential expression analyses of TEs: **A** Number of differentially expressed TEs in hybrids and backcrosses *vs* parental species by TE superfamily. **B** Expression differences (log_2FC) for the differentially expressed TE families in F1 hybrids *vs* both parental species (as deregulated TE families). **C** Venn diagrams showing the number of differentially expressed TE copies (top) and families (bottom italic) shared by generation in F1hybrids and backcrosses *vs D. buzzatii* (blue) and *D. koepferae* (green). Differentially expressed TEs maintained in each generation are shown in bold. **D** Venn diagram of the significant differentially expressed genes between *D. buzzatii* and *D. koepferae* with different TE insertions between parental species inside, or 5 kb upstream or downstream

koepferae. We detected that around 15% of the sequenced genomes (both of a size of ~ 175 Mb) were covered by putative TE families (Additional file 1: Fig. S8A). The TE content was mostly derived from Rolling Circle (RC) or LTR orders (around 9% in *D. buzzatii* and 10% in *D. koepferae*, Additional file 1: Fig. S8A). The annotated elements with the greatest number of insertions in the parental genomes were *Helitrons*, followed by *TcMar-Tc1* and *Gypsy* superfamilies (Additional file 1: Fig. S8A). We annotated a total of 35,032 TE copies in *D. buzzatii* and 29,942 in *D. koepferae*, but only 5.48% (1919 copies) and 5.66% (1695 copies) were expressed in *D. buzzatii* and *D. koepferae*, respectively.

We then studied the impact of hybridization in TEs by analyzing the expression of each TE copy coming from a parental species in F1 hybrids and backcrosses compared to the parental species they came from. Only 20% of the *D. buzzatii* copies and 12% of the *D. koepferae* copies were expressed in a parental species or a hybrid (Additional file 2: Table S20). We then analyzed the expression of these active TE copies in F1 hybrids and backcrosses. We detected more expression differences of *D. buzzatii*-TE copies in F1 hybrids than of *D. koepferae ones* (24.27% of differentially expressed TE copies *vs D. buzzatii* and 5.77% *vs D. koepferae*, two proportion Z-test, $p < 0.001$, Additional file 2: Tables S21 and S20), in both cases with a tendency towards underexpression *vs* parental species. In the backcrosses, we detected a significant decrease of expression differences *vs D. buzzatii* (Two proportion Z-test, all $p < 0.001$, Additional file 2: Tables S21 and S20) until stabilization, with a tendency towards underexpression in BC1 and overexpression in BC2 and BC3 (Additional file 2: Table S20). TE copy differential expression *vs D. koepferae* significantly increased (Two proportion Z-test, all $p < 0.001$, Additional file 2: Tables S21 and S20) by generation, always with a tendency towards underexpression.

Focusing on the superfamilies of the differentially expressed TE copies, we detected overexpression mainly of *Helitrons* in F1 hybrids *vs* both parental species, whereas superfamilies belonging to *Gypsy* and *R1* were also highlighted in the backcrosses (Fig. 6A and Additional file 2: Table S22). Underexpressed TE copies were mostly from *Helitron* and *Gypsy* superfamilies in F1 hybrids and backcrosses (Fig. 6A and Additional file 2: Table S22). Then, we analyzed in more detail the

most differentially expressed TE copies in hybrids and backcrosses *vs* parental species. We found that the most 50 overexpressed and underexpressed TE copies *vs D. buzzatii* were mainly detected in F1 hybrids (Additional file 1: Fig. S9A and B). However, we also observed strong activation in the backcrosses, specifically in BC2 and BC3. The 50 most activated TE copies *vs D. koepferae* were mainly found in F1 hybrids (Additional file 1: Fig. S10A), with also a few activations in the backcrosses. Nonetheless, the 50 most underexpressed TE copies in hybrids *vs D. koepferae* were mainly shown in the backcrosses due to the introgression of the *D. buzzatii* genome (Additional file 1: Fig. S10B).

To analyze the overall TE family expression, we compute the mean $\log_2$FC of all expressed TE copies of a TE family, and we consider as significantly differentially expressed the TE families with a mean absolute $\log_2$FC of at least 0.5. At a family level we can still detect more differences in TE family expression in F1 hybrids *vs D. buzzatii* than *vs D. koepferae* (76.65% of the TE families *vs D. buzzatii* and 19.35% *vs D. koepferae*, Additional file 2: Table S20), with a general trend towards underexpression. The number of differentially expressed TE families in the backcrosses decreased *vs D. buzzatii* and increased *vs D. koepferae*, mostly with a tendency towards underexpression (Additional file 2: Table S20). However, we detected 16 TE families that were deregulated in F1 hybrids *vs* both parental species (Fig. 6B): 15 were underexpressed (5 DNA TE superfamilies, 6 LTRs: from which 5 were *Gypsy*, one LINE, and 3 unknown) and 1 *Helitron* was overexpressed.

We then analyzed how many differentially expressed active TE copies (and TE families in italics) were differentially expressed in more than one hybrid generation (Fig. 6C), and which ones were maintained as differentially expressed in the backcrosses (curved arrow numbers). We detected that most differentially expressed TE copies in F1 hybrids *vs D. buzzatii* were not differentially expressed in the backcrosses. However, 15 of these copies were still deregulated in BC2. We also detected that most differentially expressed TE copies in F1 hybrids *vs D. koepferae* were maintained in BC1, and this number increased by generation, as it does when we focus on the family level. In addition, 114 TE copies were maintained as differentially expressed from F1 and until BC3.

TE insertions and gene expression variability in parental species

We observed that 32.44% of the DEG (820 genes) between *D. buzzatti* and *D. koepferae* were detected to have different TE insertions between parental species in flanking regions or inside (Fig. 6D). Since TEs can alter gene expression when inserted within or near genes, we analyzed whether the differences in gene expression between parental species could be related to different nearby TE insertions close (±5 kb upstream or downstream) or inside genes. To this end, we identified genes harboring species-specific insertions, such as those present in one species but absent in the other, *versus* genes lacking insertions in both species. In addition, we quantified genes exhibiting conserved insertions across both species and contrasted them with genes containing divergent TE insertions. Our results showed that when the insertion is present in one species and absent in the other, an effect on the gene expression is detected. The odds ratio (OR) is > 1 in the three gene regions, suggesting that the likelihood of having differential expression is higher in the genes with TE insertion in one species *vs* the genes without TE insertion in any of the species (Fisher's exact test, all $p < 0.05$ Additional file 2: Table S23). Upstream and downstream regions of genes with the same insertion in both parental species *vs* those having different insertions did not show statistically significant association between insertion type and expression difference (Fisher's exact test, $p > 0.05$, Additional file 2: Table S23). However, a significant association is observed when the differential insertions were located inside the gene. In this case, OR values < 1 indicate that genes with the same insertion are less likely to exhibit differential expression compared to those with different insertions. These results suggest that TE insertions between parental species might lead to a species-specific expression, supporting the importance of TE expression in gene regulation.

Discussion

Hybridization between *Drosophila buzzatii* and *D. koepferae* offers a unique opportunity to dissect the molecular basis of postzygotic isolation. Here, we combined phenotypic, transcriptomic, and TE analyses across F1 and three backcross generations to investigate the origin of hybrid sterility and how it resolves. We detected a complex landscape of misregulation that includes gene expression, TE activity, regulatory divergence, and chromosomal effects.

Phenotypic anomalies in F1 hybrids

Phenotypic assessment confirmed that F1 hybrid males exhibit significant testicular atrophy and lack of motile sperm, aligning with previous studies of hybrid sterility in *Drosophila* [3, 4]. This sterility was coupled with an increase in body size in hybrids compared to both parental species, followed by a progressive reduction in subsequent generations (BC1–BC3). The enlarged size in F1 hybrids is not new and may reflect gene overdominance, a phenomenon documented in other *Drosophila* hybrid systems [35]. It is well established that many

genes are associated with body size, and some of them have been linked to mTOR and insulin-signaling pathways [36]. We observed dysregulation of this gene category in F1 hybrids, which includes overexpression of known growth-related genes such as *Ras85D* or *RPS6*. These genes have been associated with an increase in pupal size [37] or cell proliferation [38] and increased protein synthesis [39], respectively. On the other hand, some body size-related genes showed underexpression in F1 compared to parental species. While it is true that the underexpression of most genes cannot be directly linked to the increased body size observed, at least two underexpressed genes (*Hr4* and *Dis3l2*) have been associated with size increases in *Drosophila* [39, 40], and they could contribute to the increased size observed in our hybrids. Most genes recovered the expression levels of *D. buzzatii* after the F1 generation due to the introgression of this species' genome through successive backcrosses. However, two genes—*Oamb* and *Alk*—remained overexpressed in the BC1 generation compared to both parental species. These genes are involved in pupal size [41] and growth regulation [42] in *Drosophila*. In the BC2 and BC3 generations, all genes largely recovered the expression profile of *D. buzzatii*, which may account for the gradual reduction in body size.

F1 and BC1 hybrids exhibited complete testicular atrophy and absence of spermatozoa, while subsequent generations showed partial recovery: cysts without mature sperm (BC2) or restored motility (BC3). Testicular atrophy was already reported in hybrids between these [6] and other [43] *Drosophila* species. F1/BC1 testicular atrophy could physically impede sperm release, mirroring the azoospermia in *D. willistoni* [44] and *D. mojavensis* [45] hybrid sterility. These results underscore the complexity of genetic interactions in speciation, where morphological and physiological traits synergistically limit gene flow between closely related species already pointed out and compiled by other authors [15, 46].

A widespread gene deregulation in hybrids

The transcriptomic analysis of testicular gene expression in hybrids and backcrosses between *D. buzzatii* and *D. koepferae* revealed extensive regulatory divergence and hybrid misexpression. These findings align with and expand upon recent literature on regulatory evolution and hybrid incompatibilities in *Drosophila* [2, 47, 48] and other taxa [49]. The comparison of testicular transcriptomes of the parental species *D. buzzatii* and *D. koepferae* already revealed that 21.74% of the genes were differentially expressed and involved in a wide range of biological processes, such as cellular functions or metabolism. Remarkably, more specific function-related GO terms were also enriched; for example, response to stimulus, immune system processes, developmental processes, and reproduction. These findings are consistent with previous studies showing that reproductive tissues tend to exhibit accelerated transcriptomic evolution and high levels of regulatory divergence [50, 51]. However, the percentage of DEGs was higher than reported before (5.9%) in these species in ovaries [8]. Furthermore, these expression differences are not only detected in germline tissues since similar patterns were also detected in larval and head tissues in other studies [52, 53], showing that these species have developed distinct transcriptional strategies in response to their different ecological environments.

In F1 hybrids, the percentage of DEG (~41%) is twice as high as the value observed between the parental species, with strong enrichment in metabolic, developmental, and reproductive processes, for example sperm motility and individualization of sperm. These values are higher than those obtained in hybrid females (~4%) of these same species in a previous study [8]. This result is not unexpected, given that hybrid females are fertile or semi-fertile, in contrast to males, which are completely sterile [54]. The findings obtained in hybrid males agree with previous research results showing that reproductive tissues, especially testes, are hotspots for rapid evolution of gene expression and divergence between species [27], often driven by sexual selection and reproductive isolation processes. Some metabolic pathways, especially the ones associated with oxidative phosphorylation and the TCA cycle, are downregulated in F1 hybrids, suggesting compromised energetic metabolism. Mitochondrial dysfunction and hybrid sterility were reported in plants [55] and, in *Drosophila*, a transgressive misexpression in pathways including metabolism was also observed [2]. This increased misexpression in hybrids is an indicator of hybrid dysfunction and sterility, as reported in other *Drosophila* hybrid systems, such as *D. melanogaster*×*D. simulans*. In this case, misregulation in reproductive tissues underlies postzygotic isolation [27]. Deregulated genes in our F1 hybrids were primarily underexpressed, consistent with a collapse in development and reproduction gene networks. These results are consistent with those that show extensive evidence for downregulation of various spermatogenesis genes, a common trait for sterile hybrid males of *Drosophila* [43, 56] and *Caenorhabditis* [57].

A gradual resolution of gene misregulation occurred across generations, with only 0.74% of genes differentially expressed in BC3 *vs. D. buzzatii*, indicating near-complete transcriptomic restoration by the third backcross. This result represents a novel contribution of this study since previous research has focused primarily on F1 hybrids or isolated backcrosses, without tracking regulatory changes over multiple generations

[26, 28]. This perspective offers insights into how introgression may mitigate gene misregulation and could serve as a basis for models of how to reverse hybrid incompatibility. Interestingly, expression profiles in F1 hybrids are biased toward the maternal species (*D. koepferae*), suggesting a maternal bias in F1 hybrid expression concordant with previous studies showing that maternal contributions are often central to hybrid gene expression [58, 59]. However, backcrosses showed progressive similarity with *D. buzzatii* expression, consistent with the expected genomic replacement during introgression. However, the proportion of DEG *vs D. koepferae* remained rather stable across generations (~ 24%), reflecting lasting interspecific regulatory divergence and suggesting asymmetric regulatory compatibility between nuclear genomes rather than maternal effects or cytoplasmic inheritance. Given that the cytoplasm in all backcross generations is derived from *D. koepferae*, the observed transcriptional patterns may instead be influenced by complex regulatory interactions between parental genomes and/or the introduction of the *D. buzzatii* Y chromosome, which has been shown to impact gene expression and male phenotypes in *Drosophila* [60–62]. Nevertheless, although maternal inheritance could potentially play a role, we cannot verify this hypothesis due to the unviability of the reciprocal cross between these species [54].

In F1 and BC1 hybrids, 220 DEG compared to both parental species were shared, and a substantial number of them were associated with developmental, reproduction, and metabolic functions. Only 34 persisted as deregulated (most overexpressed) in the BC2 generation, and only 13 continued to be deregulated in BC3 individuals. Among the genes whose expression levels were restored to those of *D. buzzatii* in the BC3 generation, we found genes such as *Dscam2* and *Glt* that ensure axonal tiling, synaptic specificity, and tissue integrity during development in *Drosophila* [63, 64]. Among the genes whose deregulation persisted in BC3, we found genes associated with metabolism and developmental processes such as *Ugt37E1* and *Tsp29Fa*, related to detoxification, metabolic regulation [65], and maturation of secretory granules (SGs) during development [66], respectively. Other genes, such as CG4757 and CG33120, were also deregulated until BC3 and have roles in neural/sensory functions and lipid modification processes, respectively [67]. This is in concordance with the results of other reported studies, with F2 hybrids and backcrosses showing transgressive expression of some genes [28]. Some authors have interpreted these results by suggesting that altered gene expression patterns may arise because of hybridization itself, rather than being inherently associated with hybrid incompatibilities [51].

On the other hand, our analyses indicate that DEGs in F1 hybrids and backcrosses are not randomly distributed across the genome but tend to be overrepresented on the sex chromosomes. This pattern is particularly evident in comparisons between F1 hybrids and *D. koepferae*, where the number of DEGs on the sex chromosomes was significantly higher than on the autosomes. A similar but more restricted pattern was observed in backcrosses, where only significant enrichment on the sex chromosomes compared to *D. buzzatii* was found. Differences in gene expression between the X chromosome and autosomes have been previously reported. However, patterns vary across studies, with some showing an excess of X-linked DEGs [68], others a depletion [69], and some no significant differences [70]. Overall, our results seem to support the idea that sex chromosomes might be hotspots of regulatory changes and misregulation of genes in hybrids, but additional studies will be necessary to confirm this observation.

Misregulation of reproduction-related and Y-linked genes in hybrids

Our transcriptomic analysis showed that out of the 1497 genes associated with reproductive functions (e.g., spermatogenesis, male gamete generation, and sexual reproduction), around 65% were expressed differentially in testes of F1 hybrids compared to each parental species. From these, 26% (~ 383 genes) were classified as deregulated, meaning they significantly deviated from the expression of both parental species, a strong signature of hybrid-specific misregulation. Some of these genes are well-known for their roles in spermatogenesis in *D. melanogaster*, such as *mil*, *fan*, and *Eip74EF* (Fig. 3A). For example, *mil* encodes a kinase important to sperm tail formation and fertility; mutations in this gene can cause male sterility due to defective sperm individualization [71]. *fan* encodes an important protein for sperm development [72] and *Eip74EF*, a transcription factor induced by ecdysone, is essential for later stages of spermatogenesis [73]. The consistent deregulation of these genes in both F1 and BC1 males indicates direct transcriptional disruption of the gametogenic program, supporting the idea that developmental arrest is causing the absence of sperm, as observed in these two hybrid generations [44].

By the BC3 generation, the number of deregulated reproduction-related genes was drastically reduced, with only two genes showing a slight overexpression ($\log_2$FC < 1): *Idgf3* and *dsx.* The first gene encodes an imaginal disc growth factor involved in morphogenesis and male fertility, likely through effects on testis development and epithelial remodeling [74]. *dsx* is a master sex-determination gene whose alternative splicing regulates sex-specific differentiation, including testis and accessory

gland development [75]. The persistence of *dsx* misregulation has been associated with a disruption in the sexual identity of gonadal tissues in sterile interspecific *Drosophila* males [4, 47]. Nevertheless, considering that most BC3 hybrids are fertile and show fold change values lower than 1 compared to *D. buzzatii*, this expression difference does not seem to play a significant role in the recovery of fertility.

Y-linked genes, critical for male fertility and spermatogenesis, were uniformly underexpressed in F1 hybrid testes (Fig. 4B), a trend similarly observed in *D. melanogaster*×*D. simulans* hybrids [76]. These genes include well-known fertility factors such as *kl-2*, *kl-3*, and *kl-5*, which encode axonemal dynein heavy chains required for sperm flagellar motility [77, 78]. The consistent underexpression of these genes in F1 hybrids suggests transcriptional incompatibilities that have been associated in other studies with improper chromatin remodeling or failure of the male-specific lethal (MSL) complex to engage properly with heterochromatic Y regions [79]. The expression of most Y-linked genes persisted in BC1, consistent with the male azoospermia, but was restored by BC2 vs *D. buzzatii* ($\log_2$FC < 1). The only gene that remained slightly overexpressed until BC3 was *kl-3*, indicating a possible compensatory overshoot or dosage imbalance. Overexpression of axonemal dyneins can disrupt flagellar assembly and function [80], and its persistence in backcrosses may reflect suboptimal regulatory tuning even after fertility is phenotypically restored.

Regulatory divergence as a driver of hybrid gene misexpression

Cis- and *trans*-regulatory changes have repeatedly been implicated in both evolutionary divergence of gene expression and the misregulation observed in interspecific hybrids, where incompatible regulatory interactions can emerge [34, 81, 82]. A large proportion of genes (32.87%) showed conserved expression, indicating similar regulatory architecture in the testes of both parental species (Fig. 5A). However, many genes exhibited regulatory divergence, with *cis*-regulatory divergence (10.18%) being more frequent than *trans*-regulatory divergence (8.64%), underscoring the key role of *cis*-elements in creating species-specific expression patterns. These results are in concordance with those observed in other *Drosophila* species pairs (e.g., *D. melanogaster*×*D. simulans*), where *cis*-regulatory changes tend to dominate at shorter evolutionary distances [83]. In fact, previous findings propose that *cis*-regulatory changes tend to accumulate more readily due to their localized effects and reduced pleiotropic constraints [84]. Interestingly, when we looked at how expression differed between the parental species, we found that genes with both *cis*- and *trans*-regulatory differences were significantly more likely to be differentially expressed, suggesting that regulation changes are responsible for a substantial fraction of the transcriptional differences between species. This result is slightly different from previous expression analysis in ovaries of these same species, where a higher proportion of *trans*-regulatory divergence compared to *cis*-regulation was observed [8]. These differences are likely because *cis*-regulatory effects tend to exhibit greater sex-specific expression differences than *trans*-regulatory effects [85]. Our results support models where cumulative effects of *cis* and *trans* mutations contribute to fixed expression differences in *Drosophila* [86] and yeast [87]. In F1 hybrids based on their expression patterns relative to the parental species, genes with compensatory regulation (*cis*- and *trans*-regulatory changes act in opposite directions to preserve parental expression levels) were significantly overrepresented among deregulated genes in F1 hybrids. This finding supports the idea that compensatory regulatory evolution maintains stable expression in parental backgrounds, but it is altered in hybrids, unmasking a hidden regulatory divergence [81, 86]. When examining the backcross generations, we observed that compensatory regulation was still enriched among deregulated genes, especially in BC1 and BC2.

TE deregulation in hybrids is limited but directional

The genome annotation of *D. buzzatii* and *D. koepferae* revealed that approximately 15% of each genome (~175 Mb genome size) is composed of putative TEs, in line with previous estimates for cactophilic *Drosophila* species [52, 88]. The dominant TE orders were Rolling Circle (RC) and LTR retrotransposons, reflecting a typical *Drosophila* TE landscape, where these two orders were reported as major contributors to genomic expansion in *D. melanogaster* [89] and other *Drosophila* species [88, 90]. However, approximately 20% of annotated *D. buzzatii* and 12% of *D. koepferae* TE copies were transcriptionally active in testis. A study of TE mobility in *D. melanogaster* suggested that 24 TE superfamilies were potentially active [91], which represents around 66% of the superfamilies (36) detected in this species genome [92]. However, these results are not directly comparable as our comparison between parental species was conducted at the copy level. Regarding F1 hybrids, they exhibited significantly more differential expression of *D. buzzatii*-derived TEs (24.3%) than *D. koepferae*-derived TEs (5.8%), with a general trend towards underexpression. These results are in concordance with those

obtained in hybrid females between these same species using transcriptomic data [8]. In contrast, these findings differ from previous observations with expectations based on our previous studies [11, 21] and observations of other studies in *Drosophila* [93, 94] and plants [95] where increased expression of some TEs was observed. However, these results may not be directly comparable, as the sex of the individuals analyzed and/or the methodologies employed often differ. The underexpression result challenges the prevailing assumption that hybridization universally destabilizes TE silencing mechanisms. This finding may reflect a protective overactivation of the host silencing machinery in response to genomic stress, particularly for two piRNA pathway genes: *Ago3*, *Gasz*, which were highly upregulated in F1 hybrids *vs* both parental species. Such a compensatory mechanism has also been previously observed in other piRNA pathway genes in hybrid females of these species [10]. Concordantly with the increase of piRNA production and ping-pong signal in F1 males in a previous study [21]. However, the opposite was observed in *papi*, a gene involved in proper piRNA maturation in females [96], which was underexpressed in F1 hybrids *vs* parental species. Although its role in the female germline has been extensively characterized, its function within the male germline remains largely unexplored. Despite the general underexpression of TEs observed in F1 hybrids, certain elements belonging to the *Helitron* family were overexpressed. This highlights the complex mechanisms involved in TE regulation in hybrids, which include not only piRNA-mediated control but also regulation associated with chromatin state. Notably, a previous study in hybrid females of these species reported specific TE families whose altered expression relative to the parental species could be attributed to the differential enrichment of chromatin marks [8]. While in some cases the causes of TE misregulation in hybrids are known, in others the underlying mechanisms remain elusive. For instance, in a previous study on other *Drosophila* hybrids, the observed TE overexpression patterns did not correlate with piRNA abundance [13]. Likewise, in *Saccharomyces* hybrids, changes in the efficiency of TE copy number control were observed without any detectable effect on Ty1 element stability [97, 98]. The number of differentially expressed TEs decreased across backcrosses *vs D. buzzatii*, but increased *vs D. koepferae*, reflecting genomic replacement and possible epigenetic reprogramming. TE superfamilies most affected included *Helitrons* and *Gypsy*, which are the most abundant superfamilies in these species. However, only a few TE copies from these families showed persistent activation across generations. It is worth noting that a previous study of TE expression in BC1 hybrids between these same species reported that 10.6% of the expressed TEs were deregulated in BC1 ovaries, with a predominant trend toward upregulation [21]. Overall, our findings support the notion that the deregulation of specific TE copies can persist beyond the F1 generation. This suggests that, although the initial hybrid response may be to favor silencing, long-term genomic instability or introgression events may lead to the reactivation of dormant elements. Such reactivations could be the result of disrupted piRNA targeting, chromatin remodeling, or loss of maternal piRNA imprints during successive hybrid generations [33, 48].

TE insertions shape expression patterns

One of the key results of this study is the observed association between differential TE insertions and gene expression differences between *D. buzzatii* and *D. koepferae*. When we look at TE insertions located within genes or within a ± 5 kb flanking region, we found that 32.44% of the DEG between the two species were associated with TE insertions unique to each species. This result suggests that TEs may play a non-negligible role in shaping regulatory landscapes and promoting expression divergence, likely either directly or through linked regulatory changes. This may also help explain gene misregulation in hybrids in regions with high TE load. These results are consistent with a growing number of works highlighting the ability of TEs to act as potent regulatory modifiers [99, 100]. For *D. buzzatii* and *D. koepferae*, the fact that almost one-third of DEG are associated with differential TE insertions suggests that these elements have a significant role in driving regulatory differences between the species. Interestingly, the localization of these different insertions is important, especially if they are inside genes or close to them. For example, insertions within gene bodies or untranslated regions (UTRs) can affect splicing [101–103], whereas upstream insertions more likely affect transcription initiation. The observed variability in insertion sites includes multiple mechanisms through which TEs may contribute to regulatory divergence between these species. By differentially modifying gene expression patterns, TE insertions could generate novel adaptive phenotypes or postzygotic incompatibilities, which help to keep species separate [104]. If these insertions are linked to genes involved in reproduction, development, or hybrid sterility—as suggested for some cases in this study—they might directly contribute to reproductive isolation. Still, more work including functional analysis is needed to determine which of these mechanisms affected the expression of each gene with a TE insertion within or near it.

Conclusions

This study suggests that incompatibilities of *D. buzzatii–D. koepferae* hybrids may be due to a widespread deregulation of genes associated with developmental, reproductive, and metabolic functions. However, functional validation is required to confirm the causal role of these expression changes. Overall, these results reinforce the evidence accumulated in other species, suggesting that the number of sterility genes could be large. The gradual recovery of hybrid fertility, due to the *D. buzzatii* genome introgression, could be explained by the reestablishment of the initial expression levels of most genes.

TE copies tend to be strongly regulated in F1 hybrids with only a few overexpressed copies that may persist beyond the F1 generation. This is an indication that, even though the immediate hybrid response favors silencing, long-term genomic instability or introgression events may lead to the reactivation of silent elements. This study substantially advances our understanding of the mechanisms driving reproductive isolation, TE misregulation, and the evolutionary fate of hybrid genomes.

Methods

Drosophila stocks and crosses

We used *D. buzzatii Bu 28* and *D. koepferae Ko2* inbred strains described in our previous works [8, 12, 21] and maintained by brother–sister mating for at least a decade and then by mass culturing. Due to the scarce offspring obtained in interspecific crosses, we performed 45 independent crosses, each consisting of ten *D. buzzatii* males and ten *D. koepferae* females. Subsequently, 30 backcrosses were carried out using groups of ten F_1 hybrid females and ten *D. buzzatii* males. BC_2 and BC_3 individuals were then generated through ten additional backcrosses, each involving five mating pairs (Fig. 1). All stocks and crosses were reared at 25 °C in a standard *Drosophila* medium supplemented with yeast.

Body size measurement and sperm presence

Adult flies of *D. buzzatii*, *D. koepferae*, and their hybrids, aged 6 to 8 days posthatching were photographed using a Canon PowerShot G6 camera mounted on a stereomicroscope ZEISS Stemi SV 11, with a ruler as a reference system for subsequent analysis. Using ImageJ Fiji v.154.f [105], pixel-per-millimeter scales were established for each focal distance. Flies were measured individually by taking distances from the base of the antennae to the end of the abdomen.

To perform the reproductive characterization, testes from 6 to 8 days old flies were dissected into a drop of PBS 1X and examined. Sperm was then extracted by puncturing the seminal vesicle with an insect pin and examined as well. Images were taken with a ZEISS Axiolab 5 FL-LED microscope equipped with an Axiocam 208 color camera, and a Leica MZ16 F stereomicroscope with a Leica MC170 HD camera.

DNA extraction, genome assemblies, and de novo gene and TE annotation

The *Bu28* and *Ko2* reference genomes from Oliveira et al. [52] were used in this work. As a summary, DNA was extracted using the Qiagen DNeasy Blood&Tissue kit, the genomes were sequenced using Nanopore MinIon Mk1b sequencer, base calling was performed using *Albacore* v.2.3.3 and quality control with *Nanoplot* v.1.10.2 (https://github.com/wdecoster/NanoPlot). Contig assembly was done using Flye v.26 [106], and polishing was performed with four rounds of *RACON* v1.3.2 [107] and *minimap2* v2.16 [108] contig alignments. Assembly incongruences were corrected with *samtools* v1.9.0. [109], function *faidx*; and *Gepard* v1.4.0 [110]. Super scaffolding was performed with *RaGOO* v.1.1 [111], and the chromosome-scale assemblies were submitted to the BUSCO v.5.4.3 [112] analysis with lineage *diptera_odb10* (*-l* parameter). For the allele specific expression analysis, a polished version of the Nanopore genomes using *NextPolish* and publicly available Illumina paired-end genomic reads [113] and RNA-seq reads [115] was used. Gene annotation was performed using *Liftoff* v.1.6.3 [114] with the *D. mojavensis* (GCA_018153725.1) reference genome [115] (excluding genes with descriptions as "low quality protein" and "unestablished gene function") and the *–exclude_partial*; *-a 0.5* and *–s 0.75* parameters. The annotation of previously annotated genes in an already existent *D. buzzatii* reference genome [88]was included by aligning those genes and their exons to our reference genomes using blast v36.0 [88] and keeping the positions of the best hit not overlapping any other annotated gene. Chromosome Y and piwi pathway genes were re-annotated by doing a reciprocal tblast v2.15.0 [88] of the *D. melanogaster* and/or *D. mojavensis* proteins, downloaded from UniProt [116], against the putative gene sequence in our reference genomes. A total of 5 genes in the Y chromosome and 32 genes involved in the Piwi Pathway were annotated. Finally, each contig/scaffold was associated to its chromosome by using the chromosomal location of each gene in the existent *D. buzzatii* reference genome [88] and by performing reciprocal alignments of our contigs/scaffolds with the contigs/scaffolds of the existent *D. buzzatii* reference genome [88] using pblat v2.5.1 [117].

We checked the presence of the 5 Y-chromosome genes (*kl-2*, *kl-3*, *kl-5*, *Ppr-Y*, and *ORY*) by PCR in both sexes for parental species (Additional file 1: Fig. S11).

In addition, we used *PRY* primers to confirm its migration to the autosomes in our species. We used 10–20 ng of DNA from 5 adult flies, 1X Reaction Buffer Reagent (IBIANLab), 10 mM of each dNTP (Roche), and 1.25 Taq (Ibian®-Taq DNA polymerase) units. The *rp49* gene was used as a PCR positive control using primers reported in previous studies [12].

Primers amplifying both parental species (*D. buzzatii* and *D. koepferae)* were designed: Fan-F (5′–GTTGATCGTGGAACC AGCC–3′) and Fan-R (5′–CGGGCTTAACCAGACCTA CA–3′); kl-2F (5′–ACAGACAAACACAGAACAATGC–3′) and kl-2R (5′–CTTTTACCGTTTCCGTTTTCCC–3′); kl-3F (5′–CTGACTTTGGACTTGCGGAAA–3′) and kl-3R (5′-TGTATCTGCTTGCCTTCAACT–3′); kl-5F (5′-TGG TGGCTCAGGAAAACAATC-3′) and kl-5R (5′-GCAACC TCGGAATCAGTCATTA-3′); ORY-F (5′-TTGAGGCTT CACGTATCAAGC-3′) and ORY-R (5′-TTCTTGCACTAG TTGCTCTC-3′); PpRY-F (5′-TGGCGAGACGATGCTGAT -3′) and PpRY-R (5′-AGTGTAATCCGAGCTTCATCA-3′); PRY-F (5′-ACGATTGATTCCGAAACGCCA-3′) and PRY-R (5′-ACAATACTGGGACTCACCCGC-3′).

The TE annotation used in this work was obtained from Oliveira et al. [52]. Briefly, the de novo TE annotation for each genome was analysed with EarlGrey v2.2 to predict consensus using a novel method of consensus extension [118]. Posteriorly, the consensus were submitted to CDS filtering to remove the ones with high similarity with CDSs from *D. mojavensis*. Subsequently, the consensus were classified with a custom pipeline (https://github.com/OliveiraDS-hub/Pipelines-Cactophilic-Drosophila-Species), based on similarity with previously published consensus in the Dfam database v.3.7 [119]. The annotation of TE copies was performed with *RepeatMasker* v.4.1.4 [120], and consensus with less than 3 copies with at least 80% of the consensus length were removed. In addition, the authors also removed insertions shorter than 80 nucleotides, and the ones composed by simple sequence repeats in over 50% of their length [52].

mRNA extraction, library preparation, and sequencing

Total RNA was purified using the Nucleospin RNA purification kit (Macherey–Nagel) from 20–25 pairs of testes from 6 to 8-day-old males to obtain sufficient material for library preparation. Samples were sent to Macrogen for library preparation and sequencing. Duplicate TruSeq Stranded mRNA libraries, corresponding to two biological replicates per parental species, hybrids, and backcrosses. Finally, sequencing was performed for 151bp paired-end reads using Illumina technology. We obtained 66–86 million paired-end reads for each sample, resulting in a total of 1110 million paired-end reads. The transcriptomic profiles reflect population-level averages and the potential underlying genotypic heterogeneity, which is a common experimental strategy in studies requiring sufficient tissue input for high-throughput sequencing. Transcriptomic analyses of this pooled sample of backcrossed individuals, each carrying different introgressed fragments, may mask subtle expression differences due to averaging effects, but the differences reported here are likely to represent robust and consistent changes across most individuals in the pool.

Gene and TE differential expression analyses

RNA-seq sequenced reads were trimmed using Trimmomatic software v0.39 [121], with the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. For gene expression analysis, parental trimmed RNA-Seq reads were aligned to their corresponding Bu28 and Ko2 masked (using RepeatMasker [120]) reference genomes and the de novo transcript annotations using STAR v2.7.11a [122] with default parameters. Hybrid trimmed RNA-Seq reads were aligned to a concatenation of both parental masked reference genomes and concatenated transcript annotations using STAR v2.7.11a [122], only permitting 2 multiple alignments per read (*–outFilterMultimapNmax 2*). Strand-specific read count per gene was then computed using featureCounts v2.0.6 [123] (-s 2). In hybrids, only the primary alignment of the multi-mapped reads (*–primary-M*) was counted, and therefore reads aligning to both parental species were counted only once.

For TE copy expression analysis, parental RNA-Seq reads were aligned against their reference genomes using bowtie2 v2.5.2 [124] with options to perform a sensitive local alignment search (*–very-sensitive-local*) with up to 100 alignments reported for each fragment pair (*-k 100*) and a minimum alignment score threshold that permits fragments with approximately 95% or greater sequence identity to be reported (*–score-min L,0,1.6*), parameters previously used in a similar study [125]. Then, telescope v1.0.3 [125] was used with up to 200 iterations of the maximization algorithm (*–max_iter 200*) to count reads per TE copy.

Count Tables, corresponding to genes and TEs were concatenated and used for the differential expression analyses. In this way, gene counts were used to normalize TE counts. *DESeq2* function from the R Bioconductor package DESeq2 v1.42.0 [126] was used to normalize read counts, using the median of ratios method, and read counts were modelled using a negative binomial distribution. The same package was then used to identify DEG and TE copies in hybrids and back-crosses *vs* parental species performing a Wald test [126]. *p*-values were obtained using the *results* function and were adjusted for multiple testing, using the procedure of Benjamin and

Hochberg [127] with a false discovery rate *p*-value cutoff of 0.05. log2FC was shrunken using the default and recommended *apeglm* [128] algorithm of the lfcShrink function. Genes and TE copies with an adjusted *p*-value < 0.05 and at least double difference of expression (|shrunken log2FC| > 1) were considered as differentially expressed.

GO term enrichment analyses of biological processes were performed for the significant DEG using the topGO R package v2.54.0 [129] ("weight01" algorithm and Fisher's statistic). GO belonging to each gene were obtained using biomart package v2.58.0 [130]. Because of the high impact of the *p*-value adjustment, only the top 25% significant gene ontologies having the lowest *p*-values were considered as enriched. To simplify the analysis, we manually grouped the enriched GOs by the most general (maximum ancestor) GO of biological process. To analyse TE family expression, we computed the mean log2FC of all expressed TE copies of a TE family, and we considered as significantly differentially expressed the TE families with a mean absolute log2FC of at least 0.5. We also studied TE insertions inside genes and 5 Kb upstream or downstream the genes by comparing gene and TE annotated copies using bedtools v2.31.1 *intersect* command [131]. We then quantified the number of genes exhibiting differential and conserved expression between the parental species. Subsequently, we examined genes with insertions present in one species but absent in the other, genes lacking insertions or containing insertions in both species, and those in which the insertions differ between both species.

Allele-specific expression analysis

To study allele-specific expression, trimmed RNA-Seq hybrid and backcross reads were aligned to a concatenation of the polished versions of the reference genomes and transcriptomes of the parental species using STAR v2.7.11a [122], only permitting 1 alignment per read (*–outFilterMultimapNmax 1*). Then, samtools [109] was used to filter reads that aligned against the reference genome without mismatches (*[NM]==0*) in both paired-end reads. Read count per gene was computed using featureCounts v2.0.6 [123]. To make results comparable, the same pipeline was also performed for the parental species, but aligning against their own reference genome.

Then, the *DESeq2* function was used to normalize read counts using the default method. The *results* function was used to compute adjusted *p*-values using the Benjamin and Hochberg [127] method and *apeglm* [128] algorithm was used to obtain shrunken log2FC. First, an FDR cutoff of 0.05 was used to identify DEG between parental species (P). Second, the same procedure and cutoff was used to identify genes with a different abundance of the parental and the maternal allele in hybrids (H) and were considered as genes with *cis*-regulatory divergence. Finally, significant genes in either P or H were analysed for *trans*-regulatory effects (T) by comparing the P and H ratios following the same process. Genes were then categorized in the following groups as reported in a previous publication [34].

Cis-: Significant differential expression in P and H, but no significant T.

Trans-: Significant differential expression in P, and T, but not H.

Cis+ trans: Significant differential expression in P, H, and T, when *cis-* and *trans*-regulatory differences favor the expression of the same allele.

Cis X trans: Significant differential expression in P, H, and T, when *cis-* and *trans*-regulatory differences favor expression of opposite alleles.

Compensatory: Significant differential expression in H and T, but not P. *Cis-* and *trans*-regulatory differences compensate each other, resulting in no expression differences between parental species.

Conserved: No significant differential expression in H, P, or T. Conserved regulation.

Ambiguous: Other patterns have no clear biological interpretation.

Statistical tests and visualization

Three main statistical tests were used in this article, and all were performed using R v4.3.2 [132]. The Two-proportion Z-test was used to compare the distributions of significantly differentially expressed genes and TE copies across comparisons and to compare *cis-* and *trans*-regulatory categories. The chi-square test under equal assumptions was used to detect chromosome biases. Fisher's exact test under independence assumptions was computed using a 2×2 contingency table to detect associations in regulatory divergence categories and differentially expressed genes and in genes with different TE insertions between parental species and differential expression. Finally, a paired *t*-test corrected for inequality of variances was used to compare fly size of different samples.

All the results were corrected for multiple testing using the Benjamin and Hochberg [127] method. The plots and visualization of the results were performed using the R package ggplot2 v3.4.4 [133]. KEGG pathways data integration and visualization were performed using the R package *pathview* v.1.44.0 [134].

Abbreviations

TE Transposable element
BC Backcross

DEG Differential expressed genes
RC Rolling circle
MSL Male-specific lethal
UTR Untranslated region

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-026-02566-y.

Additional file 1. The file includes 11 figures in ppt format. Fig. S1. Differential thorax pigmentation between *D. buzzatii* and *D. koepferae*. Fig. S2. Male gonad phenotype of parental species and hybrids. Fig. S3. Deregulated KEGG pathways in F1 hybrids *vs* both parental species. Fig. S4. mTOR KEGG pathway deregulation for F1 hybrids *vs* both parental species. Fig. S5. Heatmap of F1 differentially expressed genes related to an increase in body size from FlyBase *vs* both parental species. Fig. S6. Gene expression of genes from FlyBase related to spermatogenesis and sperm motility. Fig. S7. Heatmap representing gene expression of Y chromosome genes from *D. melanogaster* that migrated to the autosomes in *D. buzzatii* and *D. koepferae* species. Fig. S8. TE content in the reference genome of *D. buzzatii* and *D. koepferae*. Fig. S9. Analysis of the most differentially expressed TE copies *vs D. buzzatii*. Fig. S10. Analysis of the most differentially expressed TE copies *vs D. koepferae*. Fig. S11. Verification of gene positioning on the Y chromosome of *D. buzzatii* and *D. koepferae* species by electrophoresis.

Additional file 2. The file includes 23 tables in xls format. Table S1. Body size measures in hybrids and parental species. Table S2. Body size comparisons between samples. Table S3. Differentially expressed genes between *D. buzzatii* and *D. koepferae* in testes. Table S4. Enriched Gene Ontologies in the differentially expressed genes between parental species. Table S5. Differentially expressed genes in testes of F1 hybrids and those resulting from backcrosses. Table S6. Two Proportion Z-Test for differentially expressed genes between samples. Table S7. Categories of differentially expressed genes in testes from hybrids and backcrosses. Table S8. Enriched Gene Ontologiesin deregulated genes. Table S9. Number of F1 differentially expressed genes for every KEGG pathway. Table S10. Enriched Gene Ontologies in the differentially expressed genes shared by generation. Table S11. Enriched Gene Ontologies in the deregulated genes shared by generation. Table S12. Description of deregulated genes with an ortholog in *D. melanogaster*. Table S13. Chi-square of differentially expressed genes by chromosome. Table S14. Percentages of differentially expressed genes involved in reproduction and sterility. Table S15. Percentages of deregulated genes involved in reproduction and sterility. Table S16. Deregulated genes involved in reproduction and sterility. Table S17. Regulatory divergence between parental species. Table S18. Regulatory divergence in differentially expressed genes between parental species. Table S19. Regulatory divergence in differentially expressed genes in hybrids *vs* parental species. Table S20. Differentially expressed TEs in testes of F1 hybrids and backcrosses. Table S21. Two Proportion Z-Test for differentially expressed TEs in hybrids and backcrosses. Table S22. Superfamily of the differentially expressed TE copies. Table S23. Differentially expressed genes in *D. buzzatii vs D. koepferae* with TE insertions.

Acknowledgements
The authors wish to thank José Salvador Vicente and Francisco Marco for welcoming VC into their laboratory and for their valuable advice on sperm visualization. They also thank Javier Gutierrez García for his advice with KEGG pathways and ggplot2, and Léa Jouan for her contribution to the testis dissection.

Authors' contributions
MPGG: designed the research; MPGG, AB, VC: performed the research, the sequence and statistical analysis; DSO: performed the species genome sequencing analysis and TE annotation. MPGG, AB and VC wrote the article. CV contributed to experiment designing and the manuscript edition. All authors read, corrected, and approved the final manuscript.

Funding
This work was supported by Ministerio de Ciencia e Innovación (Spain) (grant numbers PID2024-162000NB-I00 and PID2021-127107NB-I00); Generalitat de Catalunya (Spain) (grant number 2021 SGR 00526), and by the Fondation pour la Récherche Médicale. D. S. O was supported by the Eiffel Fellowship Campus France (P769649C) and CAPES–PrInt [88887.716810/2022–00].

Data availability
The data supporting the findings of this study are available within the article and in its online Additional files. Raw illumina RNA sequencing data and associated metadata generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA1297377, with SRA accession numbers from SRR34730256 to SRR34730267 (Bodelón A. Male Sterility in Drosophila hybrids: A Multi-Generational Transcriptomic Analysis of Genes and Transposable Elements in Testes). NCBI Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1297377 (2026). The *D. buzzatii* (Bu28) and *D. koepferae* (Ko2) reference genomes used in this study are available from Oliveira, D. Genomes of cactophilic Drosophila species and their respective gene and transposable element annotations, including quality control data. Zenodo. https://doi.org/10.5281/zenodo.16501018 (2025).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Departament de Genètica i Microbiologia (Edifici C), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR5558, Université Claude Bernard Lyon1, Villeurbanne, France. [3]Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (Unesp), São José do Rio Preto, São Paulo 15054-000, Brazil.

Received: 25 August 2025 Accepted: 19 February 2026
Published online: 28 February 2026

References

1. Rego C, Matos M, Santos M. Symmetry breaking in interspecific *Drosophila* hybrids is not due to developmental noise. Evolution. 2006;60:746.
2. Go AC, Civetta A. Hybrid incompatibilities and transgressive gene expression between two closely related subspecies of *Drosophila*. Front Genet. 2020;11:599292.
3. Michalak P, Noor MAF. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. Mol Biol Evol. 2003;20:1070–6.
4. Moehring AJ, Teeter KC, Noor M. Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol. 2007;24:137–45.
5. Presgraves DC, Meiklejohn CD. Hybrid sterility, genetic conflict and complex speciation: lessons from the *Drosophila simulans* clade species. Front Genet. 2021;12:669045.
6. Naveira H, Fontdevila A. The evolutionary history of *Drosophila buzzatii*. XXI. Cumulative action of multiple sterility factors on spermatogenesis in hybrids of *D. buzzatii* and *D. koepferae*. Heredity. 1991;67:57–72.
7. Ranz JM, Namgyal K, Gibson G, Hartl DL. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res. 2004;14:373–9.
8. Bodelón A, Fablet M, Veber P, Vieira C, García Guerreiro MP. High stability of the epigenome in *Drosophila* interspecific hybrids. Genome Biol Evol. 2022;14(2):evac024.

9. Kitano J, Okude G. Causative genes of intrinsic hybrid incompatibility in animals and plants: what we have learned about speciation from the molecular perspective. Evol J Linn Soc. 2024. https://doi.org/10.1093/evolinnean/kzae022.
10. Gámez-Visairas V, Romero-Soriano V, Martí-Carreras J, Segarra-Carrillo E, García Guerreiro MP. Drosophila interspecific hybridization causes a deregulation of the piRNA pathway genes. Genes. 2020;11:1–15.
11. Vela D, Fontdevila A, Vieira C, García Guerreiro MP. A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. PLoS One. 2014;9:e88992.
12. Romero-Soriano V, Garcia Guerreiro MP. Expression of the retrotransposon Helena reveals a complex pattern of TE deregulation in *Drosophila* hybrids. PLoS One. 2016;11:e0147903.
13. Kelleher ES, Edelman NB, Barbash DA. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 2012;10(11):e1001428.
14. Lopez-Maestre H, Carnelossi EAG, Lacroix V, Burlet N, Mugat B, Chambeyron S, et al. Identification of misexpressed genetic elements in hybrids between Drosophila-related species. Sci Rep. 2017. https://doi.org/10.1038/srep40618.
15. Coyne JA, Orr HA, Otte D, Endler JA. Two rules of speciation. In: Otte D, Endler JA, editors. Speciation and its consequences. Sunderland, MA: Sinauer Associates; 1989. p. 180–207.
16. Dobzhansky T. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudobscura hybrids. Genetics. 1936;21(2):113–35.
17. Morán T, Fontdevila A. Genome-wide dissection of hybrid sterility in *Drosophila* confirms a polygenic threshold architecture. J Hered. 2014;105:381–96.
18. Labrador M, Utzet F, Fontdevila A. Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol. 1999;16(7):931–7.
19. Romero-Soriano V, Burlet N, Vela D, Fontdevila A, Vieira C, García Guerreiro MP. *Drosophila* females undergo genome expansion after interspecific hybridization. Genome Biol Evol. 2016;8:556–61.
20. Carvajal AR. A three-locus system of interspecific incompatibility underlies male inviability in hybrids between *Drosophila buzzatii* and *D. koepferae*. Genetica. 1996;98:1–19.
21. Romero-Soriano V, Modolo L, Lopez-Maestre H, Mugat B, Pessia E, Chambeyron S, et al. Transposable element misregulation is linked to the divergence between parental piRNA pathways in *Drosophila* hybrids. Genome Biol Evol. 2017;9:1450–70.
22. Ungerer MC, Strakosh SC, Stimpson KM. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biol. 2009;7:40.
23. Ungerer MC, Strakosh SC, Zhen Y. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol. 2006;16(20):R872-3.
24. Nosil P, Schluter D. The genes underlying the process of speciation. Trends Ecol Evol. 2011;26(4):160–7.
25. Cabot EL, Davis AW, Johnson NA, Wu CI. Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics. 1994;137(1):175–89.
26. Brill E, Kang L, Michalak K, Michalak P, Price DK. Hybrid sterility and evolution in Hawaiian *Drosophila*: differential gene and allele-specific expression analysis of backcross males. Heredity. 2016;117:100–8.
27. Majane AC, Cridland JM, Blair LK, Begun DJ. Evolution and genetics of accessory gland transcriptome divergence between *Drosophila melanogaster* and *D. simulans*. Genetics. 2024;227(2):iyae039.
28. Runemark A, Moore EC, Larson EL. Hybridization and gene expression: beyond differentially expressed genes. Mol Ecol. 2025;34(15):e17303.
29. Stefanini MI, Milla Carmona P, Iglesias PP, Soto EM, Soto IM. Differential rates of male genital evolution in sibling species of *Drosophila*. Evol Biol. 2018. https://doi.org/10.1007/s11692-018-9444-0.
30. Richmond MP. The role of aedeagus size and shape in failed mating interactions among recently diverged taxa in the *Drosophila mojavensis* species cluster. BMC Evol Biol. 2014;14:255.
31. Piergentili R. Multiple roles of the Y chromosome in the biology of *Drosophila melanogaster*. Sci World J. 2010;10:1749–67.
32. Brennecke J, Aravin A, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell. 2007;128(6):1089–103.
33. Parhad SS, Tu S, Weng Z, Theurkauf WE. Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. Dev Cell. 2017;43:60-70.e5.
34. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 2010;20:816–25.
35. Soto EM, Soto IM, Carreira VP, Fanara JJ, Hasson E. Host-related life history traits in interspecific hybrids of cactophilic Drosophila. Entomol Exp Appl. 2008;126:18–27.
36. Shrivastava NK, Yadav A, Rauhila N, Farand AK, Khati M, Shakarad MN. The molecular underpinnings of body size regulation: transcriptional changes in foxo and mTor in Drosophila melanogaster selected for fast development. Gene Rep. 2023;33:101841.
37. Li Y, Li S, Jin P, Chen L, Ma F. miR-11 regulates pupal size of Drosophila melanogaster via directly targeting Ras85D. Am J Physiol Cell Physiol. 2017;312(1):C71-82.
38. Karim FD, Rubin GM. Ectopic expression of activated Ras1 induces hyperplastic growth and increased cell death in Drosophila imaginal tissues. Development. 1998;125:1–9.
39. Lin JI, Mitchell NC, Kalcina M, Tchoubrieva E, Stewart MJ, Marygold SJ, et al. *Drosophila* ribosomal protein mutants control tissue growth non-autonomously via effects on the prothoracic gland and ecdysone. PLoS Genet. 2011;7:e1002408.
40. Towler BP, Pashler AL, Haime HJ, Przybyl KM, Viegas SC, Matos RG, et al. Dis3L2 regulates cell proliferation and tissue growth through a conserved mechanism. PLoS Genet. 2020;16:e1009297.
41. Imura E, Shimada-Niwa Y, Nishimura T, Hückesfeld S, Schlegel P, Ohhara Y, et al. The Corazonin-PTTH neuronal axis controls systemic body growth by regulating basal ecdysteroid biosynthesis in Drosophila melanogaster. Curr Biol. 2020;30(11):2156–65.
42. Gouzi JY, Moressis A, Walker JA, Apostolopoulou AA, Palmer RH, Bernards A, et al. The receptor tyrosine kinase alk controls neurofibromin functions in *Drosophila* growth and learning. PLoS Genet. 2011;7:e1002281.
43. Ferguson J, Gomes S, Civetta A. Rapid male-specific regulatory divergence and down regulation of spermatogenesis genes in *Drosophila* species hybrids. PLoS ONE. 2013;8:e61575.
44. Davis H, Sosulski N, Civetta A. Reproductive isolation caused by azoospermia in sterile male hybrids of Drosophila. Ecol Evol. 2020;1010(12):5922–31.
45. Reed LK, Markow TA. Early events in speciation: polymorphism for hybrid male sterility in Drosophila. Proc Natl Acad Sci U S A. 2004;101(24):9009–12.
46. Satokangas I, Martin SH, Helanterä H, Saramäki J, Kulmuni J. Multilocus interactions and the build-up of reproductive isolation: multilocus interactions in speciation. Philos Trans R Soc Lond B Biol Sci. 2020;375(1806):20190543.
47. Presgraves DC. A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. Genetics. 2003;163(3):955–72.
48. Bladen J, Cooper JC, Ridges JT, Guo P, Phadnis N. A new hybrid incompatibility locus between Drosophila melanogaster and Drosophila sechellia. Genetics. 2024;226(3):iyae001.
49. Wang RJ, Ané C, Payseur BA. The evolution of hybrid incompatibilities along a phylogeny. Evolution. 2013;67(10):2905–22.
50. Meiklejohn CD, Presgraves DC. Little evidence for demasculinization of the Drosophila X chromosome among genes expressed in the male germline. Genome Biol Evol. 2012;4(10):1007–16.
51. Mack KL, Nachman MW. Gene regulation and speciation. Trends Genet. 2017;33:68–80.
52. Oliveira DS, Larue A, Nunes WVB, Sabot F, Bodelón A, García Guerreiro MP, et al. Transposable elements contribute to the evolution of host shift-related genes in cactophilic Drosophila species. bioRxiv. 2025. https://doi.org/10.1101/2024.03.27.58702.
53. De Panis D, Dopazo H, Bongcam-Rudloff E, Conesa A, Hasson E. Transcriptional responses are oriented towards different components of the rearing environment in two Drosophila sibling species. BMC Genomics. 2022;23:515.
54. Marin I, Ruiz A, Pla C, Fontdevila A. Reproductive relationships among ten species of the Drosophila repleta group from South America and the West Indies. Evolution. 1993;47:1616–24.

55. Yamagata Y, Yamamoto E, Aya K, Win KT, Doi K, Sobrizal, et al. Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. Proc Natl Acad Sci U S A. 2010;107(4):1494–99.
56. Sundararajan V, Civetta A. Male sex interspecies divergence and down regulation of expression of spermatogenesis genes in drosophila sterile hybrids. J Mol Evol. 2011;72:80–9.
57. Li R, Ren X, Bi Y, Ho VWS, Hsieh CL, Young A, et al. Specific down-regulation of spermatogenesis genes targeted by 22G RNAs in hybrid sterile males associated with an X-chromosome introgression. Genome Res. 2016; 2016;26(9):219–32.
58. Chan WY, Chung J, Peplow LM, Hoffmann AA, van Oppen MJH. Maternal effects in gene expression of interspecific coral hybrids. Mol Ecol. 2021;30:517–27.
59. Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. Strong maternal effects on gene expression in Arabidopsis lyrata hybrids. Mol Biol Evol. 2016;33:984–94.
60. Johnson NA, Hollocher H, Noonburg E, Wu CI. The effects of interspecific Y chromosome replacements on hybrid sterility within the *Drosophila simulans* clade. Genetics. 1993;135(2):443–53.
61. Poikela N, Laetsch DR, Kankare M, Hoikkala A, Lohse K. Experimental introgression in *Drosophila*: asymmetric postzygotic isolation associated with chromosomal inversions and an incompatibility locus on the X chromosome. Mol Ecol. 2023;32(4):854–66.
62. Sackton TB, Montenegro H, Hartl DL, Lemos B. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in *Drosophila*. Proc Natl Acad Sci U S A. 2011;108(41):17046–51.
63. Millard SS, Flanagan JJ, Pappu KS, Wu W, Zipursky SL. Dscam2 mediates axonal tiling in the *Drosophila* visual system. Nature. 2007;447:720–4.
64. Olson PF, Fessler LI, Nelson RE, Sterne RE, Campbell AG, Fessler JH. Glutactin, a novel *Drosophila* basement membrane-related glycoprotein with sequence similarity to serine esterases. EMBO J. 1990;9(4):1219–27.
65. Ahn SJ, Marygold SJ. The UDP-glycosyltransferase family in *Drosophila melanogaster*: nomenclature update, gene expression and phylogenetic analysis. Front Physiol. 2021;12:648481.
66. Ma CIJ, Yang Y, Kim T, Chen CH, Polevoy G, Vissa M, et al. An early endosome-derived retrograde trafficking pathway promotes secretory granule maturation. J Cell Biol. 2020;219(3):e201808017.
67. Öztürk-Çolak A, Marygold SJ, Antonazzo G, Attrill H, Goutte-Gattat D, Jenkins VK, et al. FlyBase: updates to the *Drosophila* genes and genomes database. Genetics. 2024;227:iyad211.
68. Davis BW, Seabury CM, Brashear WA, Li G, Roelke-Parker M, Murphy WJ. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. Mol Biol Evol. 2015;32:2534–46.
69. Llopart A. The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. Mol Biol Evol. 2012;29(12):3873–86.
70. Rafati N, Blanco-Aguiar JA, Rubin CJ, Sayyab S, Sabatino SJ, Afonso S, et al. A genomic map of clinal variation across the European rabbit hybrid zone. Mol Ecol. 2018;27:1457–78.
71. Wakimoto BT, Lindsley DL, Herrera C. Toward a comprehensive genetic analysis of male fertility in *Drosophila melanogaster*. Genetics. 2004;167(1):207–16.
72. Ma Z, Liu Z, Huang X. OSBP- and FAN-mediated sterol requirement for spermatogenesis in *Drosophila*. Development. 2010;137:3755–84.
73. Urness LD, Thummel CS. Molecular interactions within the ecdysone regulatory hierarchy: DNA binding properties of the *Drosophila* ecdysone-inducible E74A protein. Cell. 1990;63(1):47–61.
74. Žurovcová M, Ayala FJ. Polymorphism patterns in two tightly linked developmental genes, Idgf1 and Idgf3, of *Drosophila melanogaster*. Genetics. 2002;162:177–88.
75. Rideout EJ, Dornan AJ, Neville MC, Eadie S, Goodwin SF. Control of sexual differentiation and behavior by the doublesex gene in *Drosophila melanogaster*. Nat Neurosci. 2010;13(4):458–66.
76. Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol. 2011;9(8):e1001126.
77. Gepner J, Hays TS. A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. Proc Natl Acad Sci U S A. 1993;90(23):11132–6.
78. Carvalho AB, Lazzaro BP, Clark AG. Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. Proc Natl Acad Sci U S A. 2000;97(24):13236–44.
79. Lemos B, Araripe LO, Hartl DL. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science. 2008;319:91–3.
80. Zur Lage P, Stefanopoulou P, Styczynska-Soczka K, Quinn N, Mali G, Von Kriegsheim A, et al. Ciliary dynein motor preassembly is regulated by Wdr92 in association with HSP90 co-chaperone, R2TP. J of Cell Biol. 2018;217(7):2583–9.
81. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics. 2005;171:1813–22.
82. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430:85–8.
83. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between *Drosophila* species. Nat Genet. 2008;40(3):346–50.
84. Emerson JJ, Li W-H. The genetic basis of evolutionary change in gene expression levels. Philos Trans R Soc Lond B Biol Sci. 2010;365(1552):2581–90.
85. Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. The roles of *cis*- and *trans*-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. Genome Res. 2014;24(1):84–95.
86. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in *Drosophila*. Genome Res. 2014;24(5):797–808.
87. Schaefke B, Emerson JJ, Wang TY, Lu MYJ, Hsieh LC, Li WH. Inheritance of gene expression level and selective constraints on *trans*-and *cis*-regulatory changes in yeast. Mol Biol Evol. 2013;30(9):2121–33.
88. Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, et al. Genomics of ecological adaptation in cactophilic Drosophila. Genome Biol Evol. 2015;7:349–66.
89. Rahman R, Chirn GW, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, et al. Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. Nucleic Acids Res. 2015;43(22):10655–72.
90. Tahami MS, Vargas-Chavez C, Poikela N, Coronado-Zamora M, González J, Kankare M. Transposable elements in *Drosophila montana* from harsh cold environments. Mob DNA. 2024;15(1):18.
91. Adrion JR, Song MJ, Schrider DR, Hahn MW, Schaack S. Genome-wide estimates of transposable element insertion and deletion rates in *Drosophila melanogaster*. Genome Biol Evol. 2017;9(5):1329–40.
92. Villanueva-Cañas JL, Horvath V, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. Nucleic Acids Res. 2019;47(13):6842–57.
93. García Guerreiro MP. Changes of Osvaldo expression patterns in germline of male hybrids between the species *Drosophila buzzatii* and *Drosophila koepferae*. Mol Genet Genomics. 2015;290:1471–83.
94. Carnelossi EAG, Lerat E, Henri H, Martinez S, Carareto CMA, Vieira C. Specific activation of an I-like element in *Drosophila* interspecific hybrids. Genome Biol Evol. 2014;6:1806–17.
95. Kawakami T, Dhakal P, Katterhenry AN, Heatherington C, Ungerer MC. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol Evol. 2011;3:156–67.
96. Liu L, Qi H, Wang J, Lin H. PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. Development. 2011;138(9):1863–73.
97. Czaja W, Bensasson D, Ahn HW, Garfinkel DJ, Bergman CM. Evolution of Ty1 copy number control in yeast by horizontal transfer and recombination. PLoS Genet. 2020;16(2):e1008632.
98. Hénault M. The challenges of predicting transposable element activity in hybrids. Curr Genet. 2021;67:567–72.
99. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42.

100. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199.
101. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86.
102. Oliveira DS, Fablet M, Larue A, Vallier A, Carareto CMA, Rebollo R, et al. ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. Nucleic Acids Res. 2023;51(18):9764–84.
103. Coronado-Zamora M, González J. Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across Drosophila natural strains. Genome Res. 2023;33(9):1541–53.
104. Serrato-Capuchina A, Matute DR. The role of transposable elements in speciation. Genes. 2018;9(5):254.
105. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5.
106. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6.
107. Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
108. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
109. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
110. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007;23:1026–8.
111. Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck F, Lippman Z, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20(1):224.
112. Simão F, Waterhouse R, Ioannidis P, Zdobnov E. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
113. Moreyra N, Almeida F, Allan C, Frankel N, Matzkin L, Hasson E. Phylogenomics provides insights into the evolution of cactophily and host plant shifts in Drosophila. Mol Phylogenet Evol. 2022;178:107653.
114. Shumate A, Salzberg S. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37(12):1639–43.
115. Kim B, Wang J, Miller D, Barmina O, Delaney E, Thompson A, et al. Highly contiguous assemblies of 101 drosophilid genomes. Elife. 2021;10:e66405.
116. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–9.
117. Kent WJ. BLAT - the BLAST-like alignment tool. Genome Res. 2002;12:656–64.
118. Baril T, Imrie R, Hayward A. Earl grey: a fully automated user-friendly transposable element annotation and analysis pipeline. Mol Biol Evol. 2024;41(4):msae068.
119. Hubley R, Finn R, Clements J, Eddy S, Jones T, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2015;44(D1):D81–9.
120. Smit AFA, Hubley R GP. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org. Accessed 26 Nov 2020.
121. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
122. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
123. Liao Y, Smyth GK, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
124. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
125. Bendall ML, de Mulder M, Iñiguez LP, Lecanda-Sánchez A, Pérez-Losada M, Ostrowski MA, et al. Telescope: characterization of the retrotranscriptome by accurate estimation of transposable element expression. PLoS Comput Biol. 2019;15:e1006453.
126. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.
127. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125:279–84.
128. Zhu A, Ibrahim J, Love M. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2018;35(12):2084–92.
129. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version 2.42.0. 2020. https://rdrr.io/bioc/topGO.
130. Durinck S, Spellman P, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4:1184–91.
131. Quinlan AR, Hall IM. BEDTools : a flexible suite of utilities for comparing genomic features. 2010;26:841–2.
132. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.r-project.org.
133. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
134. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29:1830–1.

Publisher's Note