UAB
Universitat Autònoma de Barcelona

Impact of stress in gene and Transposable Element expression in *Drosophila*

Alejandra Bodelón de Frutos

Departament de Genètica i Microbiologia

Universitat Autònoma de Barcelona

Doctoral thesis submitted for the degree of

Doctor in Genetics (PhD)

Director and tutor: Dr. Maria del Pilar García Guerreiro

2023

Impact of stress in gene and Transposable Element expression in *Drosophila*

Dissertation respectfully submitted by

Alejandra Bodelón de Frutos

To Universitat Autònoma de Barcelona for the degree of Doctor of Philosophy (PhD),
as per the Doctorate Program in Genetics

Author   *Director*

Alejandra Bodelón de Frutos   Dr. Maria del Pilar García Guerreiro

Bellaterra, 30 June 2023

"We know about the components of genomes [Transposable Elements] that could be made available for such restructuring. We know nothing, however, about how the cell senses danger and instigates responses to it that often are truly remarkable."

Barbara McClintock

The significance of responses of the genome to challenge, *Science*, 1984

ABBREVIATIONS LIST

AFLP	Amplified Fragment Length Polymorphism
AP	Apuric/apyrimidic
bp	Base pairs
DIRS	Dictyostelium intermediate repeat sequence
DNMT	DNA Methyltransferase
DSB	Double Strand Break
dsRNA	Double-stranded RNA
EN	Endonuclease
endo-siRNA, siRNA	endogenous short interfering RNA
H3K4me3	Tri-methylation of lysine 4 on histone H3
H3K9me3	Tri-methylation of lysine 9 on histone H3
H3K27me3	Tri-methylation of lysine 27 on histone H3
Hel	Helicase
HR	Homologous Recombination
HSE	Heat Sock Element
Hsp	Heat shock protein
HSR	Heat shock response
ICR	Internal complementary region
IN	Integrase
ITR	Terminal inverted repeat
Kb	Kilobase
LINEs	Long Interspersed Nucleotide Element
LTR	Long Terminal Repeat
miRNA	microRNA
MITE	Miniature Inverted–Repeat Transposable Element

Mya	Million years ago
NAHR	Non-Allelic Homologous Recombination
NHEJ	Non-Homologous End Joining
ORF	Open reading frame
PR	Protease
Rep	Replication initiator
RNase H, RH	Ribonuclease H
RT	Reverse transcriptase
piRNA	Piwi-interacting RNA
PLE	Penelope-like element
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol III	RNA polymerase III
Pol B	DNA polymerase B
Pro	Protease
PT	Polygenic threshold
SDR	Split Direct repeat
SINE	Short Interspersed Nucleotide Element
TE	Transposable Elements
TERT	Telomerase reverse transcriptase
TIRs	Terminal inverted repeats
TPRT	Target-primed reverse transcription
TSD	Target site duplication
TSS	Transcription start site
UTR	Untranslated Region
YR	Tyrosine recombinase

Index

Abstract

Transposable Elements (TEs) are repetitive DNA sequences that can move from one location to another in the host genome though a process called transposition. New TE insertions are usually detrimental for the organism and are selected against, thus disappearing from the population. However, some TE insertions can also be beneficial by developing novel functions in the host genomes and can be positively selected, and even fixated. When these changes are produced in the germline, they can be inherited in next generations, leading to the host evolution. Nonetheless, organisms as *Drosophila* also developed mechanisms involving TE-derived piRNAs to protect themselves from the negative TE effects. piRNAs are involved in post-transcriptional TE silencing in the germline, and in pre-transcriptional silencing, by adding silencing chromatin marks to TEs.

It is known that different stresses can lead to the activation of specific TE families. For example, interspecific crosses between *Drosophila buzzatii* and *Drosophila koepferae* species produce hybrids with fertility problems and an increase of transposition rates. Recent studies observed that TE activation in these F_1 hybrid ovaries could only be partially explained by a divergent piRNA pathway gene expression and piRNA pools between parental species, suggesting that other additional mechanisms could lead to the observed hybrid TE activation. Another stress controversially related to TE activation in *Drosophila* is the heat stress, which is of great importance considering the global warming due to the climate change. This stress is also known by activating several heat-induced genes, mainly heat shock proteins, and by repressing other genes. *Drosophila subobscura* is a species native of the Palearctic region, which has recently colonized South and North America and has a rich inversion polymorphism with latitudinal clines, making it of high interest to monitor temperature increases. On the other hand, the relationship between changes in Hsp70 (O chromosome heat shock protein) levels found in previous studies in non-stressful conditions, and the O arrangements has not been conclusive, making difficult at molecular level to establish a correlation between temperature and chromosomal arrangements.

In this work, we first evaluated how differences in gonadal epigenomes between parental species *D. buzzatii* and *D. koepferae* can explain differences in F_1 hybrid expression, being the first epigenomic study in *Drosophila* hybrids. Globally, we demonstrate a trend towards underexpression of genes and TE families in hybrids in addition to high epigenomic conservation between parental species and their hybrid ovaries. However, we detected some hybrid epigenomic changes in comparison to the parental species that could explain the observed hybrid gene and TE expression. We also detected additional mechanisms, such as regulatory divergence between parental species, that could contribute to parental-hybrid gene expression differences. We

subsequently studied the effect of heat stress on the transcriptome of two different geographic origin *D. subobscura* populations, being this work the first *D. subobscura* global transcriptomic study under normal and heat stress conditions. The comparison of the genomic expression between populations, in non-stressful conditions, showed differences in gene expression likely involved in the adaptation of the populations to different environments, and a higher TE activation in the Curicó colonizer population. We also found a trend towards overexpression of genes after heat stress, mainly heat shock protein genes, and a sex-specific response to heat stress. An activation of a few specific TE families in the germline of both *D. subobscura* populations after heat stress, in addition to changes in piRNA cluster production, and in the siRNA and piRNA amounts associated to specific TE families was also detected. However, we could not directly relate these small RNA changes with the observed TE activation, suggesting additional mechanisms influencing TE expression after heat stress. Globally all our results support that TE activation and repression after genomic and thermal stress in the germline is a complex phenomenon involving an interplay between multiple factors. Therefore, further studies focused on the effect of interspecific hybridization on male hybrid epigenome and the effect of heat stress on the germline epigenome, will be needed to decipher the mechanisms leading to TE activation after stress.

1.   Introduction

In this section, I will explain the theoretical basis about the impact of two different stresses, genomic and thermal stress, on the genomic expression, including genes and Transposable Elements (TEs), and the epigenome of *Drosophila*. In the first part, I will introduce the model species used in these studies: *Drosophila buzzatii*, *D. koepferae* and *D. subobscura*. The two first ones are two cactophilic sibling species, which cohabit in a few areas of South America and whose offspring originate hybrids with fertility problems and an increase in transposition rates. *D. subobscura*, is a species native from the Palearctic region that has recently colonized America and has a high inversion polymorphism related to its adaptation to temperature. In the second part, I will focus on the Transposable Elements (TEs): classification, regulation (specifically in *Drosophila*) and their consequences on the host genome, including their putative role in adaptation. Finally, I will explain the mechanisms leading to TE activation, mainly focusing on the interspecific hybridization and the heat stress.

1.1   *Drosophila* species

Drosophila melanogaster has been used as a model organism since the early 20th century, when the pioneering work of Thomas Hunt Morgan demonstrated that the genes are stored in chromosomes, which are the main unit of heredity and have a role in sex determination [1]. Nowadays, it remains one of the most used model organisms for basic research because it is relatively easy to maintain,

Figure 1: Phylogenetic relationships among different *Drosophila* species, adapted from [3].

inexpensive, has very few restrictions on their use in the laboratory and a rapid generation time with large number of offspring. In addition, approximately 75% of the known human disease genes have an ortholog in the *D. melanogaster* genome [2]. Nevertheless, in the last years basic research have also been extensively focused on other *Drosophila* species, who share most of their advantages with *D. melanogaster* and sometimes, even additional ones. In this work, we used three species: *D. buzzatii* and *D. koepferae*, belonging to the *repleta* group, and *D. subobscura*, belonging to the *obscura* group. The genus *Drosophila* is usually divided in two subgenera: the

Drosophila subgenus, where the *repleta* group belongs, and the *Sophophora* subgenus, where the *obscura* group belongs [3] (Figure 1).

1.1.1 *Drosophila buzzatii* and *Drosophila koepferae*
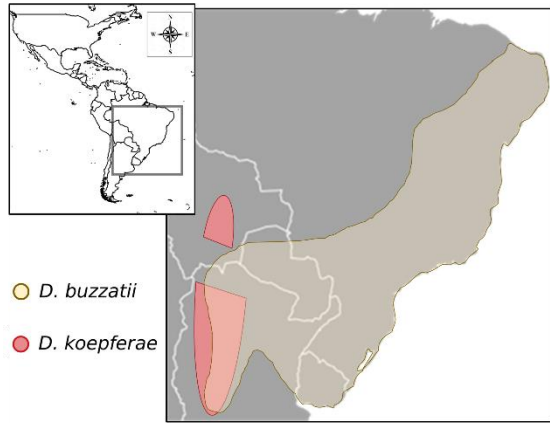
Figure 2: Distribution of *D. buzzatii* and *D. koepferae* in South America, adapted from [8].

Drosophila buzzatii and *Drosophila koepferae* are two sibling cactophilic species included in the *repleta* group of the subgenus *Drosophila* [4, 5]. This group is considered one of the most important and successful radiations in the genus *Drosophila* [5], containing approximately 100 New World endemic species, many of them found in the arid or semiarid deserts, where they use decaying cacti as breeding substrates [6–8]. The *repleta* group is divided in five subgroups: *mulleri*, *hydei*, *mercatorum*, *repleta*, and *fasciola,* which are subdivided into complexes and clusters [5]. *D. buzzatii* and *D. koepferae* are included in the *buzzatii* cluster and complex, part of the *mulleri* subgroup [4, 5]. Both species are native to the arid lands of South America, but their distribution is quite different. *D. buzzatii* seems to be native of the arid areas of the north of Argentina [9], but is widely distributed in South America, including parts of Bolivia, Paraguay and Brazil [8] (Figure 2). Additionally, this semi-cosmopolitan species has colonized the Old World a few hundred years ago, including the Mediterranean area, Macaronesian Islands (specifically Canary Islands and Madeira), East and Western Africa [9], and Australia (period 1931–1936) [10], following man-mediated dispersion of its natural host plants [9, 10] and reaching a wide distribution. *D. koepferae* is distributed in northwestern Argentina and central Bolivia, showing an hiatus in the Argentinian Prepuna and reappearing in the Bolivian Altiplano [11] (Figure 2). Thus, these two species partially overlap in the arid lands of the northwestern Argentina and the southern Bolivia [4].

D. buzzatii and *D. koepferae* feed mostly on the decaying tissues of several Cactaceae species, including prickly pears (genus *Opuntia*) and columnar cacti (*Cereoideae* subfamily) [11], which are not uniformly distributed throughout South America [12] and have differential spatial and temporal availability: the prickly pears is more abundant and ephemeral and the columnar cacti is less abundant and lasting longer [13]. Both species have been found in the nature in both types of cacti, showing a certain niche overlap [11, 13], and they do not seem differently attracted to each host in the nature [13]. Nevertheless, these species are considered 'specialist' because in the field they show specificity for a preferred host: *D. buzzatii* breeds and feeds mainly on cacti

of the *Opuntia* genus, whereas *D. koepferae* on columnar cacti, such as *Trichocereus*, *Cereus* and *Neoraimondia* [4, 11, 12]. In addition, the abundance and distribution of these species is clearly connected to the abundance of their primary host [12], and each species increases significantly the number of laid eggs [14] and emerged individuals [13] in their preferred host *vs* the other in nature and laboratory conditions. Moreover, columnar cacti and *Opuntia* cacti decaying processes have different yeast communities associated [15] and their chemical characteristics are different: most of *D. koepferae*-used cacti have higher presence of toxic alkaloids [16], whereas *D. buzzatii*-used cacti have slightly more free sugars and total fats [7] and hence it is considered a more nutritive host [17]. The effect of the toxic alkaloids in *D. buzzatii* was shown to be more harmful than in *D. koepferae* [15], including morphological changes [17–20] and a decrease of viability [19] in the former, in addition to a detoxification and stress response transcriptomic modulation [21] that gives *D. buzzatii* some transcriptional plasticity in toxic environments [22]. Nevertheless, the alkaloid tolerance observed in *D. koepferae* seems to depend on other nutritional components of the host [15], and its transcriptomic response vary by the host cacti [21]. Consequently, host selection by these *Drosophila* species seems to be related to their adaptations to exploit resources with different ecological (spatial and temporal predictability) and compositional (chemical and nutritional) properties [7]. However, it could also be related to interspecific competition in overlapping niches, which affected the performance of both species in the hosts [23].

The divergence time of *D. buzzatii* and *D. koepferae* has been estimated to be approximately 4-5 million years ago (Mya) in some studies using a few specific *loci* [6] or a *de novo* transcriptome [24]. Nevertheless, in recent studies using a large number of *loci* [25] or a complete mitogenome [8], the divergence time of these species was estimated to be 1.54 Mya and 1.84 Mya, respectively. The position of *D. koepferae* in the cluster has also been controversial: it has been selected to be closely related to *D. buzzatii* based on general external morphology [4], chromosomal inversions [26] and some molecular data-based phylogenetic analysis [6, 8]. However, it has also been described to be more closely related to other species of the *buzzatii* cluster based on male genitalia [4] and other molecular data-based phylogenetic analysis [7, 25]. Moreover, hybrids between different species of the *buzzatii* cluster can be produced, obtaining partially or completely sterile F_1 offspring [27]. In addition, introgressive hybridization between few species has also been proposed [25], reflecting the complexity of the *buzzatii* cluster.

In this thesis, we focused on the interspecific crosses between *D. buzzatii* and *D. koepferae*, explained in more detail in section **1.2.4.1**. Crosses between *D. koepferae* females and *D. buzzatii* males produce F_1 sterile hybrid males and partially fertile hybrid females [28], the other direction of the cross does not produce offspring [29]. F_1 hybrid females can be backcrossed with either parental species to increase the introgressed segments from that parental species [30]. Both

species have similar karyotypes, consisting of five pairs of autosomes, four pairs of equal-length acrocentric chromosomes and one pair of dot, and a pair of sexual chromosomes, where the X chromosome is acrocentric [4].

1.1.2 *Drosophila subobscura*

D. subobscura is a species of the *obscura* group of the subgenus *Sophophora* [31]. This species, together with the island-endemic *D. madeirensis* from Madeira, and *D. guanche,* from the Canary Islands, constitute the *subobscura* cluster [32]. *D. subobscura* is native to the Palearctic region [32] (Figure 3). This species was first encountered in Europe, where is extended all over the continent except for Iceland and the northern parts of Scandinavia [32, 33]. The eastern European frontier is mostly unknown, but this species has
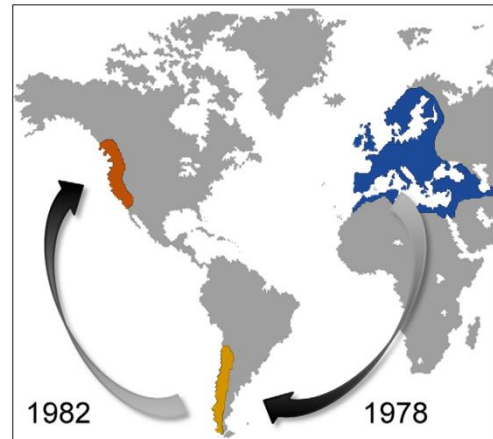
Figure 3: Distribution and colonization of *D. subobscura*, adapted from [38].

also been detected in some regions of Western Asia and North Africa [32, 34]. This species was also found in many Macaronesian Islands, such as Madeira, the Azores, and Canary Island [32]. In 1978, this species was detected for the first time in Puerto Montt (Chile) and spread rapidly in that county [34, 35] (Figure 3). *D. subobscura* was also found in 1981 in Argentina at the east of the Andes and in an isolated collection in Mar del Plata, in the Atlantic Coast [36]. In 1982, this species was found in the Pacific Northwest, where it was collected in the Cascades Range from southern Oregon to southern British Columbia [32, 37] (Figure 3). Therefore, this species is encountered from sea level to high elevation, and it is mainly found in the forest and on the edge of the forest, and hence this species can leave the forest [32, 33]. The abundance of this species differs between different geographical areas, indicating an ecological role of this species as central (multiple habitats) or marginal (rare habitats), and distribution limits could possibly be extended because of climate change [32].

The ecology of *D. subobscura* is still poorly known. Even though fruits seem to be the preferred medium, this species is considered "generalist", and their food, such as decaying fruits, fermenting sap, decaying vegetation and fungi, probably vary geographically and seasonally [32, 39]. This species seems to have a clear separation between feeding and breeding sites, and females oviposit preferentially in fruits, of which bananas are the best attractors [32, 39]. In addition, and unlike other *Drosophila* species, *D. subobscura* only mates in presence of light and is monandrous [33, 40]. Even though the optimal temperature of this species and preferred for egg laying [33] is 18ºC, remarkable differences in sensitivity to temperature can be found: females seem to have

higher heat resistance and, sometimes, also desiccation resistance, than males [41]. *D. subobscura* can survive in a long threshold of temperatures (between 3ºC and 37ºC), but maintenance at 25ºC or higher may induce male sterility and ovary damage [32]. Temperature also influences the longevity of this species [42], even though females seem to have longer lifespan than males [32], and it also influence the developmental time: the complete life cycle of this species is about 23 days at 18ºC in optimal larvae density [42]. Temperature together with humidity also influence dispersal of the species [42].

This species has five large pairs of acrocentric chromosomes, including the X chromosome (A) and four autosomes (J, U, E and O), and one small dot chromosome, without chromocenter [32, 43]. The Y chromosome is also acrocentric and seems to be totally heterochromatic [32]. This species has a rich inversion polymorphism in all five chromosomes, with more than 60 different inversions recorded in more than 85 gene arrangements [44], most of them located on the O chromosome [32]. The chromosomal polymorphisms of *D. subobscura* vary in frequency by latitudinal clines, which are practically parallels in both the Old and New World [45]. The duplication of the chromosomal polymorphism distributional patterns strongly support their adaptative nature [35]. Different researches focused on the specific environmental factors that work as evolutionary forces by affecting the inversion frequencies and establishing parallel latitudinal clines, and showed that several inversions were related mainly to temperature [44, 46, 47], but also rainfall and dryness (humidity) [44, 47, 48]. This idea was also supported by the changes in inversion frequencies detected by seasonal changes [48] or heat waves [49]. Thus, these arrangements could be classified as 'cold' (typically high-latitude arrangements) or 'warm' (typically low-latitude arrangements) adapted inversions [44, 46]. An increase of 'warm' adapted arrangements have also been described due to the increase of temperature derived from the global warming [50, 51], and even in some cases, decreasing the arrangement variability of the populations almost fixing "warm" adapted inversions [52]. Consequently, these shifts in chromosomal arrangements in response to climate change make this species of high interest to measure and monitor global warming [51]. Nonetheless, laboratory attempts to establish the specific genomic regions affected by the inversions and their adaptative significance, are still inconclusive. Studies of the inversion breakpoints showed that most of them do not seem to disrupt any obvious candidate gene for direct adaptation to temperature [53], making their adaptive value difficult to decipher.

1.2 Transposable elements

Transposable Elements (TEs) are DNA sequences that are able to move from one location to another in the host genome. They were discovered in 1940s by Barbara McClintock in *Zea mays* [54], where they represent nearly 85% of the genome [55]. They are present in most eukaryotes, compromising around 3% to 85% of the genome according to size in plants [56], 45% of the human genome [57], and around 20% in *Drosophila melanogaster* [58], and they can also be found in most prokaryotes in a lower percentage of around 0.3% [59]. McClintock's theory proposed that the mutagenic effect of TE transposition increases host genetic variability, eventually inducing genome modifications that allow adaptation to environmental change [54], acting as motors of evolution. However, to prevent deleterious effects, hosts have evolved different strategies to avoid their activation and mobilization, such as epigenomic mechanisms [60]. Nevertheless, it has been observed that certain stresses, such as genomic or thermal stresses, can activate the TEs [61]. All these ideas will be explored in the following section.

1.2.1 TE classification

The classification used in this thesis, is based on the classification proposed by Wicker et al. [62]. TEs are divided in two main classes according to their transposition mechanisms: Class I elements or retrotransposons and Class II elements or DNA transposons [63] (explained in detail below). The classes are subdivided into subclasses and then into orders according to their replication and/or integration mechanisms. Orders are divided into superfamilies, characterized by uniform general features, such as protein and non-coding domains structure. Superfamilies are subdivided into families, defined as a group of TEs with high DNA sequence conservation. Hence, whereas superfamilies have low DNA sequence conservation and only limited protein similarities between them, families are usually similar at the protein level, but with minimal DNA sequence conservation, mainly in the coding region. Each family can have different insertions (copies) in the host genome, each one corresponding to a specific transposition and insertion event. Insertions are often defined as members of the same TE family when they are longer than 80 bp and share at least 80% sequence identity over 80% of their length, which is called the 80–80–80 rule. Each TE family can be represented by a consensus sequence, which is an approximation of the ancestral sequence of the active TE, and similarity between two consensus sequences is less than 75% [64]. In some cases, families can be subdivided in subfamilies to differentiate subpopulations of TEs belonging to large families that can be clearly separated, or to distinguish between autonomous and non-autonomous elements from the same family [62]. Therefore, TEs can also be classified according to their ability to move autonomously or not. Autonomous TEs encode all the enzymatic machinery needed for their transposition, whereas non-autonomous TEs lack some or

all the machinery. However, non-autonomous TEs are sometimes still capable of mobilization in *trans* using the machinery encoded by their autonomous analogues, from which they often emerge by deletion events.
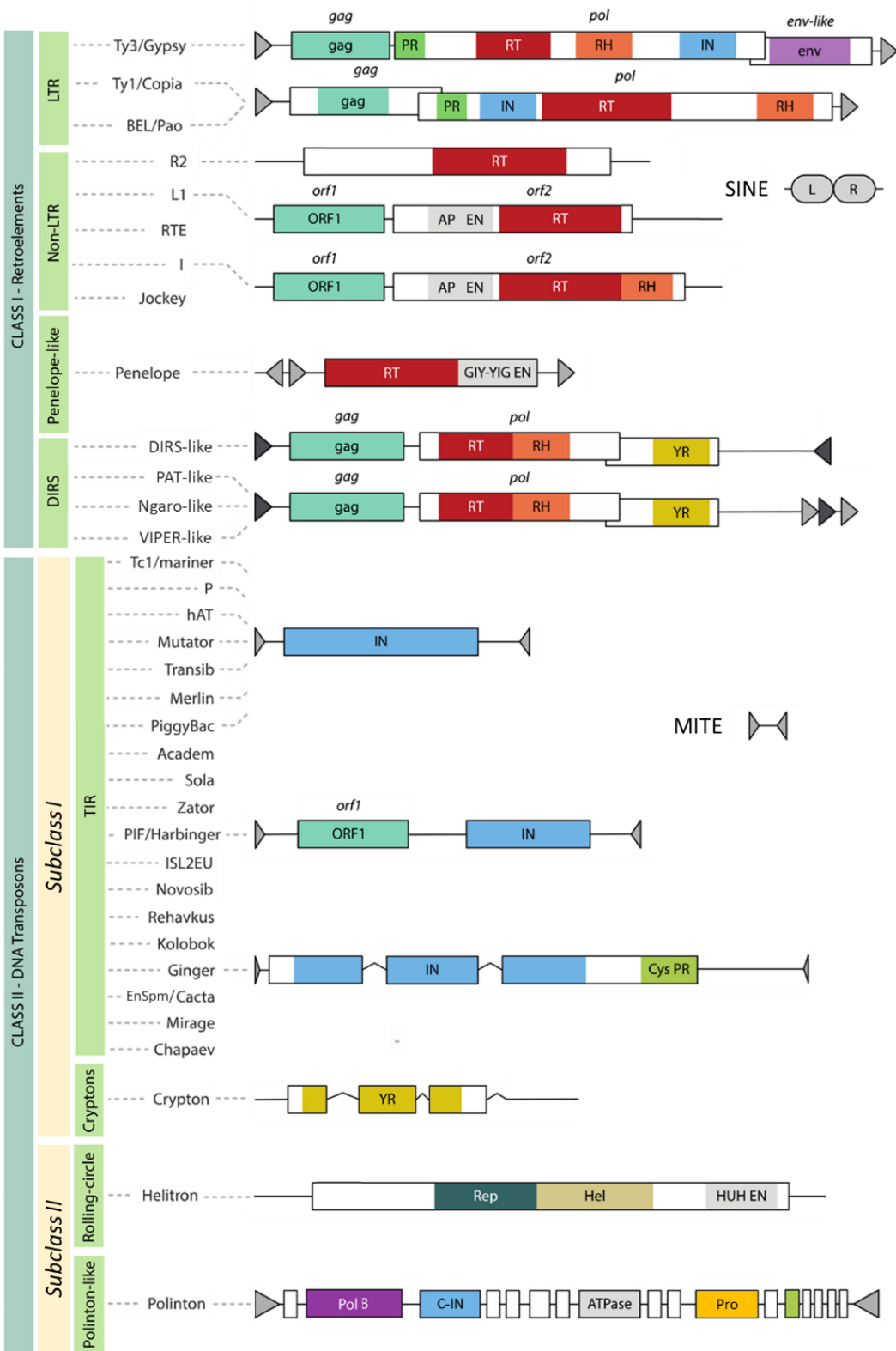
Figure 4. TE classification and characteristics, adapted from [65].
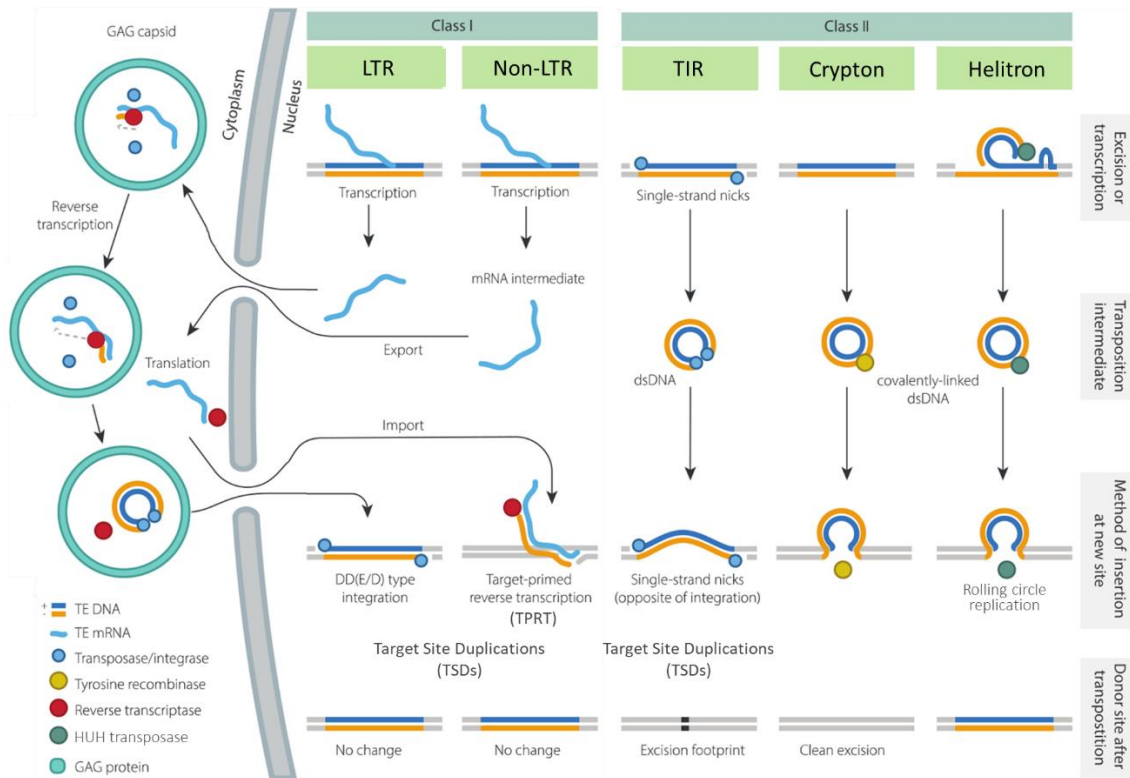
Figure 5. Transposition mechanisms, adapted from [65].

1.2.1.1 Class I elements

Class I elements or retrotransposons transpose by reverse transcription of an RNA intermediate, which is then integrated into the host genome [63]. This DNA-RNA-DNA mechanism is also called *copy-and-paste*, because the original template remains intact and a new copy is integrated in a new location, increasing the number of insertions in the host genome (Figures 4 and 5). Retrotransposons are subdivided in four main orders according to their replication and integration mechanisms [62].

Long Terminal Repeat elements

The **Long Terminal Repeat** (*LTR*) order includes elements with an structure and transposition mechanism similar to retroviruses [63]. Even though they are present in many eukaryotes [65, 66], they are the predominant order in plants, and they can extent from hundred base pairs (bp) up to a few kilobases (Kb) [62]. They have two long direct sequence repeats (LTRs) flanking the internal region that encodes for one or multiple long open reading frames (ORFs) encoding different proteins [63, 66]. The *gag* gene encodes for a structural protein that binds to the retrotransposon RNA and the *pol* gene encodes for an aspartic protease (PR) required for protein processing, a reverse transcriptase (RT), a Ribonuclease H (RNase H, RH) required for the

complementary RNA degradation, and a DDE integrase (IN) [62, 65, 66] ([Figure 4](#)). Sometimes, an additional ORF of an unknown function could be found, as well as an *envelope* (*env*)-like gene [66]. This latter gene, is usually non-functional in TEs, but when it is functional, *LTR* retrotransposons act as retroviruses, infecting other cells, as observed in some *gypsy* [67] and *ZAM* [68] families in *Drosophila*. We find two main *LTR* superfamilies based on the RT amino-acid composition and the order of the coding regions in the *pol* gene: *Pseudovirus* (usually referred as *Ty1/copia*-like) and *Metavirus* (usually referred as *Ty3/gypsy*-like) [66]. Inside this latter order and, despite their structural similarities, the *Bel–Pao* superfamily (*Semotiviruses*) is separated from the other families because it forms a distinct clade based on RT phylogenies [69] ([Figure 4](#)).

For their mobilization, the *LTR* element is transcribed from a RNA polymerase II (Pol II) promoter site located in the 5' LTR [66] as a single polycistronic RNA [65] ([Figure 5](#)). The resulting RNA is translated in the cytoplasm and the RNA molecules are packed into virus-like particle using the *gag* proteins [66]. Then, the RNA transcript is reverse transcribed by the RT: first, a host tRNA, which binds the primer-binding site (PBS) immediately adjacent to the 5' LTR, is used as a primer to produce a copy of part of the 5' LTR [70]. Second, the created cDNA binds the 3' LTR of the RNA by complementarity, and act as a primer to reverse transcribe the rest of the RNA template. Finally, the RH activity hydrolyzes the complementary RNA, except a small region, which is used as a primer for the reverse-transcription of the second strand in a similar way as the first [70]. The new cDNA is imported into the nucleus [70] where it is integrated into the genome by the IN, which produces single-stranded gaps [71]. The repair of these gaps produces small target site duplications (TSDs) [62] ([Figure 5](#)). Even though most of the retrotransposon families do not insert site-specifically, some of them have site preferences. For example, some *Ty1/copia* families in yeast preferentially target tRNA genes [72] and subtelomeres [73]. In *D. melanogaster,* some *gypsy*-like families have the preference to insert in a target DNA with specific characteristics, such as *gypsy* (5'-ATPuPyAT-3'), *Idefix* (5'-TATATA-3') [74] and *ZAM* (5'-GCGCGC-3') [75].

Non-LTR elements

This order includes Long and Short Interspersed Nucleotide Elements (*LINEs* and *SINEs*), simpler than the previous order and devoid of LTRs [63], which give them their name. Nevertheless, after the discovery of other retrotransposon without LTRs (*DIRS*-like elements, explained below), they are more appropriately called target-primed *Non-LTRs* by their transposition mechanism [71]. *Non-LTRs* are present in all Eukaryotic kingdoms, they are highly variable depending on the organism, and have a size of several Kb [62]. They are divided into five main superfamilies *L1, RTE, I, Jockey* and *R2* [76]. They usually have two ORFs ([Figure 4](#)), the first one (ORF1) encodes

for a *gag*-like protein with nucleic-acid binding ability and chaperon activity that creates ribonucleoprotein particles (RNPs) [77]. The role of the *gag*-like protein remain unclear, but it is thought to facilitate the entry of the RNA to the nucleus and the formation of a complex containing the DNA primer at the target site, and the RNA template for the initiation of the reverse transcription [78]. However, this ORF is absent in the superfamily *R2* [79]. The second ORF (ORF2) encodes a protein with both endonuclease (EN) activity, usually apuric/apyrimidic (AP) [62], and reverse transcriptase (RT) activity [71] (Figure 4). *Non-LTRs* usually do not contain RH, which is present in some members of the superfamily *I* [80]. Flanking the ORFs, there are 5' and 3' Untranslated Regions (UTRs), which despite being quite variable [81], they often contain an internal Pol II promoter at the 5' end [77] and a poly(A), A-rich or short repeat sequence at their 3' end [62, 63] that is recognized by the RT protein [81] (Figure 4). Some exceptions are the *R2* elements, which have no apparent endogenous promoter and are thought to be co-transcribed by RNA Polymerase I (Pol I) with the host rRNA repeats [82], and the mammalian *L1* element, for which the 3' UTR is dispensable for retrotransposition [79].

In general, the RNA retrotransposons are transcribed in the nucleus, exported to the cytoplasm for gene translation, and imported to the nucleus with the necessary proteins, where it is revers-transcribed and integrated simultaneously [81] using a target-primed reverse transcription (TPRT) mechanism (Figure 5). This mechanism uses a new 3' end in the insertion target site, nicked by the EN, as a primer for the reverse transcription and integration of the TE [71, 83]. The mechanisms for the synthesis of the second strand is still unclear, but the hypothesis is that there is a second nick in the other strand, which also acts as a primer and the new cDNA copy is used as a template [81]. They generally produce TSDs upon insertion [62]. The common premature termination of their reverse transcription [62, 84] can lead to 5' truncated copies [71], which prevents their propagation though the loss of sequences necessary for their transcription or transposition [63]. Even though most *Non-LTR* elements only have some target site preference, such as the preferential insertion of *LINE-1* in AT-rich regions [85], some of them are integrated into specific sequences [86]. For example, *R1* and *R2* elements, both transcribed by Pol I [87], are inserted into the 28S regions of the rRNA in a sequence dependent manner [88]. In *D. melanogaster*, the subfamilies *TAHRE*, *TART* and *HeT-A* of the *Jokckey* superfamily are associated to telomere ends, and the *gag* protein has been suggested to target these chromosome ends [89, 90]. Finally, the *SINEs* are non-autonomous *Non-LTR* elements of around 100-600 bp that contain internal promoters for Pol III and it is suggested they were derived from ancestral RNA genes [91], such as tRNA [92], 7SL RNA [93] or rRNA [94]. They use the proteins encoded by *LINEs* in *trans*, hence using the same transposition mechanisms [91]. A family of this order, called *Alu,* is one of the most abundant repetitive elements in the human genome and copies are usually located in GC-rich regions [85].

Penelope-like elements

Penelope-like elements (PLEs) were first found in *D. virilis* [95], but have been identified in other phyla, such as specific species of amoebae, fungi, plants, unicellular animals, cnidarians, rotifers, worms, fishes, arthropods and amphibia, among others [96, 97]. They have direct or inverted oriented pseudo-LTRs [98] and a single ORF encoding a RT domain, more similar to the telomerase reverse transcriptase (TERT) than to other retrotransposon RTs, and a GIY-YIG endonuclease (EN) domain, similar to intron-encoded and bacterial UvrC DNA repair endonucleases (Figure 4) [97]. They transpose by a TPRT-like mechanism [71, 96, 99] (Figure 5), and the EN has been proposed to have a certain degree of target sequence specificity, possibly for AT-rich regions [99]. Like *Non-LTR* retrotransposons, these elements frequently have 5′ truncations, being able to create non-autonomous derivates, and they also produce TSDs of variable length upon insertion [96]. Interestingly, some PLEs can retain introns after transposition [98]. This order also includes PLEs elements without EN, which sometimes include an additional ORF with an undefined function, which has been described to be telomere-associated in some species of rotifers, fungi, stramenopiles (protists) and plants [100, 101].

DIRS

Dictyostelium intermediate repeat sequence (*DIRS*) elements, also called tyrosine recombinase (YR) retrotransposons, contain a YR gene instead of an IN [102]. They have been detected in a diverse set of species, from fungi and green algae to a wide range of animals [102]. *DIRS* elements have been proposed to be separated into four superfamilies: *DIRS*-like, *PAT*-like, *Ngaro*-like, and *VIPER*-like [103]. The elements included in this order have high diversity structures and unusual characteristics, such as long overlapping ORFs, new protein-coding domains and introns [102]. Generally, they contain three ORFs: a putative *gag*-like gene, a YR and the reverse transcriptase/RNase H (RT/RH) (Figure 4) [102, 103], but some of them also include a methyltransferase (*DIRS*-like and *PAT*-like) or an hydrolase (*Ngaro*-like), both with still unknown roles [102]. They have unique terminal repeat sequences that vary in structure [65], for example *DIRS-like* elements contain long terminal repeats inverted and non-identical (ITRs), flanking an internal region containing a short non-coding sequence at the 3' end (Figure 4) [104]. This region is known as the internal complementary region (ICR), and has complementarity to different parts of both ITRs [102, 105]. The rest of the superfamilies have 'split' direct repeats (SDRs) flanking their internal region [102, 103] (Figure 4). SDRs are usually two pairs of complementary direct repeats (A and B) in an alternating pattern (usually A_1 - internal region - B_1 - A_2 - B_2) [102, 105]. Their transposition mechanisms is not fully understood, but after the reverse transcription of their mRNA, the paring between complementary regions of the TE seem to be involved in the synthesis of a full length copy element, which is then inserted into the genome by a site-specific

recombination mediated by YR and without producing TSDs [71, 104, 106]. This integration seems to be targeted into a trinucleotide site-specific for each particular YR [102].

1.2.1.2   Class II elements

Class II elements or DNA transposons are a very disparate group of elements present in eukaryotes and prokaryotes that move using a single or double-stranded DNA intermediate, and therefore their mobilization is directly from DNA to DNA [107]. They have a transposition mechanism of *cut-and-paste*, based on their direct excision from their location and their insertion in a new genomic region [63]. Although this transposition mechanism by itself does not generate an increase in the copy number, except for *Helitrons*, it can be increased by preferentially transposing from a replicated to a not-replicated site during host DNA synthesis [108]. They are divided in two subclasses according to the number of DNA strands that are cut during their mobilization [62]. The first subclass includes DNA transposons that break both DNA strands for transposition and are subdivided in two orders: *TIRs* and *Cryptons*. The second subclass includes DNA transposons poorly characterized, which cleave only a single DNA strand, and are divided in *Helitrons* and *Maverick* or *Polinton*.

TIRs

TIRs are a diverse order of DNA transposons consisting of two short terminal inverted repeats (TIRs) flanking a coding region, usually a single ORF encoding a recombinase called transposase [65]. In most cases, this transposase contains DDE motifs (*D:* aspartic and *E:* glutamic residues) in its catalytic domain [65] (Figure 4). They are widely expanded in different organisms and their exact transposition mechanism varies between superfamilies and according to the transposase type. Briefly, the transposase catalyses the first DNA strand break by an hydrolysis of the phosphodiester bonds, which joins two successive nucleotides of the DNA, close to the end of the TIRs [71, 109]. The mechanism involved in the second strand break varies according to the transposase. Finally, the new 3'-OH ends, which are protected by the transposase, are inserted into the new target site by a staggered trans-esterification reaction, also performed by the transposase [71, 109] (Figure 5). Upon insertion, repair of the gaps flanking the insertion produces TSDs whose size varies depending on the element's superfamily. Some superfamilies also show insertion site preferences: next to TA dinucleotides for *Tc1/Mariner*, TAA for *PIF/Harbinger*, TTAA for *piggyBac* [62] and for *Kolobok* [110], CCGG for *Ginger* [111], AT-rich tetranucleotides for *Sola* [112], and GC-rich for *Transib* [113] superfamilies. These elements can also suffer deletions of their internal coding sequence, producing non-autonomous elements called Miniature Inverted–Repeat Transposable Elements (MITEs). These small elements mainly

consist of two TIRs flanked by TSDs inserted in AT-rich regions and are widely represented in diverse taxa at a high copy number [114].

Crypton

The **Crypton** order includes poorly known DNA transposons, only present in a few eukaryotes [115]. They contain a long gene with several introns that encodes for a Tyrosine Recombinase (YR) [115] ([Figure 4](#)). They lack TIRs but they are flanked by short direct repeats produced by recombination during their insertion. It has been proposed that their transposition involves an excision from the host genome by recombination mediated by YR and using the short direct repeats. Then, the circular double stranded DNA molecule is inserted in the target site by recombination between the circular molecule and the host target DNA [40] ([Figure 5](#)).

Helitron

Helitrons are abundant in a wide range of eukaryotic lineages, although they are more frequent in animal and plant genomes [116]. They replicate using a rolling-circle replication [117], are devoid of TIRs and do not generate TSDs after their insertion [116]. *Helitrons* usually have conserved 5'-TC and CTRR-3' motifs (where R is a purine) and a short palindromic sequence of around 16 to 20 nt, a few nucleotides from the 3′ end, with the potential to create a short hairpin structure [117]. Most *Helitrons* are non-autonomous [118], but autonomous copies encode a Rep/Helicase (Rep/Hel, where Rep: replication initiator) protein formed by two conserved tyrosine residues (Y2-type HUH transposase) and a 5'-to-3' DNA helicase domain (Hel) [116] ([Figure 4](#)). It has been suggested that for their 'peel and-paste' transposition mechanism, the HUH transposase makes a single strand break at the 5' end of the TE, that leads to a DNA synthesis in the 3' end of the host DNA while producing a circular TE DNA by the Hel domain [117] ([Figure 5](#)). The Hel is then paused at the 3' end of the TE by the hairpin structure, and the HUH transposase produces the second strand break to free the single strand DNA circle, that will be inserted in the target site producing a new copy [117]. However, some studies in maize suggest that some *Helitrons* do not increase their copy number, and they use a *cut-and-paste* mechanism [65]. In general *Helitrons* are inserted at AT target sites [117].

Maverick/Polinton

The **Maverick/Polinton** order is also called self-synthesizing transposons because they encode for their own polymerase [65]. They are the most complex DNA transposons in eukaryotes and are poorly characterized. They have been found in several eukaryotes, such as fungi and some animals, including insects [119]. They are long (15-20 Kb) and the coding region is flanked by long TIRs [119] usually with 5'-AG and TC-3'termini [120]. Autonomous *Polintons* can encode

for more than 10 different proteins, often virus-like proteins, including: a protein-primed DNA polymerase B (Pol B), a retroviral-like integrase called c-integrase (C-IN), adenoviral-like protease (Pro) and a putative ATPase (ATP) [120] ([Figure 4](#)). It has been suggested that during the transposition the element is excised through a single strand cleavage by the C-IN, followed by the synthesis of DNA by the Pol B and facilitated by the ATPase [120]. They usually produce a 6 bp TSD upon insertion [119]. Some elements of this order can also be non-autonomous [120]. Because many *Polinton* elements also encode for a capsid-like protein, they could also be considered endogenous viruses [121].

1.2.1.3 TE landscape in the *Drosophila* species studied

The genomic content of DNA repeats (mostly based on TE content) can vary between *Drosophila* species. Indeed, in a study of 26 *Drosophila* species, the content of repeats ranged from 4.65% in *D. busckii* to 30.80% in *D. suzukii* [122], and it has been detected to be approximately 20% in *D. melanogaster* [58]. Moreover, TE content significantly correlates with genome size [122]. In addition to interspecific differences, intraspecific TE landscape variation has also been described in several *Drosophila* species [58]. All TE classes are present in the *Drosophila* genomes, but generally the most abundant are the *LTR* retrotransposons and *LINEs* [122]. Nonetheless, *DIRS* and *SINEs* have not been found yet in *Drosophila* [58].

At this time, we have a reference genome for the three species studied in this thesis. However, whereas the *D. buzzatii* and the *D. subobscura* genomes were already available in 2014 [123] and 2019 [124] respectively, the reference genome of *D. koepferae* has just been published in 2022 [21]. The TE fraction in *D. buzzatii* genome was estimated to be around 11%, of which the most abundant order is the *Helitron* [125]. In the case of *D. koepferae*, the repetitive content in the genome has not been properly performed [21]. However, the estimated TE classes and proportions of recent and active elements in *D. koepferae* were similar to *D. buzzatii* when using transcriptomic data (9.8% of the transcriptome in *D. buzzatii* and 8.8% in *D. koepferae*) [24]. Additionally, gonad TE expression was mainly represented by retrotransposons: *LINEs* were the most expressed category followed by *LTRs* [24]. Regarding *D. subobscura*, the cursory estimation of the TE content made from the reference genome was 7.69%, where the highest percentage corresponded to the TE Class II [124]. However, prior to this thesis, no proper study of its repetitive content and TE expression had ever been performed.

1.2.2 TE Regulatory mechanisms

Even though some transposition events can create genetic variability, being a source of biological innovation and regulatory elements, most of them are detrimental for the host (explained in

section [1.2.3]) and are selected against. TE mobilization in the germline results in a TE frequency increase in the next generation, regardless of the possible disadvantages produced in the hosts, increasing TE propagation [126]. Natural selection can act on the new TE insertions, fixating the adaptive ones and having an impact on the species evolution [127]. Conversely, transposition in the soma is not transmitted to the offspring, reducing the fitness of the individual organisms and leading to an evolutionary dead-end [126]. For this reason, some authors suggested that TEs evolve to be silent in somatic cells but active in the germline [126]. However, while some authors support that somatic mobilization of TEs is more common than previously thought in *Drosophila* and other organisms [127], other studies in *Drosophila* suggest the opposite [128]. For example, the relaxation of the epigenetic control in the early development of the germline from the primordial germ cells, gives an opportunity for TE activation [129]. In addition, an example of TE repression in somatic cells is the *P* element of *Drosophila*: its mRNA retains the third intron in these cells, which prevents the transposase translation, but it is spliced out in the germline [130]. On the other hand, TE transposition in the germline can also produce sterility in the organisms [131, 132], restricting their evolutionary transmission. Hence, the balance between TE expression, sufficient to promote their amplification in the host genomes, and TE repression, sufficient to avoid a massive host fitness disadvantage that would counteract the benefit of TE expansion in the host genome, is the basis for TE survival. Therefore, the focus of TE research has been performed in the germline or the first stages of embryo development, moments when transposition can be transmitted to the next generation [127].

TEs can scape negative selection by evolving some target site specificity to direct their integration in locations of the genome where they minimize the host damage [133] (explained in section [1.2.1]). For example, *LINE-1* are preferentially inserted in AT-rich regions [85], which may restrict their insertion in genes, usually GC rich. A poor TE-encoded protein transcription and suboptimal translation [134] has also been proposed for some AT-rich TEs [135] in some mammal and plants [135] due to the relationship between high GC content, and increase of mRNA levels [136] and protein translation [137], but this relationship was not found in *Drosophila* TEs [135]. TEs have also evolved some characteristics to control their expansion, such as an accumulation of mutations in their TE copies, which avoids or decreases their transposition [138]. For example, the target-primed reverse transcription of *Non-LTRs* often stops prematurely [62, 84], leading to 5' truncated and inactive copies [71]. Some TEs have evolved self-regulatory mechanisms to control their own number of insertions [138]. For example, the overproduction of a transposase above a certain threshold can reduce the overall level of transposase activity, as observed in mammalian cells in the *piggyBac Tc1/mariner* transposons [139]. They can also contain a mutated transposase which impairs the activity of the wild-type enzyme, reducing the transposase activity, as described in the *Tc1/mariner Mos1* [140].

In addition, and to minimize the negative effects of TE mobilization, organisms have also evolved potent restriction mechanisms to combat TE amplification, acting as a first defense against new TE invasions [141]. Unified control mechanisms have been found in almost all organisms, which include a pre-transcriptional mechanism, based on avoiding mRNA transcription, and a post-transcriptional mechanisms, based on TE mRNA silencing or degradation by small RNAs with homology to the target [141]. The pre-transcriptional TE silencing mechanisms include: i) DNA methylation by DNA Methyltransferases (DNMTs), as observed in organisms with large genomes (vertebrates, flowering plants and some fungi) [141] ii) TE-associated histone modifications, which mainly include H3K9me3 in plants and animals [142]. Besides, the post-transcriptional TE control mechanisms are based on the interaction between proteins of the Argonaute family, with two different small non-coding RNA classes: i) endogenous short interfering RNAs (endo-siRNA), mainly involved in TE silencing of fungi and plants, but also in animal somatic cells [60] ii) Piwi-interacting RNAs (piRNA), specific of animals and mainly involved in TE silencing in the germline [127], but also in the somatic cells of some organisms [142] and in the regulation of some genes [142]. The main processes are similar between organisms with specific peculiarities in each case [143]. In addition, pre-transcriptional and post-transcriptional TE silencing mechanisms are also inter-connected: siRNAs in plants and piRNAs in animals have been associated to changes in the DNA methylation and addition of repressive chromatin marks, which increase the TE silencing in these species [60, 142]. Here we will explain in more detail the main regulation pathways in *Drosophila*.

<u>Regulatory mechanisms in *Drosophila*</u>

In *Drosophila*, the main mechanisms for TE regulation based on small RNAs in both somatic and germ cells [58] are the piRNA and the siRNA pathways. TE silencing in somatic cells is mainly based on endogenous short interfering RNAs (endo-siRNAs), which a are 21-nt long RNAs [144]. However, this pathway is also active in the germline working in the same way as in the soma [145]. The piRNA pathway is the main TE silencing pathway in the *Drosophila* germline [146], but it is also active in the surrounding cells that support oogenesis and the ovarian somatic follicle cells [147]. Nonetheless, this pathway shows differences between those two tissues. piRNAs are 23-30 nt long RNAs and are associated with the Piwi (P-Element induced wimpy testis)-Argonaute subfamily proteins, which inspired the name [58]. Both piRNAs and siRNAs in this organism have been involved in post-transcriptional TE silencing by histone modifications [143, 148], which include enrichment of TE H3K9me3 and H3K27me3 heterochromatic marks [149]. DNA methylation in this organism is almost completely absent [150].

Regarding the piRNA pathway, the piRNA biogenesis starts with long piRNA precursors transcribed mainly from discrete regions in the genome called piRNA clusters or piRNA-

producing *loci* [151]. The first discovered cluster in *Drosophila* was the heterochromatic *locus* flamenco (*flam*), located on the pericentromeric heterochromatin of the X chromosome [152]. This *locus* controls the expression of several TEs, mainly *gypsy*, *ZAM* and *Idefix* in ovarian somatic cells [152]. Nevertheless, due to the improvement of the whole genome sequencing techniques, other piRNA clusters have been identified. For example, 142 piRNA clusters covering around 3.5% of the assembled genome are annotated in *D. melanogaster* [151]. Most of these clusters are located in pericentromeric and telomeric heterochromatin regions of all chromosomes, and contain mainly defective TE sequences that lost their ability to transpose [151]. However, some clusters were also detected in euchromatic regions [151, 153], or generated from active TE insertions [153]. Their sizes vary substantially by cluster and the 15 largest clusters were detected to produce up to 70% of the total piRNAs of this species [151]. piRNA clusters evolve rapidly, only 70%–80% of them are detected to be conserved within *Drosophila* species and almost no similarities are shared between closely related species [154]. In *Drosophila* and other arthropods [155], piRNA clusters can be 'uni-stranded', which generate piRNA precursors from only one genomic strand, or 'dual-stranded' which are transcribed from both DNA strands [142, 151]. Uni-stranded piRNA clusters possess their own promoter, have a similar mechanism to canonical mRNAs [156] and are the main ones in ovarian somatic cells [147], except the cluster *20A*, also expressed in the germline [157]. Dual-stranded clusters lack a clear promoter, contain H3K9me3 chromatin mark, related to transcription repression [158], and produce the majority of piRNAs in the germ cells of *Drosophila* [151, 159].

Regarding differences between sexes, even if most piRNA clusters are dual-stranded [160] and shared between *D. melanogaster* females and males (except those on the Y chromosome), the piRNA production is different [161]. piRNA pathway has been extensively and mainly studied in ovaries because in testes is less active, expressing less piRNAs and lower expression levels of piRNA pathway genes [162]. However, it is known that the Rhino–Deadlock–Cutoff protein complex is required for dual-stranded piRNA cluster expression in both sexes [160], suggesting the same or a similar mechanism. A recent study in *D. melanogaster* questioned the role of piRNA clusters in TE control, because the simultaneous deletion of the three most highly expressed germline piRNA clusters (*42AB, 38C,* and *20A*) did not induce TE reactivation or increase in transposition, even though there was a strong reduction of piRNAs amounts [163]. The only cluster known to be essential for TE repression is the uni-stranded piRNA cluster *flam* [151, 164, 165]. This suggests that most piRNAs involved in TE silencing, are obtained from numerous TE insertions dispersed across the genome [163]. This idea is supported by the discovery, in the same species, of novel TE insertions inducing more frequently *de novo* piRNA production from the TE flanking regions than from the piRNA clusters [166].

Finally, even though this section will focus on the impact of the piRNA pathway in TE regulation, a growing number of studies found that piRNA can also modulate gene expression. An important example is the X-linked Stellate (*Ste*) genes in *D. melanogaster*, present in several copies, which encode the β-subunit of the Casein kinase II [167]. These genes are repressed in testes by piRNAs derived from the homologous Y-linked Suppressor of Stellate (*Su(Ste)*) locus [168]. *Ste* derepression produces an accumulation of abundant crystalline aggregates in spermatocytes and meiotic disorders, thus preservation of their silencing is essential for male fertility [169]. In addition, the 3' untranslated region (UTR) of the protein-coding gene traffic jam (*tj*), encoding a Maf transcription controlling gonad morphogenesis in the same species, also produces piRNAs [170]. In the gonadal somatic cells, this gene seems to activate Piwi expression and supply piRNAs to Piwi to regulate the expression of other genes, such as Fasciclin 3 (*Fas3*), a cell adhesion molecule [170]. In addition, the 3′ UTR of *c-Fos*, a proto-oncogene influencing many cell and developmental processes, recruits a Piwi complex that produces primary piRNAs, which causes instability and degradation of the gene-encoding mRNA [171].

1.2.2.1 Regulatory mechanisms in the *Drosophila* germline

In *Drosophila* early embryos, both piRNAs and Piwi-family proteins (Piwi, Aub and Ago3), are inherited from the maternal germline, and are necessary to establish the piRNA pathway in the adult germline [172]. These maternally inherited piRNAs and Piwi-family proteins, together with the embryonic expression of Piwi and other piRNA pathway proteins, control transposon expression in early development stages and deposit the repressive mark H3K9me3 on dual-stranded piRNA clusters [158, 173], necessary to start their transcription. In addition, the expression of Piwi during early phases of embryogenesis in the developing somatic cells prevents TE expression in the embryo and helps to establish its epigenetic landscape [174].

piRNA cluster transcription

The dual-stranded piRNA cluster transcription (Figure 6) generally starts with the DNA-binding protein Kipferl (CG2678), which binds to guanosine-rich DNA motifs and recruits Rhino, an heterochromatin protein 1 (HP1) homologue, to specific *loci* [175]. However, some Rhino-dependent dual-stranded piRNA clusters do not need the recognition of a specific DNA sequence by Kipferl [175]. Then Rhino, recognizes and binds to H3K9me3 chromatin marks of dual-stranded piRNA clusters [159] and recruits Deadlock (Del) and Cutoff (Cuff) proteins to form the Rhino–Deadlock–Cutoff (RDC) complex [157, 158, 176]. This trimeric complex, without the need of promoter sequences [142], activates the cluster transcription via a non-canonical

mechanism: Del recruits the transcription initiation factor IIA subunit 1 (TFIIA-L), called Moonshiner (moon), which recruits its co-factor TFIIA-S and the TATA-box binding protein (TBP)-related factor 2 (TRF2) [177]. This TFIIA pre-initiation complex allows to initiate a Pol II dependent transcription [177]. Because the H3K9me3 chromatin mark is not strand-specific, transcription occurs randomly from both genomic strands, possibly originating precursors of several sizes [178]. The Cuff protein of the RDC complex, which is similar to a transcription termination factor, binds
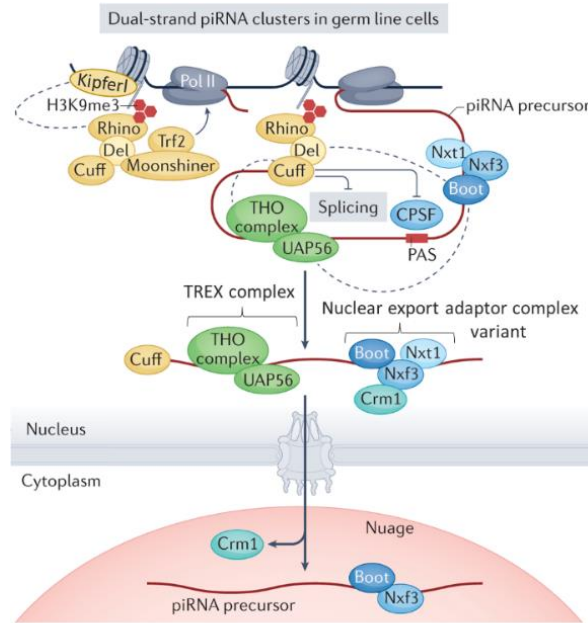
Figure 6. Dual-stranded piRNA cluster transcription, adapted from [148].

nascent piRNA precursor transcripts and contributes to their transcription, processing and stability [158, 179]. Cuff inhibits 3′ end cleavage and polyadenylation of piRNA precursor transcripts by interfering with the recruitment of the cleavage and polyadenylation specificity factor (CPSF) complex [180], and it also suppresses piRNA precursor splicing by recruiting the transcription and export (TREX) complex, which contains Hel25E (a putative RNA helicase), and the THO complex [176, 181]. It has been proposed that Cuff also binds the 5'-monophosphate of the piRNA precursors transcripts, protecting them from their degradation by the exonuclease Rat1 [180]. Finally, dual-stranded piRNA precursor transcripts, which are long, un-capped, un-spliced, and devoid of a poly(A) tail [182], are exported to the cytoplasmic perinuclear ribonucleoprotein (RNP) granules called 'nuage' [183].

piRNA processing

The piRNA precursors are subsequently subjected to a processing and maturation mechanism in the germline: the Ping-Pong amplification cycle (Figure 7A) and the Phasing (Figure 7B), resulting into mature piRNAs [143]. The Ping-Pong amplification cycle is carried out in the 'nuage' and produces secondary piRNAs. The Phasing is carried out in the interface between the 'nuage' and the mitochondrial outer membrane and produces primary piRNAs [143]. For the Ping-Pong cycle initiation, maternally inherited Aub proteins already loaded with maternal piRNAs are needed [172]. Aub is predominantly loaded with antisense piRNA precursors that often have an uracil in their 5' end, which binds preferentially to an adenine at the first position of its target [184] (Figure 7A). The complex Aub-piRNA binds to complementary sense piRNA

precursor transcripts or TE mRNAs [143], slicing them between the 10th and 11th nt from the 5′ end of the piRNA guide and generating a new monophosphorylated 5′ end in the target [151, 185]. Specifically, the Krimper protein binds to loaded Aub and unloaded Ago3 [186], forming the Ping-Pong piRNA processing complex [187] and ensuring the close proximity of these two proteins, so that Ago3 is loaded with sense piRNA given by Aub [188]. The DEAD-box ATP-dependent RNA helicase Vasa facilitates the transfer of the new 5' end, produced by Aub to Ago3 [182, 189], in a complex called Amplifier, which seems to also involve the protein Qin/Kumo (Qin) [189]. Qin inhibits homotypic Aub–Aub interactions, leading to efficient heterotypic Aub–Ago3, which ensures that antisense piRNAs predominates *vs* sense piRNAs [190, 191]. The Hsc70/Hsp90 chaperone machinery, including Hsc70-4, Hsp90, Hsp70/Hsp90 organizing protein (Hop) and Droj2 (DnaJ-like-2) [192], together with the co-chaperone shutdown (Shu) [193], are also required to load piRNAs onto Ago3 [192, 193] (Figure 7A). Then, the final 3' end of the piRNA loaded into Ago3 can be produced by: the trimming of the 3' end of the piRNA fragment by the exoribonuclease Nibbler (Nbr) [194] or by an unknown exonuclease, recruited by Papi [195, 196], or to a much less extent, though a cleavage by Zucchini (Zuc) [143, 197], supported by Minotaur (Mino) [198]. Finally, the 3' end is 2′-O-methylated by the methyltransferase Hen1 to produce a mature piRNA [199] and which protects the piRNAs from degradation [200].

Then, Ago3, which is mainly loaded with sense piRNAs with an adenine at the 10th position from the 5' end, binds to antisense TEs or piRNA cluster transcripts, slicing them between the 10th and 11th nt from the 5′ end of the piRNA guide and producing a new monophosphorylated 5′ end [185] (Figure 7A). The new fragment, which starts with a uracil, is loaded onto Aub with the help of a yet unidentified protein (Vasa or similar) [189]. Consequently, there is an overlap of the first 10 nt of the mature piRNA, produced by Aub, and the first 10 nt of the mature piRNA, produced by Ago3 [151, 185]. The 3' end of the piRNA precursor loaded into Aub can be produced by Nbr or an unknown exonuclease recruited by Papi [195, 196] and the subsequent 2′-O-methylation by Hen1 (Figure 7A). However, the 3' end is predominantly produced by Zuc together with the production of Phased primary piRNAs associate with Piwi (explain below in more in detail, Figure 7B). Aub and Ago3 continue this reciprocal cycle to constantly produce antisense piRNAs and amplify the amounts of TE repressing piRNAs (Figure 7A). Additional proteins localized in the 'nuage' are required for an effective Ping-Pong amplification, but their functions are not well described yet. For example, Spindle-E (Spn-E), a putative DExH-box RNA helicase has been proposed to make the RNAs more accessible for Aub loading [201]. Tejas (Tej) [202] and its paralog Tapas [203], have been proposed to interact with Spn-E and Vasa in the 'nuage', and in synergy, they seem to work as a core component for the 'nuage' assembly, controlling its dynamics and organization [203, 204].

Figure 7. piRNA germline processing, adapted from [148].

The Phased piRNA biogenesis starts with the cleavage of a long piRNA precursor (antisense TE sequence or piRNA cluster transcript) by the Ago3 complex [195] (Figure 7A). Then, Aub binds the long 5′-monophosphorylated piRNA precursor [195], as does Armitage (Armi) [205]. Armi (an ATP-dependent helicase) translocates the Aub-long piRNA precursor complex from the 'nuage' to the outer membrane of the mitochondria [205, 206], working as a bridge between the Ping-Pong cycle and the Phased piRNA production [205] (Figure 7B). In addition to the explained proteins, Brother of Yb (BoYb) and Sister of Yb (SoYb), two proteins homolog to the soma-specific protein female sterile (1) Yb (fs(1)Yb or Yb), are also essential for the Primary piRNA biogenesis [207]. BoYb and SoY are believed to replace the function of Yb in the germline (explained in detail in section 1.2.2.2) [207]. Vreteno (vret) protein interacts with SoYb and

BoYb, Armi and Phased piRNAs-loaded proteins, stabilizing these complexes [208]. Then, Armi together with the loaded Aub, binds to mitochondrial membrane proteins Gasz and Daedalus (Daed), maintaining this complex close to the phospholipase Zuc [209], which assisted by Mino [198] cleaves the long piRNA precursors to produce the 3′ end of the Aub-loaded piRNA [195, 197]. Then, the Aub-piRNA complex can be translocated again to the 'nuage' to continue the Ping-Pong cycle, making the Ping-Pong cycle and the Phased piRNA biogenesis closely interconnected. In addition, Piwi proteins repeatedly bind the remaining of the long piRNA precursor transcript [195, 197, 210] to guide Zuc in its cleavage into consecutive fragments [197, 210], producing an strings of head to-tail piRNAs [205] (Figure 7B). Piwi has a preference to bind RNA fragments starting with an uracil whereas Zuc preferentially cleaves them immediately upstream the uracil [197], producing an strong uracil bias in the 5' end of the Phased piRNAs [151, 211, 212]. Then, Papi has been proposed to assists the 3′ end trimming of the Piwi-bounded piRNAs [213] by an unknown protein. Finally, the consecutive chain of piRNAs are loaded onto Piwi, methylated by Hen1 [199], and the Piwi-piRNA complexes are imported into the nucleus [214]. The Piwi-loading process seems to be facilitated by the earlier described Hsc70/Hsp90 machinery, together with Shu [215]. The Piwi protein in *Drosopihla* lacks of slicer activity, and therefore the silencing function of the Piwi–piRNA complex is based on avoiding TE mRNA translation by binding and TE transcription by inducing repressive chromatin marks (explained below) [216, 217]. Alternatively, and to a lesser extent, mature Phased piRNAs can also be loaded into Aub to be included into the Ping-Pong cycle [195]. Therefore, whereas the Phased production of primary piRNAs from cleaved piRNA cluster transcript generates diversity of the piRNA pool to silence a wide range of TEs by avoiding their translation [195], the Ping-Pong cycle increases this pre-existing piRNA pool decreasing the TE mRNA amounts [218].

piRNA dependent transcriptional TE silencing

In addition to the post-transcriptional TE silencing (explained above), Piwi is also involved in the pre-transcriptional TE regulation. In the nucleus, the Piwi-piRISCs (piRNA-induced silencing) complex scans nascent TE transcripts, including also protein-coding mRNAs, and these transcripts bind to the guide piRNA by complementarity [219] retaining these transcripts in the nucleus [220] and inducing a co-transcriptional silencing [182]. This process occurs through the action of the protein Asterix/Gtsf1 (Astx) that interacts with Piwi [221] and recruits Panoramix/Silencio (Panx) proteins [222] (Figure 8). The latter forms a trimeric complex with a gonad-specific variant of the nuclear RNA export machinery [220, 223–225], which binds to the TE mRNA in a Piwi-dependent manner and reinforces the association of Piwi with the target [225], improving the co-transcriptional TE silencing [223] and the TE transcript retention in the

nucleus [220]. In addition, there is a dimerization of Panx-nuclear RNA export machinery complexes by the protein dynein light chain LC8/Cut up (Ctp), which interacts with Panx [226], to forms molecular condensates that have been proposed to be required for heterochromatin formation [227]. Moreover, Panx recruits silencing heterochromatic effectors to the *loci* to perform the pre-transcriptional silencing [223]. The SUMO E3 ligase Su(var)2-10 is recruited into the complex [228] and, together with additional proteins, recruits the H3K9
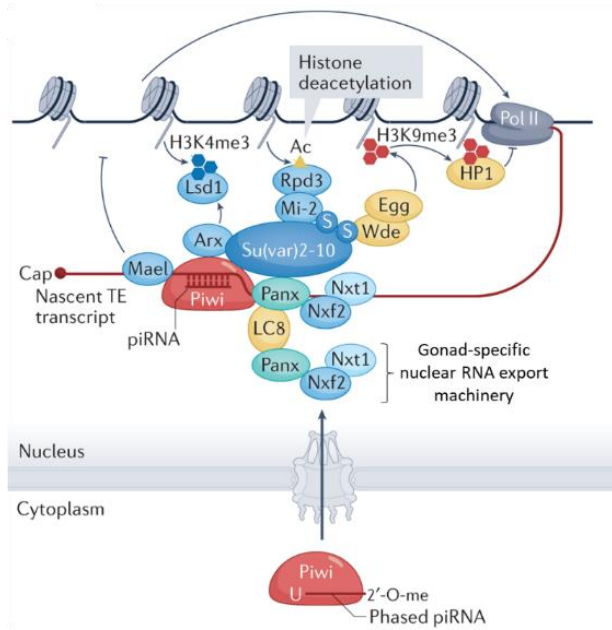
Figure 8. piRNA dependent pre-transcriptional TE silencing, adapted from [148].

methyltransferase Eggless/SetDB1 (Egg) to the chromatin target [229], which establishes H3K9me3-marked heterochromatin in the *loci* [230]. This is followed by the chromatin compaction by the heterochromatin protein 1 (HP1a) [231]. Su(var)2-10, together with additional proteins, also recruits the histone deacetylase Rpd3 to remove H3K9 acetylation [232] and the Lsd1-CoREST complex to remove histone H3K4me3 [233]. Finally, Maelstrom (Mael), which is guided by the Piwi-piRISC complex to TE transcripts and dual-strand piRNA clusters, suppresses the canonical Pol II transcription of TEs outside and within piRNA clusters in the germline, allowing the Rhi-dependent non-canonical dual-strand piRNA cluster transcription (Figure 8).

1.2.2.2 Regulatory mechanisms in the *Drosophila* somatic tissues

The main regulatory mechanism to silence TEs in *Drosophila* somatic cells is the siRNA pathway, because Aub and Ago3 are not expressed in these tissues [165]. Notwithstanding, the piRNA pathway is also present in the ovarian somatic follicle cells and has been associated to the control of some *gypsy*-like retrotransposons to prevent the infection of the oocyte [58, 234]. In the ovarian somatic cells only Piwi is expressed, and therefore only the Phased piRNA biogenesis but not the Ping-Pong cycle amplification takes place [165]. Interestingly, somatic piRNA pathway has also been described in a few other selected somatic tissues, such as adult fat body [235], brain [236] or in head and thorax samples [145].

siRNA pathway

This pathway begins with the synthesis of long double-stranded RNAs (dsRNA), precursors of endo-siRNAs, through three distinct mechanisms: i) bidirectional transcription of a unique TE insertion ii) the interaction of two complementary transcripts derived from distinct TE insertions iii) the formation of an extended hairpin structure, for example by the complementarity of two inverted repetitive elements [237, 238] (Figure 9). Afterwards, the enzyme Dicer-2 (Dcr2), together with its co-factor Loquacious-PD (Loqs-PD) [239] cleaves the long dsRNAs into siRNAs duplexes. Dcr-2 and the dsRNA-binding-domain containing protein R2D2, which together form the RISC-loading complex (RLC), load the duplexes onto Argonaute 2 (Ago2) to form the pre-RNA-induced silencing complex (RISC complex). In this
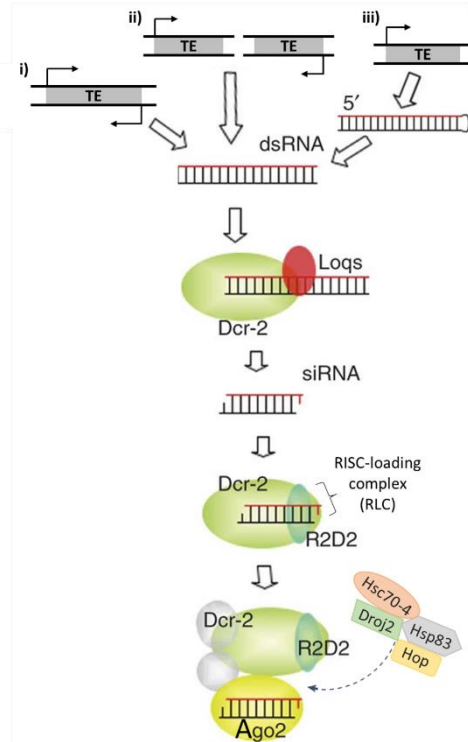
Figure 9. siRNA pathway, adapted from [243].

step, the Hsc70/Hsp90 chaperone machinery, including Hsc70-4, Hsp83, Hop, and Droj2, is required to specifically load small RNA duplexes into the RISC complex [240] (Figure 9). Once loaded with the siRNA duplex, Ago2 cleaves the non-incorporated siRNA strand, liberating the single-stranded guide [241]. Then, protein Hen1 methylates the 3' end of the siRNA, resulting in a 2'-O-methyl group, which may protect the siRNAs from degradation [242]. The mature complex binds TE transcripts, due to their complementarity to the siRNA guide, and Ago2 cleaves them [58]. The endo-siRNA biogenesis pathway in the germline seems to operate in the same way as in the soma [145], but more research is needed to know if additional proteins are also involved. Finally, TE-derived endo-siRNAs have also been proposed to be involved in heterochromatin formation in *Drosophila* somatic tissues [148]. However, this idea and the pathway behind this function have not extensively been studied in this organism.

piRNA pathway in ovarian somatic cells

The best described somatic piRNA cluster, representative for the uni-stranded piRNA clusters, is the flamenco *locus* (*flam*) [143], which is the main source of piRNAs in the ovarian somatic cells [244]. This large cluster contains many defective TE sequences (mainly of *gypsy* [152], *Idefix* and *ZAM* elements) [151, 245], and it is conventionally transcribed in a long precursor transcript that directly produce antisense piRNAs that will target TE mRNAs [151, 152, 164, 246]. For their

transcription, uni-stranded piRNA clusters have their own defined promoters [156, 247] enriched in H3K4me2 [157], commonly found in *Drosophila* active promoters [149], and devoid of H3K9me3 [157], considered an epigenetic silencing mark [149]. They are transcribed by the Pol II [156] [182], and in the case of *flam*, the promoter lacks a TATA-box, but instead it contains a consensus initiator motif (Inr) and a downstream promoter element (DPE) [156] (Figure 10). The transcription of *flam* requires the recognition of a binding site by the transcription factor Cubitus interruptus (Ci) [156]. The long *flam* transcript is alternatively spliced, generating a high diversity of piRNA precursor transcripts, all containing the first exon at their 5' end [156]. This first exon, together with the second one, seems to contain *cis*-regulatory signals

Figure 10. Uni-stranded piRNA cluster transcription, adapted from [148].

required for the binding of the Yb protein enabling the piRNAs production from downstream sequences [248]. Additionally to the canonical splicing, piRNA precursor transcripts from uni-stranded clusters are 5′ capped and 3′ polyadenylated, as canonical mRNAs produced by Pol II [143, 182, 249]. The classical exportin complex, together with the exon junction complex (EJC) [250], Hel25E [250], and a pair of nucleoporins [251], mediates the export of these transcripts to the Yb bodies. The nucleoporins interact first with the *flam* exportin complex and then with the cytosolic protein Yb (the core component of the Yb body), working as a bridge for the nuclear export to the Yb bodies [251] (Figure 10). The Yb bodies are perinuclear ribonucleoprotein (RNP) granules [252–254] enriched in Yb protein and other proteins involved the piRNA pathway, usually located in the perinuclear region close to the mitochondria [255, 256].

In the nucleus, piRNA precursors from various uni-stranded piRNA clusters have been proposed to accumulate, before their export, into a single nuclear structure called 'Dot COM' located near the Yb-body [250, 257, 258]. When *flam* transcripts are exported from the nucleus to the cytoplasm, they also seem to gather together creating another *flam* piRNA precursor accumulation called *flam* bodies, which overlap with Yb bodies [250, 258, 259]. These accumulations seem to vary in a spatio-temporal way, for example in early oogenesis, *flam* transcripts are rapidly transferred to the cytoplasm and 'Dot COM' cannot be visualized, whereas

as during oogenesis the cytoplasmic piRNA precursor delivery decreases together with a decrease in Armi and Yb protein levels, which likely leads to a drop in piRNA production [260]. Additionally, it has also been suggested that in ovarian somatic cells *flam* downstream transcript regions act as a scaffold for Yb body assembly together with Yb and other piRNA pathway proteins [250, 255, 258] and reduces the piRNA production from protein-coding mRNAs [255, 258].

Figure 11. Primary piRNA production (Phasing, A) and piRNA-dependent TE silencing (B) in *Drosophila* ovarian somatic cells, adapted from [148].

After their exportation to the cytoplasm, the piRNA precursors are processed in the interface between the Yb bodies and the mitochondria [255]. Yb recognizes specific sequences of the *flam* uni-stranded piRNA precursors [259, 261] and recruits protein Armitage (Armi) [253], which binds the downstream region of the precursor [262] (Figure 11A). These precursors are then cleaved by an unknown endonuclease to generate their 5′ ends [143]. At the same time, Armi recruits and binds unloaded Piwi [252] in the Yb bodies, and the latter is loaded with piRNA precursor transcripts though their 5′ end. When the complex is assembled (Armi–Piwi–pre-piRNA), it is translocated to the outer mitochondrial membrane by Armi [262], where the precursor piRNAs are processed into mature piRNAs by Zuc in the already explained Phased biogenesis (section **1.2.2.1**) (Figure 11A). In addition to the described proteins, vret and SoYb are

28

also required for the primary piRNA biogenesis in ovarian somatic cells, and the former interacts with the Armi-Yb complex and also with SoYb [207, 208]. In the nucleus, Piwi also silences TEs in the same way as in the germline (section **1.2.2.1**). However, in somatic cells Mael is not necessary for the dual-stranded piRNA cluster transcription (Figure 11B). This protein, while interacting with Piwi, reduces the occupancy of the SWI/SNF chromatin remodeling complex and Pol II on TE promoters through binding to Brahma (Brm), the core adenosine triphosphatase of the chromatin remodeling complex [222].

1.2.3   Impact of TEs in the host genomes

The presence of TEs in the genomes of almost all living organisms has different consequences. It is known that differences in haploid genome size are often poorly correlated with organism complexity or number of genes, due to the presence of repetitive DNA in the genomes, including TEs, which is known as the "C-value paradox" [263]. Therefore, TE content seems often to roughly correlate with the genome size [264, 265]. For instance, the genome of the fungi *Encephalitozoon intestinalis*, one of the most compact in eukaryotes (2.3 Mbp), is completely devoid of TEs [266], whereas the lungfish *Neoceratodus forsteri*, one of the biggest eukaryote genomes (43.000 Mbp), contains around 90% of TEs [267]. The increase in copy number of TEs by transposition is also correlated to a significant expansion in genome size [268] mainly in plants [269], but also in other organisms, such as mammals [270] and *Drosophila* species [122]. In addition, the reduction of genome size can also be related to the loss of TE sequences in organism [264], through processes such as recombination between target site duplications [271]. Hence, the organism genome size is the result of a balance between insertions and loss of TEs [264]. Moreover, TE base composition usually has a bias towards Adenosine and Thymine, and often different from the average composition of the host genes [135]. Thus, the base composition of a genome with high TE percentages will depend on the GC content of the TEs [134]. Even though this fact is less studied, a positive correlation between the genomic base composition and the TE base composition was found in fish [272]. Additionally, high recombination regions tend to be GC-rich, whereas AT-rich regions usually have low recombination rates [273]. The insertions of numerous TEs in low recombination regions usually leads to a higher AT percentage [134, 274].

Furthermore, transposition events can cause mutations, for example TE insertions within protein-coding exons can disrupt genes by changing the open reading frame (ORF), which results in a truncated transcript or an aberrant protein [275]. They can also disrupt splicing sites, being a source of exon-skipping, exonization (creation of new exons), or producing new polyadenylation signals. TE insertions in introns could also lead to the same effect by disrupting or introducing novel splicing sites [275]. New TE insertions in protein coding genes have been associated to human diseases. For example, a *LINE-1* (*L1*) element insertion into an exon of the factor VIII

gene was associated to hemophilia A [276]. Different *Drosophila* mutants with changes in their eye phenotype and pigmentation appeared after the insertion of the *LTR* element *flea* into a *Notch* intron [277]. In addition to their insertional mutagenesis, *cut-and-paste* TEs can also be deleterious by an imprecise excision from their original place [278]. One example is the imprecise excision of the DNA transposon *Tc1* in *Caenorhabditis elegans* from the *unc-54* gene, encoding for the myosin heavy chain B (MHC B), which creates a new 5' donor splice site [279]. In addition, the AT-rich base composition of some TEs [135] could also impact the surrounding genes [134]. For example, insertions of *L1* or *Alu* sequences within genes can introduce new polyadenylation signals [280].

TEs can also alter gene expression when their insertions disrupt or originate new gene regulatory regions, such as promoters, enhancers or repressors [275]. In *Drosophila*, an insertion of the DNA transposon *HMS-Beagle* within the TATA box region of the cuticle protein 3 (CP3) gene is responsible for the gene inactivation [281]. TEs can also contain binding sites for transcription factors (TFs). For example, around 20% of the transcription factor binding sites (TFBSs) in two types of humans and mouse cells are located within transposable elements, mainly *LTR* elements [282]. Furthermore, TE insertions near genes could also influence their epigenetic regulation [283]. For instance, the DNA methylation of a non-autonomous Mutator–like DNA transposon (*MuLE1*) inserted in the *Dfr-B* gene promoter region, can spread to the nearby regions, repressing the gene expression and producing variegated flowers [284]. In *Drosophila*, the heterochromatin formation (H3K9me3) assisted by piRNA-mediated TE silencing can spread into flanking genomic regions, affecting the expression of nearby genes [285].

The transposition process can also produce Double Strand Breaks (DSBs), which can be repaired by Homologous Recombination (HR), Non-Homologous End Joining (NHEJ) or Non-Allelic Homologous Recombination (NAHR). These processes could generate genome rearrangements, such as inversions, translocations, deletions, or duplications [286]. In addition, the repetitive nature of the TEs and their sequences (such as TSDs) in the genomes, can also produce inter- and intrachromosomic rearrangements by ectopic recombination [278]. For example, Homologous and Non-Homologous recombination mediated by *Alu* elements produce deletions and insertions in the adenomatous polyposis coli (APC) tumor-suppressor gene, leading to the hereditary colorectal cancer syndrome familial adenomatous polyposis (FAP) [287]. The inversion *j* in the chromosome 2 of *D. buzzatii* has been produced by recombination between opposite-oriented copies of the *foldback* (*P*-superfamily) DNA transposon *Galileo* [288].

Nonetheless, other TE features can also impact the organisms as is the case of TE transcript intermediates that can trigger the innate immune response [138]. The TE machinery or the TE mobilization process can also move non-TE related sequences [289] in a process called DNA

transduction. This process is a relatively common phenomenon in the human genome by *L1* [289]. In addition, the deviation of the retrotransposon enzymatic machinery to non-TE transcribed DNA, could provoke a 'retroposition' of a transcribed gene inducing the formation of new copies with pseudogenes features that are usually not expressed [290]. Finally, they can also give rise to new repetitive DNA classes, such as microsatellites, minisatellites, and satellite DNA [286]. For example, non-autonomous *Non-LTR* retrotransposons *SINE* are known to originate the Satellite 1 in *Xenopus Laevis* [291].

Despite TEs often provoke negative impacts in the host genomes, they can occasionally be favorable for the organisms. For example, in the peppered moth (*Biston betularia*), the insertion of a DNA transposon, *carbonaria*, within an intron of the *cortex* gene increases its expression giving the black coloration of this organism during the industrial revolution (industrial melanism) [292]. In addition, other TEs have been described to produce tissue-specific gene expression in different organisms. For example, in humans, the expression of genes proximal to tissue-specific hypomethylated TE sequences was strongly correlated to TE hypomethylation [293]. Moreover, new exons derived from *Alu* sequences, that are usually alternatively spliced, can be included in the gene transcript in a tissue-specific manner [294]. TEs also contribute to the developmental transcriptome in *Drosophila* by giving alternative promoters to some genes, which can be differentially regulated according to developmental context [295].

The beneficial TE insertions can also be adopted by the host genomes to serve novel functions, in a process called domestication [296]. In *Drosophila,* chromosome terminal repeats consist of sets of telomere-specific *LTR* retrotransposons of three families, *HeT-A, TART* and *TAHRE*, which elongate the telomers using a targeted transposition to chromosome ends by their 3' oligo(A) tails (target-primed reverse transcription) [297]. In this same organism, retrotransposons are also involved in the formation of centromeres, and thus in the chromosome architecture [298]. TEs have also been proposed to be involved in the mammalian placenta origin and the diversification of placental tissues [299] by the expression regulation of genes located nearby an insertion of the *MER20* DNA transposon in the endometrium [299]. In addition, new genes can also evolve from TE proteins. For example, the telomerase reverse transcriptase (TERT), which maintains telomere length in replicating cells, has evolved from an ancient *L1* retrotransposon insertion [300]. The RAG recombinases, involved in the V(D)J recombination, which produce a highly diverse repertoire of immunoglobulins, have also been proposed to be evolved from the transposase genes of an ancient "RAG transposon" [301, 302].

Therefore, when a new TE insertion occurs, it is exposed to natural selection and genetic drift [303]. Most TE insertions are strongly deleterious and are rapidly eliminated from the population by purifying selection [138]. Insertions with little or no effects (neutral) in the organisms can be

lost or fixed, depending on the efficiency of selection and genetic drift to eliminate them from the population, which vary between species [138]. The new TE insertions can also accumulate mutations, losing their transposition ability, and then, not being able to propagate in the host genomes [303]. Nonetheless, some TE insertions are favorable, and can be positively selected in the host genomes, being fixed and/or domesticated [303], and thus driving genome evolution. Therefore, even though TEs have been described to be mainly deleterious for the organisms in many kinds of ways, there is no doubt that TEs also provide new genetic variation that can help the organisms to evolve.

1.2.4   Factors leading to TE activation

When we focus on TE expression, spontaneous TE transpositions have been detected in both prokaryotes [304] and eukaryotes [305], but its rate is usually low. However, sporadic transposition bursts, characterized by the movement of numerous TE sequences through the genome, during a short evolutionary time, could happen spontaneously [306, 307]. These transposition bursts are usually associated with stressful conditions, and different types of stress have been proposed to increase transposition rates [61]. For this reason, Barbara McClintock proposed that TEs could be activated by different challenges, altering the host genomes by creating new variation in the organisms, which might have an adaptive function that allows the host to survive these challenges and evolve [54].

One of the stresses leading to TE activation is the environmental stress, which includes abiotic (temperature, irradiation or chemicals) and biotic (infections by virus or pathogens) stress [61]. For example, heat stress and chemicals activate the *gypsy*-like retrotransposon *MAGGY* in a plant pathogenic fungus [308]. In addition, the viral injection of particles of the avian Rous Associated Virus type 2 (RAV-2) induce somatic transposition of the *LTR* retrotransposon *mdg1* in *Drosophila* [309]. The other stress leading to TE activation is the genomic stress, which includes any influence that may disrupt the stability of the genome [61]. For example, crosses between species [310–312] or between different strains of the same species [131, 313] have been described to increase TE transposition in many organisms. It is worth to mention the hybrid dysgenesis, in which intraspecific crosses of two different *Drosophila* strains with males carrying a specific TE and females lacking it, can show aberrant phenotypes, including high mutation rates, chromosomal rearrangements, male recombination, and sterility [314], caused by an increase of TE transposition. TE repression maintenance across *Drosophila* generations depends on the maternal deposition of piRNAs in the offspring (see section **1.2.2.1**). Therefore if the maternal individual lacks a specific TE family, it will be unable of transmitting the corresponding piRNAs to their progeny [172], and this TE will be activated. This phenomenon has been described in different *Drosophila* species and in a wide range of TE families, such as the *P* [131], *I* [132] and

hobo [315] elements in *D. melanogaster*, the *P* element in *D. simulans* [316], and *Polyphemus* [317], *Ulysses*, *Paris*, *Helena* and *Penelope* [95, 318] in *D. virilis* intraspecific crosses. Nonetheless, some "tolerance" or "resistance" to hybrid dysgenesis was described in *D. melanogaster* [319, 320] and *D. simulans* [316], but could not be related to maternal deposited piRNAs, indicating that other mechanisms could suppress this phenomena, such as different DNA repair gene efficiency [320], or paternally inherited copy number [321]. Moreover, other types of stress can also activate TE transposition. For example, TE activation during normal aging, which could induce age-associated diseases, has been observed in different species and has been related to a loss of TE silencing mechanisms by age [322]. In section **1.2.4.1** and **1.2.4.2** we will focus on the stresses used in this thesis: interspecific hybridization and heat stresses, respectively.

1.2.4.1   Interspecific Hybridization

Any influence that disrupts the genome stability can be considered genomic stress [61]. Interspecific hybridization is the genomic stress produced by the reproduction between parental individuals from different species, producing F_1 hybrids heterozygous for all *loci* except for the hemizygous sex chromosomes [323]. The F_1 hybrids can be backcrossed (BC) with one of the parental species, obtaining a BC_1 progeny with both homozygous and heterozygous *loci*, of which the homozygous *loci* come exclusively from the backcrossed parent, increasing the genomic contribution of this ancestor to the hybrid genome [324]. This incorporation of the alleles from one species into the hybrid results in an introgressed genotype [325], which can be repeated by multiple generations of backcrosses. In this way, hybridization generates novel allelic combinations in the offspring that can lead to intermediate genomic characteristics between parental species or similar to one of them [326], but as well to novel genomic characteristics not found in the parental populations, such as transgressive gene expression: over or underexpression relative to both parental species [327]. Those new characteristics can be involved in novel adaptations in fast developing or new environments [326, 328]. Nevertheless, the mixture of two distinct genomes and epigenomes can also lead to hybrid incompatibilities, which results in a defective hybrid offspring, usually sterile, and with genetic instabilities, or even to inviable offspring [324]. Therefore, hybrids have unique genetic backgrounds that impact their gene expression [329, 330], genomic instability and sterility [310, 329, 331], but also their transposition rates [314].

The most popular concept of species, the "biological species", is described as "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" [332, 333]. Therefore, for speciation it is critical to prevent gene flow between species, and this is achieved by reproductive isolation barriers [334]. These isolating barriers are traditionally classified as those acting before zygote formation or pre-fertilization, called pre-

zygotic barriers, such as geographical and ecological barriers or gametic incompatibility; and those acting after zygote formation or fertilization, called post-zygotic barriers, such as zygote mortality or hybrid incompatibilities, including hybrid inviability and sterility [334]. Hence, to be able to hybridize, species must be able to overcome these isolation barriers, which in the nature is a quite frequent phenomenon estimated to occur in around 25% of plant species and 10% of animal species [335]. In addition, natural interspecific hybridization has been related to the origin of new species in many genera, called hybrid speciation. This phenomenon has been extensively reported in plants. For example, three sunflower species of the genus *Helianthus* (*H. anomalus*, *H. deserticola* and *H. paradoxus*) have been originated by the hybridization of *H. annuus* and *H. petiolaris* [336]. Even though hybrid speciation is uncommon in animals, it has also been described. For example, hybridization between the butterfly species *H. melpomene* and *H. cydno* originated the *Heliconius heurippa* species, which has an intermediate wing color and pattern between its parental species [337]. In some cases, the new hybrids have been clearly associated with a shift to a new resource, which is the case of hybrids between two host-specific species of tephritid fruit flies that originate into a new host-specific species [338]. Speciation has also been proposed to be the origin of new species in some vertebrates, such as the Darwin's finch "Big Bird" linage by hybridization of *Geospiza fortis* and *G. conirostris* [339] or the dolphin *Stenella clymene* by hybridization of *S. coerueloalba* and *S. longisrostris* [340].

As introduced before, the fusion of two distinct genomes and epigenomes in the hybrid have different impacts on the progeny. Interspecific hybrids can present new phenotypic variation, including phenotypical changes compared to both parental species, such as in plant [341] and animal [342] hybrids; intermediate phenotypes between parental species, as in *Drosophila* [343] hybrids, or new phenotypic combinations with characteristics from each parental species, as observed in animal [344, 345] hybrids. Interspecific hybrids can also present fertility problems, as described in animals [342, 344], plants [346] and *Drosophila* [329, 347], and genomic instability and rearrangements [310, 331]. They show novel gene expression patterns, as described in plant [341], animal [348, 349] and *Drosophila* [329, 330] interspecific crosses, in some cases including expression changes of piRNA pathway genes [329, 348]. TE transposition activation have been also reported in interspecific hybrids of several organisms, such as plants [312], yeast [311], and animals [348], including mammals [310] and different *Drosophila* species, such as *D. melanogaster* and *D. simulans* [350], *D. arizonae* and *D. mojavensis* [330] and *D. buzzatii* and *D. koepferae* [24] and *D. virilis* and *D. littoralis* [351].

Even though the effect of hybridization on the hybrid epigenome has been less studied, an impact has been described as well. For example, new histone methylation patterns have been observed in animals [310, 342] and plant [341] hybrids, and changes in chromatin marks have been detected in plant hybrids [341]. Changes in small RNAs amounts have also been reported in

hybrids in comparison to parental species in plant [352] and animal [344, 345] hybrids, including of piRNA amounts in animal hybrids [348]. In *Drosophila* hybrids, changes in piRNA levels, in comparison to parental species, have been detected extensively in crosses between *D. melanogaster* and *D. simulans* [350], *D. melanogaster* and *D. mauritiana* [329], *D. arizonae* and *D. mojavensis* [330] and *D. buzzatii* and *D. koepferae* [24], and they have been related to TE activation in a similar way as hybrid dysgenesis.

D. buzzatii-D. koepferae hybrids

Hybridization has been extensively studied in *D. buzzatii* and *D. koepferae* species. The sterility of *D. buzzatii-D. koepferae* male hybrids depends on the size of the introgressed fragments in autosomes [347, 353], with a required minimum threshold of approximately 25–30% of an average autosome [30, 354] that affects different specific epistatic sterility factors [355]. Nevertheless, small fragment introgressions of the X chromosome also seem to lead to hybrid sterility [30, 353] and unviability [30, 356] independent of the fragment size. Therefore, a model of several genetic factors inducing sterility in *D. buzzatii-D. koepferae* hybrids, depending on the introgressed segment and starting at a size threshold [347, 357] was proposed: the polygenic threshold (PT) model [357]. In addition to sterility, *D. buzzatii-D. koepferae* hybrids and backcrosses display developmental distortions, such as death at the pupal stage or abdominal region abnormalities [358]. They also show asynapsis [347] and high frequencies of new chromosomal rearrangements [331]. This last finding suggested a putative role of interspecific hybridization on TE mobilization, inducing new rearrangements in the progeny [359]. This idea was first supported by the increase of transposition of *Osvaldo* retrotransposon in *D. buzzatii-D. koepferae* hybrids [360, 361]. Later a genome-wide study, using Amplified Fragment Length Polymorphism (AFLP) markers, showed that hybridization increased genomic instability by an increase of transposition rates of *Osvaldo, Helena* and *Galileo* [362]. Hence, a significant and global effect of hybridization on TE instability was found, whereby the transposition bursts depended on the individual (hybrid) and the TE family. TE mobilization was also observed to impact the genome sizes of the hybrid females, producing an expansion that was not observed in hybrid males [363]. These results lead to the hypothesis that hybridization could contribute to a relaxation of the mechanisms controlling TEs, such as the piRNA pathway in the germline. A study of the retrotransposon *Helena* in *D. buzzatti-D. koepferae* hybrids revealed that its expression was not altered in hybrid somatic tissues, but was repressed in the germline by an increase of its piRNA levels in F_1 testes [364]. In F_1 ovaries, *Helena* was shown to be additively expressed between parental species as well as the associated piRNA levels, but they showed a Ping-Pong signal deficiency [364]. Lately, a genome-wide study analyzed the impact of hybridization in TE expression and piRNA amounts in the *D. buzzatii-D. koepferae* hybrids

germline. The authors found that divergent sequences and production of piRNA pathway proteins, and differences in piRNA pools between parental species could only partially explain the observed instabilities and TE activation in F_1 hybrid ovaries and BC_1 backcrosses [24]. However, the results observed in F_1 ovaries were not the same than in F_1 testes, showing that hybridization has a sex-biased impact, and that an increase in piRNA production in F_1 testes could explain the detected TE underexpression [24]. In another study, differences in the piRNA pathway gene expression in ovaries of the same F_1 hybrids were also detected, reinforcing the hypothesis of an involvement of the piRNA pathway in TE activation in *D. buzzati-D. koepferae* hybrid females [365]. However, all these studies together concluded that other additional mechanisms could also be involved in the ovary TE activation observed in these F_1 hybrids.

1.2.4.2 Heat shock stress

In the last years, global warming has been an important issue that affects the climate, rises the global temperature and makes extreme heat waves more frequent [366]. These changes promote heat-associated illnesses in humans, such as cardiovascular, pulmonary, psychiatric, and neurological diseases, among others [367], and aggravate pathogenic diseases [368]. They also have an impact in other organisms, such as those used for food production, for example in agriculture [369], livestock [370] and fishes [371]. Therefore, understanding the impact of high temperatures [367] and heat waves [49] on the organisms, their adaptation to the new environment [51] and the limits of their adaptation, is a key to understand our future.

After a heat stress, the cells activate the Heat Shock Response (HSR) to protect the organism from this stress. This response is triggered by a small increase of temperature, even in organisms living at extreme temperatures [372], or after exposure to other stresses, such as anoxia [373], oxidative stress [374], ischemia [375] and chemicals of several classes [376]. The HSR is universal [377]: it is found in both prokaryotes and eucaryotes, and in practically all tissues. This phenomenon was discovered in the polytene chromosomes of *Drosophila*, where just few minutes after a heat shock a new puffing pattern was detected [378]. These puffs were associated with high rates of RNA synthesis, and often coincided with a regression of the puffs present before the stress [379, 380]. The activated puffs were related to the synthesis of new proteins [381], called Heat Shock Proteins (Hsps) [382], and the time of regression of these heat shock puffs and the re-activation of the old puffs depended on the intensity of the stress, which allows an estimation of the damaging effects of heat in the organism [383]. Regarding what is known about the HSR in *D. subobscura*, the heat stress also increases the activity of certain puffs related to the synthesis of Hsp proteins and decreases the puffs observed prior to the stress [384]. The same puffs were also activated after heat stress in different strains (independently of the genetic background) [380] and during anoxia [373], showing the universality of this response. These puffs, corresponding to

the main Hsps, were also activated after heat stress, with some differences, in the closely related species *D. guanche* and *D. madeirensis* [385, 386]. However, differences in puff activation after a heat stress depend on the time of development and the temperature of stress [380], being this latter optimal at 31-34ºC in *D. subobscura* [380, 385, 386].

Molecular mechanism of the Heat Shock Response activation

The molecular mechanism leading to this response has also been deeply studied. Under normal conditions, the Heat Shock Factor (Hsf), which binds Hsp promoters, is repressed in its inactive form in monomers [387] by constitutively expressed Hsp [388], specifically in a multichaperone complex formed by Hsp90 [389, 390], Hsc70/Hsp70 [388, 391], in synergy with the co-chaperon DroJ1 (Hsp40 in other organisms) [388]. In *Drosophila,* there is only one Hsf, called Hsf1 [392], which has different alternatively spliced isoforms with different transcriptional activities [393], and its inactive form is generally located in the nucleus in non-stressful conditions [394]. After a stress, there is an accumulation of unfolded proteins that activates the HSR [395] where Hsps bind to the unfolded proteins, leading to the liberation and de-repression of the pre-existent Hsf1 [388]. Free monomeric Hsf1 trimerizes [387] and binds with high affinity to the Heat Sock Elements (HSEs) [392] that are consensus sequences formed by alternating and multiple repetitions of nGAAn or nTTCn, where n is any nucleotide [396], located at the 5' regulatory region upstream of the TATA box of the Hsp and other heat-shock induced genes in eukaryotes [397]. The number of repetitions of the HSE vary by Hsp and species, but often contain at least 3 continuous repetitions [398]. In non-stressful conditions, in most Hsp and heat shock induced genes of *Drosophila* [397], the Pol II initiates transcription, but pauses elongation proximal to the promoter, a few nucleotides downstream of the transcription start site (TSS) [399] and this process is followed by the incipient mRNA 5' cap formation [399]. In addition, the recruitment and stabilization of Pol II also needs other factors, such as the transcription factor GAGA-associated factor (GAF), which binds to GA present in the promoter and keeps the promoter region open and free of nucleosomes [400]. The release of Pol II to transcribe the heat shock induced genes is at a low level during normal non-stress conditions [399]. Nonetheless, after a heat stress, the trimerized Hsf1s already bound to HSEs is hyper-phosphorylated [401], and recruits quickly, directly and indirectly, coactivators and other factors affecting the chromatin structure and composition [402]. This recruitment promotes the release of Pol II, and therefore the transcript elongation [397], producing elevated levels of transcripts of Hsps and heat shock induced genes [388]. When the available Hsps increases, free Hsp70 in synergy with DroJ1 form a complex with Hsf1 [388], interfering with the integration of the latter with the transcription machinery and inhibiting the HSE-containing gene expression in a feedback loop [403]. During recovery, Hsf1 interacts with constitutively expressed Hsp, and acquires a monomeric inactive form, becoming

repressed [388]. This feedback loop avoids the accumulation of Hsp after stress, which has been described to be detrimental for cells at normal temperature [404]. In *Drosophila*, heat stress-induced genes in a Hsf1-independent manner have also been detected [405], but they have a lower heat shock induction in comparison to the Hsf1-dependent ones [397]. In this organism, it has also been proposed that Hsf1 regulates other genes, besides the known heat shock genes, during non-stress conditions [405]. Finally, the heat shock dependent repression of highly expressed genes that occurs after a heat stress has been described to be Hsf1-independent in *Drosophila*, and it seems to be regulated during the transcription initiation and Pol II recruitment step, being the promoter-proximally paused Pol II reduced [397]. Nonetheless, even though the heat shock response has been extensively studied, there are still processes that are unknown, such as how the gene repression works.

Heat shock protein families in *Drosophila*

Hsps, also known as chaperones, are involved in many functions, in normal conditions and under stress, such as folding newly synthesized proteins, re-folding or degrading denaturalized proteins by stress, dissociating toxic protein aggregates, transporting proteins into subcellular compartments, preserving and restructuring the cytoskeleton, assembling multiprotein complexes and enzymes and regulating cell-cycle, signaling pathways and receptors [406]. Under stressful conditions, Hsps proteins protect the organisms from toxic effects, such as protein unfolding and loss of functional conformation and protein unspecific aggregations [407, 408], providing the organisms a temporary tolerance to the stress [377]. These proteins are subdivided into families, named according to their relative molecular weight in kilodaltons and their sequence conservation, and they are conserved among living organisms [409]. In *Drosophila*, the major heat shock protein families consist of: Hsp70, Hsp90, Hsp100, Hsp60, small Hsps, Hsp40 and other co-chaperones and chaperonins [406]. The *Hsp70* gene family contains ATP-dependent chaperones [406], from which the most important are Hsp70 and the heat shock cognate (Hsc) proteins. *Hsp70* is low expressed at normal temperatures but highly heat-induced [404], and Hsc genes, highly homologous to *Hsp70*, are constitutively expressed and, in general, not heat-induced [410]. Therefore, this family is critical for protein folding in both non-stress (newly synthesized) and stress (refolding) conditions. They perform their functions with the support of a co-chaperone family called atypical Heat Shock Proteins 70, and in synergy with Hsp40/DnaJ-type cochaperones [411]. The Hsp40 family, is a family of co-chaperones formed by many diverse proteins [406] which contain the J-domain domain [412]. They assist other Hsps, mainly Hsp70s, to perform their functions, for example by stimulating the ATPase activity of chaperone proteins or stabilizing their interaction with substrate proteins [412]. Some of them, such as Droj1, are constitutively expressed but also heat-inducible [388]. The Hsp90 family includes the essential

protein Hsp83, which is constitutively expressed at high levels but is also heat-inducible [413] and is involved in many cellular processes beyond ATP-dependent protein folding, such as the piRNA pathway [414] and epigenetic modifications [415].

In *Drosophila,* the only one member of the Hsp100 family detected was ClpX. This protein is an ATP-dependent chaperone that folds protein in the mitochondria [416]. The Hsp60 family (also called chaperonins) contains two groups. The first group catalyzes ATP-dependent protein folding, also in the mitochondria, and is supported by the heat-inducible co-factor family called heat shock protein 10 chaperonins [417]. The second group contains paralogous subunits that assemble in a multi-subunit ring complex, called the TCP-1 Ring Complex (TRiC) or the Chaperonin Containing TCP-1 (CCT) complex [418]. They assists proteins to perform ATP-dependent folding of newly synthesized proteins in the cytosol [419]. Finally, the small Hsp (sHsps) family includes ATP-independent molecular chaperones small in molecular weight [420] and located at different cellular compartments [421]. They are among the most upregulated Hsps after stress, but they also have tissue and developmental time-specific constitutive expression [422]. Most of them are clustered in a 12 Kb section of the 67B region on the 3L of *D. melanogaster* [423], but some of them are located in other chromosomal regions [422]. They prevent heat-induced protein aggregations and can maintain proteins in a re-foldable state [421]. In addition to the above-mentioned proteins, there are additional proteins and cofactors that help Hsps to perform their functions and most of them are upregulated after heat stress. This group includes proteins such as: *stv*, a Hsp70 co-chaperone during stress recovery [424] and the Hsp70/Hsp90 organizing protein homolog Hop, which enable Hsp90 protein binding activity [414, 425].

Regarding the Hsps in *D. subobscura,* two members (Hsp70 and Hsp68) of the Hsp70 gene family were detected on the O chromosome and both have two copies in this species [426]. They are the Hsps with the highest heat-inducible expression [385, 386]. Nevertheless, some low basal protein levels of Hsp70 have been described in this species [427]. Studies about the influence of *D. subobscura* inversions on the Hsp70 gene expression are inconclusive. Basal Hsp70 expression has been described to be comparable between *D. subobscura* populations with two different arrangements on the O chromosome, being higher in females than in males [428]. On the contrary, different basal Hsp70 protein levels were detected between populations with other two O arrangements [427], and could be explained by differences in gene regulatory elements [428], but the expression differences disappeared after heat shock [427]. Other Hsps were also detected in this species, such as a single copy of the Hsp83 gene, located on the J chromosome [386, 429], which is constitutively expressed in particular developmental stages and is also heat inducible at low heat temperatures (24ºC), with a maximum induction and synthesis at 31-34ºC [430]. However, Hsp83 induction decreased at 37ºC, temperature at which this protein

is not produced [430]. Finally, four genes (Hsp23, Hsp28, Hsp22 and Hsp26) of the small Hsp family were detected on the J chromosome [386, 429]. These genes seem to be highly conserved between species, but with some evolutionary changes regarding their duplications [386]. Finally, other not annotated puffs were also induced in the sexual chromosome A during heat stress [386].

Effect of heat stress in *Drosophila* and other organisms

Even though the activation of Hsp genes and genes related to the stress response is universal in all organisms, other genes have also been described to be activated and repressed after heat stress. These genes are involved in different functions, such as transcription regulation [431], transport and metabolism [431, 432], signaling [432, 433], apoptosis [432], and cell cycle [431, 434], among others. Moreover, these additional deregulated genes seem to depend on different variables, such as the severity [431] and duration of the heat stress [435], the time of recovery [432], the stress resistance [349], the organism [349, 431–433], the sex [436], the tissue [434, 435], and even the biological replicate [437]. Heat stress has also been described to affect the body energy reserves by decreasing body fat stores [438], the reproductive system by inducing sterility [439], and the memory [440], the development and the immune system [441]. Heat stress disturbs the concentrations of free radicals, through an increase of Reactive Oxygen Species (ROS) production, which result in cellular and mitochondrial oxidative damage though oxidative stress [442]. In addition, it interrupts RNA splicing [443], so major Hsp genes are usually devoid of introns, being an exception the Hsp83 gene in *Drosophila* [443].

The heat stress has also been associated with an increase of transposition of some specific TE families in different organisms, such as plants [444, 445], fungi [446], nematodes [447] and human cells [448]. This stress also impacts the production of miRNAs and siRNAs in plants [449], and in some cases siRNA associated to TEs [450]. An impact of this stress on the small RNA expression of other organisms [451–453] has also been detected. Heat stress also impacts the epigenome by affecting chromatin marks in animals [454] and plants [445], and by decreasing substantially the genome-wide DNA methylation in plants [455], including the TEs. In some cases, heat shock-induced epigenomic changes can be inherited by the progeny, leading to an inherited thermotolerance memory. Indeed, inherited changes in histone marks [456, 457] and in small RNAs [458, 459] in some plants and animals, and in DNA methylation in this latter [457] were reported.

When we focus on *Drosophila*, results are ambiguous: heat stress increases the expression of specific TE families [460, 461], but a lack of mobilization was also found in others [462, 463]. Changes in the expression of specific small RNAs were also detected, such as changes in miRNAs [464, 465] and siRNAs [466, 467]. This stress also affected the expression of specific piRNAs

[464, 468], the production of specific piRNA clusters [468] and the Ping-Pong signal [464]. Nevertheless, no changes on the piRNAs levels [192] or on the Ping pong signal [468] were found in other studies. In any case, in none of the studies the changes in piRNA levels observed could be clearly correlated to changes in TE expression [192, 464, 468]. Therefore, it was suggested that Hsp70, present at high levels after heat stress, interact with the Hsc70-Hsp90 complex, involved in the Ago3 loading step of the piRNA pathway in *Drosophila* [192]. This interaction was proposed to increase the TE expression by displacing this Ago3 complex, critical for the piRNA-mediated TE repression, from the 'nuage' to the lysosomes, for its degradation [192]. In addition, changes in chromatin marks after a heat stress were also found in *Drosophila* tissue cells [469], but in the germline the H3K9me3 and H3K27me3 levels did not significantly changed after a heat shock in seven studied TEs [192]. Nonetheless, a heat shock stress in this organism during early embryogenesis also disrupted the heterochromatin formation, and produced changes that were inherited by the next generation [470]. All together the results in *Drosophila* show that the effect of heat stress in the genome is still unclear, and more research should be performed.

2.   Objectives

2.1   General objectives

In the 80s, Barbara McClintock already proposed that different stresses could lead to TE activation. Nowadays, many studies support this hypothesis, and different stresses have been described to activate TE expression in many different organisms. Nonetheless, the magnitude of expression changes is not fully known and the mechanisms leading to TE deregulation are usually hard to decipher due to the complexity of the TE regulatory mechanisms. The stress is known to affect gene expression in many ways, including epigenomic modifications. However, the effect of stress in gene and TE expression can also depend on the genomic background, making difficult to extract general conclusions about how stress affects the organism genomic expression.

Differences in gene expression and increases in transposition rates, have been reported in interspecific F_1 *D. buzzatii-D. koepferae* hybrids. A previous study detected a differential TE expression in the hybrid ovaries based on incompatibilities in the piRNA pathway between parental species: an accumulated divergence in piRNA pathway gene expression, and different piRNA pools with higher piRNA levels in *D. koepferae* and a stronger Ping-Pong signal in *D. buzzatii*. Nonetheless, these incompatibilities could only partially explain the differential TE expression observed in F_1 hybrid ovaries, suggesting that the divergence in piRNA pathway proteins, involved in chromatin remodeling, could also influence TE deregulation. In addition, the specific molecular mechanisms leading to the observed gene expression in *D. buzzatii-D. koepferae* hybrids were still unstudied.

On the other hand, *D. subobscura* species is known by having chromosomal inversion latitudinal clines, suggested to be associated to temperature. We ignore the general effects of the heat stress on the transcriptome of this species, including both gene and TE expression, and specifically in the germline. Previous molecular studies on the Hsp70 (located on the O chromosome) expression do not reach a consensus: changes in protein levels were observed between individuals bearing the same O arrangement, but in another study identical mRNA expression was observed in individuals bearing different O arrangements. These results, together with the fact that the effect of heat stress on TE expression and piRNA amounts does not appear to offer conclusive results in other *Drosophila* species, suggest that further genome-wide studies are required to assess this issue.

Therefore, the main goal of this thesis is to gain insights into the molecular mechanisms leading to TE differential expression under stress, and its effect in gene expression. First, we focused on the inheritance of the *D. buzzatii* and *D. koepferae* epigenome in the F_1 hybrid ovaries,

and its influence on their gene and TE expression. We also studied additional mechanisms, such as divergence between parental species, involved in the observed hybrid gene expression. Second, we focused on the effect of heat stress on the gene and TE expression of two *D. subobscura* populations (Madeira and Curicó) bearing the same O rearrangement. Subsequently, we focused on the influence of heat stress on the small RNA amounts (siRNAs and piRNAs) associated to TEs, together with the piRNA cluster production. Finally, we assessed the impact of the observed small RNA changes in the TE expression under heat stress.

2.2 Specific objectives

The specific objectives of this thesis are described below:

- To assess the effect of *D. buzzatii* and *D. koepferae* interspecific hybridization on the whole transcriptome and epigenome of F_1 hybrid ovaries (**section 3.1**), focusing on:
  - o Describing the chromatin landscape of H3K4me3, H3K9me3 and H3K27me3 around the genes in *D. buzzatii, D. koepferae* and F_1 hybrid ovaries.
  - o Characterizing the contribution of the chromatin landscape to the gene and TE expression in *D. buzzatii, D. koepferae* and F_1 hybrid ovaries.
  - o Analyzing gene and TE expression in F_1 hybrid ovaries and the mechanisms leading to their inheritance, by:
    - Studying gene and TE expression in hybrids *vs* parental species.
    - Assessing hybrid chromatin enrichment differences in genes and TEs in comparison to the parental species.
    - Evaluating the effect of chromatin enrichment changes in hybrids *vs* parental species on the detected gene and TE expression changes.
    - Analyzing the effect of the regulatory divergence between parental species on gene expression inheritance in hybrids.
    - Assessing the effect of the piRNA amounts on TE expression inheritance in hybrids.
- To assess the effect of heat stress on the whole transcriptome of two *D. subobscura* populations (**section 3.2**), focusing on:
  - o Characterizing the differences in gene and TE expression between two different *D. subobscura* populations.
  - o Annotating *de novo* the TE landscape of this species.
  - o Assessing the production of small RNAs in ovaries and testes of *D. subobscura*.
  - o Detecting *de novo* the piRNA clusters in this species.
  - o Analyzing how heat stress affects gene expression in *D. subobscura*, by:

- Studying gene expression differences after heat stress and its effect on heat-stress responsive genes and piRNA pathway genes.
- Assessing the similarities of the heat shock response between sexes and populations of *D. subobscura*.
- Analyzing how heat stress affects TE expression in *D. subobscura* and the mechanisms leading to this effect, by:
  - Determining TE expression changes after a heat stress.
  - Analyzing the effect of heat stress on the piRNA and siRNA amounts targeting TE families and on the production of piRNA by clusters.
  - Determining if changes in piRNA and siRNA amounts are related to TE expression changes after heat stress.

3.   Results

The results of this Thesis have been divided in two chapters corresponding to two different publications, here classified in two sections. The **3.1** section comprises a study published in an open access peer-reviewed journal (published version in Annex **8.5**), while the **3.2** section study is currently under journal revision.

3.1 High Stability of the Epigenome in *Drosophila* Interspecific Hybrids

Genome Biology and Evolution, Volume 14, Issue 2, February 2022

https://doi.org/10.1093/gbe/evac024

This section is a verbatim reproduction from the following published paper:

Bodelón A, Fablet M, Veber P, Vieira C, García Guerreiro MP. High Stability of the Epigenome in Drosophila Interspecific Hybrids. Genome Biol Evol. 2022 Feb 4;14(2):evac024. doi: 10.1093/gbe/evac024. PMID: 35143649; PMCID: PMC8872975.

High stability of the epigenome in *Drosophila* interspecific hybrids

Alejandra Bodelón[1], Marie Fablet[2,3], Philippe Veber[2], Cristina Vieira[2] & Maria Pilar García Guerreiro[1*]

[1] Grup de Genòmica, Bioinformática i Biologia Evolutiva. Departament de Genètica i Microbiologia (Edifici C). Universitat Autònoma de Barcelona. 08193 Bellaterra, Barcelona, Spain

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

[3] Institut Universitaire de France

[*] Author for Correspondence: mariapilar.garcia.guerreiro@uab.es

Abstract

Interspecific hybridization is often seen as a genomic stress that may lead to new gene expression patterns and deregulation of transposable elements (TEs). The understanding of expression changes in hybrids compared to parental species is essential to disentangle their putative role in speciation processes. However, to date we ignore the detailed mechanisms involved in genomic deregulation in hybrids. We studied the ovarian transcriptome and epigenome of the *Drosophila buzzatii* and *Drosophila koepferae* species together with their F_1 hybrid females. We found a trend towards underexpression of genes and TE families in hybrids. The epigenome in hybrids was highly similar to the parental epigenomes and showed intermediate histone enrichments between parental species in most cases. Differential gene expression in hybrids was often associated only with changes in H3K4me3 enrichments, while differential TE family expression in hybrids may be associated with changes in H3K4me3, H3K9me3, or H3K27me3 enrichments. We identified specific genes and TE families, which their differential expression in comparison to the parental species was explained by their differential chromatin mark combination enrichment. Finally, *cis-trans* compensatory regulation could also contribute in some way to the hybrid deregulation. This work provides the first study of histone content in *Drosophila* interspecific hybrids and their effect on gene and TE expression deregulation.

Keywords: Epigenome, *Drosophila*, interspecific hybrids, deregulation, transposable elements, histone methylation

Significance Statement

The genomic stress caused by hybridization between different species is a cause of gene and transposable element deregulation. Histone modifications were often associated to genomic deregulation, but little is known to date about their role in *Drosophila* hybrid anomalies, and their elucidation is essential to unravel the role of these phenomena in speciation. Using transcriptomic and epigenomic analyses, we found that even though the epigenome of the parental species is highly conserved in hybrids, some changes detected in histone marks were associated to gene and TE deregulation. This work provides the first study of histone modification in *Drosophila* hybrids and its putative impact on gene a TE expression.

3.1.1   Introduction

Interspecific hybridizations are a source of phenotypic variation [341, 343], sterility [347, 471, 472] and genomic instability [347, 362, 473] in hybrid offspring. These phenomena are induced by the fusion of two different parental genomes and epigenomes, which act as a genomic stressful factor in the hybrid genome [61]. This integration leads to deregulation of gene expression in

hybrids, which has been described in *Drosophila* [24, 365, 471, 472, 474] and other organisms [341, 475]. For example, the misexpression of genes involved in spermatogenesis in *Drosophila* male hybrids [471, 472] has been related with their sterility. Deregulation of gene expression was also observed in hybrid females of *D. buzzatii* and *D. koepferae* [24, 365] and in other *Drosophila* hybrids [474]. Gene expression in *Drosophila* is controlled at different levels, including epigenetic modifications [149] and *cis-* and *trans*-regulatory elements, among others. Divergence between parental regulatory elements may play an important role in gene expression deregulation in hybrids [476]. For example, differences in the *cis* and *trans* parental regulatory elements were related with misexpression in *Drosophila* hybrids [477].

In addition to gene expression disruption, a growing number of evidence in *Drosophila* suggest that transposable elements (TEs) could also play an important role in the hybrid anomalies. For example, increases of transposition rates and/or misexpression were observed in interspecific hybrids between *D. buzzatii* and *D. koepferae* [361, 362, 364, 478] and other *Drosophila* species [330, 350]. In *Drosophila* germline transposition is controlled at two levels: transcriptional silencing, by heterochromatinization processes, and post-transcriptional by piwi-interacting RNAs (piRNAs) involved in transposon transcript degradation. In this sense, the causes of TE release in hybrids have been the matter of recent research focused on the study of disruption of the germline mechanisms involved in TE regulation, basically those affecting piRNA production [24, 330, 350]. Recent discoveries pointed to a dysfunction of piRNA pathway in *D. melanogaster-D. simulans* hybrids [350], which would have a deficient global piRNA production, whereas a similar or higher piRNA production was observed in *D. buzzatii-D. koepferae* ovaries [24]. Moreover, in the later study a high proportion of the TEs overexpressed in hybrids did not have associated piRNAs in parents or hybrids, which suggests a more complex deregulation network, at least, in these hybrids.

Less is known about the behaviour of epigenomes in interspecific *Drosophila* hybrids, which is of fundamental interest in evolutionary biology. Histone modification is a post-translational chemical modification of histone proteins. In *Drosophila,* the histone marks H3K4me3 are associated to activation of gene expression, whereas H3K9me3 and H3K27me3 are associated to gene repression [149]. H3K4me3 marks are abundant in euchromatin, and H3K9me3 and H3K27me3 take part of constitutive and facultative heterochromatin respectively [479, 480]. Moreover, it was also shown that TEs are enriched in the H3K9me3 and H3K27me3 heterochromatic marks, suggesting a putative co-localization of these silencing marks in these sequences [149]. In plants, histone modification occurring during interspecific hybridization has been associated to gene expression variation [341, 481]. However, in *Drosophila*, little is known about the role of histone modification on gene expression and their consequences on hybrid

incompatibility and TE deregulation. Studies on *D. melanogaster-D. simulans* hybrids suggest that adaptive divergence of heterochromatin proteins is an important force driving the evolution of genes involved in hybrid incompatibility [482]. Our previous work on *D. buzzatii-D. koepferae* hybrids suggested the existence of interacting phenomena, including incompatibilities of the piRNA pathway, due to a functional divergence between parental species, as one of the causes responsible of TE mobilization (Romero-Soriano et al. 2017; review Fontdevila 2019). However, we cannot disregard other putative mechanisms in these species, including histone modifications causing changes in TE expression.

To unravel the causes of this hybrid incompatibility, we investigated the expression profiles and the patterns of three histone marks: H3K4me3, H3K9me3 and H3K27me3, using RNA-sequencing (RNAseq) and chromatin immunoprecipitation and deep sequencing (ChIPseq) in the genome of the parental species (*D. buzzatii* and *D. koepferae*) and their hybrids. *D. buzzatii* and *D. koepferae* are two cactophilic sibling species, belonging to *repleta* group [483], that diverged approximately 4-5 million years ago [6, 24, 484]. Crosses between *D. buzzatii* males and *D. koepferae* females result in F_1 sterile males and fertile females [29], even though a few cases of partial fertility with atrophy in one of the ovaries is also observed [28]. Hybrid females can be backcrossed with *D. buzzatii* males [28, 485]. In this system, we can only obtain hybrids from crosses between *D. koepferae* females and *D. buzzatii* males, the reciprocal cross does not produce offspring [29].

We found that both genes and TE families detected as differentially expressed in hybrid ovaries in comparison to *D. buzzatii* and *D. koepferae,* tended to be mostly underexpressed. In contrast, we found a high conservation of the parental chromatin mark patterns in hybrid genes and TEs, showing intermediate levels between the parental species. Nevertheless, we could associate some changes in gene and TE family expression in hybrids with their corresponding histone modifications *vs* parental species.

3.1.2   Results

Deregulation of gene and TE expression in hybrids

We analyzed and compared the ovarian transcript amounts of *D. buzzatii* and *D. koepferae* and their F_1 female hybrids. We found that, out of a total of 13,621 protein coding genes and 658 TE families, 5.92% and 29.64%, respectively, were differentially expressed between the parental species (Supplementary file 1: Table S1). Similarly, hybrids also showed a lower percentage of differentially expressed genes, in comparison to the parental species (4.57% *vs D. buzzatii* and 3.99% *vs D. koepferae,* Figure 1A), than TE families (22.95% *vs D. buzzatii* and 24.16% *vs D.*

koepferae, Figure 1B). Gene expression in hybrids was more similar to the maternal species *D. koepferae* than to the paternal species *D. buzzatii* (*Z-test, p* = 0.018, Supplementary file 1: Table S2), but the number of differentially expressed TE families in hybrids was similar in comparison to both parental species (*Z-test, p* = 0.603, Supplementary file 1: Table S2).

Fig.1 Differential expression analyses of genes and TEs: *(A)* Number of differentially expressed genes and *(B)* TE families in hybrids *vs* parental species of the total of 13,621 genes and 658 TE families. Colors indicate gene expression categories. *(C)* Expression categories in hybrids *vs* parental species. *(D)* Differential gene expression analyses in hybrids *vs D. buzzatii* (green) and *D. koepferae* (blue). Positive log2FC values correspond to genes more expressed in hybrids. The genes showing the 20 highest log2FC values and displaying an ortholog in *D. melanogaster* are shown. Genes common to both comparisons are in bold. *(E)* Violin plots representing the distribution of differentially expressed orders of TEs in hybrids *vs* parental species. Points indicate the log2FC of each TE family and for each comparison. The percentages of differentially expressed TE families per order are framed. The expected (total) TE family percentages are at the bottom. Red-dashed lines indicate the log2FC threshold for significancy (±1).

Globally, we could classify the differentially expressed genes and TEs in hybrids in three categories (Figure 1C): i) *D. koepferae-* or *D. buzzatii*-like expression, which corresponded to the majority of differentially expressed genes (Figure 1A) and TE families (Figure 1B). ii) Overexpressed or underexpressed in hybrids in comparison to both parental species, which were considered as deregulated in hybrids (Figure 1C). We observed a deregulation trend towards underexpression in hybrids compared to both parental species: 145 genes and 33 TE families underexpressed *vs* 29 genes and 12 TE families overexpressed (Figure 1A and 1B). iii) Additive expression which included differentially expressed genes and TE families in hybrids with an intermediate expression between both parental species. Only a small number of genes (29 and 40 genes) were included in this category (Figure 1A). Regarding the TEs, the additive expressed TE families were mostly underexpressed in comparison to *D. buzzatii,* and overexpressed in comparison to *D. koepferae* (Figure 1B). Both results were consistent with the differences between the parental species, where the mean TE family expression was higher in *D. buzzatii* than in *D. koepferae* (mean log2FC= 0.27 or 1.21-fold increase in *D. buzzatii vs D. koepferae*), and opposite for the genes (mean log2FC= -0.23 or 0.85-fold decrease in *D. buzzatii vs D. koepferae*; Supplementary file 1: Table S1).

We found that, even though there was a general bias towards underexpression of genes (Figure 1D) and TE families (Figure 1E) in hybrids, those detected as overexpressed exhibit the highest differences of expression compared to parental species. In fact, the values of mean log2FC were 2.53-3.01 (FC=5.78-8.06) for genes and 3.33-3.56 (FC=10.06-11.79) for TE families overexpressed in hybrids *vs D. koepferae* and *D. buzzatii* respectively, whereas they decreased to -1.62 -1.72 (FC=0.33-0.30) for genes and to -1.85 -1.74 (FC=0.28-0.30) for TE families underexpressed in hybrids vs *D. koepferae* and *D. buzzatii*, respectively. Regarding the genes, most of the overexpressed shown in Figure 1D, were involved in metabolism (small molecules or protein), development, signalling and stimulus response, reproduction and transportation, and cell organization (Supplementary file 2).

An interesting example was the *no hitter gene* (nht), highlighted in bold in Figure 1D, which was overexpressed in hybrids compared to both parental species. This gene is involved in spermatogenesis and regulation of gene expression and may be related to female hybrid sterility [474]. Moreover, as expected, most of the genes with higher differences between hybrids and the parental species (gene names labelled in Figure 1D) were at the same time detected as the most different between the parental species (green and blue gene names in Supplementary file 3: Figure S1a).

Regarding underexpressed genes, the Gene Ontology (GO) analysis revealed a shared enrichment for 11 terms in hybrids *vs* both parental species (green GO in Supplementary file 2)

and were mainly related to developmental processes, cell adhesion and reproduction (i.e. gonad development). Furthermore, the fact that most of them had functions related with reproduction, reinforces the idea of a putative role in hybrid fertility loss [472]. In the case of the overexpressed genes in hybrids, only two enriched GO terms were shared with both parental species (green GO in Supplementary file 2), and were linked to metabolic and cellular process. Additional GO terms, related to the same or other processes (such as response to stimulus and biological regulation, among others) were enriched in the differentially expressed genes, but only in comparison to one of the parental species (Supplementary file 2).

To detect a putative location-bias of the differentially expressed genes in hybrids in comparison to the parental species, previously reported in some *Drosophila* hybrids [472, 486], we studied their distribution per chromosome. We found that the number of differentially expressed genes across chromosomes in hybrids compared to both parental species (Supplementary file 3: Figure S2a), was different from the random expectation (*Chi-squared test, p < 0.001*, Supplementary file 1: Table S3). This result is due to a higher number of differentially expressed genes on the dot chromosome 6 (1.93% *vs D. buzzatii* and 2.21% *vs D. koepferae*) than the expected (0.54%). When this chromosome was excluded from the test, the differences become non-significant (*Chi-squared test, D. buzzatii: p = 0.320* and *D. koepferae: p = 0.443*, Supplementary file 1: Table S3).

Concerning the TEs, we studied the distribution and expression of the differentially expressed TE families, per order, in hybrids in comparison to the parental species. We used the Repbase classification [487] where TEs were divided in *LTR*, *LINE* and *DNA* orders. Due to their particular replication mechanism, we considered the *RC/Helitron* as a group apart from the *DNA* order. As shown in Figure 1E, the number of differentially expressed TE families by order was similar to the random expectation considering their relative proportions in the genome (*Chi-squared test, p = 0.183* for *D. buzzatii* and *D. koepferae,* Supplementary file 1: Table S3). The most extreme underexpression values were observed in *LINE* and *LTR* families in hybrids *vs D. koepferae*. In contrast, these families together with *RC*, were the most overexpressed in hybrids *vs D. buzzatii*. Last, the *LTR* order was the one where some families showed extreme overexpression and underexpression values in hybrids *vs* both parental species. When we went deeper into the superfamilies of elements belonging to each order (Supplementary file 3: Figure S2b), highly expressed elements in hybrids included mainly *Gypsy* (*LTR*), *Helitron* (*RC*) and Jockey (LINE) elements in comparison to *D. buzzatii* and, Gypsy (LTR), hAT (DNA), TC1Mariner (DNA) elements compared to *D. koepferae*. Some of these TEs were detected as the most differentially expressed between the parental species (green and blue TE names in Supplementary file 3: Figure S1b). Consistently with what was found in previous works [24], the

most highly overexpressed TEs in hybrids *vs* both parental species belonged to Gypsy subfamily (Supplementary file 3: Figure S2b).

Epigenetic mark landscapes are conserved and contribute to gene and TE expression

In *Drosophila*, the constitutive heterochromatin is enriched in H3K9me3 epigenetic mark, whereas H3K27me3 and H3K4me3 are associated to facultative heterochromatin and euchromatin, respectively [488]. We investigated the landscape of epigenetic marks in the parental species, unknown to date, and in their hybrids. Analyses of the distribution of histone marks in genes and their surrounding regions (Figure 2A, B) showed that H3K4me3 euchromatic mark is enriched around the start codon (SC) and throughout the coding sequence in hybrids and their parental species, which is similar to the well-described *D. melanogaster* species. H3K9me3 and H3K27me3 epigenetic marks were mainly depleted in the gene body in all species (Figure 2A, B).

We then studied the contribution of these histone marks to the gene and TE family expression in the parental species *D. buzzatii*, *D. koepferae* and their hybrids, using the linear model (RNA ~ K4 + K9 + K27 + Input) described in Material and Methods. We found that 62% (r^2–adjusted) of gene expression variation was explained by the additive linear relationship with log-transformed histone mark enrichments, with a p-value lower than 2.2×10^{-16} in all species (Supplementary file 1: Table S4). In the case of TE expression, the adjusted r^2 reached values from 61% in hybrids, to 75% in *D. koepferae* (Supplementary file 1: Table S4).

Globally, we detected that the size effect, representing the contribution of each chromatin mark to the gene expression according to our model, was similar in hybrids and parental species (Figure 2C), being H3K27me3 the chromatin mark with the greatest contribution to gene expression. H3K4me3 was strongly positively associated with gene expression showing coefficient values from 1.59 in *D. koepferae* to 1.75 in hybrids, whereas H3K27me3 was strongly negatively associated: from -2.24 in *D. koepferae*, to -2.56 in hybrids (Supplementary file 1: Table S4). H3K9me3 was the one with the least positive contribution to gene expression (from 0.35 in *D. koepferae* to 0.5 in *D. buzzatii*).

Regarding TEs, the contribution of H3K4me3 and H3K9me3 to the expression was similar in hybrids and parental species (Figure 2D). H3K4me3 was positively associated with TE expression, with coefficient values from 0.96 in *D. koepferae* to 1.58 in *D. buzzatii* (Supplementary file 1: Table S4), whereas H3K9me3 was negatively associated (-1.04 in *D. buzzatii* to -1.16 in hybrids), as expected. Contrary to the pattern observed in genes, H3K27me3 was the chromatin mark with the lowest contribution to TE expression and its contribution

depended on the species: positively associated in *D. buzzatii* and hybrids (0.38 and 0.71 respectively) and negatively in *D. kopferae* (-0.33).

Fig.2 Global epigenetic patterns in *D. buzzatii, D.koepferae* and hybrids and their contribution to expression: *(A)* Average levels of the histone marks H3K4me3, H3K9me3 and H3K27me3 over all annotated gene sequences (exons and introns) and the ±2kb surrounding regions. *(B)* Heatmaps showing the density scores of each histone mark in *D. buzzatii, D. koepferae* and their hybrids. White regions in the Heatmaps indicate missing data. SC: start codon, EC: end codon. *(C, D)* Mean linear effects of H3K4me3, H3K9me3 and H3K27me3 enrichments on RNA read counts (log transformed) in *(C)* genes and *(D)* TE families. Colours indicate the species. The 95% confidence intervals are indicated.

Maintenance of the parental enrichment of epigenetic marks in the hybrid genome

We computed read counts corresponding to enrichments in H3K4me3, H3K9me3 and H3K27me3, in the gene bodies and TE families from hybrids and their parental species. We found that only a small percentage of all genes (1.40-2.83%) were differentially enriched in hybrids

compared to any parental species and chromatin mark (Figure 3A). A lower number of differentially H3K4me3 and H3K27me3 enriched genes was detected in hybrids when they were compared to *D. buzzattii* than to *D. koepferae* (*Z-test, p < 0.001*, Supplementary file 1: Table S5). Regarding TE families, the hybrid *vs* parents comparisons revealed very similar patterns for H3K9me3 enrichments (1.52 to 1.37% of differentially enriched TE families) while they were the most contrasted for H3K27me3 enrichments (9.27 to 6.53% differentially enriched TE families, Figure 3B). However, the number of differentially enriched TEs in hybrids was similar to one or another parental species (*Z-test, p = 0.817* H3K4me3 and H3K9me3; *p = 0.198* H3K27me3; Supplementary file 1: Table S5). Moreover, H3K9me3 was the chromatin mark with more similar patterns between hybrids and the parental species for both genes and TE families, which was consistent with the similar patterns of this chromatin mark when the parental species were compared (Supplementary file 1: Table S6). On the contrary, in TE families H3K27me3 was the mark with the largest differential enrichment in hybrids *vs* the parental species, being at the same time the most contrasted between the parental species (Supplementary file 1: Table S6).

In general, most of the differentially enriched genes and TEs in hybrids have *D. buzzatii* or *D. koepferae-* like chromatin mark levels. However, we also observe a high number of genes displaying histone mark enrichments outside of the range of parental values, especially for H3K4me3 —48 less enriched and 62 more enriched in hybrids *vs* both parental species (Figure 3A). In the same way, a high number of TE families show intermediate patterns of enrichment between the parental species (Figure 3B).

We studied the distribution across chromosomes of the differentially enriched genes for each epigenetic mark in hybrids in comparison to each parental species and we did not find significant differences in any case (*Chi-squared test, p > 0.05* in all cases, Supplementary file 1: Table S7). Globally, we observed that the extreme changes in chromatin mark levels were towards an increase in hybrids in comparison to the parental species (Supplementary file 3: Figure S3). For example, some extreme changes of the euchromatic mark H3K4me3 were observed in chromosome 5 in comparison to *D. koepferae,* and of the euchromatic mark H3K27me3 in comparison to *D. koepferae* (chromosomes 4 and 5) and *D. buzzatii* (chromosome 2).

When we studied the orders of the differentially enriched TE families in hybrids *vs* parental species, we did not find differences (*Chi-squared test, p > 0.05* in all cases, Supplementary file 1: Table S7) for any chromatin mark (Figure 3C). If we focus on the specific enrichment of epigenetic marks, three TEs belonging to *DNA*, *LINE* and *LTR* families showed the most extreme values of H3K4me3 in hybrids *vs D. koepferae* (ranging from 6 to 9). However, *RC* and unknown TEs showed values of this chromatin mark similar to both parental species, being only a small percentage differentially enriched. We also found a high increment of H3K27me3 in

one *LTR* element in comparison to *D. koepferae* (~6), while the changes were less extreme in comparison to *D. buzzatii*. As reported for gene enrichment, only small differences of H3K9me3 amounts were observed in hybrids *vs* parental species.
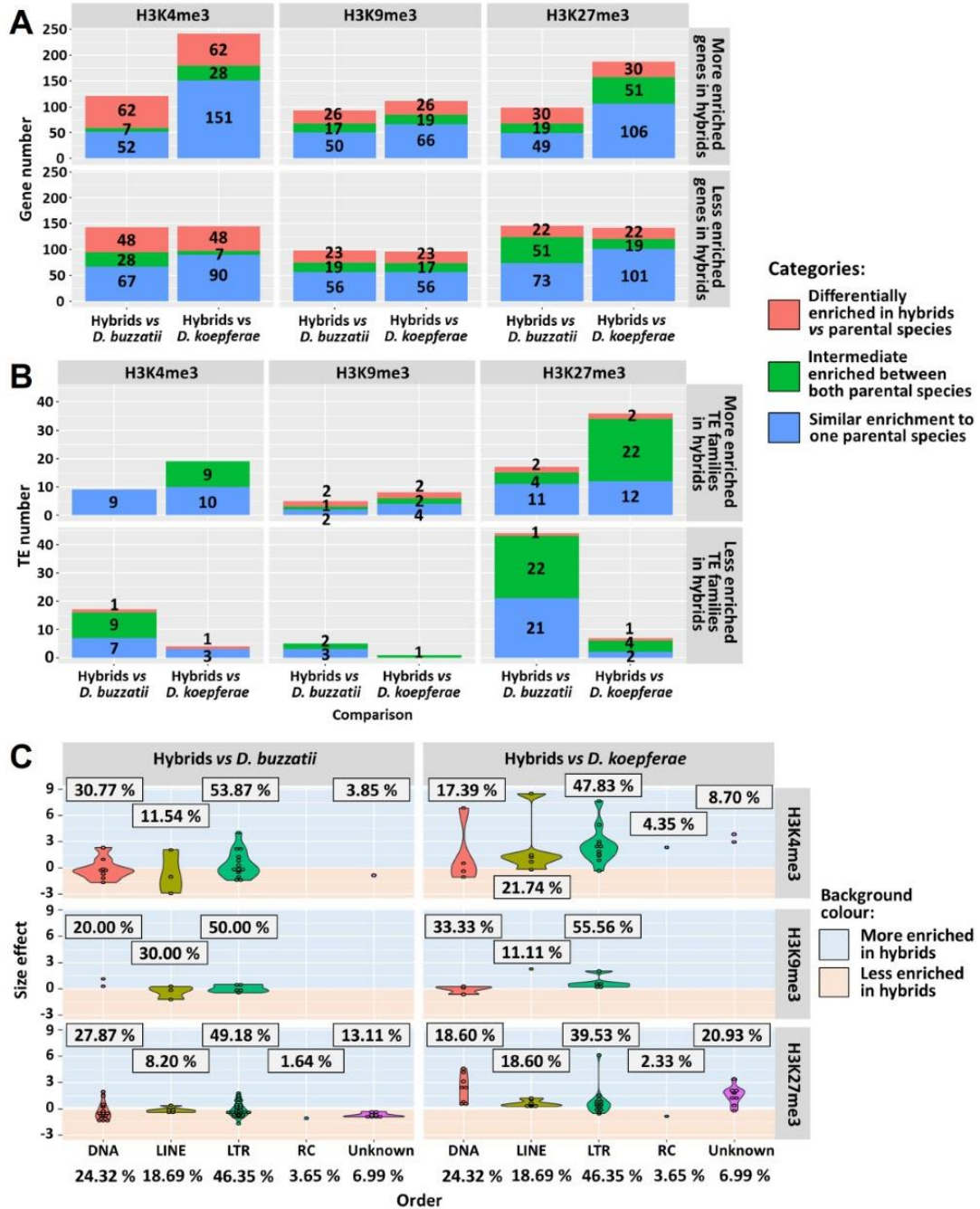
Fig. 3 Comparison of epigenetic marks in hybrids and parental species: *(A)* Number of differentially enriched genes and *(B)* TEs of the total of 13,621 genes and 658 TE families in H3K4me3, H3K9me3 or H3K27me3 chromatin marks in hybrids *vs* the parental species *D. buzzatii* and *D. koepferae* respectively. Colours indicate the different categories of enrichment. *(C)* Violin plots representing the distribution of differentially H3K4me3, H3K9me3 and H3K27me3 enriched TE orders in hybrids *vs* parental species. Points indicate the size effect. The percentages of differentially enriched TE families per order are framed. The expected (total) TE family percentages are at the bottom.

Small changes in the hybrid epigenome affect their gene and TE expression

We analyzed the association between gene and TE expression changes in hybrids in comparison to the parental species and the corresponding changes in chromatin marks (Figure 4A-L, and Supplementary file 1: Table S8) using the whole set of genes and TE families (significant and non-significant). Regarding the genes, we found that changes in the enrichment of the heterochromatic marks H3K9me3 and H3K27me3 were not associated with expression changes in hybrids in comparison to the parental species (*Fisher's exact test, D. buzzatii*: *p = 0.763* both chromatin marks; *D. koepferae*: *p = 0.970* and *p = 0.155* respectively, Supplementary file 1: Table S8). However, changes in H3K4me3 in hybrids seemed to be associated with gene expression changes (*Fisher's exact test, p ≤ 0.001*: comparison with both parental species, Supplementary file 1: Table S8). Indeed, genes that were underexpressed in hybrids compared to *D. buzzatii* also showed a reduced H3K4me3 enrichment. However, and unexpectedly, genes that were underexpressed in hybrids compared to *D. koepferae* also displayed an increased H3K4me3 enrichment. This apparent inconsistency could be explained by the differences in enrichment of this epigenetic mark between parental species: the genes enriched in H3K4me3 in hybrids *vs D. koepferae* often corresponded to those more enriched in *D. buzzatii* than in *D. kopferae* (blue colour, Figure 4A), the opposite was also observed.

TE families, as genes, showed an intermediate inheritance of the chromatin marks in hybrids: TEs more enriched in a chromatin mark in hybrids in comparison to one parental species tended to be less enriched when they were compared to the other parental species (Figure 4G, I, K). In addition, we found an association between TE expression changes in hybrids in comparison to the parental species and the corresponding changes in the epigenome. For example, an impoverishment of the heterochromatic mark H3K9me3 was found in many underexpressed TEs in hybrids compared to both parental species (*Fisher's exact tests, p < 0.001*, Supplementary file 1: Table S8). In the same way, a high TE number showing a decrease of H3K4me3 or H3K27me3 chromatin marks, were underexpresed in hybrids in comparison to *D. buzzatii* (*Fisher's exact test,* H3K4me3: *p < 0.001* and H3K27me3: *p=0.002*, Supplementary file 1: Table S8). These results were opposite to those found in hybrids compared to *D. koepferae*, where an increase of H3K27me3 was observed in underexpressed elements (*p < 0.001*, Supplementary file 1: Table S8). To determine if there was an association between chromatin marks in TEs, we checked the correlation between the euchromatic mark H3K4me3 and the other two, in hybrids in comparison to the parental species (Supplementary file 3: Figure S4). We found that the increase of H3K4me3 in hybrids was associated with an increase of H3K9me3 and H3K27me3 (*Linear model, p ≤ 1.669* x 10^{-15} in all cases, Supplementary file 1: Table S9) and the opposite, with some TEs highly expressed and extremely enriched in H3K4me3 in hybrids (red dots in the upper part right of

Supplementary file 3: Figure S4). When we focus on the genes considered as differentially expressed and differentially enriched in hybrids, we observe that the epigenetic mark associated with the most extreme expression changes in hybrids (absolute log2FC values ranging from 4 to 6) in comparison to both parental species was H3K4me3 (Figure 4M). Changes in the remaining chromatin marks were related with less extreme expression changes (absolute log2FC values from 1 to 3). Regarding the TEs showing significant differences in enrichment and expression in hybrids *vs* parental species (Figure 4N), we observed two TE families whose increase in H3K4me3 was associated to an extreme overexpression in hybrids in comparison to *D. buzzatii* (log2FC from 6 to 11). Additionally, increases in the number of H3K27me3 marks in hybrids in comparison to *D. kopferae* were likely related to extreme overexpression in hybrids (log2FC from 3 to 6). Chromatin mark changes in the remaining TEs were related with less extreme expression changes (absolute log2FC values from 1 to 3).

If we focus on the expression and chromatin enrichment of specific genes (Figure 4M) in hybrids, we observe some examples whose expression is explained by the combination of different chromatin marks. For example, the *CG3902* gene, involved in oxidation-reduction processes, was underexpressed in hybrids in comparison to *D. buzzatii* and less enriched in H3K4me3 but more in H3K9me3. The gene *ptc*, involved in different processes including gonad development, was underexpressed in hybrids in comparison to *D. koepferae* and more enriched in H3K4me3 and H3K27me3. In other cases, the expression changes in genes (Figure 4M) and TEs (Figure 4N) could be explained by the content of a unique epigenetic mark when hybrids are compared to one or another parental species. For example, the gene *Nepl4* (Figure 4M), involved in protein processing, had a low H3K4me3 content in hybrids *vs* both parental species and was underexpressed in hybrids. On the other hand, the gene named *3* (Figure 4M) and the unknown$_1$ TE (Figure 4N) were detected as underexpressed and less enriched in H3K4me3 in hybrids in comparison to *D. buzzatii*, but overexpressed and more enriched in this chromatin mark in comparison to *D. koepferae*. Finally, LTR$_2$ and DNA$_3$ were enriched in H3K27me3 in comparison to one parental species, but the opposite in comparison to the other (Figure 4N).
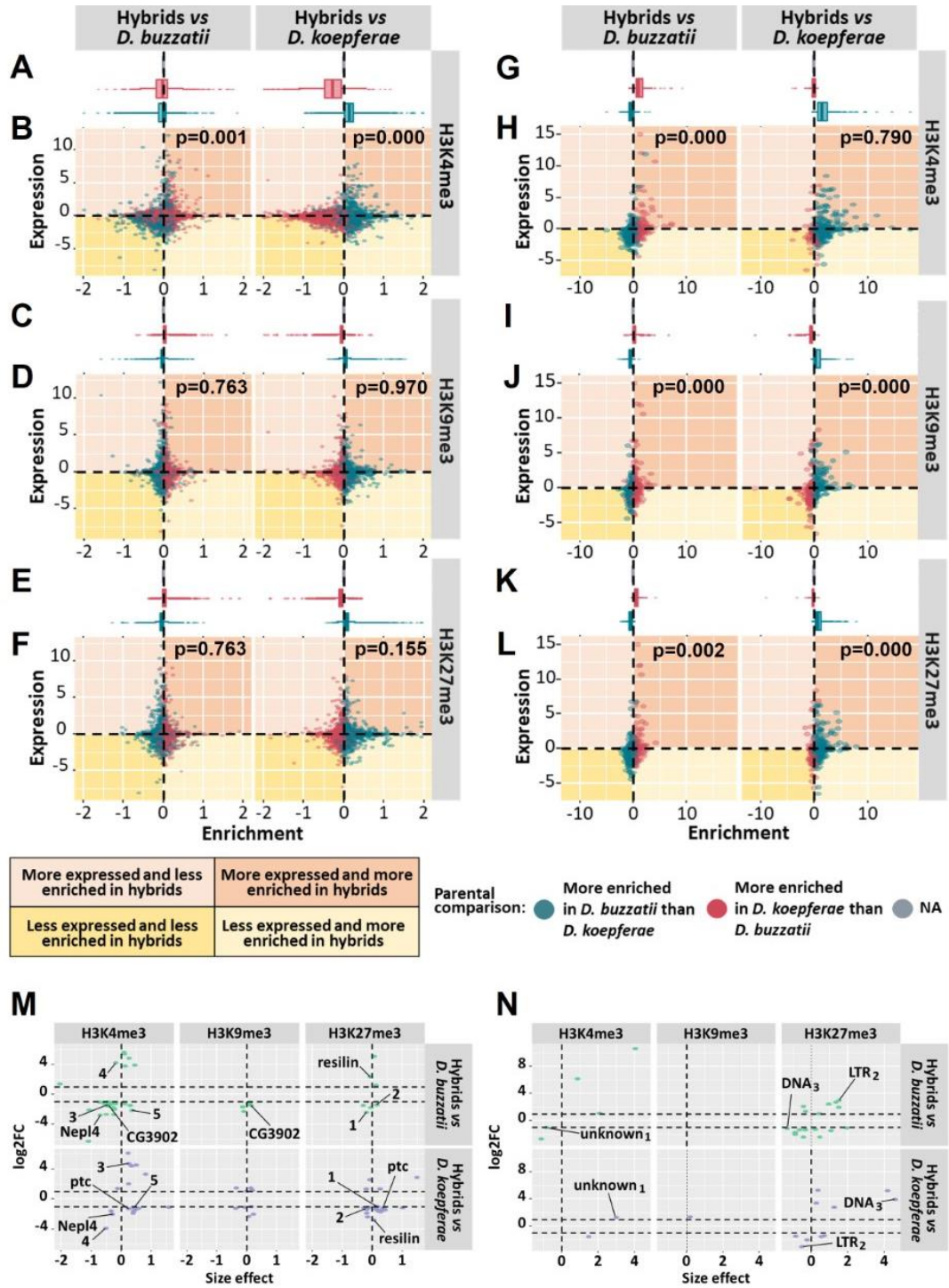
Fig. 4 Association of expression changes with the corresponding chromatin mark changes: *(A-L)* Differential expression (log2FC) and differential enrichment (size effect) of H3K4me3, H3K9me3 and H3K27me3 in all genes *(B, D, F)* and TE families *(H, J, L)* (significant and non-significant) in hybrids *vs D. buzzatii* and *D. koepferae* respectively. Top plots represent the distribution of the differential enrichment between parental species in genes *(A, C, E)* and TE families *(G, I, K)*. Colours represent the results of the differential enrichment between parental species, and values that cannot be computed are included in the NA category. Fisher's exact test p-values are shown in the right corner. Gene outliers are not included in the Figure *(A-F)*. *(M, N)* Log2FC and size effect representation of significant differentially expressed and enriched genes *(M)* and TEs *(N)* in hybrids *vs* the parental species. Significant genes *(M)* in at least two comparisons are named using their *D. melanogaster* ortholog or, otherwise, a number. The TE families *(N)* common to at least two comparisons are marked, indicating their category and a subscript.

Additional mechanisms influencing hybrid deregulation

The divergence in *cis-* and *trans-* regulatory elements was proposed as one of the causes of expression differences between *Drosophila* species, as well as of gene deregulation in their hybrids [477, 489, 490]. To detect regulatory divergence between *D. buzzatii* and *D. koepferae* and its putative effects on hybrid expression, we compared the relative allele expression in hybrids (H), the differential expression between parental species (P) and the ratio between these two metrics (T). We categorized the genes in different classes, as in [489] and described in Material and Methods. As shown in Figure 5A, most of the genes (57.65%), categorized as conserved, do not have regulatory divergence between parental species. However, we found slightly more genes showing evidence of *trans*-regulatory divergence (6.83%) than *cis*-regulatory divergence (6.24%) (*Z-test*, $p = 0.049$, Supplementary file 1: Table S10). We also studied how the regulatory divergence influences expression differences between parental species. We found more differentially expressed genes between parental species with *trans*-regulatory divergence (*Fisher's exact test*, $p < 0.01$, Supplementary file 1: Table S11) than expected, while no differences to what was expected were observed in genes with *cis*-regulatory divergence (*Fisher's exact test*, $p = 0.128$ for *cis*-regulatory, Supplementary file 1: Table S11). Finally, we studied how the regulatory divergence observed in parents influenced the inheritance of gene expression in hybrids. We analyzed genes with regulatory divergence between *D. koepferae* and *D. buzzatii*, which were differentially expressed in hybrids *vs* parental species, using the expression categories previously described (Figure 1C). We found a higher number of differentially expressed genes with *trans*-regulatory divergence than with other regulatory classes in all expression categories, except those of genes deregulated in hybrids. In this latter category, we also observed a significantly higher number of genes with compensatory regulation (*cis-* and *trans*-regulatory differences compensate each other) than in the other expression categories (additive or *D. buzzatii/koepferae*-like) (*Fisher's exact test*, Additive expression: $p = 0.039$, *D. buzzatii* and *koepferae* -like expression: $p < 0.001$, Supplementary file 1: Table S12).

To gain insight into other mechanisms affecting TE deregulation in hybrids, we examined the expression of piRNA pathway genes considering their role in germline development, epigenetic regulation and TE silencing. We observed that *aubergine* (aub) and *Sister of yellow body* (SoYb) were deregulated towards underexpression in hybrids in comparison to both parental species (Supplementary file 3: Figure S5). The chromatin mark levels were similar to the parental species, with only a few genes differentially enriched, mostly in comparison to only one parental species. Globally, differentially enriched genes decreased H3K4me3 levels in hybrids in comparison to the parental species. *Yellow body* (Yb), which was detected as overexpressed in hybrids *vs* both parental species, was intermediate enriched in H3K27me3 in hybrids *vs* both

parental species, whereas *vreteno* (vret), which had an additive expression in hybrids *vs* parental species, was less enriched in H3K4me3 in hybrids *vs* both parental species. *Brother of Yellow body* (BoYb) and *armitage* (armi), which were detected as overexpressed in hybrids in comparison to one parental species, were more enriched in H3K4me3 and less enriched in H3K9me3 respectively in comparison to the same parental species. However, most differentially enriched genes were not detected as differentially expressed (Supplementary file 3: Figure S5).
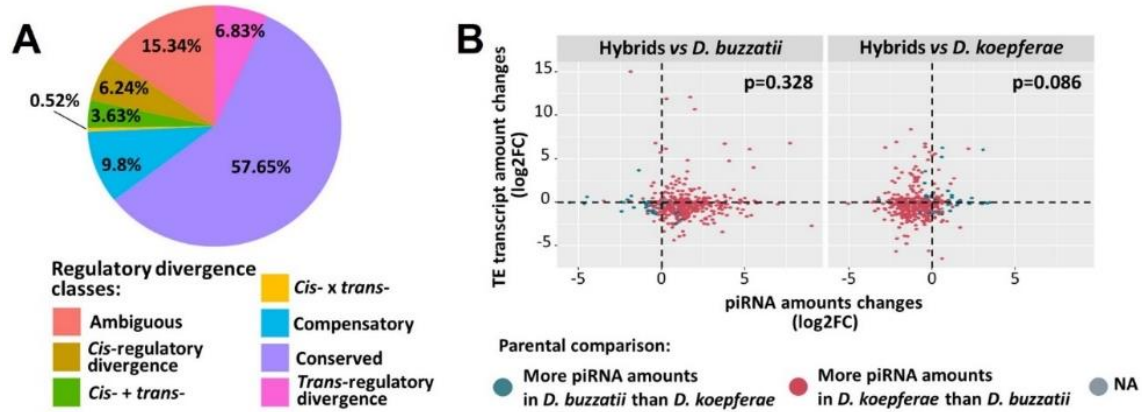
Fig. 5 Additional factors affecting gene and TE expression respectively: *(A) cis-* and *trans-* regulatory divergence between parental species. The pie chart represents the percentage of total genes categorized by regulatory divergence class. *(B)* Differential expression (log2FC) values of TEs and piRNA amounts (log2FC) in hybrids *vs D. buzzatii* and *D. koepferae* respectively. Colours represent the results of the piRNA amount changes between parental species. Values that cannot be computed are included in the NA category. Fisher's exact test p-values are shown in the right corner.

We next studied the association of piRNA amounts and TE expression in hybrids and parental species. We observed that, on average, parental species showed differences in piRNA amounts: TE families were associated with more piRNAs in *D. koepferae* than in *D. buzzatii* (mean log2FC of -2.14 in *D. buzzatii vs D. koepferae* comparison, Figure 5B), as observed in our previous work [24]. Hybrids exhibited an additive pattern of piRNA amounts between parental species: less piRNA amounts than *D. koepferae* and more than *D. buzzatii*. Additionally, as shown in Figure 5B, the differences in piRNA amounts were not associated to the changes in TE transcript amounts in hybrids in comparison to the parental species (*Fisher's exact test, p* = 0.328 for *D. buzzatii* and *p* = 0.086 for *D. koepferae,* Supplementary file 1: Table S13). There were also no differences when TE classes were analyzed separately (*Fisher's exact test, p* = 0.732 for Class I and *p* = 0.564 for Class II, Supplementary file 1: Table S13).

3.1.3 Discussion

Bias to underexpression of genes and TEs in hybrid females

The analysis of the total ovarian transcriptome of parental species showed that 5.92% and 29.64 % of the protein-coding genes and TE families, respectively, were differentially expressed between *D. buzzatii* and *D. koepferae* parental species. These differences are similar to those obtained in other studies with other species of the *repleta* group [330]. When the transcriptome of F_1 hybrids and their parental species were compared, around ~4% of genes and ~23% of TEs from hybrids were differentially expressed compared to any parental species. A greater deregulation of TEs than of genes was already observed in a previous work with *D. melanogaster-simulans* artificial hybrids [350], but the percentages were different: 0.7% of genes were deregulated and 12% of TE families were overexpressed in comparison to both parental species (no data provided about underexpression). In contrast, ~78% of differentially expressed genes of *D. melanogaster-simulans* female hybrid *vs* their parental species were found in other studies [474]. These discrepancies could be due to the different methodological approaches used along with the different genetic background of the *Drosophila* stocks.

On the other hand, in our hybrid females most of the differentially expressed genes tended to be underexpressed, as observed in previous studies in plants [491] and in *Drosophila* females [474]. The underexpression seems also to be the rule in *Drosophila* hybrid males between species of the *melanogaster* group [471, 472], which have been associated to male sterility. In our case, even if most F_1 hybrid females are fertile, some are partially fertile [28], which could explain that Gene Ontology terms of our underexpressed genes are associated to developmental processes, cell adhesion and reproduction. Instead, only a few genes were extremely overexpressed in hybrids, for example the *no hitter gene*, which is involved in spermatogenesis processes. The overexpression of male-specific genes in female hybrids is not new, and has been attributed to a failure in the mechanisms controlling the expression of these genes in females [474].

If we focus on the resemblance of gene expression in hybrids *vs* parental species, we found a bias towards genes that resemble more to one of the parental species, being the number of genes having an additive expression between both parental species low, as observed in previous studies with *D. melanogaster-simulans* hybrids [474]. Moreover, the number of genes in hybrids sharing similar transcript amounts with *D. koepferae* was higher than with *D. buzzatii*. Maternal effects were pointed out in *Arabidopsis lyrata* [492], *Xenopus* [475] and in coral [493] intraspecific hybrids. Since only one direction of cross can be performed in our case, it is difficult to attribute these results unequivocally to maternal effects.

When the distribution of the derepressed genes in F_1 hybrid females along chromosomes was considered, we did not find an overrepresentation of deregulated genes in any chromosome, except in chromosome 6 (dot). This bias could be explained if this chromosome was different between our parental species, which is reinforced by the highest rate of molecular evolution found in this dot chromosome in the closely related species *D. mojavensis* [494]. No bias was found in a previous *Drosophila* study [471]. However, an X-chromosome bias of differentially expressed genes was found in other studies performed in *Drosophila* hybrid males [472, 486]. The faster evolution of X-linked genes [495] proposed to explain the differential expression between X and autosomal genes has, however, been questioned by other authors [496] who consider that the homozygous autosome effects in reproductive isolation, approach those of X chromosome.

We showed that 6.84% of TE families were completely deregulated in F_1 hybrid females, in comparison to both parental species, and had a trend towards underexpression. These results, obtained using a different approach (normalization of TE counts using gene counts) and updated analysis tools, contrast with our previous results where the number of TE families upregulated slightly exceeded that of downregulated [24]. However, a few TE families belonging to the Gypsy superfamily showed values of expression very high compared to parental species, concordantly with previous results in these species where an increase of transposition and expression of the *Osvaldo* element was shown [361, 362, 478]. Results on TE expression in hybrids reported in the literature are very heterogeneous, finding cases of overexpression of specific elements by RT-qPCR [478, 497] or by RNAseq [350]. However, examples of underexpression in hybrids, affecting most [498] or a high percentage of TEs were also reported [499], but usually results of underexpression, if they exist, are poorly discussed. Finally, no evidence of TE reactivation was found in natural hybrid lineages of *Saccharomyces*, suggesting that other factors like population structure and hybrid genotype are major determinants of TE content [500]. The finding of underexpressed TE families in our hybrids highlights that regulation of some TEs exits in a way. Since cases of overexpressed TE families were also observed, we suggest that deregulation processes could be closely related to the TE family and the genetic background of species involved in the hybridization processes.

The epigenomic landscapes of parental species and hybrids are similar to other *Drosophila* species

Gene and TE deregulation in *D. buzzatii* and *koepferae* hybrids were previously described in the literature [24, 343, 362, 364, 365, 478], as well as in other hybrids of *Drosophila* [472, 482, 501] and of other organisms [502–504]. To get insight about these genomic deregulation mechanisms, we studied the epigenomic landscape of a euchromatic mark (H3K4me3) and two heterochromatic marks (H3K9me3 and H3K27me3) in *D. buzzatii*, *D. koepferae* and their F_1

hybrids, and we found a high similarity across these species. Our results do not globally differ from those of the *D. melanogaster* species epigenome described in the literature and are consistent with the reported high conservation of the active chromatin epigenome landscapes across *Drosophila* species [505]. As reported in other studies [149, 488, 506, 507], we found the H3K4me3 active chromatin mark located at the 5' ends of actively transcribed genes and a depletion of H3K9me3 and H3K27me3 in the gene body. H3K9me3 was previously described to be enriched in promoters but depleted in 5' transcribed regions of active genes [149, 488], whereas the H3K27me3 was under-represented in the gene body [149, 488, 508]. However, both were reported to be enriched in transposons and repetitive sequences in *Drosophila* [149, 507] and other organisms [509, 510].

We next investigated whether the genomic expression was explained by the epigenetic marks, finding that globally the gene expression was positively correlated with the active mark and was negatively correlated with the H3K27me3 repressive mark, concordantly with previous studies in *Drosophila* [149] and plants [481]. Correlations between the epigenomic landscape and gene expression in *Drosophila* have been previously described in other studies [149, 511]. According to our model, 62% of gene expression was explained by epigenetic marks, being H3K27me3 the one that most influence gene expression, followed by H3K4me3 as observed in other *Drosophila* studies [149]. H3K4me3 was positively associated with gene expression, consistent with its enrichment in actively transcribed genes found in other works [149, 488, 506, 507]. H3K27me3 was found to be negatively correlated to mRNA levels, consistent with its reported function of binding target for Polycomb repressive complex 1 (PRC1) [149, 488, 507, 512]. Surprisingly, the epigenetic mark H3K9me3, usually associated to heterochromatin regions [149], was positively associated with the gene expression, even if its contribution is very low. Increases of this epigenetic mark, along with other euchromatic ones, have been described in intronic and 3' end regions of some active heterochromatic genes in some stages of *D. melanogaster* [512, 513].

The epigenetic marks found on TEs explained more than 60% of their expression, being H3K9me3 and H3K4me3 the most influential, with a negative and positive association to TE expression, respectively. H3K9me3 together with H3K27me3, are known to be abundant in TEs and are associated to their silencing [507, 510] and H3K4me3 to their activation in *Drosophila* [149, 512] and other organisms [510]. The different association of H3K27me3 with some species (positively correlated with TE expression in *D. buzzatii* and hybrids) could be explained by the different chromatin marks associated to each TE copy inside the same family.

Hybrids exhibited limited changes in histone marks

We studied differences in the enrichment of three chromatin marks (H3K4me3, H3K9me3 and H3K27me3) in genes and TEs between the parental species *D. buzzatii*, *D. koepferae* and their hybrids, constituting a pioneer study in the *Drosophila* genus. We found that most of the genomic regions in hybrids show similar histone modification patterns compared to the parental species. This is consistent with the maintenance of the parental histone modification patterns, as well as their additive inheritance in intraspecific [503, 514, 515] and interspecific [341] plant hybrids described in the literature. However, we also observe that 1.40-2.83% of genes and 1.37-9.27% of TE families are differentially enriched in hybrids compared to any parental species. H3K4me3 seems to be related with most gene expression changes, consistent with the findings in rice and maize hybrids [481, 504] where gene expression changes were correlated to the enrichment of this epigenetic mark.

In the case of TEs, TE family expression changes in hybrids in comparison to *D. buzzatii* were related to their corresponding enrichment changes in H3K4me3, H3K9me3 and H3K27me3, whereas they were related only to H3K9me3 and H3K27me3 enrichment changes when hybrids were compared to *D. koepferae*. These results are in disagreement with previous results reported in interspecific *Arabidopsis* hybrids, where changes in H3K9me2 and H3K27me3 do not coincide with the TEs having their expression changed [516]. These differences could be explained by the small number of differentially expressed TEs found in their hybrids and the use of a different organism and methodology. Because in our study we cannot distinguish individual TE copies, they were analyzed at a family level (658 TE families), meaning that changes in expression or chromatin mark amounts are the result of the addition of the different copies per family. Indeed, intraspecific and interspecific variations in histone marks were observed in *Drosophila* TE copies [517], indicating that specific epigenetic modifications in TE individual copies in our hybrids could go unnoticed. Further inspection showed that increases of H3K4me3 were associated with increases of H3K9me3 and H3K27me3 in hybrids. Changes in the epigenetic status in individuals submitted to other stress, were already observed in *Bari-Jheh* TE: a H3K9me3 dominant pattern turned to increases in H3K4me3, H3K9me3 and H3K27me3 enrichments [518].

Other mechanisms may have a role in hybrid genome deregulation

Even though in hybrids we found a relationship between epigenetic status and expression changes in genes and TEs, the histone modification patterns did not account for the whole genome deregulation. To get more insight into this aspect, we studied the regulatory divergence between *D. buzzatii* and *D. koepferae*, and its contribution to gene expression differences between them and their hybrids. In general, we found that most genes did not have regulatory divergence

between parental species, which is consistent with the small percentage (5.92%) of differentially expressed genes found between *D. buzzatii* and *D. koepferae*. We also detected more *trans*-regulatory than *cis*-regulatory divergence between these species, as well as a higher number of differentially expressed *trans*-regulated genes. These results are in concordance with the high *trans*-regulatory divergence reported between *D. melanogaster* and *D. sechellia* species [489] and the association of this class with gene differential expression between parental species, but opposite to other studies in *D. melanogaster* and *D. simulans* [490, 519]. When we studied how the regulatory divergence affects the gene inheritance in hybrids, we found that most of the differentially expressed genes in hybrids *vs* parental species had *trans*-regulatory divergence. These results are in contrast to what was reported in the additive expressed genes, where *cis*-regulatory changes were more frequent than *trans*-regulatory changes [520]. Finally, we also found a high number of deregulated genes in hybrids with *cis-trans* compensatory evolution, which was also previously reported in *D. melanogaster* and *D. simulans* hybrids and considered an important cause of hybrid deregulation [477]. The differences reported, both between previous works and our own, could be due to use of different *Drosophila* species, different methodologies and the use, in this work, of ovarian tissues *vs* whole adults.

We as well examined the expression of piRNA pathway genes for their role in germline TE silencing and we found that *aubergine* and *Sister of yellow body* were both underexpressed, whereas *yellow body* was overexpressed in hybrids *vs* both parental species. Results of expression of these genes showed the same general trend as observed in a previous study with these species using a different analysis approach of RNAseq data [24]. Nevertheless, overexpression of three additional piwi pathway genes in the same hybrids was detected by RT-qPCR in our previous work [365], suggesting that the activation of these genes could be a primary response to the hybridization stress. Discrepancies found between RT-qPCR and RNAseq results highlight the different sensibilities of these two techniques. In addition, we found that the chromatin mark levels of the piRNA pathway genes was similar to the parental species, with only differences in a few genes, which is consistent to the general trend observed in other genes. In addition, we found that the piRNA amount changes were not associated with TE expression changes in hybrids, since we found an intermediate inheritance of piRNAs between the parental species in the hybrid, being *D. koepferae* the parental species with the highest amount, as observed in our previous work [24].

We suggest that changes in transcript amounts in hybrids are either the result of the enrichment/impoverishment of a specific mark or the disproportion in the active/repressive mark content, together with other mechanisms such as *cis-trans* compensatory regulation. Three histone marks, considered as relevant to the expression in *Drosophila*, were considered in this work, but we cannot discard the effect of other epigenetic marks like H3K9ac highly positively correlated

to gene expression [149]. Other factor that could affect gene and TE deregulation is the asynapsis, frequently observed in our hybrids and reported in other *repleta* hybrids [347]. Asynapsis, is known for influencing the *trans* regulation of *Ultrabithorax* gene alleles in *Drosophila* [521], for contributing to mouse intersubspecific hybrids infertility [522] and to the female meiotic losses in mammals [523].

3.1.4 Conclusions

The hybridization between *D. buzzatii-koepferae* species promotes a higher deregulation in TE families than in genes, both towards underexpression. The epigenome of the parental species is in general highly preserved in the hybrids, but some changes of the parental chromatin landscape are also observed in hybrids and are associated with their new gene and TE family expression patterns. Finally, *cis-trans* compensatory regulation could also be involved in expression deregulation of some genes. The present study has contributed to a better understanding of the mechanism affecting genomic deregulation in hybrids. Nevertheless, we cannot discard additional mechanisms, resulting from the incompatibility of the two different paternal genomes in the hybrids, which could interact forming a complex gene network and contribute to the deregulation patterns observed. This and the fact that this study is limited to hybrid females, makes that additional studies are necessary to go deeper into a better knowledge of the regulatory mechanisms and the factors involved in hybrid male sterility.

3.1.5 Material and Methods

Drosophila stocks and crosses

We used *D. buzzatii* Bu 28 and *D. koepferae* Ko2 inbred strains described in our previous works [362–364, 478]. Both strains were maintained by brother-sister mating for at least a decade and then by mass culturing. We performed 10 different interspecific crosses of 10 *D. buzzatii* males to 10 *D. koepferae* females, due to the scarce offspring obtained in interspecific crosses. All stocks and crosses were reared at 25ºC in a standard *Drosophila* medium supplemented with yeast.

Chromatin preparation

Ovaries of 5 day-old *Drosophila* females were dissected in PBT (1X phosphate-buffered saline and 0.2% Tween 20). We performed chromatin extraction from two biological replicates of 50 ovaries per sample of parental species and 70 ovaries for hybrids. Samples were resuspended in Buffer A1 (HEPES 15 mM, Sodium Butyrate 10 mM, KCl 60 mM, Triton ×100 0,5%, NaCl 15 mM) plus 1.8% formaldehyde, homogenized with a dounce tissue grinder (15 times) and incubated for 10 minutes at room temperature. The crosslink was then stopped with glycine to a

final concentration of 125 mM. Samples were subsequently incubated 3 minutes, kept on ice and washed 3 times with Buffer A1. We then added 0.2 ml of lysis buffer (HEPES 15 mM, EDTA 1 mM, EGTA 0.5 mM, Sodium Butyrate 10 mM, SDS 0.5%, Sodium deoxycholate 0.1%, N-Lauroylsarcosine 0.5%, Triton x100 1%, and NaCl 140 mM) and incubated 1-2 hours at 4 °C. After the lysis process, we sonicated the samples using Biorruptor® pico sonication device from Diagenode: 32 cycles of 30 s ON/30 s OFF, for parental species, and 35 cycles of 30 s ON/30 s OFF for hybrids. The sheared cross-linked chromatin was recovered from the pellet after a spin step at 10,000g 4 minutes at 4ºC.

To check DNA size, samples were previously de-crosslinked with NaCl 5M, boiled 15 min and treated with 1 µl of 10mg/ml RNAse A (37ºC, 30 min). They were purified with phenol-chloroform, precipitated in absolute ethanol, washed in 70% ethanol, resuspended in 20 µl of DEPC and run in a 1.5% agarose gel.

For the immunoprecipitation step, Magna ChIPTM G chromatin immunoprecipation Kit (Millipore) was used together with antibodies against H3K4me3 (Abcam; ab8580), H3K9me3 (Abcam; ab8898), and H3K27me3 (Abcam; ab6002). We separated 20 µl of chromatin for input and the remaining 180 µl were distributed in three aliquots where 1 µl of each antibody plus 20 µl of magnetic beads and dilution buffer up to 530 µl were added. Samples were incubated overnight at 4 °C in an agitation wheel. After the beads were washed with Low salt buffer, High salt buffer, LiCl buffer, and TE buffer. Separation of chromatin from the beads was performed using 0.5 ml of CHIP elution buffer. 1 µl of Proteinase K was then added to each sample and incubated at 62 °C for 2 hours in a shaker at 300 rpm. Samples were then purified with the columns provided by the kit and stored at −20 °C. ChIP enrichment was quantified by real-time PCR of a well-known genomic region enriched for the different histone marks studied: rp49 for H3K4me3, kkv for H3K9me3, and light for H3K27me3. The following gene-specific primers were used: rp49-forward: 5'-GTCGTCGCTTCAAGGGCCAAT-3', rp49-reverse: 5'-ATGGGCGATCTCACCGCAGTA-3', kkv-forward: 5'-TAATCCAGCCACGCCCATTT-3', kkv-reverse: 5'-CCCAACGTTTGCATTGCTGA-3', light-forward: 5'-CGAGTACAAAATGAATAGCTCCG-3', light-reverse: 5'-GCGGTTCTCCTCAATGAT-3'.

Chromatin Sequencing

Duplicate Truseq ChIP libraries, corresponding to two biological replicates per sample, were performed by Macrogen Inc., Seoul, Korea. Sequencing was carried out using an Illumina Hiseq4000. We obtained 22-34 millions of paired-end reads for each sample, resulting in a total of 659.9 millions of paired-end reads.

ChIPseq visualization patterns

ChIPseq sequenced reads were trimmed using Trimmomatic software v0.39 [524] and aligned to the *D. buzzatii* genome downloaded from the *Drosophila buzzatii* Genome Project web page (http://dbuz.uab.cat, last accessed January 7, 2015) using Bowtie2 v2.3.5.1 [525]. For the alignment, the default parameters of the *--very-sensitive-local* mode with two extra-modifications to increase the sensibility *(--local -D 20 -R 3 -L 20 -N 1 -p 12 --gbar 3 --mp 5,1 --rdg 4,2 --rfg 4,2)* were used to reach the highest percentage of aligned reads with both parental species and their F_1 hybrids. Reads with map quality score lower than 30 and unmapped reads were filtered using Samtools v1.10 [526] and excluded from further analysis. Deeptools v3.3.2 [527] was used to visualize the enrichment of each chromatin mark around the start codon (SC) and the end codon (EC). First, *bamCompare* was used to normalize the ChIPseq samples by depth using the RPKM method and by the input (control). Finally, the read density values were computed using *computeMatrix* and visualized with *plotHeatmap*.

Gene alignments

RNAseq reads from Romero *et al.* [24] were treated the same way as ChIPseq reads to ensure that results were comparable. Gene sequences (only body region: from the start codon to the end codon) from *D. buzzatii* were obtained with *getfasta* of Bedtools v2.29.2 [528] using the genome sequence and its gene annotations (Guillén et al. 2014). Gene sequences were masked using RepeatMasker v4.1.1. (http://www.repeatmasker.org) and the Repeat Masker from the *D. buzzatii* browser (http://dbuz.uab.cat, last accessed January 7, 2015). A total of 37 genes were completely masked and excluded from further analysis, and a total of 13,621 protein coding genes were included in the reference sequences.

To ensure that there was no bias due to the use of only the *D. buzzatii* reference genome, a *de novo* transcriptome for each parental species was created using Trinity v2.9 [530] and the corresponding RNAseq. *D. buzzatii* and *koepferae* transcriptomes were then aligned against all the ChIPseq inputs (Supplementary file 4: Alignment_inputChIPSeq). We also randomly selected 40 genes amongst Trinity outputs, and computed nucleotide divergence between parental species using the Mega software [531] and the Jukes-Cantor model [532] (Supplementary file 4: Nucleotide_Divergence). Both, the similar alignment percentages and the low average divergence, indicate that bias of using *D. buzzatii* as reference genome, if any, is marginal. Nevertheless, we decided to be conservative and use only the protein coding genes. Trimmed RNAseq and ChIPseq reads were aligned to the *D. buzzatii* masked gene sequences using Bowtie2 v2.3.5.1 [525] and the parameters explained above. eXpress v1.5.1 [533] with the default options and then an additional online EM round (to increase the accuracy) was used to quantify read

71

counts for each gene. All isoforms of a gene were considered together. We used rounded effective counts for the following steps.

TE alignment

TE RNAseq [24] and ChIPseq read counts were analyzed using the TEcount module of the TEtools pipeline [534] (available at https://github-com/l-modolo/TEtools). First, the manually constructed TE library containing consensus TEs from both *D. buzzatii* and *D. koepferae*, described in a previous work [24], was aligned with the RNAseq and ChIPseq reads, using Bowtie 2 v2.2.4 [525] with the most sensitive option and keeping a single alignment for reads mapping to multiple positions (--very-sensitive). Read counts were computed per TE family, adding all reads mapped on copies from the same family. Count Tables corresponding to genes and TEs were concatenated and were used for the differential expression and enrichment analyses. Genes counts were used to normalize TE counts, following the guidelines of TEtools pipeline [534].

Differential expression and enrichment analyses

The statistical analyses were performed using R v4.0 [535]. The visualization of the results was performed using the R package ggplot2 v3.3.2 [536]. The *DESeq2* function from the R Bioconductor package DESeq2 v1.28.1 [537] was used to normalize read counts, using the default method, and to model the read counts using a negative binomial distribution.

For RNAseq, the DESeq2 package was used to identify differentially expressed genes and TE families while performing a Wald test [537]. The p-values were adjusted for multiple testing using the procedure of Benjamin and Hochberg [538] with an FDR cutoff of 0.05, and were obtained using the *results* function from the DESeq2 package. The logarithmic 2 fold change (log2FC) were shrunken using the default and recommended *apeglm* algorithm [539] of the *lfcShrink* function. Genes with an adjusted p-value lower than 0.05 and at least double of expression in one species above the other (absolute shrunken log2FC > 1) were considered as differentially expressed. Gene Ontology term enrichment analyses of biological processes were performed for the underexpressed and overexpressed significant genes using the topGO R package v.2.42.0. [540] ("weight01" algorithm and Fisher´s statistic) and the ortholog genes in *D. melanogaster* obtained from the *Drosophila buzzatii* Genome Project web page (http://dbuz.uab.cat, last accessed January 7, 2015). The obtained Fisher's exact test p-values were not adjusted but, from the total gene ontologies with p < 0.05, only the top $\frac{1}{3}$ with the lowest p-values were considered as enriched.

Regarding the ChIPseq data, a handmade script was used to perform the differential enrichment analysis between hybrids and both parental species. Briefly, the 'Regularized log'

transformed counts obtained using DESeq2 for each chromatin mark were considered and analyzed using the following linear model for each gene: RldKm ~ RldKi + species (RldKm: log transformed counts for the considered histone mark, RldKi: log transformed counts for the input). We considered contrasts between pairs of species modalities (*D. buzzatii*, *D. koepferae*, hybrid). Because of the high number of tests and the lack of extreme p-values, the Benjamin and Hochberg [538] adjustment of the p-values had a high effect in our results. For this reason, the p-values were not adjusted and, from the genes and TE families with p < 0.05, the rounded $\frac{1}{3}$ of the top genes and TE families with the lowest p-values were considered as differentially enriched.

The significant genes were assigned to their respective chromosomes following the scaffold assignation to chromosomes obtained from a previous work [123] to detect potential chromosome-biases of differentially expressed genes. Additionally, the order of the significant TE families was analyzed to detect possible order-biases in our significant TEs.

Within each genome, we quantified the contribution of histone mark enrichments on transcript amounts using the following linear model on log-transformed read counts for each gene and TE family: RNA ~ K4 + K9 + K27 + Input. In addition, we tested whether changes in RNAseq counts in hybrids were associated with changes in ChIPseq counts using Fisher's exact tests on 2x2 matrices: genes and TE families were classified as displaying a positive or negative log2FC of RNAseq counts in hybrids *vs* parents and as displaying an increase or decrease in histone mark enrichment in hybrids *vs* parents.

Allele-specific expression analysis

We created a the *de novo* transcriptome with the RNAseq of both parental species, together with the hybrids using Trinity v2.9 [530]. A SuperTranscript [541] was then created as a general reference transcriptome. The GATK pipeline for variant calling [542] was used to detect differences between the general reference transcriptome and each parental species. The VCF files were filtered following the GATK guidelines, including a coverage depth of at least 10, and the exclusion of variants only present in one replicate, which were considered assembly errors. These variants were replaced in the general reference transcriptome using the FastaAlternateReferenceMaker GATK tool [542] to create a reference transcriptome for each parental species.

HISAT2 [543] with the "no-softclip" option was used to align the RNAseq from hybrids to both parental reference transcriptome. Then, CompareBams [544] of Jvarkit was used to compare the alignments and FilterSamReads from Picard [545] to filter-out the reads aligning in different position when the data were aligned to each parental species. Samtools v1.10 [526] was used to remove multimapped and unmapped reads and BamTools v.2.4.0 [546] to keep only reads

that align without mismatches. Bedtools multicov [528] and a manually updated GTF were used to count reads aligning to each gene.

The reference transcriptomes were annotated using BLAT v.35x1 [547] with the parameters --minIdentity = 80 and --maxIntron = 75000 and the gene sequences of *D. buzzatii*, keeping the best match with an overlap of at least 50%. Subsequently, statistical analyses were performed using R v4.0 [535] and DESeq2. The *DESeq2* function was used to normalize read counts, using the default method. First, the *results* function was used to compute the logarithmic 2 fold change and the adjusted p-values using the Benjamin and Hochberg [538] method. A FDR cutoff of 0.05 was used to identify differentially expressed genes 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 analyzed for *trans*-regulatory effects (T) by comparing the P and H ratios with the same *results* function. Genes were then categorized in the following groups as reported in a previous work [489]:

- *Cis-* only: Significant differential expression in P and H, but no significant T.
- *Trans-* only: Significant differential expression in P, and T, but not H.
- *Cis- + trans-*: Significant differential expression in P, H and T. *Cis-* and *trans-*regulatory differences favor expression of the same allele.
- *Cis-* x *trans-*: Significant differential expression in P, H and T. *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 with no clear biological interpretation.

Piwi pathway genes and piRNA amount study

Additionally, reciprocal tblast v2.10.1. [548] of the *Drosophila melanogaster* piRNA pathway proteins from UniProt [549] and the reference gene sequence list were performed. A total of 31 proteins were detected and associated with *D. buzzatti* genes. The *Argonaute 3* gene (*ago3*), not included in the reference genes, was manually included in the list from our previous results [365].

Small RNAseq raw reads from our previous work [24] were used to study the piRNA regulatory data in the TE expression study. Using *PRINSEQ lite* [550], we isolated 23-30 nt-long reads and considered them as piRNAs. For normalization purposes, we also isolated 20-23 nt-long reads and searched for microRNA sequences: low-quality reads were removed using UrQt

[551], then trimmed 20-23 nt reads were aligned to the masked genes using Bowtie V.1.3.0 [552] and keeping a single alignment for reads mapping to multiple positions. These reads were considered to correspond to microRNAs and counts were computed using eXpress v1.5.1 [533]. TE counts among piRNAs were computed using the TEcount module of TEtools [534]. piRNA counts were then normalized so that the sum of microRNA counts is the constant across samples.

Regarding TE sequences, RNAseq and CHIPseq data were integrated following the same approaches and processes as for genes. To get insight into the chromatin mark combination, we study the association between the euchromatic mark H3K4me3 and the heterochromatic marks (H3K9me3 and H3K27me3) in hybrids *vs* each parental species.

Statistical Tests

Three main statistical tests were used in the paper and were performed using the R v4.0 [535] program:

- *Two proportion Z-test* was used to compare distributions of significant sequences (genes and TE families) across comparisons and *cis-* and *trans*-regulatory classes.
- *Chi-square test* under equal assumption was used to detect chromosome-biases and TE-category-biases.
- *Fisher's exact test* under independence assumption was computed using a 2x2 contingency Table to detect associations in: gene and TE expression (log2FC) and chromatin mark enrichment (size effect); TE euchromatic mark (H3Kme3) and heterochromatic marks (H3K9me3 and H3K27me3); *trans-* and *cis*-regulatory divergence and differences of expression between parental species; compensatory and remaining classes and gene deregulation in hybrids, and TE expression and piRNA quantities, in hybrids *vs* each parental species independently.

The results were corrected for multiple testing using the Benjamin and Hochberg [538] method.

Availability of data and materials

The data is included in the article and in its online supplementary material. Raw ChIPseq data generated from this article have been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA796032: accession numbers from SRR17535990 to SRR17536013.

contribution in the DESeq pipeline and Mariana Galvao Ferrarini for her advises in Gene Ontology analyses. Finally, we wish to thank Lars Ootes for his advices with the allele-specific expression analyses two anonymous reviewers for their valuable comments on the manuscript.

Author contributions

AB: Data analysis, interpretation, and paper writing: MPGG and CV: Conception, design, interpretation and paper writing; MF: Data analysis, interpretation and paper writing. PV: Data analysis. All authors read and approved the final manuscript.

Supplementary Material

Supplementary material is available online and can be downloaded in this link and also is available in Annex **8.1**:

Supplementary file 1.xlsx: Statistical analyses. Annex **8.1.1**.

Supplementary file 2.xlsx: Gene Ontology analyses. Annex **8.1.2**.

Supplementary file 3.pdf: Additional Figures. Annex **8.1.3**.

Supplementary file 4.xlsx: Alignment bias and divergence between parental species. Annex **8.1.4**.

3.2 Impact of the heat stress on the Transposable Elements and small RNA expression in *Drosophila subobscura*

This chapter is a verbatim reproduction from the last version of the revied paper.

Impact of heat stress on Transposable Element expression and derived small RNAs in *Drosophila subobscura*

Alejandra Bodelón [1], Marie Fablet [2,3], Daniel Siqueira de Oliveira [2,4], Cristina Vieira [2], Maria Pilar García Guerreiro [1*]

[1] Grup de Genòmica, Bioinformática i Biologia Evolutiva, Departament de Genètica i Microbiologia (Edifici C), Universitat Autònoma de Barcelona, Spain.

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

[3] Institut universitaire de France, France.

[4] Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (Unesp), São José do Rio Preto, São Paulo, 15054-000, Brazil

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

Abstract

Global warming is forcing insect populations to move and adapt, triggering adaptive genetic responses. Thermal stress has been described to alter gene expression, repressing the transcription of active genes, and inducing that of others, such as those encoding heat shock proteins. It has also been related to the activation of some specific Transposable Element (TE) families. However, the actual magnitude of this stress on the whole genome and the factors involved in these genomic changes are still unclear. We studied mRNAs and small RNAs in gonads of two populations of *Drosophila subobscura,* which is considered a good model to study adaptation to temperature changes. In control conditions, we found that a few genes and TE families were differentially expressed between populations, pointing out their putative involvement in the adaptation of populations to their different environments. Under a heat stress, we did not observe huge expression changes, but sex specific changes in gene expression together with a trend towards the overexpression of genes, mainly related to the heat shock response were detected. The heat shock only affected the expression of some specific TE families, mainly retrotransposons, siRNAs and piRNAs amounts derived from some TE families, along with the piRNA production from some piRNA clusters. Nevertheless, changes in small RNA amounts could not be clearly correlated to the observed TE expression changes, indicating that other factors as heat-shock dependent chromatin modulation, could also be involved. This work provides the first whole transcriptomic study including genes, TEs and small RNAs after a heat stress in *D. subobscura.*

Keywords: *Drosophila*, stress, transposable elements, piRNAs, RNAseq, heat shock stress

Significance

Global warming provokes intense heat waves affecting the organism genomes. Usually, heat stress alters gene expression, but the effect on transposable Element (TEs) activity and their control mechanisms, involving small RNAs, is not clear. Here we studied how the thermal stress affects the gonadal transcriptome of *Drosophila subobscura* and we found that changes on the expression of specific TE families were not always coupled with their derived small RNAs, indicating that other factors should also be involved. This work provides the first whole genome expression study in *D. subobscura*, including TE expression after a heat stress, and provides a framework for future studies on the thermal effects on the epigenome and their consequences for organisms.

3.2.1   Introduction

Temperature is a well-known stressful factor that can alter gene expression in *Drosophila* [553] and other species [436]. High temperatures have been observed to repress the transcription of

several genes and activate that of others, mostly those related to the stress response [554]. When the organisms return to optimal temperatures their cellular activities are gradually restored at rate depending on the severity of the stress [553]. The most important group of genes upregulated after a heat stress and other stressful factors are those coding heat shock proteins (Hsps) [555]. Hsps are molecular chaperones that modulate the structure and folding of other proteins [556] and play important roles in transport, signal pathways, and activation of enzymes and receptors [557]. These functions provide the organism a temporary enhanced tolerance to stress [555]. Even though the heat shock proteins have been widely studied [554], less is known about the global effects of high temperatures on the genomes [432].

Some authors also associated the heat stress to an increase of transposition of some specific Transposable Element (TE) families in *Drosophila* [460, 558, 559], such as *copia* [560] or *412* [561], but a lack of mobilization was found in other cases [463, 562]. In *Drosophila* somatic cells*,* some TEs are silenced by endogenous small interfering RNAs, endo-siRNAs [144, 563, 564]. In the ovarian somatic and germ cells, transposition is mainly controlled by Piwi-interacting RNAs (piRNAs) [151, 165, 565, 566], even though the siRNA pathway is also active [563]. Most of gonadal piRNAs are produced from specific TE enriched *loci,* called piRNA clusters, that cover around 3.5% of the *Drosophila melanogaster* genome [151]. These loci are transcribed from one or two DNA strands into long piRNA precursors, which once processed give rise to primary piRNAs [165, 566]. In the germline, primary piRNAs are amplified in a Ping-Pong amplification cycle, producing secondary piRNAs, which magnifies the TE repression [151, 185], a process that has been observed to be more relevant in females than in males [162]. On the other hand, it is also known that the heat shock proteins Hsp70 [192] and Hsp83 [567] in *Drosophila* are involved in piRNA biogenesis and their functional alterations cause TE transpositions in germ cells. However, the effects of heat shocks on the piRNA amounts and consequently on the TE activity remain ambiguous. For example, in *D. melanogaster* the heat shock treatments caused strain-specific modulation of the expression from certain piRNA-clusters and changes of the piRNA levels targeting some specific TE families [468], but not others [192]. However, no correlation between changes in piRNA levels and TE transcripts has been observed under heat stress [192, 468].

Drosophila subobscura is a species of the *obscura* group of the subgenus *Sophophora* [31]. It is native from the Palearctic region, broadly distributed in Europe, and in the late 1970s and early 1980s has invaded areas of South and North America [47]. This species has a rich inversion polymorphism, with more than 65 identified inversions, most of them located on the O chromosome [124]. Because some of them have adaptive roles, showing repeaTable spatiotemporal patterns in frequencies related to temperature [46], they can be used to monitor global warming [51]. However, little is known about the mechanisms responsible for such patterns

nor how the response to heat stress may be influenced by the genetic background of the populations under study. In fact, the different basal protein levels of Hsp70 (located on the O chromosome) detected in two populations with different O arrangements, disappeared after heat stress [427]. In contrast, comparable basal Hsp70 mRNA levels in populations bearing different O chromosomal arrangements were detected in another study [428]. The aim of this work is to study how heat stress affects the whole genome expression of two different background *D. subobscura* populations, Madeira and Curicó, coming from a native Palearctic region and from a Chilean coloniser region respectively, and both bearing the same O chromosome arrangement. This study was performed in gonads because germline changes are inherited and the piRNA pathway is active in this tissue [151, 165, 565, 566] and the effect of heat stress on these tissues have not been deeply studied yet. We studied the expression profiles and the quantities of small RNA derived from TE families in ovaries and testes. We found gene expression differences between the gonads of Madeira and Curicó that could be related to population adaptation to different environments. When populations were submitted to a heat stress, an impact on gene expression, with a trend towards overexpression, was observed in genes involved in the stress response, such as Hsps. The heat stress did not globally impact the amounts of TE-derived small RNAs, but it changed the expression of only some specific TE families, usually population and tissue specific, in addition to the small RNAs (siRNAs and piRNAs) derived from some specific TE families and the piRNA production for some piRNA clusters. Additionally, we did not find a clear influence of these small RNA changes in TE activation, suggesting that other factors, such as changes in the epigenome, could be involved in TE activity.

3.2.2   Results

Differential expression between populations in control conditions

To compare the expression levels prior to stress (control conditions) between distinct adapted populations, we studied the gonadal transcriptome of males and females from two *D. subobscura* populations: one from the native Palearctic region (Madeira) and another one from a colonized American region (Curicó). The two populations had identical chromosomal arrangements in all chromosomes (A_2, J_1, E_{st}, O_{3+4}), except on U (U_{1+2} arrangement in Madeira and U_{1+8+2} in Curicó) (supplementary Figure S1). All these inversions are frequent in each region [52, 568]. We found more genes differentially expressed between populations in testes (9,68%) than in ovaries (4.23%), supplementary Table S1 and S2, Two proportion Z-test, p < 0.001. Overall, the gene expression was more similar between populations (Spearman's rank correlation coefficient ρ = 0.974 and 0.970 in females and males respectively, p < 0.001 in both cases, supplementary Table S3) than between sexes (ρ = 0.611 and 0.568 in Curicó and Madeira respectively, p < 0.001 in

both cases, supplementary Table S3). The Gene Ontology (GO) Enrichment Analysis on differentially expressed genes between populations showed enrichment in GOs involved in general processes (cellular, metabolic, and multicellular organismal processes, biological regulation, signaling and response to stimulus) in all samples, but also in interesting GO that could be related to the adaptation of these populations to their habitat (supplementary Figure S2). For example, *Cyp12d1-p*, encoding a protein involved in the response to the insecticide dichlorodiphenyltrichloroethane (DDT), was more expressed in Curicó males and Madeira females and both copies of *Hsp70* were more expressed in Madeira population (Figure 1*A* and *B*). However, some GOs were exclusively enriched in one population, for example, that of the immune system processes, which was only enriched in Curicó females. Although, it does not mean that other genes belonging to this GO category were not differentially expressed in other populations, as is the case of *Tnpo* gene much more expressed in Madeira than in Curicó in both sexes (Figure 1*A* and *B*). In the same way, GOs related to rhythmic processes were only enriched in Curicó populations, an example of a gene belonging to this GO is *BTBD9* that is involved in the circadian cycle (Figure 1*A* and *B*). We also detected an enrichment of growth GOs when males of both populations were compared (supplementary Figure S2). Moreover, even though an enrichment of reproduction GOs was not observed, the gene *Ag5r2*, involved in multicellular organism reproduction, was highly expressed in Madeira males (Figure 1*B*). We analyzed if the genes with differences in expression between populations were the same in both sexes, and we noticed that only 7% were shared by males and females (Figure 1*E*).

TE expression differences between populations in gonadal tissue were studied using a list of *de novo* annotated TEs in the *D. subobscura* reference genome [124], as described in Material and Methods. We found that 12.83% of the reference genome was covered by putative TE families (supplementary Figure S3*A*): 1.09% corresponded to DNA elements, 2.14% to retrotransposons (1% to LTRs, 1.14% to LINEs) and 9.59% were included in the unknown category. Excepting this last TE category (62,836 insertions), the elements with the greatest number of insertions were the *Helitrons*, followed by *jockey* and *gypsy* superfamilies (supplementary Figure S3*B*). When we compared the TE expression between the two populations, we have almost twice as many TEs more expressed in Curicó than in Madeira in both sexes (supplementary Table S4). However, we found a similar percentage of TE families showing different expression between populations in ovaries (8.09%) and in testes (9.88%), supplementary Table S4 and S5, Two proportion Z-test, p = 0.265. The differentially expressed TEs belonged to different classes: DNA class (*Polinton, mariner, Transib* or *Helitron*), and retrotransposons, (*copia, gypsy, BEL, Penelope* or *jockey* superfamilies), Figure 1*C* and *D*. When we analyzed if the differentially expressed TE families were the same in both sexes, we noticed that 14% were shared by males and females (Figure 1*F*). Globally, we found that TE expression is fairly similar between populations (ρ =0.968 and 0.974,

p < 0.001, supplementary Table S6) and between sexes (ρ =0.837 and 0.828 in Curicó and Madeira respectively, p < 0.001, supplementary Table S6), unlike what was observed when the expression of genes was compared (more differences between sexes).
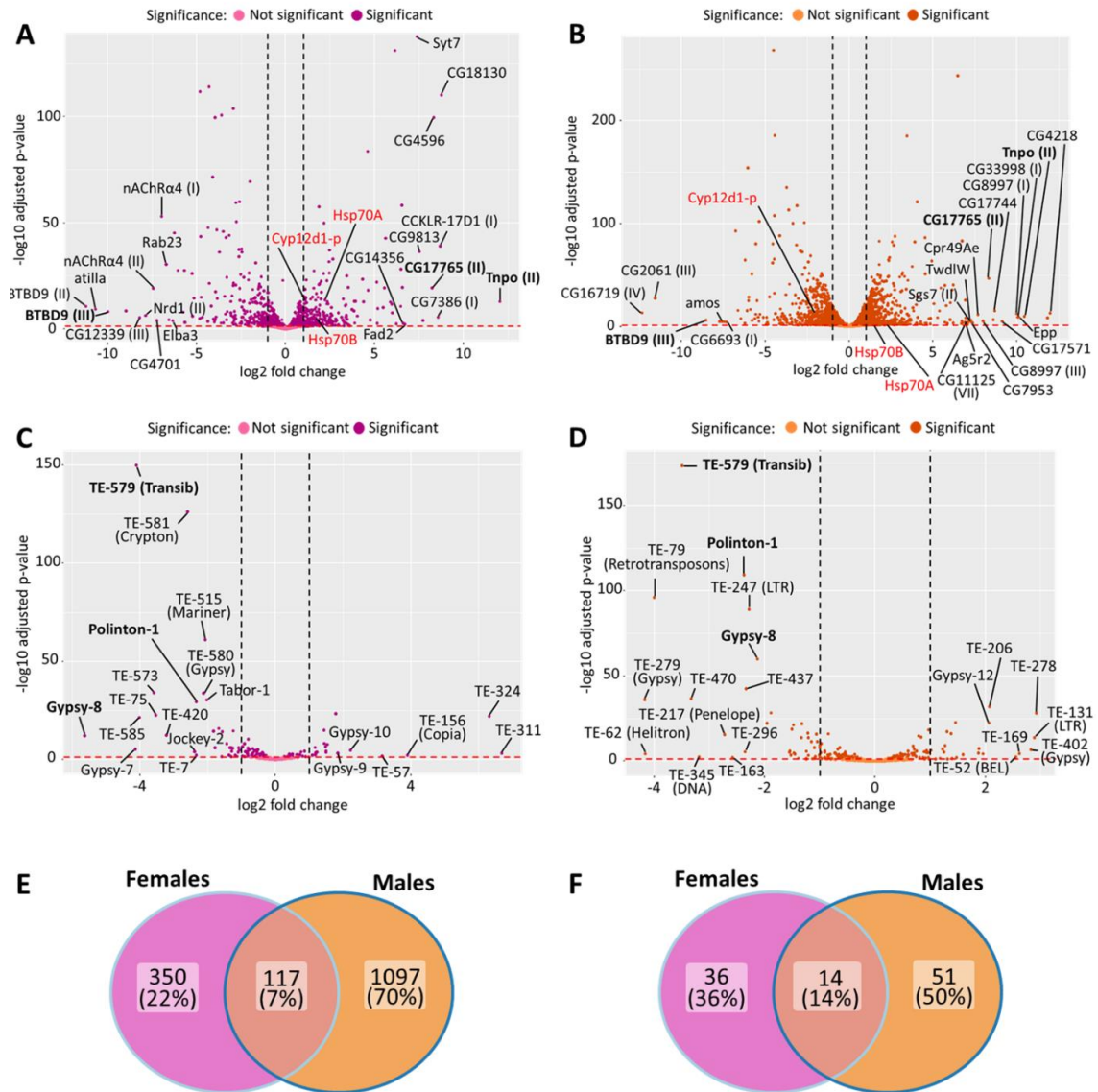
Fig. 1. Changes in expression in gonads between populations. (*A-D*) Differences of expression in gene (*A-B*) and TE families (*C-D*) in Madeira *vs* Curicó populations: females (pink, *A-C*) and males (orange, *B-D*). Positive log_2FC values correspond to genes and TEs more expressed in Madeira. The names of genes showing the 20 highest log_2FC values, and displaying and ortholog in *D. melanogaster*, are shown in *A-B*. Genes and TE families whose differential expression is common in females and males are in bold. In red, other genes mentioned in the text. *Hsp70* is called as characterized in *D. subobscura* [428]. (*E-F*): Venn diagram showing the number of differentially expressed genes (*E*) and TE families (*F*) shared by both sexes when the two populations were compared.

Sex-specific gene expression response after a heat shock

We then studied how the heat stress affected the ovaries and testes of Madeira and Curicó populations by comparing their gonadal transcriptome under heat shock *vs* control conditions, and we found that from 0.78% to 1.22% of the total expressed genes changed their expression after a

heat stress in both populations and sexes (supplementary Table S7). Whereas we observed a similar number of genes with changes in expression in females from both populations (Two Proportion Z-test, p = 0.210, supplementary Table S8), a higher number of differentially expressed genes was observed in Madeira males compared to Curicó (Two Proportion Z-test, p = 0.004, supplementary Table S8), suggesting a stronger impact of heat stress in the males of this population. Moreover, a trend to overexpression *vs* underexpression was observed (Figure 2*A-D* and supplementary Table S7), exhibiting that the effect of heat stress in gonads is different from that found in other tissues, where gene underexpression is usually the rule after heat stress [432]. As expected, the genes showing the highest overexpression in most samples corresponded mainly to the heat shock protein family (Figure 2*A-D*), whereas underexpressed genes were involved in other biological functions. Differentially expressed genes after a heat stress were mainly enriched in general GOs involved in cellular, developmental, metabolic, and multicellular organismal processes, as well as biological regulation (supplementary Figure S4). When we compare the GOs of overexpressed genes *vs* underexpressed, we observed a greater representation of genes of response to stimulus, signaling, biological regulation, locomotion, rhythmic processes, and immune system that are overexpressed. However, the genes involved in cellular and metabolic processes tended to be more represented in the underexpresed category, in addition to reproduction in males (supplementary Figure S4). We also found some genes involved in processes, such as histone chromatin remodeling, that changed their expression only in one sex after stress, for example the *lid* gene (encoding a histone demethylase) was overexpressed after a heat stress in females (data not shown).

We also looked for common differentially expressed genes after a heat shock (Figure 2*E*) between populations or sexes. We found a similar percentage of genes with changes in expression unique for each population-sex, except in Madeira males (nearly twice as many differentially expressed genes). However, we observed much more genes with changes in expression shared between populations, considering individuals of the same sex (29 in females and 25 in males, Figure 2*E*), than those shared between sexes within a population (2 genes in Curicó and 1 in Madeira, Figure 2*E*). Additionally, we observed that 18 differentially expressed genes after a heat stress were shared between populations and sexes. The detailed study of these last genes showed they were all overexpressed, in different levels depending on the sex and populations (Figure 2*F*). Most of them corresponded to heat shock genes (*Hsp67Ba*, *Hsp67Bc*, and both copies of genes *Hsp68*, *Hsp27* and *Hsp70*) or related to the heat response, such as *stv*, all highly expressed after the heat stress (Figure 2*F*). Additionally, we found genes related to gene expression, such as *CG12071* (negative regulation of transcription by RNA polymerase II) and *CG6511* (positive regulation of transcription). Finally, we found genes involved in several functions, such as *culd* (photoreceptor-cell enriched transmembrane protein), *Nach* (Sodium Channel family of proteins),

Jon74E (proteolysis), *ref(2)P* (autophagy activation by ubiquitinated proteins), *CG10357* (enables lipase activity) and *CG12947* and *CG8620* (coding for proteins with unknown functions).
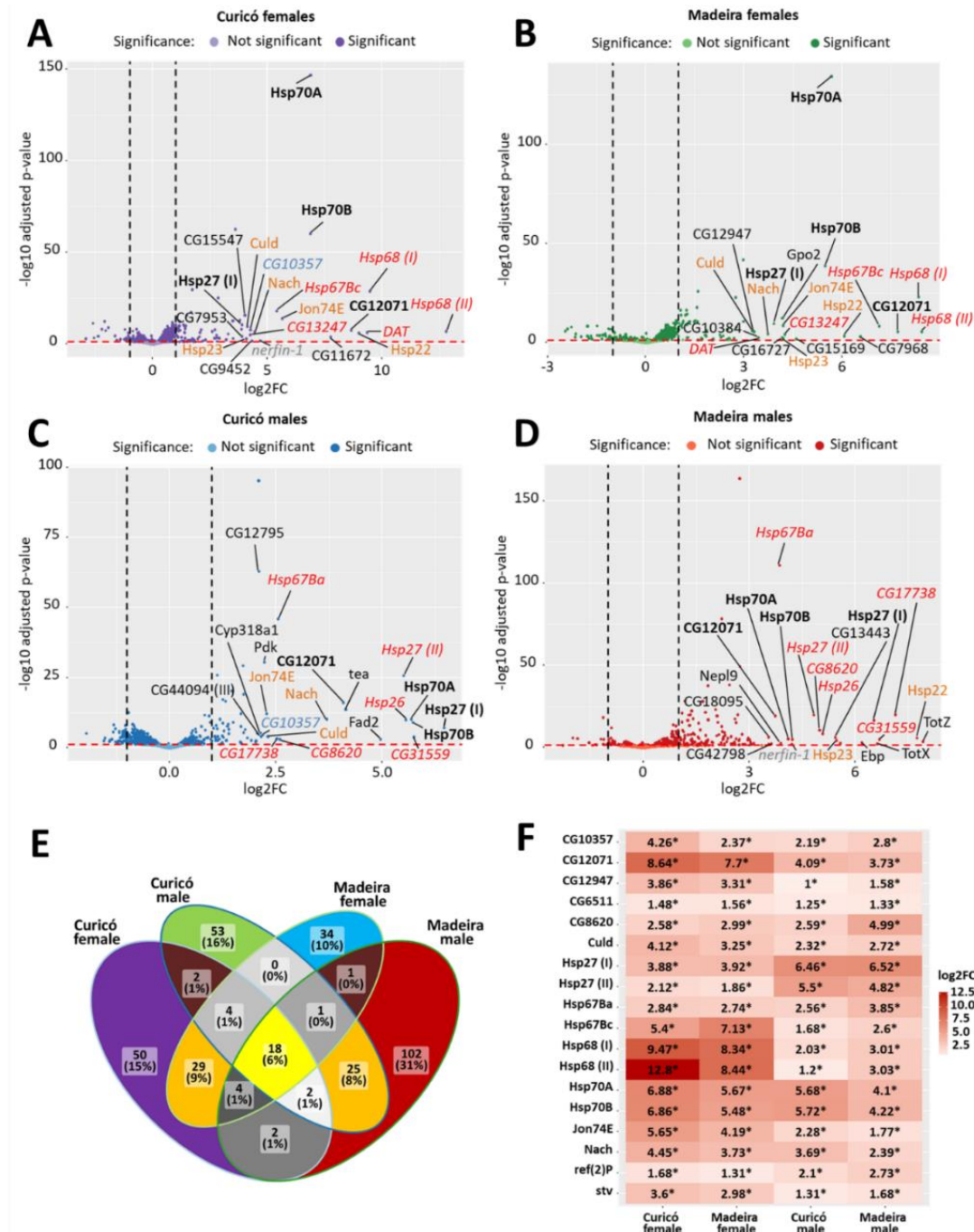
Fig. 2. Changes in gene expression in gonads after a heat shock. (*A-D*) Differential gene expression analysis in heat shock *vs* controls in Curicó females (purple, *A*), males (blue, *C*), and Madeira females (green, *B*) and males (red, *D*). Positive log$_2$FC values correspond to genes more expressed after a heat shock *vs* control. Only the names of genes showing the 20 highest log$_2$FC values and displaying and ortholog in *D. melanogaster* are shown. *Hsp70* is called as characterized in *D. subobscura* [428]. Genes common to all comparisons are in bold, to three comparisons in orange and to two are in italic (red color if shared by sex, blue if shared by population and grey if other combination). (*E*) Venn diagram showing the number of differentially expressed genes shared between populations and sexes. (*F*) Heatmap of the 18 differentially expressed genes after a heat shock shared in all populations (Curicó and Madeira) and sexes (males and females).

Heat shock effects on genes of heat response and Piwi pathway

To gain insight into the heat shock response, we studied in depth the heat shock genes encoding for proteins having an important role on stress response. In Figure 3*A*, we observed that most of the heat shock genes were overexpressed after a heat shock in the gonads of all populations and sexes. Some of them were significantly overexpressed in a similar rate in all samples (*Hsp67Ba*, both copies of *Hsp70*, and *Hsp83*). In the same way, *Hsp23* and *Hsp22* were also overexpressed in all conditions except in Curicó males, likely due to the low count in this sample. In contrast, the overexpression of other heat shock genes varies according to sex: both copies of *Hsp68* and *Hsp67Bc* copy were much more overexpressed in females than in males, and the opposite for *Hsp26* and both copies of *Hsp27* genes. Underexpression of the heat shock genes was also observed according to population: *Hsp10* and *Hsp60C* genes were underexpressed in Curicó females and *Hsp60B* in males, whereas *Hsp67Bb* was overexpressed in females from the same population. Regarding the heat shock cognate genes, constitutively expressed without a stress stimulation [569], we found less changes in expression than those observed in the heat shock protein genes. We detected an overexpression of one copy of *Hsp70-3* and *Hsp70-4* in males, *Hsc70Cb* in both males and Curicó females, *Hsp70-5* in Curicó females, and finally an underexpression of *Hsp70-1* in Curicó males.

Subsequently, we studied the changes in expression of other genes encoding for other heat response proteins (Figure 3*B*). We observed overexpression of *stv* (encodes a *Hsp70* co-chaperone during stress recovery) and *DnaJ-1* (encodes a heat-shock protein co-factor), in both cases the expression tends to be higher in females than in males. In Madeira population we observed overexpression of *l(2)efl* (a family of small heat shock protein genes). *Gp93* (encodes a heat shock protein *Hsp90* family) and *Stip1* (enable *Hsp90* protein binding activity) genes were only overexpressed in testes. In ovaries we also detected an overexpression of *Rme-8* gene, which belongs to the heat shock protein 40/*Dnaj* co-chaperones involved in the regulation of border follicle cell migration. Other genes (mostly from the Chaperonin Containing TCP-1 complex group) changed their expression after a heat shock only in one population or sex, but mainly in Curicó males. Finally, the Heat shock factor (Hsf), which binds heat shock promoter elements (HSE) and activates heat shock protein transcription, was not overexpressed after a heat stress.

We examined the expression of the piRNA pathway genes for their role in germline TE silencing. As observed in Figure 3*C*, the heat shock does not have effect on the expression of these genes in testes. Regarding ovaries, we observed few expression changes: overexpression of *piwi* and *hop* in both populations, and overexpression of *tapas* and underexpression of *Hel25E* in Curicó. In Madeira population overexpression of *tej* and *qin* and underexpression of *ci* are observed. All these gene expression changes had low $\log_2$FC values ($|\log_2$FC$| <1$), except *BoYb*

which expression is four-fold higher after a heat sock in the females of Curicó *vs* control (log$_2$FC of 1.97). It is interesting to highlight that other genes belonging to heat shock protein group (*Hsp83, Hsc70-4* and *Hsp70*), which changed their expression, have also an involvement in the piRNA pathway [192].
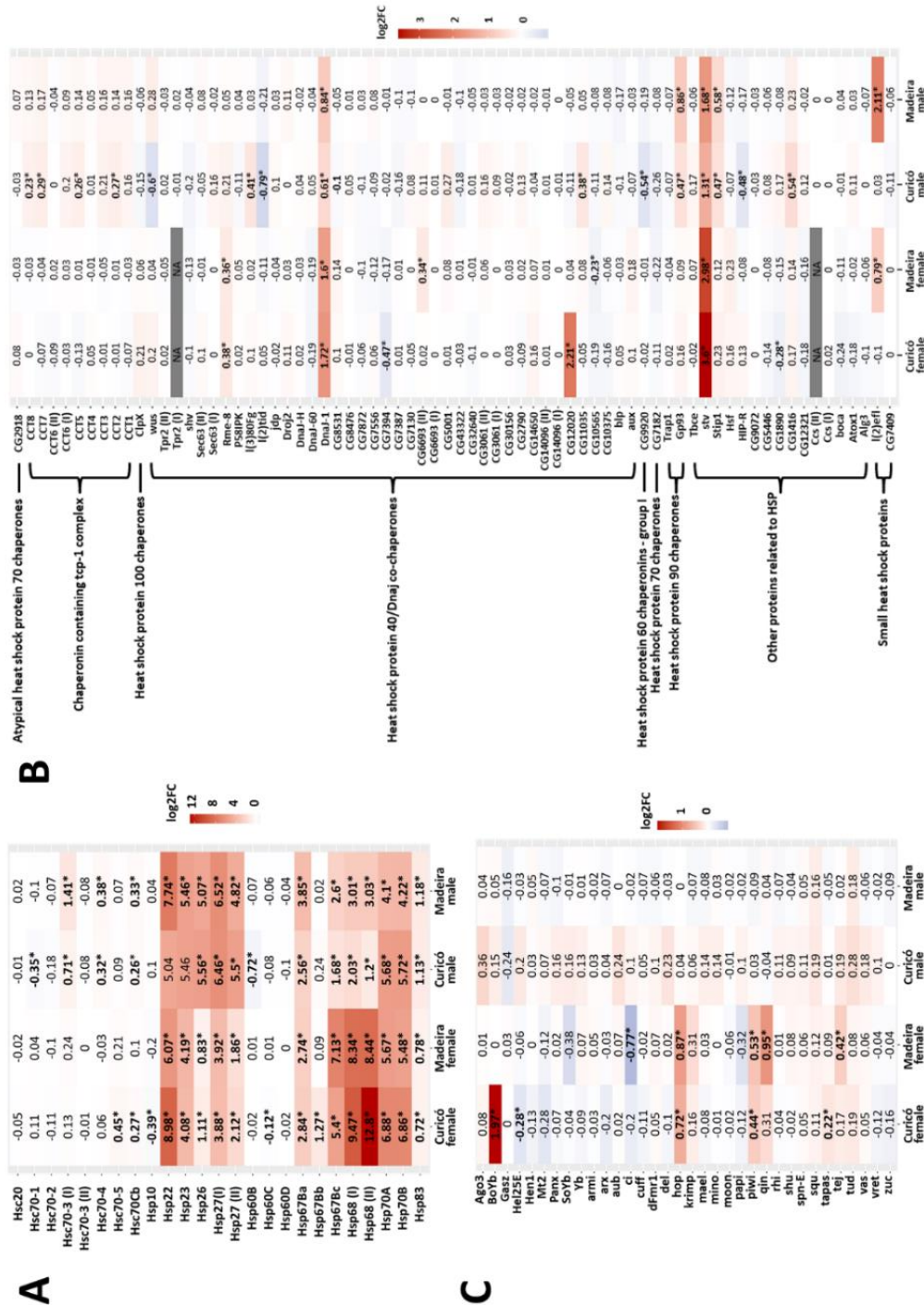
Fig. 3. Expression changes in gonads of specific gene groups in heat-shocked samples *vs* controls (log$_2$FC). (*A*) Heat Shock Protein genes. (*B*) Other genes related with the heat shock response grouped by Gene Group, as described in FlyBase. (*C*) piRNA Pathway Genes. Significant values (p < 0.05) are indicated with an asterisk and in bold. Colors indicate the values of the differences in gene expression (log$_2$FC). Values that could not be computed, due to low number of reads, are shown in grey (NA).

Only a few TE families change in expression after a heat shock

To detect if the heat stress activates TE expression in *D. subobscura*, we compared the TE transcription levels after a heat shock *vs* control in testes and ovaries of Madeira and Curicó populations. We found that from 0.77% to 1.75% of the expressed TE families changed their expression after a heat stress (supplementary Table S9), this percentage is similar both between sexes and between populations (Two Proportion Z-test, all p ≥ 0.45, supplementary Table S10). We observed a trend of TE families towards the overexpression (0.46-1.75% overexpressed *vs* 0-0.31% underexpressed, supplementary Table S9, and Figure 4*A-D*). After annotation of the differentially expressed TEs, we noticed that most of them were retrotransposons, mainly from the *gypsy* superfamily (Figure 4*A-D*). We also found that most of the TEs with changes in expression were specific of a population or sex (Figure 4*A-E*), but some differentially expressed TE families were only shared by sex (1 in females and 3 in males, Figure 4*E*). For example, a *gypsy* family was underexpressed in ovaries (Figure 4*A*, *B* and *E*) and other 3 TE families, one *gypsy* and two unknown, were overexpressed in testes (Figure 4*C*, *D* and *E*). However, only one overexpressed TE (*gypsy* superfamily) was commonly overexpressed in both sexes and populations (bold name in Figure 4*A-D* and 4*E*) after the heat shock. Finally, when we focused on the 23 differentially expressed TE families of Figure 4*F*, we detected some TEs with significant differences in expression in other comparisons but in a lower magnitude ($|\log_2 FC| < 1$). For example, *Polinton-1* and *TE-441* were also activated in all samples after this stress (Figure 4*F*).
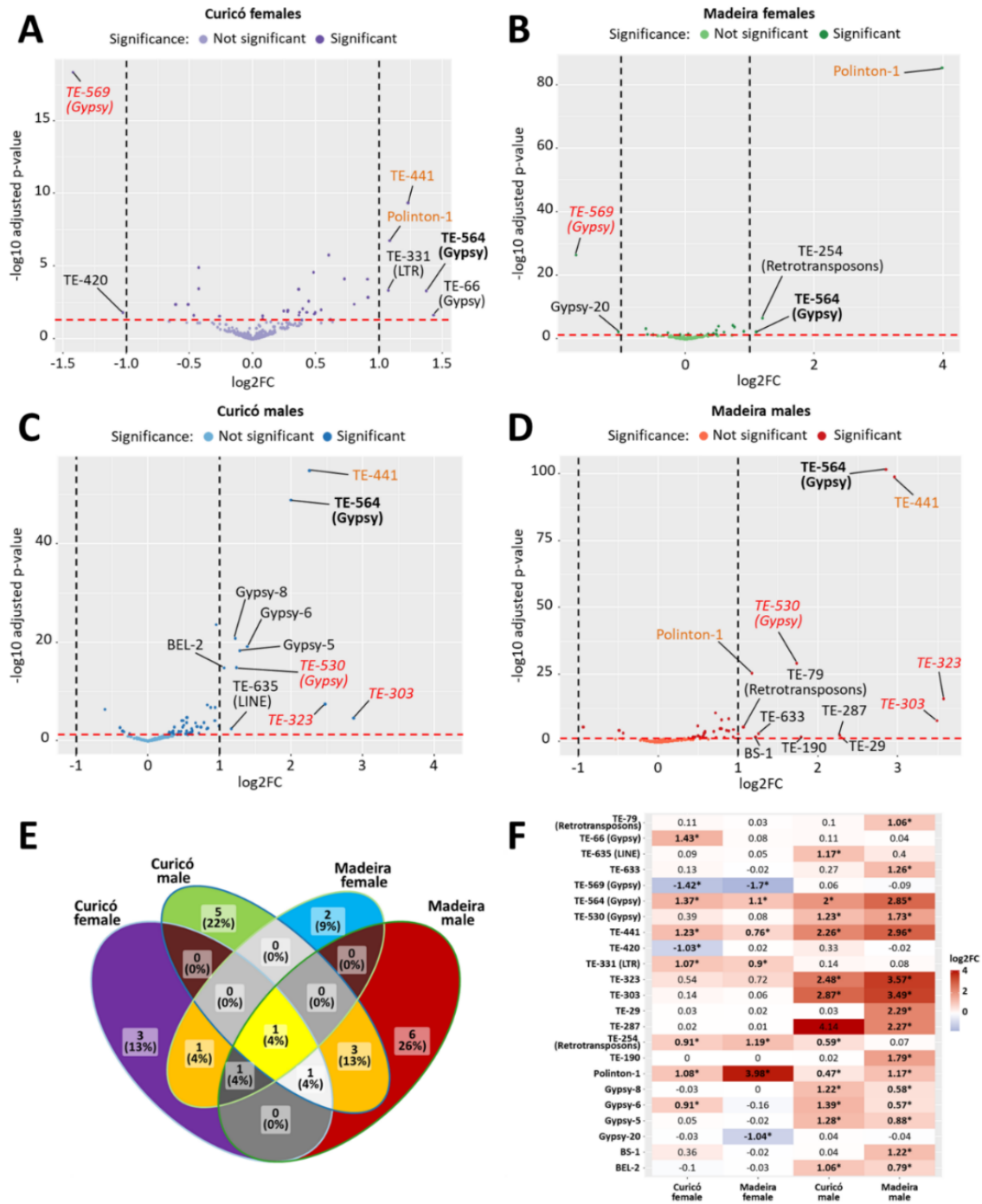
Fig. 4. Changes of TE expression in gonads after a heat shock. (*A-D*) Differential TE expression analysis in heat shock *vs* control in Curicó females (purple, *A*), males (blue, *C*), and Madeira females (green, *B*) and males (red, *D*). Positive log$_2$FC values correspond to TE families more expressed after a heat shock. The differentially expressed TE family or superfamily names are displayed. TEs common in all comparisons are in bold, orange in three comparisons and the ones shared by sex are in italic and red. (*E*) Venn diagram showing the number of differentially expressed TEs shared between populations and sexes. Brown regions are TE families shared by population, orange shared by sex, and yellow shared in all comparisons. (*F*) Heatmap of the 23 differentially expressed (|log$_2$FC| > 1) TE families after a heat shock in at least one population or sex. The superfamily of the unknown TEs, if annotated, is shown in brackets.

89

Impact of the heat stress on the small-RNA amounts

To study if the heat stress affected the small-RNA amounts in gonads, we first annotated the piRNA clusters in the reference genome of *D. subobscura* [124] to have an overview of their distribution along chromosomes. We used the software proTRAC [570], which predicts and analyzes genomic piRNA clusters based on mapped piRNA sequence reads, as explained in Material and Methods. We identify a total of 85 TE piRNA clusters using piRNAs from the ovaries of both populations (with a minimum overlap of 80%) at control conditions, covering a total of 1.29% of the reference genome and being 67 of them bidirectionally transcribed (supplementary Table S11, and data S1 available at https://figshare.com/s/f632c12591271effbf79). Most of these piRNA clusters were located close to the centromere, and most of the annotated ones had insertions mainly from the *gypsy* superfamily (Figure 5*A*). In addition, fifteen of the twenty clusters producing most of the piRNA in each population were the same in Madeira and Curicó (highlighted lines in supplementary Table S11), and they produced more than 70% of the total piRNA reads mapped in all clusters (74.03% in Madeira and 76.56 % in Curicó). The TE piRNA clusters that did not overlap at least 80% between populations or were detected only in one population were considered as unique of that population, and we found 58 unique clusters in Madeira and 73 in Curicó (supplementary Tables S12 and S13, and data S2 and S3, available at https://figshare.com/s/ac3d1dc5f9eb0dd3f2dc and https://figshare.com/s/80cce38193134e9305b2), covering a 0.80% and 0.84% of the reference genome, respectively. We then checked the production in testes and in control conditions of the 85 clusters identified in ovaries, and only 31 clusters in Madeira and 45 in Curicó seem to be expressed, being only 26 expressed in both populations and sexes.

We then studied the impact of the heat stress on the production of the 85 common clusters detected in ovaries (supplementary Table S11) and we found that this stress seems to increase the production of piRNAs for some clusters and to decrease for others. The general comparisons of piRNA production per cluster before and after heat stress, considering both kind of clusters, pointed to an effect of the piRNA cluster production only in Curicó females because in this sample most of clusters tend to decrease their piRNA production after stress (supplementary Table S14). However, when the clusters producing more piRNAs after heat shock, were separated from those with less piRNAs, we observed significant differences in all comparisons (one-tailed Wilcoxon signed-rank test, all $p \leq 0.001$, supplementary Tables S15 and S11). The effect of heat shock on piRNA cluster production tended to be different in males and females: more clusters that decrease piRNA production in females and the opposite in males (supplementary Table S15).
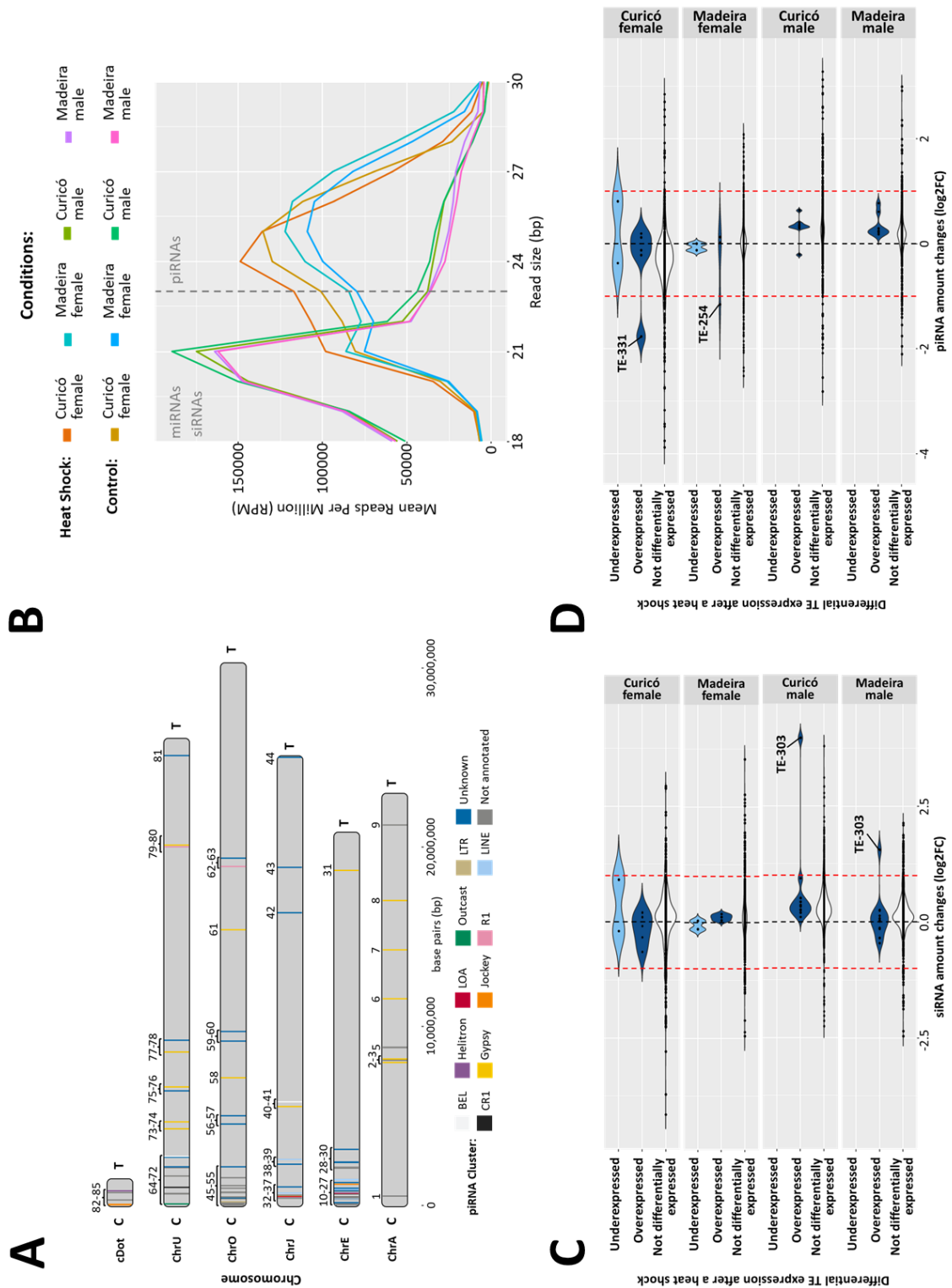
Fig. 5. Overview of piRNA clusters, small RNA quantities and their changes after a heat stress in gonads. (*A*) Location of the piRNA clusters by chromosome in the reference genome (C: centromere and T: telomere) and its annotation based in the main TE superfamily/class present. (*B*) Mean reads per million (RPM) of small RNAs in all samples. (*C-D*) Differential small RNA expression analyses in heat shock *vs* control by population and sex. Positive log$_2$FC values correspond to TE families with a higher amount of siRNAs (*C*) and piRNAs (*D*) after a heat shock. TE families are grouped by their differential expression after stress. Red-dashed lines show the log$_2$FC value considered as significant. The names of the differentially expressed TE families with significant differences in small RNAs are shown.

We then studied the global amounts of the small RNAs (small interfering RNA (siRNAs) and Piwi-interacting RNAs (piRNAs)) due to their role in the TE silencing in the *Drosophila* germline. An overview of the quantity of small RNAs, normalized by RPM, present in our samples (Figure 5*B*) showed higher quantities of piRNAs and lower of miRNAs and siRNAs in females than in the males, as expected due to the higher efficiency of this pathway in ovaries *vs* testes [162]. However, we could not find a significant difference in the total amount of piRNAs and siRNAs produced before and after a heat shock (T-test, all $p \geq 0.157$, supplementary Table S16). We then analyzed the impact of the heat stress on the small RNAs, normalized using miRNAs, associated to each TE family, and we noticed changes in their amounts after a heat stress (points outside the red-dashed lines in Figure 5*C* and *D*). We then checked if TE expression changes could be explained by the siRNA and piRNA amount changes after the heat stress, and we did not find a strong association between them, except in Curicó males that exhibited a trend to more TE expression and more piRNA amounts respectively (Fisher's exact test, supplementary Table S17 and Figures S5 and S6). In addition, changes in small RNAs mainly affected non-differentially expressed TE families (Figure 5*C* and *D*). When we focused on the differentially expressed ones, we noticed that most of these TE families had little changes ($|log_2FC| < 1$) in their small RNA amounts (both for siRNA and piRNA). Except the male overexpressed *TE-303* family, that showed an increase of siRNAs after a heat shock, and *TE-331* and *TE-254* families, overexpressed in the females of Curicó and Madeira populations respectively, that exhibited a decrease in their piRNA amounts after a heat stress (names in Figure 5*C* and *D*).

We finally analyzed the impact of the heat stress on the Ping-Pong signal. First, we compared the number of TE families with and without Ping-Pong signal in control *vs* stress conditions, and we could not detect a global impact of the heat shock in the Ping-Pong signal for the studied TE families (Chi-Square, all $p > 0.979$, supplementary Table S18). We then focused on the specific TE families whose Ping-Pong signal changed in control *vs* heat shock conditions, and we did not find a global tendency. In fact, the 18 TE families with changes on this signal were unique for each population-sex, except the *TE-637*, detected in Curicó females and Madeira males (supplementary Figures S7 and S8). In Curicó females, 5 out of 7 TE families only show Ping-Pong signal after stress (supplementary Figure S7). Nevertheless, 4 out of 5 TE families and all 3 TE families in Madeira females and Curicó males respectively only show Ping-Pong signal in control conditions (supplementary Figures S7 and S8). Finally, in Madeira males this signal is only detected in control conditions in one half of the TE families (out of 4 TE families, supplementary Figure S8). However, the differences in the Ping-Pong signal upon heat shock are very small and not significant, with some specific exceptions mainly unique by population-sex.

3.2.3   Discussion

Adaptative expression differences in two *Drosophila subobscura* populations

We studied the gene expression in the germline of the Madeira and Curicó populations of *D. subobscura* in control conditions and we found that, when populations were compared, 4.23% of the genes in females and 9.68% in males were differentially expressed. Contrasting percentages of gene expression differences have been reported when populations of different *Drosophila* species were compared, and our results showed a higher percentage [571], similar [572] or even lower [573] than what was previously described. However, the differences observed could be due to different methodology, *Drosophila* species, population origins and/or tissues analyzed. We also noticed more gene expression differences between testes than between ovaries, as was previously observed when two closely related *Drosophila* species from the *repleta* group were compared [574]. Regarding the function of the differentially expressed genes between populations, in addition to general GOs involved in many processes, we detected an enrichment of gene functions previously reported when two populations from different origins were compared, such as biological regulation [571]; signaling and response to stimulus [571], such as response to chemicals used in pesticides [572, 575] or temperature [575]; immune system processes [571, 573]; growth [573], such as body size [576]; rhythmic processes, such as the circadian rhythm [571]; and reproduction [571, 572]. These GOs could be related to the adaptation of *D. subobscura* to new environments during its colonization from the Palearctic region to America [45], where Madeira and Curicó are located respectively. However, because of the possible effects of the Madeira island colonization from the continent populations [577], it is difficult to have expectations of the differences between these two populations and compare our results with other studies. Additionally, inversion polymorphisms, especially on O chromosome, in this species have also been related to local adaptation to environment [51]. In fact previous studies in populations bearing different O arrangements showed differences in the Hsp70 protein amounts in control conditions [427] but invariable mRNA expression were detected in another study [428]. Our populations only differ on the U chromosomal arrangement (with both U_{1+2} and U_{1+8+2} inversions related to a "warm" thermal adaptation [47]), but they carry identical O chromosomal arrangement, suggesting that other factors as the genetic background could also be affecting the gene expression.

Regarding the TE content, we found that around 12.83% of the *D. subobscura* reference genome was covered by putative TEs, with a similar percentage of *DNA*, *LTR* and *LINE* elements. These results slightly differ from the previous reported annotation of repetitive content in the same reference genome [124], where more DNA TEs than the rest were found. These differences can be explained by the *de novo* annotation performed in this study, using a different methodology,

which allowed us to identify new TE families. In *D. subobscura* the families with highest insertion numbers were *Helitron*, *jockey* and *gypsy*. *Helitron* and *gypsy* superfamilies were also the most represented in other *Drosophila* colonizing species, for example *Drosophila suzukii* [578]. When the TE expression was compared in our populations, we detected almost twice as many TE families more expressed in Curicó than in Madeira. Differences in TE expression between populations are not uncommon and have been described in other *Drosophila* species [534]. Here the increase of TE expression in Curicó could be explained by the colonization of the American continent by *D. subobscura*. In fact, an increase of insertion site frequencies in this species [579] and of TE insertions, in other species [578, 580], have previously been observed in colonizing populations. Taking together the results of TE and gene expression, globally most of the differentially expressed genes and TEs between populations were unique by sex. However, low percentages of sex-specific differentially expressed genes between populations have been reported in other studies [573, 575]. These differences could be explained by the different tissues analyzed: we used gonads, which have much more differences in expression between sexes than other somatic tissues [581], including a different efficiency in TE regulation [161]. These results are also supported by the detection of more similarities in gene expression between populations than between sexes.

<u>Response to heat stress in the germline</u>

We found from 91 to 155 genes with changes in expression after a heat stress in our species, with a slight trend towards the overexpression. These results disagree with those of previous studies in *D. melanogaster,* where more expression changes and a trend towards the underexpression was observed [432, 582, 583]. However, these results are not actually comparable because some of them used different experimental methodologies: a longer heat stress [582] and a gradual increase of temperature [583]. In fact, it is believed that during a heat shock there is repression of genes not directly involved in heat shock control, which avoids misfolded protein accumulation, prioritizes the expression of the heat shock proteins, and protects the individual [584]. We have to highlight that most of previous heat shock studies were performed in whole flies [432, 582, 583], having little information on the real effect in germinal tissues. Thus, the differences observed between our results, performed in the germline, and the previous ones could be a consequence of the tissue-specific impact of this stress. It is possible that the stress affects the germline in a different way, for example the oogenesis processes [585], the sperm function and viability [586], or even the recovery once the stress disappears. Notwithstanding, a similar expression of the majority of genes in *Drosophila* ovaries and whole bodies was suggested in a previous work [464]. It is necessary to note that in this study we cannot completely discard a small bias in the fraction of differentially expressed genes due to the use of two biological replicates. In

addition, the detected gene activation in our study could be explained by the overexpression after heat stress in testes and ovaries from both populations of the *CG6511* gene, involved in positive regulation of transcription, whereas the underexpression detected could be related with the overexpression of *CG12071*, involved in the negative regulation of transcription by RNA polymerase II.

We found that the heat shock proteins were among the genes showing the highest overexpression after a heat stress, in agreement with what has been previously described in the literature [432, 582]. We also detected overexpression in all populations and sexes of genes that have been previously reported, such as *stv* [582, 583] and *ref(2)P* [582, 583]. We as well noticed other common genes with changes in expression whose functions were related to the heat response, such as gene expression [432, 583]. The overexpression of genes related to proteolysis and autophagy activation by ubiquitinated proteins, could also be a response to the accumulation of unfolded proteins in the cell related to this stress [553], and a path to destroy them. In addition, differentially expressed genes after a heat stress were enriched in general processes, as well as other previously described GOs involved in thermal adaptation, such as biological regulation [583]; developmental [583]; and immune system processes [582, 587]; rhythmic processes, such as circadian function [587]; and response to stimulus, such as response to heat or stress [432, 583]. We also noticed overexpression of genes related to locomotion, which could be related to the loss of coordination characteristic during heat stress together with the recovery of the locomotion functions once it disappears, and which is supported by the early temperature failure of *D. subobscura* [588] related to a shorter recovery time [589]. Finally, an enrichment of underexpressed genes involved in reproduction was found in males and may be explained by the impact of this stress in the fly fertility [586]: male sterility in *D. subobscura* can be induced at 25°C [32].

Unsurprisingly, we found differentially expressed genes after a heat shock unique for each population-sex, as already observed in control conditions when populations and sexes were compared. We also detected a noTable amount of sex-specific differentially expressed genes after heat stress. This could also be related to the differential thermal adaptation observed between males and females of *Drosophila* in previous studies [590], the higher heat resistance observed in *D. subobscura* females [41] and the sex-specific response after a heat stress observed in other species [436, 591]. It could also be influenced by the differential expression of genes involved in chromatin remodelling in females but not in males. These sex-specific expression differences also affected genes encoding some heat shock proteins. Differences in the induction of a few small heat shock proteins in testes and ovaries were previously described [592], but our findings did not completely match theirs: while they could not detect a heat stress protein induction of *Hsp23* in testes and *Hsp27* in both ovaries and testes, we detected an increase of mRNA amounts. We did

not find a strong impact of the heat stress on the heat shock cognate genes, as expected, due to their constitutive and non-heat inducible expression [569], or other genes related to the heat response. We also could not detect an overexpression of the Heat shock factor (*Hsf1*) either, contrary to what was observed in a previous study [583]. However, this transcription factor is already present as an inactive monomer without stress and trimerizes during stress to bind to the Hsp promoters [584], making its overexpression not necessary to perform its function. Regarding the piRNA pathway genes, we could only find a small impact of the heat stress in females, opposite to the strongest impact observed in *Drosophila* during genomic stress induced by hybridization [365] and after a 48h housing at 29 °C [464]. The only exception was *BoYb*, essential for the primary piRNA pathway in the germline [207] and four times more expressed in Curicó females after the heat shock, which could be involved in the piRNA cluster production decrease detected in Curicó ovaries. In addition, the heat shock genes involved in this pathway (*Hsp83, Hsc70-4* and *Hsp70*) [192] changed their expression, except for *Droj2* [192]*,* suggesting that the piRNA production could be affected in some way in all populations-sexes even though other genes of the pathway were not.

Limited association between TE activation and small RNA amount changes after heat stress

We used a whole transcriptomic approach to study the impact of the heat stress on the TE expression in *Drosophila*, and we found a moderate impact of this stress on the TE expression in our species, with an activation from 0.46% to 1.75% (3 and 12 respectively) of TE families in at least one population. We have to highlight that the actual effects of heat stress on the TE activity did not seem very clear: increases of transcription [192, 468, 560] or transposition [460, 558] of some specific TE families were detected in some cases and absence of transposition induction in the same or other TE families [463, 562] were also reported in *Drosophila*. These results, together with the differences in transposition observed between TE families and even between individuals of the same stock after a heat stress [460], pointed out the importance of the genomic context [463] or even the TE families studied. This could also explain why we only detected three common TE families overexpressed in different degrees in all population-sexes, the rest changing their expression uniquely in one population and sex. When sexes were compared, we noticed more activation after heat stress in males than in females as described in a previous study [192], and which could be explained by the higher TE expression in testes than in ovaries in control conditions and the differences in TE regulation between *Drosophila* sexes [162]. The higher TE activation in males *vs* females could also be related to the presence of the Y chromosome in the former, known for its high abundance of TEs in comparison to other chromosomes [593]. In addition, we detected more differentially expressed TE families shared between sexes than between populations. We also found two TE families that decrease their expression after a heat

stress in ovaries from both populations in concordance with a previous study in *Drosophila* ovaries [468]. In addition, we detected that most of the families with changes in expression were retrotransposons, mainly from the *gypsy* superfamily or family according to previous studies on retrotransposons in *Drosophila* [192, 460, 468, 558, 560].

We then annotated the TE piRNA clusters in the *D. subobscura* reference genome, and we found 85 common clusters in the females of both populations covering around 1.29% of the reference genome, and 58-73 cluster unique of Madeira and Curicó, covering around 0.80-0.84% of the reference genome respectively. This total percentage of around 2.10% of the genome annotated as piRNA clusters is a bit smaller than the 3.5% found in *D. melanogaster* [151]. However, differences in the detection method, as well as the higher TE percentage of *D. melanogaster* [594], and its lower genome compaction [595], in comparison to *D. subobscura* [124], can explain these differences. Nevertheless, the total number of piRNA clusters per population was similar to *D. melanogaster* (142 piRNA clusters) [151]. Consistently with what was previously observed in *D. melanogaster*, most of the piRNA clusters found in this study were dual-stranded [147, 163] and located in the pericentromeric area [147, 151], which was also the richest TE insertion region in the reference genome of *D. subobscura* [124]. We found that 15 of the most producing piRNA clusters shared in both populations already produced more than 70% of the total piRNAs, as in other *Drosophila* species [151, 163]. Additionally, we detected that slightly less than half of the piRNA clusters identified in Madeira and Curicó, were unique for each population, which could be explained by the fast piRNA cluster evolution suggested in previous studies [163]. piRNA cluster differences within species were also found in other studies but in a lesser extent (20-30%) [154]. Nonetheless, the higher percentage detected in our study could be explained by the difference in the detection method used here in comparison to the other study [154]. On the other hand, we detected less piRNA clusters expressed in males, which is consistent with the lower piRNA production in testes and the detection of more active piRNA clusters in ovaries [162]. We additionally found that the heat stress had a general impact on the expression of the TE piRNA clusters leading to a general decrease of piRNA production in Curicó females. In all other samples, a significant increase of piRNA production was observed in some clusters and a decrease in others, which agrees with the impact in the production of some piRNA clusters previously found after heat stress in *Drosophila* [468]. However, considering that no changes in piRNA levels were detected globally, further explorations are necessary to fully understand the production of piRNAs by piRNA cluster.

When we studied the small RNA expression in ovaries and testes, we found higher amounts of piRNAs in females than in males, as expected according to the previously results reported in *Drosophila* [162], considering the higher efficiency of the piRNA pathway in the ovaries. Even though we could not find global significant differences in the total amount of

piRNAs and siRNAs produced before and after a heat shock, in contrast to other genome-wide small RNAs study [464], we noticed changes in the amount of both small RNAs targeting specific TE families. The effect of the heat stress on the small RNAs in *Drosophila,* has not been extensively studied and in the cases where it has been, results were different. For example, significant changes of piRNA amounts of some specific TE families were detected in a previous study after a heat shock [468], whereas no significant effect in the piRNAs targeting other set of TE families was shown in another [192]. In addition, a transcriptome-based small RNA study found temperature-dependent changes of most transposon-derived piRNAs [464]. However, these studies were performed in *D. melanogaster* and there were also differences in the heat stresses performed in this study and the previous ones: short heat stress [468] and three heat shock repetitions [192] or 48h housing at 29°C [464]. The effect of the heat stress on the siRNAs has been even less studied: in *Drosophila,* the only available study reported an increase of the siRNAs associated to some genes [466] whereas in plants an increase of siRNA amounts targeting some genes and a decrease of the amounts targeting other genes has been reported [596]. Regarding the influence of the small RNA amount changes on the TE expression, we could not find a clear association of both siRNA and piRNA amount changes and TE expression changes. Studies of the effect of heat stress in the siRNAs targeting TE families and their influence in TE expression has not been performed yet. Nonetheless, in other studies detecting changes in piRNA amounts after heat stress no correlation between changes in piRNA levels and TE transcripts was already reported [464, 468]. This could be expected by the complexity of the piRNA-mediated TE regulation, which includes both a transcriptional and post-transcriptional silencing, making difficult to detect a direct correlation between TE expression and TE silencing mechanisms [597]. However, we detected some cases of small RNAs and TE expression changes after stress: two overexpressed TE families with a significant decrease of their piRNA amounts in females belonging to different populations, the overexpression of a TE family with higher amounts of siRNAs in males from both populations and a positive general association between piRNA amount changes and TE expression in Curicó males, being this last one previously observed in control conditions [598]. Finally, we studied the Ping-pong signal in our samples, and we detected signature in both females and males. Even though this signal has been more studied in females, it has also been detected in testes [162], and piRNAs with Ping-Pong signature seems to be even more abundant in *Drosophila* spermatogonia [599]. When we studied the effect of the heat shock on the Ping-pong signal, we could not find a general effect, finding only 18 TE families with changes on this signature. This result is in concordance to what was observed by other authors, where no significant general effect of the heat shock on this amplification cycle was observed in the studied TE families, with some exception [468]. However, other study detected an increase of this signal after 48h *Drosophila* housing at 29ºC [464], which could show an impact in this pathway after a longer stress.

Even though we noticed that changes in expression of few specific TE families could be explained by their small RNA amount changes after heat stress, we did not detect this clear association globally, which was in a concordance with other studies [464, 468] and highlight that other mechanisms could also be involved in this activation. For example, changes in the epigenome affecting TE expression under the genomic stress, have been observed previously [600]. Specifically, *Hsp83* was described to be involved in epigenetic modification [415], as well as other changes in the epigenome have been observed after a heat stress [553, 601]. Nonetheless, other study shown that H3K9me3 and H3K27me3 did not significantly change after a heat shock in seven TEs studied in the germline [192]. They propose that the interaction of Hsp70 with the chaperone-Ago3 complex in the germline induce the displacement of all factors to the lysosomes resulting in a functional collapse of piRNA biogenesis, that could contribute to TE activation. Finally, the location of TEs in the heat shock promoters have been described in different heat shock proteins in *Drosophila* [602, 603]. The insertion of TEs near these genes could result in an increase of their transcription after a heat stress. For example, three small TE sequences have been detected next to the promoter of the *Hsp83* gene in both of our populations (supplementary text S1) showing that these insertions are also present in our species. Nevertheless, these TEs were not activated after a heat stress. Further work in the germline under heat stress needs to be performed to clarify the mechanism for TE activation, in addition to the role an impact of TE expression in stress adaptation.

3.2.4 Material and methods

Drosophila Stocks and crosses

We founded isofemale lines from two *D. subobscura* stock populations, one from an original population from Madeira island (Portugal) and another one from a colonizer population from Curicó (Chile). Flies were laboratory-maintained by mass culturing in a standard *Drosophila* medium supplemented with yeast, a 12:12 light/dark cycle and the optimal temperature of 18ºC for this species [46]. 20 generations later, crosses of 25 males x 25 females changing the medium every 3-4 days were performed to control larval density. In order to determine the inversion polymorphism of the lines under study, males were individually crossed to three virgin females from the cherry-curled recessive marker strain (*ch-cu*). This strain is homokaryotipic for the standard arrangement in all chromosomes, except the O chromosome, which is homozygous for the O_{3+4} arrangement [124].

Chromosomal inversions

For the line inversion identification, we first incubated clean and dry slides in an 3X SSC/1X Denhardt's solution for 2 hours and 30 minutes in a water bath at 65ºC. Slides were cleaned with distilled water and immersed in ethanol/acetic acid (3:1) at room temperature for 20 minutes. After air dry, slides were stored at 4ºC until use. The coverslips were siliconized with Repel-Silane (Amersham) solution for 10 seconds, washed in ethanol for a few seconds and distilled water. Third instar larvae salivary glands were dissected in NaCl 0.8%, immersed in acetic acid 45% for one minute and fixed in a 1:2:3 solution (lactic acid: water: acetic acid) for 4-8 minutes. Preparations were observed under a phase contrast microscope.

Stress treatment and gonad dissection

Heat shock stress experiments were performed in 5-day-old virgin males and females placed in sealed empty vials and immersed in water-baths at 32°C for 60 minutes. Then, they were kept at 18ºC for 30 minutes before gonadal dissection in PBT (1× phosphate-buffered saline, 0.2% Tween 20). Non-heated samples were maintained at the optimal temperature of 18ºC and then 5-day-old individuals dissected. All gonad samples were frozen in liquid nitrogen and stored at -80ºC until RNA extraction.

mRNA and small RNA extraction, Library Preparation, and Sequencing

Total RNA was purified from 20-25 pairs of testes and 6-9 pairs of ovaries per sample with the Nucleospin RNA purification kit (Macherey-Nagel). Samples were sent to GenomEast for library preparation and sequencing. Duplicate TruSeq Stranded mRNA libraries, corresponding to two biological replicates per sample and temperature condition, were prepared. Finally, a paired-end sequencing was performed for 100 bp read length using the Illumina HiSeq4000 technology. Adapter dimer reads were removed using DimerRemover (available in https://sourceforge.net/projects/dimerremover/). We obtained 47-117 million paired-end reads for each sample, resulting in a total of 1,192 million paired-end reads.

For the small RNAs, a manual extraction of total RNA from 40-45 pairs of testes and 6-10 pairs of ovaries per sample, was performed using QIAzol (QIAGEN) and the phenol-chloroform method. Small RNA separation, library preparation and sequencing were performed by Fasteris SA. Small RNAs of 18-30 nucleotides were purified using polyacrylamide gels and then the libraries were prepared (2 replicates/sample) using the Illumina TruSeq small RNA kit and a specific treatment anti 2S. Finally, a single-end Illumina Sequencing was performed for 50 bp reads using NextSeq500. Bases that correspond to the adapters were removed using Trimmomatic v0.32 [524], using the trimming options seedMismatches: 2,

palindromeClipThreshold:30 and simpleClipThreshold: 5, and inserts were sorted in separate sequence files according to their size. We obtained 25-45 million single-end reads for each sample, resulting in a total of 459 million single-end reads.

De novo annotation of TE families and piRNA clusters in the reference genome

RepeatModeler [604] and EDTA [605] with default parameters were used to annotate TE families *de novo* in the *D. subobscura* reference genome [124], detecting 313 and 542 consensus sequences respectively. Consensus sequences smaller than 100 bp (37 sequences) were removed, and a MegaBLAST [606] of all sequences against themselves was performed. Consensus sequences were clustered as the same TE family when an identity and an overlap of at least 80% between sequences of the cluster was detected. Promiscuous sequences belonging to many clusters (33 sequences) were removed. A total of 785 consensus sequences of 702 different putative TE families were obtained. All consensus sequences were masked using a slow search of RepeatMasker v4.1.2 [607] with the --*norna* and --*nolow* parameters and as a custom library the RepBase [608] database including all described TE *Drosophila* sequences. With these results, we annotated the TE consensus families: more than 80% of the sequence was masked by a specific TE family, or the superfamilies: percentage of 50%. We then used RepeatMasker v4.1.2 [607] with the same parameters above and the list of TE consensus *de novo* annotated as the custom library, to detect all TE insertions in the reference genome. To merge insertions of the same TE, the script "One code to find them all" [609] with the --*unknown* and --*insert* 100 options was run. Then, all insertion sequences were included in a fasta file, and the class, family and superfamily of each insertion in a rosette file, as required by the TEcount module of TETools [534].

For the first global piRNA cluster annotation both replicates of the same population were merged and only female samples (higher piRNA production [162]) and control conditions were used, to avoid possible expression changes due to the heat treatment. proTRAC [570] software was used for piRNA cluster detection. It applies a sliding window approach to detect *loci* that exhibit high sequence read coverage and then analyzed them with respect to typical piRNA and piRNA cluster characteristics to ensure high specificity [570]. We followed the software recommendations: TBr2_collapse.pl script to remove redundant sequences; TBr2_duster.pl to remove low-complexity reads and sRNAmapper v1.0.5 [610] for piRNA mapping in the reference genome keeping only the best alignments. Reallocate.pl (available at https://www.smallrnagroup.uni-mainz.de/software/reallocate.pl) was run to allocate read counts of multiple mapping sequences according to the genomic region transcription rate. Finally, the identification of the clusters was performed using proTRAC v2.4.4 [570] with the specific options of --*pdens 0.05 --clsize 5000 --pimin 23 -1Tor10A 0.3 -clstrand 0.5*, the RepeatMasker output of the TE insertions and the transcriptome annotation in the reference genome. These specific

options were previously used to identify clusters in other *Drosophila* species [163]. To compare the annotated clusters in each population, the regions were converted in a bed file using *convert2bed* of BEDOPS v2.4.38 [611] and the bedtools v2.29.2 [528] *intersect* command. We annotated the clusters as common to both populations if their coordinates overlapped at least 80%, and the rest were considered as population specific. The same *intersect* command was used to study their overlap with exons (using the transcriptome annotation) and TE family insertions, and clusters were manually curated considering their overlap with genes and TE families. They were also annotated if most of the insertions were from a specific superfamily or class (more than 50% of TE insertions in the cluster). In order to compare the results between samples, the number of mapped reads per cluster was extracted from the proTRAC output, using the Reads Per Million (RPM) normalization. Although the global clusters were annotated in females in control conditions, cluster detection was also performed in males and under heat stress conditions in both sexes (coordinates overlapping at least 80% with the annotated clusters), and mapped reads per cluster were extracted from the proTRAC output.

Gene and TE differential expression analyses

RNAseq sequenced reads were trimmed using Trimmomatic software v0.39 [524], with the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. To study gene expression, trimmed reads were aligned to the masked *D. subobscura* reference genome [124] using STAR v2.7.9a [612] with the *--quantMode* to count the reads per gene using *htseq-count* [613] and giving the transcript annotation available in the browser. TE expression was then analyzed using the TEcount module of the TEtools pipeline [534]. First, the RNAseq data was aligned with the fasta file including all TE insertions in the reference genome of *D. subobscura* using Bowtie2 v2.2.4 [525] with the most sensitive option and keeping a single alignment for reads mapping to multiple positions (*--very-sensitive*). Read counts, per TE family, were computed adding all reads mapping on copies from the same family. Count Tables, corresponding to genes and TEs, were concatenated and then used for the differential expression analyses. In this way, gene counts were used to normalize TE counts, following the guidelines of TEtools pipeline [534]. *DESeq2* function from the R Bioconductor package DESeq2 v1.34.0 [537] was used to normalize read counts, using the median of ratios method, and read counts were modeled using a negative binomial distribution.

DESeq2 v1.34.0 [537] was also used to identify differentially expressed genes and TE families between populations (Madeira *vs* Curicó) and conditions (heat shock *vs* control) performing a Wald test [537]. The p-values were obtained using the *results* function from the DESeq2 package and adjusted for multiple testing, using the procedure of Benjamin and Hochberg [538] with an FDR cutoff of 0.05. The $\log_2$FC was shrunken using the default and

recommended *apeglm* algorithm [539] of the *lfcShrink* function. Genes with an adjusted p-value lower than 0.05 and at least double difference of expression, between the evaluated conditions (|shrunken log$_2$FC| > 1), were considered as differentially expressed. GO term enrichment analyses of biological processes were performed for the significant differentially expressed genes using the topGO R package v2.46.0 [540] ("*weight01*" algorithm and Fisher's statistic). GO belonging to each gene were obtained using eggNOG [614]. Because of the high impact of the p-value adjustment, only the top $\frac{1}{4}$ 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 GO of biological process. In addition and to avoid false positive TE annotation, differentially expressed TE families were also analyzed with TEAid v4.28.21 software [615] (available at https://github.com/clemgoub/TE-Aid) and online BLAST (*nblast* and *blastx*) [548].

Finally, to confirm heat shock and piRNA pathway gene annotation and/or annotate the ones not previously annotated, a reciprocal *tblast* v2.10.1. [548] of the *D. melanogaster*, *D. pseudobsucra* or *D. subobscura* (if available) proteins, downloaded from UniProt [549], was performed against the putative gene sequence in the reference genome. A total of 17 heat shock, 8 heat shock cognate and 34 genes involved in the Piwi pathway were annotated in *D. subobscura*.

Small RNA analyses

The miRNAs were first annotated in *D. subobscura* aligning each miRNA precursor annotated in *D. pseudoobscura* available in miRBase [616], with the *D. subobscura* reference genome using *blat* [547]. The best alignment (*pslReps*), was then retrieved and converted in bed format (*convert2bed* of BEDOPS v2.4.38 [611]) and a fasta file with all sequences in our reference genome (*getfasta* of bedtools v2.29.2) was obtained. We then mapped the small RNA-seq reads of 20-23 nucleotides to the miRNA fasta sequences annotated in our reference genome, using Bowtie v1.3.0 [617] with the most sensitive option (-*S*) and keeping a single alignment for reads mapping to multiple positions. Counts were computed using eXpress v1.5.1 [533]. To ensure that miRNA production was not affected by stress, the miRNA aligned reads were normalized by Reads Per Million (RPM), and their amounts, before and after stress were compared, founding that the differences were not significant (T-test, all p ≥ 0.705, supplementary Table S19). Small RNA data was then analyzed with the TEcount module of the TEtools pipeline [534], reads of 21 nucleotides (siRNAs) and 23-30 nucleotides (piRNAs) were aligned to our custom TE library using Bowtie v1.3.0 [617] with the most sensitive option (-*S*), keeping a single alignment for reads mapping to multiple positions, computing read counts per TE family and adding all reads mapped on copies of the same family. Finally, counts were normalized using miRNA total counts, and the differences in small RNA amounts in heat shock samples *vs* control was computed

manually calculating log$_2$FC. TE families with 2-fold differences in their piRNA/siRNA amounts were considered significant.

Finally, we studied the Ping-Pong cycle only in the TE families with enough piRNA amounts, which are the ones with a sum of piRNA count in all samples higher than the length of the TE family consensus, for each sex independently. The longest insertion of a TE in the reference genome was considered as the consensus. With this methodology, we selected only 219 families to detect the Ping-Pong signal. We then aligned our piRNA reads (23–30 nucleotide) against these TE family consensus using Bowtie v1.3.0 [617] (-*S* option) and we checked for the presence of 10-nt-overlapping sense–antisense read pairs using the *signature.py* pipeline [618]. The mean z-scores and overlapping pair reads for each overlap between the two replicates of each sample was computed. We then selected the samples with a number of overlapping pairs equal or greater than 50, to avoid miscalculations by low read count. We considered that there was Ping-Pong signal when the z-score for the 10-nt overlap was larger than 2 and for the rest of overlaps equal or lower than 2.

Statistical Tests and visualization

Five main statistical tests were used in this article and all were performed using R v4.1.3 [535]. The Two proportion Z-test was used to compare the distributions of significant genes and TE families across comparisons. The Spearman's rank correlation was used to compare general gene and TE expression between sexes and populations. The Wilcoxon signed rank test was used to compare the cluster production of piRNAs before and after a heat stress (two and one-tailed). The T-test was used to compare the normalized aligned count reads in heat shock *vs* control for small RNAs. The Fisher's exact test under independence assumption was computed using a 2×2 contingency Table to detect associations in TE expression changes in heat shock *vs* control and the corresponding small RNA amount changes. Finally, the Chi-square was used to compare the number of TE families with and without Ping-Pong signal in control conditions and after heat stress. All the results were corrected for multiple testing using the Benjamin and Hochberg [538] method and the plots and visualization of the results were performed using the R package *ggplot2* v3.3.5 [536].

Acknowledgments

Data Availability

Illumina RNA and small RNA sequencing data are accessible in the links: https://uab-my.sharepoint.com/:f:/g/personal/1361248_uab_cat/EuIkxF516UpKuD1Iy0QVlAcBKJcnNmLsUaqUsYS2JDdHUQ?e=56qJ7 and https://uab-my.sharepoint.com/:f:/g/personal/1361248_uab_cat/EsKqIVkvE1tNjKzdPhlak8QB4vk92nMMP1Uhld6BoDbSYw?e=bOTBVN. They will be deposited in the Sequence Read Archive after acceptance. proTRAC outputs for annotated piRNA clusters are available in the accompanying repositories: data S1 containing shared piRNA clusters in Madeira and Curicó at https://figshare.com/s/f632c12591271effbf79, data S2 containing piRNA clusters annotated only in Madeira at https://figshare.com/s/ac3d1dc5f9eb0dd3f2dc and data S3 containing piRNA clusters annotated only in Curicó at https://figshare.com/s/80cce38193134e9305b2).

Supplementary Material

The supplementary data is available in Annex **8.2**:

Supplementary Tables. Numeric results and statistical tests and piRNA cluster description of shared and unique clusters in Madeira and Curicó. Annex **8.2.1**.

Supplementary Figures and text: file containing all supplementary Figures and text containing *Hsp83* gene sequence in Madeira and Curicó. Annex **8.2.2**.

4. Discussion

4.1 *D. buzzatii-D. koepferae* interspecific hybridization

4.1.1 Differences in gene and TE expression between *D. buzzatii* and *D. koepferae* species

Genes

We compared the gene expression in ovaries between two cactophilic sibling species *D. buzzatii* and *D. koepferae* (section **3.1.2**) and we detected that 5.92% of the annotated genes were differentially expressed (see Table **S1** in Annex **8.1.1**). Differences in gene expression between *Drosophila* species were already reported in previous studies: 78% of the expressed genes showed divergent expression between *D. melanogaster* and *D. sechellia* (divergence time ~1.2 millions of years ago (Mya)) [489], whereas this percentage decreased to 43.7% between *D. melanogaster* and *D. simulans* (divergence time ~2.5 Mya) [474]. Nonetheless, a comparison of the transcriptome of two *repleta* species, *D. mojavensis* and *D. arizonae*, showed that approximately 6% of the genes were differentially expressed in ovaries (divergence time 0.6-1 Mya) [330, 574]. Although the divergence time is related to the differences in expression, other factors such as the tissues analyzed, and the methodology may affect the results. In fact, the first two studies used whole flies [474, 489] whereas ovaries were used in the others [330, 574]. The methodologies were also different: RNA-seq in three of them [330, 489, 574] and microarrays in the other [474]. In the same way all studies applied different bioinformatic approaches for differential expression analyses [330, 474, 489, 574]. The transcriptomic differences between our parental species are in line with the 6% obtained between other species of the *repleta* group, above mentioned [330, 574], and our methodology and approach is also more similar to that study than to the others.

When we focus on the most differentially expressed genes between *D. buzzatii* and *D. koepferae* (see Fig. **S1A** in Annex **8.1.3**), most of them are involved in transport (*CG8791* and *CG44098*), metabolic processes (*CG17999* and *CG15343*), nervous system processes (*Nrx-1*), immune response (*SP2353*), developmental processes (*cyr*, *Fbp2*, *Hs3st-B* and *Doc1*), gene expression and protein metabolism regulation (*CG4810* and *Phlpp*) or behavior and circadian rhythm (*DAT*). Coincidentally, genes with similar functions were differentially expressed in previous studies between *D. mojavensis* and *D. arizonae* [330, 574], and in both cases these genes could be related to the adaptation of these species to new environments. In addition to these functions, we also found differentially expressed genes between *D. buzzatii* and *D. koepferae* related to xenobiotic transport and metabolism, such as solute carrier transporters (*CG8791* and *CG44098*), glycosyltransferases (*Ugt86Dd*) and oxidoreductases (*CG15343* and *Fbp2*).

Differences in the expression of genes involved in xenobiotic transport and metabolism, were also reported in a recent study in these species when third-instar larvae were reared in primary and secondary cacti hosts containing different chemical (alkaloid) and nutritional (dead-yeast extract) compound modifications [21]. Indeed, the detected differences in expression between these two species involved approximately 7-10% of the genes [21], being the value very similar to that found in our work. Nonetheless, in our study both species were fed with a *Drosophila* standard medium, and it is known that in *D. koepferae* the viability declines and the developmental time increases in experimental media *vs* cacti media [15]. This idea is also supported by the greater transcriptional plasticity detected in *D. buzzatii vs D. koepferae* after the exposure to the different nutritional and chemical components, and hence *D. koepferae* is considered more a specialist [15]. Interestingly, we also detected an overexpression of the Alkaline phosphatase 10 (*CG10592*) in *D. koepferae* in comparison to *D. buzzatii* (see Fig. S1A in Annex **8.1.3**). This enzyme catalyzes the dephosphorylating, by hydrolysis, of molecules such as alkaloids [619], and its activity is increased by the consumption of these compounds [620]. Therefore, the differences in expression of these genes could be related to the adaptation of these species to their specific hosts [4, 7, 11, 12], specifically the adaptation of *D. koepferae* to the use of columnar cacti as a preferred host, rich in alkaloid compounds [15].

We also studied the differences in the chromatin landscape between our parental species, for the euchromatic mark H3K4me3 and the heterochromatic marks H3K9me3 and H3K27me3 [149]. We found that 2.50% of the genes showed differences between the parental species in H3K4me3, 1.67% in H3K9me3 and 2.70% in H3K27me3 (see Table S6 in Annex **8.1.1**). We concluded that the chromatin landscape of both parental species was similar and also similar to that of *D. melanogaster* [149]. A high chromatin landscape conservation between *D. melanogaster* and *D. miranda* species (divergence time ~30 Mya) [505] and between *D. melanogaster* and *D. simulans* species (divergence time ~5 Mya) [621] have also been detected. These studies found higher conservation of active chromatin marks compared to heterochromatic marks [505, 621], which is in agreement with the slightly higher differences of H3K27me3 found in this work. However, in contrast to our results, the most different mark between *D. melangoaster* and *D. miranda* species was the heterochromatic mark H3K9me2 [505]. In addition, in the same study they found that around 26% of the genes with species-specific expression had also divergence in their chromatin profiles [505], suggesting that the chromatin mark differences detected between our parental species could also contribute to the expression differences. In fact, in our study, 62% of the gene expression in both *D. buzzatii* and *D. koepferae* could be explained by the detected chromatin marks, being H3K4me3 positively correlated to gene expression, H3K27me3 strongly negatively correlated and H3K9me3 slightly positively correlated.

In addition to the chromatin landscape, another mechanism that could lead to changes in gene expression between *D. buzzatii* and *D. koepferae* is the gene regulatory divergence (see Annex **8.3**). We detected that ~ 58 % of the genes did not have regulatory divergence between parental species, which could explain the low percentage of gene expression differences between them (see Table S10 in Annex **8.1.1**). Moreover, *trans*-regulatory divergence between species, mainly produced by structural or expression changes of transcription factors (TF), are expected to affect the expression of multiple genes [622, 623]. However, *cis*-regulatory differences, produced by changes in promoters or enhancers, affect only the transcription a specific gene by modifying the binding site of the TF [622, 623]. Additionally, *trans*-regulatory mutations are observed to be more frequent than *cis*-regulatory ones because the size of the mutational target is larger in the former, being the latter more likely fixed in the population because they usually affect the expression of fewer genes [623, 624]. Therefore, as observed in *C. elegans* [625], *trans*-regulatory mutations can be selected against in natural populations, and hence, it was hypothesized that *cis*-regulatory changes between species were favored over *trans*-regulatory changes to drive gene expression evolution [623]. A greater proportion of *cis*-regulatory differences contributing to gene expression differences, were detected between *D. melanogaster* and *D. simulans* species [490, 519, 623]. However, we noticed that genes with *trans*-regulatory divergence had more expression differences between our species than genes with *cis*-regulatory divergence (see Table S11 in Annex **8.1.1**), as was previously reported between *D. melanogaster* and *D. sechellia* [489]. The detection of more *trans-* than *cis-* regulatory divergence between *D. melanogaster* and *D. sechellia* [489] was proposed to be related to a smaller population size of *D. sechellia* in comparison to other *Drosophila* species [489], making natural selection less efficient in this species. As far as our species are concerned, the larger historical effective population size of *D. buzzatii* compared to *D. koepferae* [626], together with the narrow distribution of *D. koepferae* [11] (section **1.1.1**), leads us to expect less *cis*-regulatory divergence than between species that have maintained larger population sizes.

The higher detection of more *trans*-regulatory divergence than *cis-* regulatory divergence leading to gene expression differences in our study could also be related to the experimental approach, such as biases due to environmental variation in expression or other sources of noise, as observed in other studies [627]. In addition, differences in regulatory divergence between tissues have also been described in hybrids of *D. willistoni* subspecies [628] and *D. mojavensis* populations [627]. Therefore, our results using only ovaries, and other studies, using whole flies [489], might not be completely comparable. The gene regulatory network is really complex, and we also have to consider that molecular interactions between regulatory components can affect the expression of genes in a different way [477], complicating the study of the regulatory divergence.

In addition, to the above mentioned higher contribution of *trans*-regulatory divergence than *cis*-regulatory divergence to gene expression differences between *D. buzzatii* and *D. koepferae*, the highest percentage of regulatory divergence between parental species was attributed to the category called compensatory (9.80% of the genes, see Table S10 in Annex **8.1.1**). This category is based on the coevolution of *cis-* and *trans*-regulatory divergence in each species, compensating each other and having the same expression between species (see Annex **8.3**). Even though these findings are in contrast to some studies in *Drosophila* [489, 519], genes showing evidence of both *cis-* and *trans*-regulatory changes were described to be mainly compensatory in another study, and the number of genes included in this category increased with divergence time between the species under study [629]. The explanation is that the fixation of these compensatory changes by stabilizing selection, maintain the optimal gene expression in both species [477]. As a consequence, the regulation of gene expression will diverge more rapidly than gene expression itself [629], leading to an accumulation of compensatory divergent changes [477]. Our results support this idea. Finally, to link the changes in the chromatin landscape between *D. buzzatii* and *D. koepferae* with their regulatory divergence, the binding of a TF to a gene not only depends on the *cis*-regulatory divergence between parental species, but also on the chromatin accessibility [630]. Therefore, the changes in chromatin landscape between parental species could also contribute to *cis*-regulatory divergence-like expression changes.

Transposable elements

Regarding the differences in TE expression between *D. buzzatii* and *D. koepferae* (section **3.1.2**), 29.64% of the annotated TEs had differences in expression between parental species, with 15.35% of them more expressed in *D. buzzatii* than in *D. koepferae* (Table S1 in Annex **8.1.1**). These percentages of differentially expressed TEs in ovaries were similar to those found in other two *repleta* species *D. mojavensis* and *D. arizonae* [330, 574]. TE expression in *D. buzzatii* and *D. koepferae* gonads is mainly represented by LINE retrotransposons [24]. However, the most differentially expressed TEs between these species were *LTR* retrotransposons (most of them belonging to the *gypsy* superfamily or family) and some DNA TE families (see Fig. S1B in Annex **8.1.3**).

We also detected that TE families usually had more piRNA amounts in *D. koepferae* than in *D. buzzatii* (Fig. 5B of section **3.1.2**). Nonetheless, the number of TEs more expressed in *D. koepferae vs D. buzzatii* was only slightly higher, which could be explained by the higher levels of Ping-pong signature observed in *D. buzzatii* [24]. The differential expression of mainly *gypsy* TE families between parental species could also be related to a different silencing of these TEs between parental species.

Finally, we also detected changes in the chromatin mark H3K4me3 in 5.47% of the TEs, H3K9me3 in 1.98% and H3K27me3 in 9.73% of the TEs between *D. buzzatii* and *D. koepferae* (Table S6 Annex **8.1.1**). As explained before, the piRNA pathway is also involved in transcriptional TE silencing by stablishing H3K9me3 heterochromatic marks in TEs [230] and also by removing active chromatin marks, such as H3K9 acetylation [232] and H3K4me3 methylation [233]. This mechanism is conserved in *Drosophila* species, and it could be the reason for a high conservation of the H3K9me3 chromatin mark between parental species, in comparison to the other studied chromatin marks. Additionally, a recent study found that TE repressive epigenetic marks can be extended to their adjacent regions [631]. The strength of the TE-mediated epigenetic effects in adjacent genes was found in all studied species, but its strength seemed to depend on different factors as: TE class/family, length or insertion location and the genetic background of the organism [631]. In addition, the TEs producing such epigenetic effects seem to be selected against across multiple species [631]. Therefore, less genes showing changes in H3K9me3 between populations would be expected (see Table S6 in Annex **8.1.1**), as observed in the present work.

In addition, the chromatin mark H3K27me3 has also been detected in TEs, usually co-localized together with H3K9me3 [149]. However, this mark is also known to be enriched in Polycomb/Trithorax response elements (PRE) [149, 632], which are motifs located upstream the genes that maintain their active or silent transcriptional state for many cell generations as determined earlier in the development [632]. PRE evolution is extraordinarily dynamic between *Drosophila* species [632], suggesting high differences in this chromatin mark between species. These dynamic intergenic regions could include TEs, contributing to the high H3K27me3 differences detected in TEs between *D. buzzatii* and *D. koepferae*. However, we could not detect so many H3K27me3 differences in genes between parental species because we only performed the enrichment analysis in the gene body, and this chromatin mark was found to be depleted in the coding region of the genes in our species (see Table S6 in Annex **8.1.1**) and in *D. melanogaster* [149].

4.1.2 Impact of interspecific hybridization on gene expression

We detected differences on expression in 4.57% and 3.99% of the annotated genes in hybrids *vs D. buzzatii* and *D. koepferae* respectively (section **3.1.2**). Even though more differences in gene expression were found in previous studies in *D. melanogaster-D. simulans* [474] and *D. melanogaster-D. sechellia* [489] female hybrids when compared with parental species, similar percentages to ours were detected in *D. mojavensis-D. arizonae* hybrid ovaries [330, 574]. It is necessary to note that former studies followed a different methodology to ours [474, 489], whereas the latter followed a more similar one [330, 574] (explained in detail in **4.1.1**).

Additionally, the lower gene expression differences found in our hybrids *vs* parental species than in other hybrids is supported by a conserved chromatin landscape detected in our hybrids (from 1.4% to 2.8% of genes changed their chromatin enrichment in hybrids *vs* parental species, see Fig. 3A in section **3.1.2**).

Differentially expressed genes in hybrids were divided into three categories according to their expression in comparison to both parental species (Fig. 1C in section **3.1.2**). Genes belonging to a additively expressed category follow an additive genetic model, which predicts that gene expression in hybrids is inherited equally from both parental species [492, 633]. We detected a low percentage of genes included in this category, as previously detected in other *Drosophila* hybrids [474, 489, 574]. Moreover, *cis*-regulatory changes observed between parental species seem to have additive effects on hybrid gene expression more often than *trans*-regulatory changes [489, 520]. Therefore, the low percentage of *cis*-regulatory divergence found between *D. buzzatii* and *D. koepferae* supports the low percentage of additively expressed genes in hybrids (see Table S10 in Annex **8.1.1**). Nonetheless, we could not find *cis*-regulatory differences between parental species in the additively expressed genes in hybrids (see Table S12 in Annex **8.1.1**), suggesting that their intermediate expression could be related to the general additive chromatin mark inheritance observed in our hybrids (section **3.1.2**). One example is the gene *3* in Fig. 4M, section **3.1.2**, underexpressed and less enriched in H3K4me3 in hybrids in comparison to *D. buzzatii*, but overexpressed and more enriched in this chromatin mark in comparison to *D. koepferae*.

The rest of differentially expressed genes in hybrids showed a non-additive expression and were divided in two categories: *D. koepferare* or *D. buzzatii*-like expression, or deregulated (Fig. 1C in section **3.1.2**). In our case, as in some other *Drosophila* hybrids [489, 574], most genes were included in the *D. koepferare* or *D. buzzatii*-like expression category (see Fig. 1A in section **3.1.2**), and they can follow two inheritance models. The first model is the parental effect model, which supports that hybrid gene expression is regulated by parental mechanisms, being similar to the maternal or the paternal species [492, 633]. We detected more genes with *D. koepferare*-like than *D. buzzatii*-like expression (see Table S2 in Annex **8.1.1**) in the hybrids. This tendency could be the result of genomic imprinting [634]. However, the hybrid epigenome was more like *D. buzzatii* for H3K4me3 and H3K27me3 than to *D. koepferae* (Fig. 3A in section **3.1.2** and Table S5 in Annex **8.1.1**), which agrees with studies in *D. melanogaster*, where no effects of imprinting could be found in adult files [634, 635]. Another mechanism that could lead to more *D. koepfere*-like expression in our hybrids includes cytoplasmic contributions of the egg and sperm to the zygote [634]. It is known that some genes in *Drosophila* are maternally deposited in the egg, affecting gene expression in a *trans*-acting manner, which can be propagated through development affecting the adult flies [634]. One example are the Piwi-family proteins, maternally deposited in *Drosophila*'s early embryos [172]. When we focus on the Piwi pathway genes, we

observed that Piwi and Ago3 were only differentially expressed in hybrids *vs D. buzzatii* (see Fig. S5 in Annex **8.1.3**), which could support the involvement of the maternally deposited proteins in the *D. koepfere*-like expression. However, because the opposite direction of crosses (*D. buzzatii* male x *D. koepferae* female) do not produce offspring [29], this hypothesis cannot be verified.

The second model producing *D. koepferare* or *D. buzzatii*-like expression is a dominance genetic model, in which gene expression in hybrids is determined by a dominant allele, and therefore is more similar to one of the parental species [492, 633]. This model of inheritance has been described in other *Drosophila* hybrids: *D. arizonae*-like expression inheritance was more common in *D. mojavensis-D. arizonae* hybrids [574] and *D. sechellia*-like expression in *D. melanogaster-D. sechellia* hybrids [489]. Genes with *trans*-regulatory divergence between parental species have been observed to show higher degrees of dominance inheritance [520]. Therefore, our findings of more *trans*-regulatory divergence between parental species agrees with more *D. koepferare* or *D. buzzatii*-like gene expression in hybrids. In concordance, we also found a higher proportion of *trans*-regulatory divergence in the genes showing *D. koepferare* or *D. buzzatii*-like expression in hybrids than in other categories (see Table S12 in Annex **8.1.1**). In addition, even though the chromatin landscape was in general conserved in hybrids (Fig. 3A in section **3.1.2**), some changes in chromatin marks in comparison to only one parental species were detected, which could also explain the *D. koepferare* or *D. buzzatii*-like expression of some genes. One example is the gene *CG3902* in Fig. 4M section **3.1.2**, detected as underexpressed in hybrids in comparison to *D. buzzatii*, and less enriched in H3K4me3 but more enriched in H3K9me3.

We also detected an important number of genes (Fig. 1A in section **3.1.2**) with an expression outside the ranges of the parental species, called deregulated. This category includes genes overexpressed (overdominance) or underexpressed (underdominance) in comparison to both parental species [492, 633], with a tendency towards underdominance in this study. We found a significant number of deregulated genes with *cis-trans* compensatory regulatory divergence between parental species (see Table S12 in Annex **8.1.1**), which has been previously described in other *Drosophila* hybrids [477, 489]. This compensatory regulatory divergence, which seems to be common in *Drosophila* [629] and between *D. koepferae* and *D. buzzatii* (see Table S10 in Annex **8.1.1** and explained in **4.1.1**), is based on the coevolution of *cis*-regulatory elements in one species and *trans*-regulatory elements in the other one, giving a similar gene expression in both. When these two alleles meet in the same genome, they can show a deregulated gene expression [477] (see Annex **8.3**). Therefore, the increase of genes with compensatory regulatory divergence between parental species over divergence time (explained in **4.1.1**), can also increase the proportion of genes showing misexpression in F_1 hybrids [629]. This idea is supported by the detection of less gene deregulation in *D. mojavensis-D. arizonae* hybrid ovaries, species having a lower divergence time than our parental species, with a slight tendency to

overdominance [330, 574]. However, in contrast to this hypothesis, more deregulation was detected in *D. melanogaster* and *D. simulans* hybrids [474] and in *D. melanogaster* and *D. sechellia* hybrids [489], both with a tendency towards underdominance. Nonetheless, the last two studies used a different methodology than our study (as explained in **4.1.1**). In addition to regulatory divergence, the chromatin landscape could also contribute to the misexpression of some specific genes. For example, the underexpression of *Nepl4* in hybrids *vs* both parental species seemed to be related to its loss of H3K4me3 chromatin marks (Fig. 4M in section **3.1.2**). Interestingly, a higher number of deregulated genes in F_1 hybrids were observed when originated from two specialists *Cyprinodon* pupfishes species than from generalists [636], suggesting that the deregulation observed between *D. koepferae* and *D. buzzatii* could also be the result that both are considered specialist [4, 11, 12]. Even though here we discuss the simple models of inheritance, the regulatory networks are not simple, and there are interactions between factors that can also affect the gene hybrid inheritance, such as positive or negative feed-back loops [633].

Our hybrids showed deregulated expression of only two Piwi Pathway Genes: Aubergine (*Aub*) and Sister of Yellow Body (*SoYb*), which were underexpressed and overexpressed respectively (see Fig. S5 in Annex **8.1.3**). *Aub* is involved in both the Phased piRNA biogenesis [205] and the Ping-Pong amplification [151] in the germline. Therefore, we would expect that its underexpression would impact the piRNA amounts in the hybrid germline. Nonetheless, we could not find a global decrease of piRNA production in the ovaries of those hybrids [24], in contrast to what was observed in other *Drosophila* artificial hybrids [350]. The piRNA amounts associated to TEs in our ovary hybrids were in general additively inherited between parental species (see Fig. 5B in section **3.1.2**), as well as the Ping-Pong signature [24], suggesting that the expression level of *Aub* in the hybrids was still high enough to perform its function. We also detected an underexpression of *SoYb* in hybrids, whose protein is an homologs of Yb [207, 208]. In the germline, its function can be compensated by BoYb, who together with SoYb, replaces the function of the soma-specific Yb [207]. In our study, *BoYb* was also underexpressed in hybrids *vs D. koepferae*, which could explain the slightly decrease of piRNA levels in hybrids *vs* this parental species. We also found differences in expression of an important number of other piRNA pathway genes, but only in comparison to one parental species: *spn-E*, *piwi*, *qin*, *papi*, *Panx*, *krimp*, *Hen1*, *Gasz*, *del*, *ci*, *BoYb*, *armi* and *Ago3*, or additively expressed: *vret*, *vas squ*, *moon* (see Fig. S5 in Annex **8.1.3**). These expression differences could also influence in some way the piRNA production and the Ping-Pong signature of the F_1 hybrid ovaries. Finally, the observed changes of expression of these genes could not be correlated to the observed changes in chromatin marks. Nonetheless, the high divergence of these proteins between *D. koepferae* and *D. buzzatii* previously described in a previous study [24], could explain the differences in gene expression in hybrids *vs* at least one parental species.

4.1.3 Impact of interspecific hybridization on TE expression

We found differences in expression in hybrids in 22.95 % of the TE families *vs D. buzzatii* and 24.16 % *vs. D. koepferae*, and 6.84% of the TE families had an expression out of the ranges of both parental species, and were considered deregulated (Fig. 1B in section **3.1.2**). A tendency towards the underexpression of TE families was detected, but the overexpressed families showed a higher relative difference in expression in hybrids *vs* parental species (see Fig. 1E in section **3.1.2** and Fig. S2b in Annex **8.1.3**). In contrast to our results, more overexpression of deregulated TE families was observed in *D. melanogaster–D. simulans* artificial female hybrids [350]. On the other hand, a smaller percentage (2%) of deregulated TE families was detected in a recent study in *D. arizonae-D. mojavensis* hybrids [574]. The authors suggested that longer divergence time between parental genomes may lead to a greater rate of TE deregulation [574], an idea that is supported by our study whose species present a longer divergence time than *D. arizonae-D. mojavensis*.

Regarding the mechanisms leading to TE deregulation in *D. buzzatii-D. koepferae* F_1 hybrid ovaries, a study proposed that it is the result of several interacting phenomena [24]. One of them is a partial global failure of the piRNA pathway in hybrid ovaries, which means that the high divergence in the piRNA pathway genes observed between *D. buzzatii* and *D. koepferae* is responsible for the hybrid TE deregulation [24]. Most piRNA pathway genes were differentially expressed in hybrid ovaries *vs* at least one parental species (see Fig. S5 in Annex **8.1.3**), which agrees with the high differences at both the sequence and the expression level found in these genes between *D. koepferae* and *D. buzzatii* [24]. Therefore, the mixture of divergent piRNA pathway genes in the hybrid could lead to a less efficient silencing through this pathway, which could lead to the observed TE deregulation (Fig. 1B in section **3.1.2**). Another phenomenon is the interaction of two different piRNA pools in the same genome [24]. It has been described that piRNA production strategies between the parental species are different: more piRNA production is detected in *D. koepferae*, whereas *D. buzzatii* has more Ping-Pong signature [24]. We detected that F_1 hybrid ovaries showed more piRNA amounts than *D. buzzatii* and less amounts than *D. koepferae* in most TE families (see Fig. 5B in section **3.1.2**), but the Ping-Pong signature in hybrids showed intermediate levels between parental species [24]. F_1 hybrid ovaries may have inherited different TE copies from each parental species, at sequence and/or at copy number. Therefore, the differences in piRNA amounts and Ping-Pong cycle in hybrids *vs* parental species could be insufficient to silence the new inherited TE copies, and hence, could lead to the deregulation of some specific TE families. This idea, which could lead to an increase in TE copy number in hybrids, can be supported by the fact that genome expansion after F_1 generation has been detected in *D. buzzatii-D. koepferae* hybrid ovaries [363].

Even though any order tended to be more differentially expressed than others when considering the composition of the genome (see Table S3 in Annex **8.1.1**), the highest overexpressed TEs in *D. buzzatii-D. koepferae* hybrids belonged to the *gypsy* superfamily (see Fig. S2B in Annex **8.1.1**). This finding agrees with the activation of TEs belonging to this superfamily in previous studies in these hybrids [361, 362]. Moreover, the elements belonging to this superfamily were the most differentially expressed between parental species, and hence possibly, the ones with more copies (sequence or number) or regulatory mechanism differences between *D. koepfere* and *D. buzzatii*. Although we observed overexpression of some TEs in F_1 hybrid ovaries, we found that most TE families were underexpressed (Fig. 1B in section **3.1.2**). This could be explained by the fact that global piRNA production in hybrids is more similar to *D. koepferae,* the parental species with higher piRNA levels, than to *D. buzzatii* [24]. This idea, together with the similar TE landscape detected between parental species [24], could explain a higher silencing of specific TE copies in hybrid ovaries. Finally, we also found changes in the chromatin landscape of some TE families for all three studied chromatin marks (Fig. 3B in section **3.1.2**) in hybrids *vs* parental species that could be related to TE expression changes (Fig. 4G-L,N in section **3.1.2**). Therefore, the TE expression observed in *D. buzzatii-D. koepferae* hybrid ovaries seem to be the result of the interaction of different phenomena, contributing to different or the same TE families, which complicates the interpretation of the mechanisms leading to TE deregulation in hybrids.

4.2   Gene and TE expression in two *D. subobscura* populations

4.2.1   Differences between populations in control conditions

Genes

In section **3.2.2**, we compared the gene expression in ovaries and testes between Madeira and Curicó populations of *D. subobscura*, and we detected that 4.23% of the expressed genes in ovaries and 9.68% in testes were differentially expressed (see Table S1 in Annex **8.2.1**). The percentage of differentially expressed genes fit inside the wide range of results found in the literature, when populations of different *Drosophila* species were compared. In all reported cases, genes with differences in expression could be related to the adaptation of these populations to different environments. Additionally, it is interesting to focus on some examples mentioned in section **3.2.2**. We noticed that both copies of *Hsp70* gene were more expressed in Madeira than in Curicó population, in both sexes (Fig. 1A-B in section **3.2.2**). Interestingly, a study proposed that low-latitude populations are more likely exposed to occasionally high temperatures than those located in high latitudes, and therefore in these populations high basal protein Hsp70 expression is beneficial regardless its negative effect in non-stressful conditions [427]. However, when we

116

focus on our population origins, temperature and humidity vary more throughout the year in Curicó than in Madeira, being both located in almost equivalent latitudes (Curicó in 34°58'58" S and Madeira in 32° 22' 17.9976" N) (see Annex **8.4**). We could think that other factors as chromosomal arrangements could affect this gene expression. Indeed different basal Hsp70 protein levels were detected between populations with two different O chromosome arrangements [427], but comparable basal Hsp70 mRNA levels were detected in other populations with other different O arrangements [428]. In our case, different Hsp70 mRNA levels were observed between our populations, despite both having the same arrangement in the O chromosome (O_{3+4}), highlighting a different reason for *Hsp70* differential expression, such as their adaptation to their habitat's annual average temperature, higher in Madeira than in Curicó (see Annex **8.4**).

Another interesting example of gene differentially expressed is *Cyp12d1-p*, which encodes a mitochondrial cytochrome P450, involved in the response to the insecticide dichlorodiphenyltrichloroethane (DDT) [637], and detected to be more expressed in Curicó males than in Madeira but the opposite was observed in females (Fig. 1A-B in section **3.2.2**). Interestingly, this gene was overexpressed in DDT resistant *D. melanogaster* strains in basal conditions [638] and higher basal levels of this protein were detected in males than in females, together with a more drastic induction in females when exposed to specific insecticides [639]. Therefore, the sex-dependent differences in basal *Cyp12d1-p* expression detected between our populations may reflect a population-specific sex-dependent adaptation to the exposure of insecticides. The gene *BTBD9*, which regulates brain dopamine levels and the circadian sleep/wake cycle [637, 640], was more expressed in both sexes of Curicó than Madeira (Fig. 1A-B in section **3.2.2**). Loss of this gene reduces dopamine levels and disrupts sleep in addition to increase waking and motor activity [640]. Light has been described to regulate dopamine negatively and also to modulate the timing of sleep [641]. Consequently, differences in the sleep/wake cycle between populations could be related to their adaptation to a different number of sun hours: the average sun hours in a year are higher in Curicó than in Madeira (see Annex **8.4**).

Transposable elements

In section **3.2.2**, we also detected that 8.09% of expressed TEs in ovaries and 9.88% in testes were differentially expressed between the Madeira and the Curicó populations, with almost twice as many TEs more expressed in Curicó than in Madeira in both sexes (see Table S4 in Annex **8.2.1**). We discussed that the TE overexpression in Curicó *vs* Madeira could be related to the fact that the former is a colonizing population: Curicó is located in South America, recently colonized by this species [34, 35]. An increase of insertion site frequency in this *Drosophila* species [579] and more TE insertions in other species [578, 580] have been described in colonizing populations.

Nonetheless, it is true that the colonization of the Madeira island from the continent's populations [577] could also influence its TE content: colonizers populations can suffer of founder effect followed by genetic drift [579]. However, the activation of TEs in stressful environments and their role in adaptation could contribute to the higher TE expression in Curicó considering that the Madeira island is located in the Palearctic region, where this species is native from [32]. One example supporting this idea is the increase of frequency in 13 TEs in *D. melanogaster*, during its expansion out of Africa into Europe, which are related to its adaptation to the new temperature climates [642]. In addition, it is interesting to note that the gene Transportin (*Tnpo*) was detected to be much more expressed in Madeira than in Curicó in both sexes (Fig. 1A-B in section **3.2.2**). This gene, in addition to being involved in immune system processes, it is also involved in Hedgehog signaling regulation by mediating the import of the transcription factor encoded by Cubitus interruptus (*ci*) into the nucleus [637]. The transcription of the *flam* uni-stranded piRNA cluster requires the recognition of a binding site by this transcription factor [156]. Therefore, an increase of the expression of *Tnpo* could be related with a higher piRNA production in the ovarian somatic cells of Madeira, and therefore a more efficiently TE silencing in this population. This idea is consistent with the detection of *Tnpo* in a screening performed in *Drosophila* ovaries as a gene involved in the piRNA pathway and required for proper TE silencing [643]. It is also supported by the fact that the *flam* cluster silences TEs mainly belonging to the *gypsy* superfamily [151, 152, 245], and these TEs are the ones with more differences in expression between our populations (Fig. 1C-D, in section **3.2.2**).

4.2.2 Impact of heat stress on gene expression

In section **3.2.2**, we detected that from 0.78% to 1.22% of the expressed genes changed their expression after a heat stress, with a tendency towards overexpression (see Table S7 in Annex **8.2.1**). The HSR is based on the activation of the Hsp genes [382] and the repression of other genes [379, 380], to avoid an accumulation of misfolded proteins [584]. Following this idea, a general underexpression was observed in other *Drosophila* heat-shock experiments [432, 582, 583], but not in our study. As argued before, most studies have not been performed in gonads, indicating a putative tissue-specific response to this stress: heat stress has been described to impact fly oogenesis [585] and sperm function and viability [586]. In fact, the Heat Shock Factor (Hsf1) in *Drosophila*, in addition to increase the expression of Hsp genes after heat stress and activate the HSR [392], it also controls the expression of other genes during non-stressful conditions [405]. Some of these genes are involved in reproduction, such as gamete generation, oogenesis and sexual reproduction, and in development, such as cell differentiation and embryonic development [405]. Indeed, Hsf1 is known to be required for oogenesis and early larval development in *Drosophila* non-stressful conditions [644], and the expression of some

small Hsps is detected without stress in testes of this organism and to be developmentally regulated during oogenesis in ovaries [422]. The involvement of Hsf1 in reproduction and cell differentiation supports our suggestion of a possible tissue-specific response to the heat stress in the germline. In fact, a previous study detected a different transcriptional response to heat stress in Kc cells and in 3rd instar larvae [405], suggesting that some genes could be transcriptionally induced by Hsf1 during heat stress only in specific developmental stages and/or tissue types [405].

Hsps are the ones that tend to be most affected by temperature stress. Two copies of *Hsp70* and *Hsp68* [426, 428], a single *Hsp83* copy [386, 429] and four different small Hsps genes (*Hsp23*, *Hsp28*, *Hsp27* and *Hsp26*) [386, 429] were previously reported in *D. subobscura*. Our reannotation of these genes in the *D. subobscura* reference genome is in concordance with these findings, detecting an additional *Hsp27* copy (Fig. 3A in section **3.2.2**). We detected a general activation of the main Hsps in *D. subobscura* after heat stress, as expected. Our heat stress was performed at 32ºC and the optimal puff activation in this species, after heat stress, was detected at 31-34ºC [380, 385, 386]. We detected that the most activated Hsp genes, in all our populations and sexes, were both copies of *Hsp70*, as expected due to its highly heat-induced expression [404]. Both copies of the *Hsp68* (mainly in females) and the small Hsps (mainly in males) were also part of the Hsps highly activated after stress. All of them have been previously reported in the literature to be heat inducible [382]. In addition, the small Hsps have been previously detected to be constitutively expressed and also heat-inducible at different levels and in different cell types of testes and ovaries [422], which could lead to differences in their activation between tissues. We also detected some overexpression of *Hsp83* after heat stress, as expected, because this protein is constitutively expressed but also heat-inducible [413], reaching the highest induction in *D. subobscura* at 31-34ºC [430]. On the other hand, we did not detect a strong activation of the heat shock cognate genes, as expected due to their constitutive not heat-induced expression [410]. A high activation in all populations and sexes of other heat shock response related genes, such as *Dnaj-1* and *stv*, together with *ref(2)P* and *CG6511*, (Fig. 2F in section **3.2.2**), was detected and have been described to have target sequences for Hsf1 [405], which could explain their overexpression.

Regarding the impact of the heat stress on the piRNA pathway genes, we could only detect a small impact in females (Fig. 3C in section **3.2.2**), where an overexpression of *hop* and *piwi* was observed in both populations. The former, called Hsp70/Hsp90 organizing protein homolog (*Hop*), enables Hsp90 protein binding activity [414, 425], and works together with the Hsc70/Hsp90 chaperone machinery to load piRNAs onto Ago3 [192, 193] in the piRNA pathway, and small RNA duplexes into Ago2 [240] in the siRNA pathway. In addition, Hsf1 was detected to target *Hop* [405], explaining its overexpression. Piwi is involved mainly in the Phased primary piRNA production [195, 197, 210], but also in the piRNA dependent transcriptional TE silencing

by its exportation to the nucleus [228, 645]. Therefore, the overexpression of *hop* and *piwi* could impact the piRNA and siRNA production, as observed for some TE families (see Fig. 5C-D in section **3.2.2**). In addition, the overexpression of *piwi* could also influence the chromatin landscape after heat stress. The rest of piRNA pathway genes showed low differences in expression detected in only one population. The only exception to this last genes group is *BoYb*, which is believed to replace the function of the soma-specific Yb in the germline [207], and showed high overexpression in Curicó. However, its specific function is still not well described.

4.2.3 Impact of heat stress on TE expression

We detected a moderate impact of heat stress on TE expression, with a range from 0.77% to 1.75% (see Fig. 4A-D in section **3.2.2**) of the TE families changed their expression after heat stress in ovaries and testes from two *D. subobscura* populations. We also detected a tendency towards the overexpression of TEs, most of them belonging to the *gypsy* superfamily. The machinery involved in the repression of the inactive form of the Hsf1 under normal conditions, includes a multichaperone complex formed by the constitutively expressed Hsp90 [389, 390] and Hsc70/Hsp70 [388, 391] in synergy with the co-chaperone DroJ1 [388]. Under heat stress, these constitutively expressed Hsps will bind unfolded proteins, which leads to the liberation and de-repression of Hsf1 and therefore the transcription of the Hsp genes [388]. A similar machinery, together with additional proteins, is also required in the siRNA pathway to load the siRNA duplexes into Ago2 [240], and in the piRNA pathway to load piRNAs onto Ago3 [192, 193] and to form active Piwi-piRISCs complexes [215, 414, 646]. We would expect that under heat stress, the chaperone function of these Hsps will be prioritized to the detriment of their participation in the siRNA and piRNA pathways. We found an impact of heat stress in the siRNA and piRNA amounts associated to some TE families (see Fig. 5C-D in section **3.2.2**), which support this hypothesis, in line with two other studies in *Drosophila* where they also found the same effect of heat stress in the piRNA amounts [464, 468]. Interestingly, another heat stress study could not find a decrease in the piRNA production of some TE families [192], but they detected a reduction of the piRNA loading into Ago3 by the displacement of the chaperone-Ago3 machinery to the lysosome for degradation through its interaction with Hsp70, resulting in a TE activation [192].

In contrast, and considering that Ago3 is involved in the piRNA amplification loop [151, 185], we could only detect an effect of heat stress on the Ping-Pong cycle of a few TE families, and not a general effect on this signal (Figure S7 and S8 in Annex **8.2.2**). Our findings are in concordance with a previous study which could not find an effect of heat stress on the Ping-Pong cycle after 1 hour recovery [468]. However, another study detected an increase of this signal after 48h *Drosophila* housing [464]. Interestingly, the different effect of heat stress in the Ping-Pong cycle, detected in the studies, could be related to an effect of the heat shock on the piRNA pathway

depending on the recovery time, in addition to differences in the time and strength of the heat stress. In fact, in one of these previous studies [468] they found that Hsp70 was localized in the 'nuage' and interacted with Ago3 and other chaperones 1 hour after the heat shock, but the displacement of this machinery to the lysosome was detected one day and two days after the heat shock. An increase of TE transcripts, and the recovery of the machinery into the 'nuage' was detected after three days [192]. Finally, we could not detect a correlation between changes in piRNA levels and TE transcripts after heat stress, as was already reported [464, 468]. This surprising result could be related to the fact that most piRNAs involved in TE silencing are produced from numerous TE insertions dispersed across the genome, and not from specific piRNA clusters [163, 166]. Therefore, it is possible that ours and other analyses did not distinguish between piRNAs produced by piRNA clusters or piRNAs produced by TE insertions. In addition, changes in chromatin marks after a heat shock have been observed in *Drosophila* cells [469], suggesting that epigenetic modifications could also influence TE expression after this stress.

4.3   Technical limitations and future perspectives

In this work we focused on the study of how two different stresses (interspecific hybridization and heat stress) affected gene and TE expression in three non-model species: the two cactophilic siblings *D. buzzatii* and *D. koepferae*, and *D. subobscura.* Whereas an extensive amount of information is available for the model organism *D. melanogaster,* less is known about the rest of the species belonging to the *Drosophila* genus. Regarding the two cactophilic species used in this work, only the reference genome of *D. buzzatii* was available [123] when our hybridization study was performed. Therefore, permissive alignment options were needed to equalize the percentage of RNA-Seq and ChIP-Seq reads from both species aligning to the same *D. buzzatii* reference genome. Additionally, the chromatin mark landscape was only studied in gene coding regions and not in promoters, due to the higher conservation of the former between species in comparison to other regions. Additionally, to overcome the lack of a *D. koepferae* reference genome, a new allelic-specific expression pipeline was developed and adapted to our data. It is also interesting to note that the TE content of both species was not deeply studied before [125, 647] and we used the manually curated TE library produced in a previous study [24]. Moreover, in the experiments with interspecific hybrids (RNA-Seq, small RNAs and ChIP-Seq), samples of pooled ovaries were used due to the greater ease in obtaining material compared to testes from sterile hybrid. However, since differences in chromatin states between species are roughly the same as between sexes [505, 621], it would also be interesting to study the male chromatin landscape, their inheritance in hybrids and its contribution to hybrid sterility [28]. In the same way, the study of how regulatory divergence between parental species influences male hybrids sterility, as

previously observed in other *Drosophila* hybrids [648], could also be an important issue to address in the future.

Regarding the *D. subobscura* results, even if a reference genome was available, the TE sequence annotation was insufficiently in depth, and we had to make a *de novo* TE identification. Differentially expressed genes after a heat stress depend on different variables: the severity [431] and duration of the heat stress [435], the time of recovery [432], the stress resistance [349], the organism [349, 431–433] and the sex [436], the tissue [434, 435], and even the biological replicate [437]. In addition, we ignore how heat stress could influence the general amounts of mRNA or small RNAs in *Drosophila.* For example, it is known that the median mRNA half-life shortens as the temperature increases in yeast [649]. The aforementioned set of variables make difficult to compare our study with previous ones. Additionally, Hsp83 is involved in epigenetic modifications [415] and changes in chromatin marks after heat stress have been observed in *Drosophila* tissue cells [469]. However, H3K9me3 and H3K27me3 levels in seven TEs did not significantly change in the germline after heat stress [192]. Therefore, it would be interesting to study how this stress affects the chromatin mark landscape in both genes and TEs in the germline. Moreover, it is known that the epigenomic changes, produced by heats stress during early embryogenesis in *Drosophila* can be inherited by the next generation [470]. Hence, to study the transcriptome and epigenome of future generations after heat stress, and its role in species thermotolerance should be address in the future.

Moreover, it is known that there are differences in the content, location [579] and expression [534] of TEs between populations and strains of *Drosophila*. The use of the general reference genome instead of that of our strains or populations, led us to study family TE expression rather than copy expression. The sequencing of the genome of our strains and populations is now underway and it will allow us to identify the exact location of the TEs in these genomes and, hence, perform a more accurate TE expression and chromatin landscape analysis at the copy level under both stresses. We also detected a differential expression of genes related to xenobiotic transport and metabolism between *D. buzzatii* and *D. koepferae*, as well as genes related to xenobiotic metabolism and temperature between *D. subobscura* populations. An enrichment of TEs in the close proximity to xenobiotic-metabolizing genes [650], Hsps [602] and genes targeted by transcription factors related to the stress-response [651] were detected in *D. melanogaster*. Therefore, it would be interesting to study how TE location affect the differential expression of these genes, in addition to the stress, and the putative evolutionary role of TEs in the expression differences. Finally, we could also expand the studied region of chromatin marks and see the effect of stress also on the gene promoters.

4.4 Impact of stress in the *Drosophila* genome

One of the objectives of this thesis was to study how two different types of stress (interspecific hybridization and heat stress) affected the gene expression. During hybridization, the fusion of two distinct genomes and epigenomes into F_1 hybrids [323] have an strong repercussions in the progeny, such as gene expression in an unpredictable direction in hybrids [652] or outside the expression ranges of the parental species [492, 633] (explained in **1.2.4.1**). In addition, many mechanisms and interactions lead to the observed hybrid gene expression (discussed in **4.1.2**). However, the heat stress is a disruption that affects a genome in a specific moment, leading to a repression of gene expression [379, 380] and activation of certain genes, such as Hsps [382] (explained in **1.2.4.2**). Therefore, and as expected, we detected a higher impact of interspecific hybridization in F_1 hybrid ovaries gene expression (see section **3.1.2**) than heat stress in ovaries and testes of *D. subobscura* (see section **3.2.2**). In addition, differentially expressed genes caused by the stresses were involved in different Gene Ontologies: whereas the deregulated genes in hybrids were mainly involved in metabolic, cellular and developmental process, cell adhesion and reproduction (see section **3.1.2**), we detected genes deregulated after heat stress also involved in response to stimulus, signaling, locomotion, rhythmic processes, and immune system processes (see section **3.2.2**). These differences are related to the main effect of these stresses: *D. buzzatii-D. koepferae* hybrids display sterility [28] and some developmental distortions [358], whereas during heat stress gene expression is activated mainly to protect the individual to survive this challenge [377]. However, some similarities between both stresses have also been described. For example, changes in Hsps gene expression were also observed in *D. mojavensis-D. arizonae* hybrids, but towards underexpression [653].

The main goal of this thesis was to study how stress affected TE expression. The mobilization of specific TE families under different stresses was extensively reported in *Drosophila*, but seemed to depend on the TE family, the host background and the environmental conditions [61]. New Generation Sequencing techniques (NGS) provided a better tool to examine the effect of stress in the whole genome and not only in specific TE families. A recent review based in *Drosophila* whole-genome studies summarizes that TE activation is not general, therefore stress can activate none or just a few TE families or even repress some of them [654]. Our results support this view, in which not all TEs are activated by stress, but some of them are strongly impacted (see section **3.1.2** and **3.2.2**). This activation, even though if it is of only one TE family, could have a strong effect in the genome, as it is for example observed in the hybrid dysgenesis phenomenon [131, 132, 315]. Our detection of underexpression of some TE families under both stresses, but mainly after interspecific hybridization, also support this idea [654].

We also detected that the highest overexpressed TEs after both stresses were mainly *gypsy* elements (see section **3.1.2** and **3.2.2**). The above mentioned review also found that stress predominantly activates TE families of the *LTR* order [654], and they suggest as the reason that *LTR* is the most represented order in *Drosophila* [654]. Nonetheless, in our species the overrepresentation of *LTR* elements does not seem to be the rule (see section **1.2.1.3**). Another proposed explanation is that *LTRs* are more likely to possess and retain their ancestral *cis*-regulatory activities, usually present in their *LTR* sequences, which sometimes are similar to preserved motifs required for the activation of stress-responsive genes [654]. The presence of these motifs can provoke TE activation in response to particular stimulus, even in partial copies, while influencing the expression of adjacent genes [654]. Regulatory sequences could be one of the reasons leading to our TE activation after interspecific hybridization and heat stress: specifically, TE insertions in proximity to differentially expressed genes in hybrids or Hsps and other heat-induced genes.

Regarding the effect of TE activation in the host genomes, we know that the presence of TEs in the organisms has heterogeneous impacts (see section **1.2.3**): even though they are mainly negative for the organisms, examples of beneficial impact and domestication has also been described. Therefore, Barbara McClintock proposed that the formation of new species after hybridization could be produce by genomic modifications and reorganizations provoked by TEs [54]. The mobilization of TEs in F_1 ovary hybrids after interspecific hybridization, which compromise the sterility and genomic stability of the hybrids, can hence be a source of pre-zygotic and post-zygotic barriers for the formation of new species, including the adaptation of hybrids to new habitats, among others [655]. McClintock also proposed that different genomic shocks could lead to TE activation, which increases the genomic variation of the organisms in order to survive the challenges, which may have an adaptive value in the host genomes [54]. In the same way, new TE insertions close to stress-responsive genes after a heat stress, such as Hsp70, could also lead to a better thermotolerance of a species [603].

Finally, regarding the mechanisms leading to TE activation in the germline, TE deregulation after interspecific hybridization and heat stress do not seem to be explained only by one mechanism, but a mixture of phenomena. We detected that TE activation depended on the effect of the stress in the different TE regulatory mechanisms, including the piRNA pathway genes, the piRNA and siRNA amounts targeting each TE family and the epigenome. In addition, the complexity of the TE regulation, based on both a transcriptional and post-transcriptional silencing mechanism, makes difficult to detect a direct correlation between TE expression and their silencing mechanisms [597]. Moreover, new discoveries regarding the TE silencing in the germline are being revealed, indicating that there are still things that we ignore. Therefore, the

study of the phenomena leading to TE activation is complicated, and even though every discovery makes us closer to a proper understanding of this mechanism, there is still a lot of work to do.

5. Conclusions

1) *D. buzzatii* and *D. koepferae* present differentially expressed genes that might be related to their ecological adaptation.

2) *D. buzzatii* and *D. koepferae* interspecific hybridization promotes deregulation of some TE families and genes in F_1 ovaries, both with a tendency towards underexpression.

3) The epigenome of *D. buzzatii* and *D. koepferae* is globally highly preserved between parental species and in their hybrids.

4) Interspecific hybrids show only some changes in the chromatin landscape compared to parental species, which can be associated with the new observed expression patterns of gene and TE families.

5) Gene deregulation in hybrids is related to incompatible interactions of *cis-* and *trans-*regulatory divergence between *D. buzzatii* and *D. koepferae*.

6) TE deregulation in hybrids is related to incompatibilities in piRNA pathway gene expression and piRNA pools between *D. buzzatii* and *D. koepferae*.

7) The *D. subobscura* populations from Madeira and Curicó present differentially expressed genes, such as immune system and insecticide resistance genes, likely involved in their adaptation to different habitats.

8) Annotation of TE-associated piRNA clusters in *D. subobscura* ovaries showed that they are population-specific. The most common among populations are generally dual-stranded and located in pericentromeric regions.

9) The O chromosome gene *Hsp70* is differentially expressed in non-stressful conditions in the Madeira and Curicó populations, carriers of O_{3+4} arrangement.

10) The heat stress response is sex-specific in the germline of *D. subobscura*, with a tendency towards gene overexpression, mainly of heat shock protein genes, suggesting a tissue-specific effect to this stress.

11) A few specific TE families, specific of sex and population, were activated after heat stress.

12) There is not a global impact of heat stress in the small RNA levels, but changes in the siRNA and piRNA amounts associated to specific TE families and in the production of specific piRNA clusters, were detected.

13) There is not a clear correlation between TE activation after heat stress and changes in small RNA amounts, suggesting that additional mechanisms, such as changes in the epigenome, could also be involved in TE expression.

14) *Gypsy* is the most differentially expressed TE superfamily in ovaries of interspecific hybrids and in testes and ovaries of the *D. subobscura* populations after heat stress, even though it is not the most abundant TE superfamily in the genome of these species.

15) Gene and TE deregulation mechanism both in ovaries of interspecific hybrids and in gonads of *D. subobscura* populations after heat stress, is a multifactor and complex phenomenon.

6.   Bibliography

1. Frezza G, Capocci M. Thomas Hunt Morgan and the invisible gene: the right tool for the job. Hist Philos Life Sci. 2018;40:31.

2. Barbara H. Jennings. Drosophila a versatile model in biology and medicine. materials Today. 2011.

3. Rane R V, Pearce SL, Li F, Coppin C, Schiffer M, Shirriffs J, et al. Genomic changes associated with adaptation to arid environments in cactophilic Drosophila species. BMC Genomics. 2019;20:52.

4. Fontdevila A, Pla C, Hasson E, Wasserman M, Sanchez A, Naveira H, et al. Drosophila koepferae: a new member of the Drosophila serido (Diptera: Drosophilidae) superspecies Taxon1. Ann Entomol Soc Am. 1988;81:380–5.

5. Durando CM, Baker RH, Etges WJ, Heed WB, Wasserman M, DeSalle R. Phylogenetic analysis of the repleta species group of the genus Drosophila using multiple sources of characters. Mol Phylogenet Evol. 2000;16:296–307.

6. Oliveira DCSG, Almeida FC, O'Grady PM, Armella MA, DeSalle R, Etges WJ. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the Drosophila repleta species group. Mol Phylogenet Evol. 2012;64:533–44.

7. Hasson E, De Panis D, Hurtado J, Mensch J. Host Plant Adaptation in Cactophilic Species of the Drosophila buzzatii Cluster: Fitness and Transcriptomics. J Hered. 2019;110:46–57.

8. Moreyra NN, Mensch J, Hurtado J, Almeida F, Laprida C, Hasson E. What does mitogenomics tell us about the evolutionary history of the Drosophila buzzatii cluster (repleta group)? PLoS One. 2019;14:e0220676.

9. Fontdevila A, Ruiz A, Alonso G, Ocaña J. Evolutionary history of drosophila buzzatii. I. Natural chromosomal polymorphism in colonized populations of the old world. Evolution. 1981;35:148–57.

10. Barker JSF, de M. Sene F, East PD, Pereira MAQR. Allozyme and chromosomal polymorphism of Drosophila buzzatii in Brazil and Argentina. Genetica. 1985;67:161–70.

11. Hasson E, Naveira H, Fontdevila A. The breeding sites of Argentinian cactophilic species of the Drosophila mulleri complex (subgenus Drosophila-repleta group). Rev Chil Hist Nat. 1992;65:319–26.

12. Soto EM, Goenaga J, Hurtado JP, Hasson E. Oviposition and performance in natural hosts in cactophilic Drosophila. Evol Ecol. 2012;26:975–90.

13. Fanara JJ, Fontdevila A, Hasson E. Oviposition preference and life history traits in cactophilic Drosophila koepferae and D. buzzatii in association with their natural hosts. Evol Ecol. 1999;13:173–90.

14. Fanara JJ, Hasson E. Oviposition acceptance and fecundity schedule in the cactophilic sibling species Drosophila buzzatii and D. koepferae on their natural hosts. Evolution. 2001;55:2615–9.

15. Soto IM, Carreira VP, Corio C, Padró J, Soto EM, Hasson E. Differences in tolerance to host cactus alkaloids in Drosophila koepferae and D. buzzatii. PLoS One. 2014;9:e88370.

16. de Oliveira AJB, Machado MFP da S. Alkaloid production by callous tissue cultures of Cereus peruvianus (Cactaceae). Appl Biochem Biotechnol. 2003;104:149–55.

17. Fanara JJ, Mensch J, Folguera G, Hasson E. Developmental time and thorax length differences between the cactophilic species Drosophila buzzatii and D. koepferae reared in different natural hosts. Evol Ecol. 2004;18:203–

14.

18. Carreira VP, Soto IM, Fanara JJ, Hasson E. A study of wing morphology and fluctuating asymmetry in interspecific hybrids between Drosophila buzzatii and D. koepferae. Genetica. 2008;133:1–11.

19. Corio C, Soto IM, Carreira V, Padró J, Betti MIL, Hasson E. An alkaloid fraction extracted from the cactus Trichocereus terscheckii affects fitness in the cactophilic fly Drosophila buzzatii (Diptera: Drosophilidae). Biol J Linn Soc. 2013;109:342–53.

20. Padró J, Carreira V, Corio C, Hasson E, Soto IM. Host alkaloids differentially affect developmental stability and wing vein canalization in cactophilic Drosophila buzzatii. J Evol Biol. 2014;27:2781–97.

21. 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.

22. De Panis DN, Padró J, Furió-Tarí P, Tarazona S, Milla Carmona PS, Soto IM, et al. Transcriptome modulation during host shift is driven by secondary metabolites in desert Drosophila. Mol Ecol. 2016;25:4534–50.

23. Werenkraut V, Hasson E, Oklander L, Fanara JJ. A comparative study of competitive ability between two cactophilic species in their natural hosts. Austral Ecol. 2008;33:663–71.

24. 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.

25. Hurtado J, Almeida F, Revale S, Hasson E. Revised phylogenetic relationships within the Drosophila buzzatii species cluster (Diptera: Drosophilidae: Drosophila repleta group) using genomic data. Arthropod Syst Phylogeny. 2019;77:239–50.

26. Oliveira CC de, Manfrin MH, Sene F de M, Jackson LL, Etges WJ. Variations on a theme: diversification of cuticular hydrocarbons in a clade of cactophilic Drosophila. BMC Evol Biol. 2011;11:179.

27. Machado LPB, Madi-Ravazzi L, Tadei WJ. Reproductive relationships and degree of synapsis in the polytene chromosomes of the Drosophila buzzatii species cluster. Braz J Biol. 2006;66:279–93.

28. Marin I, Fontdevila A. Stable Drosophila buzzatii -Drosophila koepferae hybrids. J Hered. 1998;89:336–9.

29. 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.

30. Naveira H, Fontdevila A. The evolutionary history of Drosophila buzzatii. XII. The genetic basis of sterility in hybrids between D. buzzatii and its sibling D. serido from argentina. Genetics. 1986;114:841–57.

31. Collin JE. Note: Drosophila subobscura sp. n. male, female. J Genet. 1936;33:60.

32. Krimbas CB. Drosophila subobscura: Biology, Genetics and Inversion Polymorphism. Hamburg: Verlag Dr. Kovac; 1993.

33. Buzzati-Traverso AA, Scossiroli RE. The "Obscura Group" of the genus Drosophila. In: Demerec MBT-A in G, editor. Academic Press; 1955. p. 47–92.

34. Brneie D, Prevosti A, Budnik M, Monclus M, Ocana J. Colonization of Drosophila subobscura in Chile I. First population and cytogenetic studies. Genetica. 1981;56:3–9.

35. Prevosti A, Serra L, Ribo G, Aguade M, Sagarra E, Monclús M, et al. The colonization of drosophila subobscura in chile. II. Clines in the chromosomal arrangements. Evolution. 1985;39:838–44.

36. Prevosti A, Serra L, Aguadé M, Ribo G, Mestres F, Balañá J, et al. Colonization and Establishment of the Paleartic Species Drosophila Subobscura in North and South America. In: Fontdevila A, editor. Evolutionary Biology of Transient Unstable Populations. Berlin, Heidelberg: Springer Berlin Heidelberg; 1989. p. 114–29.

37. Beckenbach AT, Prevosti A. Colonization of North America by the European Species, Drosophila subobscura and D. ambigua. Am Midl Nat. 1986;115:10–8.

38. Rezende E, Balanyà J, Rodríguez-Trelles F, Rego C, Fragata I, Matos M, et al. Climate change and chromosomal inversions in Drosophila subobscura. Clim Res. 2010;43:103–14.

39. Begon M, Shorrocks B. The feeding- and breeding-sites of Drosophila obscura Fallén and D. subobscura Collin. J Nat Hist. 1978;12:137–51.

40. Maynard Smith J. Fertility, mating behaviour and sexual selection in Drosophila subobscura. J. Genet. 1956, 54, 261-279. J Genet. 2005;84:17–35.

41. Erić K, Patenković A, Erić P, Davidović S, Veselinović MS, Stamenković-Radak M, et al. Stress resistance traits under different thermal conditions in Drosophila subobscura from two altitudes. Insects. 2022;13:1–17.

42. Orengo DJ, Prevosti A. Preadult competition between Drosophila subobscura and Drosophila pseudoobscura. J Zool Syst Evol Res. 1994;32:44–50.

43. Philip U, Rendel JM, Spurway H, Haldane JBS. Genetics and karyology of Drosophila subobscura. Nature. 1944;154:260–2.

44. Menozzi P, Krimbas CB. The inversion polymorphism of D. subobscura revisited: Synthetic maps of gene arrangement frequencies and their interpretation. J Evol Biol. 1992;5:625–41.

45. Ayala FJ, Serra L, Prevosti A. A grand experiment in evolution: the Drosophila subobscura colonization of the Americas. Genome. 1989;31:246–55.

46. Rego C, Balanyà J, Fragata I, Matos M, Rezende EL, Santos M. Clinal patterns of chromosomal inversion polymorphisms in Drosophila subobscura are partly associated with thermal preferences and heat stress resistance. Evolution. 2010;64:385–97.

47. Galludo M, Canals J, Pineda-Cirera L, Esteve C, Rosselló M, Balanyà J, et al. Climatic adaptation of chromosomal inversions in Drosophila subobscura. Genetica. 2018;146:433–41.

48. Rodríguez-Trelles F, Alvarez G, Zapata C. Time-series analysis of seasonal changes of the O inversion polymorphism of Drosophila subobscura. Genetics. 1996;142:179–87.

49. Rodríguez-Trelles F, Tarrío R, Santos M. Genome-wide evolutionary response to a heat wave in Drosophila. Biol Lett. 2013;9:20130228.

50. Orengo D-J, Prevosti A. Temporal changes in chromosomal polymorphism of drosophila subobscura related to climatic changes. Evolution. 1996;50:1346–50.

51. Balanyá J, Oller JM, Huey RB, Gilchrist GW, Serra L. Global genetic change tracks global climate warming in Drosophila subobscura. Science. 2006;313:1773–5.

52. Madrenas R, Balanyà J, Arenas C, Khadem M, Mestres F. Global warming and chromosomal inversion adaptation in isolated islands: Drosophila subobscura populations from Madeira. Entomol Sci. 2020;23:74–85.

53. Karageorgiou C, Tarrío R, Rodríguez-Trelles F. The cyclically seasonal Drosophila subobscura inversion O(7) originated from fragile genomic sites and relocated immunity and metabolic genes. Front Genet. 2020;11:565836.

54. McClintock B. The significance of responses of the genome to challenge. Science. 1984;226:792–801.

55. June C, Page SEEL, Pasternak S, Liang C, Zhang J, Fulton L, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2012;326 June:1112–5.

56. Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. Genomics Inform. 2014;12:87.

57. Correa M, Lerat E, Birmelé E, Samson F, Bouillon B, Normand K, et al. The transposable element environment of human genes differs according to their duplication status and essentiality. Genome Biol Evol. 2021;13:1–24.

58. Mérel V, Boulesteix M, Fablet M, Vieira C. Transposable elements in Drosophila. Mob DNA. 2020;11:1–20.

59. Brown NL, Evans LR. Transposition in prokaryotes: transposon Tn501. Res Microbiol. 1991;142:689–700.

60. Buchon N, Vaury C. RNAi: A defensive RNA-silencing against viruses and transposable elements. Heredity. 2006;96:195–202.

61. García Guerreiro MP, Guerreiro MPG. What makes transposable elements move in the Drosophila genome? Heredity. 2012;108:461–8.

62. Wicker T, Sabot F, Hua-van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.

63. Finnegan DJ. Eukaryotic transposable elements and genome evolution. Trends Genet. 1989;5 C:103–7.

64. Kapitonov V V., Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet. 2008;9:411–2.

65. Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet. 2020;54:539–61.

66. Havecker ER, Gao X, Voytas DF. The diversity of LTR retrotransposons. Genome Biol. 2004;5.

67. Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev. 1994;8:2046–57.

68. Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, et al. Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. J Virol. 2000;74:10658–69.

69. Cook JM, Martin J, Lewin A, Sinden RE, Tristem M. Systematic screening of Anopheles mosquito genomes yields evidence for a major clade of Pao-like retrotransposons. Insect Mol Biol. 2000;9:109–17.

70. Curcio MJ, Lutz S, Lesage P. The Ty1 LTR-retrotransposon of budding yeast, Saccharomyces cerevisiae. Microbiol Spectr. 2015;3:1–35.

71. Curcio MJ, Derbyshire KM. The outs and ins of transposition: From MU to kangaroo. Nat Rev Mol Cell Biol. 2003;4:865–77.

72. Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, et al. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell. 1993;73:1007–18.

73. Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, Voytas DF. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol. 2001;21:6606–14.

74. Nefedova LN, Mannanova MM, Kim AI. Integration specificity of LTR-retrotransposons and retroviruses in the Drosophila melanogaster genome. Virus Genes. 2011;42:297–306.

75. Faye B, Arnaud F, Peyretaillade E, Brasset E, Dastugue B, Vaury C. Functional characteristics of a highly specific integrase encoded by an LTR-retrotransposon. PLoS One. 2008;3:e3185.

76. Kapitonov V V, Tempel S, Jurka J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene. 2009;448:207–13.

77. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Pérez JL, Moran J V. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. Microbiol Spectr. 2016;3:1–63.

78. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol. 2001;21:467–75.

79. Moran J V, Holmes SE, Naas TP, Deberardinis RJ, Boeke JD, Kazazian HH. High Frequency Retrotransposition in Cultured Mammalian Cells. Cell. 1996;87:917–27.

80. Metcalfe CJ, Casane D. Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements. 2014;:1–13.

81. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob DNA. 2010;1:1–12.

82. George JA, Eickbush TH. Conserved features at the 5' end of Drosophila R2 retrotransposable elements: Implications for transcription and translation. Insect Molecular Biology. 1999;8:3–10.

83. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site : a mechanism for Non-LTR retrotransposition. Cell. 1993;72:595–605.

84. Novikova OS, Blinov AG. Origin, evolution, and distribution of different groups of non-LTR retrotransposons among eukaryotes. Russ J Genet. 2009;45:129–38.

85. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.

86. Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet. 2017;18:292–308.

87. Ye J, Eickbush TH. Chromatin structure and transcription of the R1- and R2-inserted rRNA genes of Drosophila melanogaster. Mol Cell Biol. 2006;26:8781–90.

88. Fujiwara H. Site-specific non-LTR retrotransposons. Microbiol Spectr. 2015;3:MDNA3-0001–2014.

89. Rashkova S, Karam SE, Kellum R, Pardue M-L. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. J Cell Biol. 2002;159:397–402.

90. Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martín-Gallardo A, Villasante A. TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. Molecular biology and evolution. 2004;21:1620–4.

91. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity. 2011;107:487–95.

92. Gilbert N, Labuda D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci U S A. 1999;96:2869–74.

93. Labuda D, Sinnett D, Richer C, Deragon JM, Striker G. Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. J Mol Evol. 1991;32:405–14.

94. Kapitonov V V, Jurka J. A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol. 2003;20:694–702.

95. Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A. 1997;94:196–201.

96. Arkhipova IR. Distribution and phylogeny of Penelope-like elements in eukaryotes. Syst Biol. 2006;55:875–85.

97. Evgen'ev MB, Arkhipova IR. Penelope-like elements - A new class of retroelements: Distribution, function and possible evolutionary significance. Cytogenet Genome Res. 2005;110:510–21.

98. Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB. Retroelements containing introns in diverse invertebrate taxa. Nat Genet. 2003;33:123–4.

99. Pyatkov KI, Arkhipova IR, Malkova N V, Finnegan DJ, Evgen'ev MB. Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. Proc Natl Acad Sci U S A. 2004;101:14719–24.

100. Gladyshev EA, Arkhipova IR. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. Proc Natl Acad Sci U S A. 2007;104:9352–7.

101. Craig RJ, Yushenova IA, Rodriguez F, Arkhipova IR. An ancient clade of Penelope-like retroelements with permuted domains is present in the green lineage and protists, and dominates many invertebrate genomes. Mol Biol Evol. 2021;38:5005–20.

102. Poulter RTM, Butler MI. Tyrosine Recombinase Retrotransposons and Transposons. Microbiol Spectr. 2015;3:MDNA3-0036–2014.

103. Ribeiro YC, Robe LJ, Veluza DS, Dos Santos CMB, Lopes ALK, Krieger MA, et al. Study of VIPER and TATE in kinetoplastids and the evolution of tyrosine recombinase retrotransposons. Mob DNA. 2019;10:34.

104. Cappello J, Handelsman K, Lodish HF. Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell. 1985;43:105–15.

105. Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio A V. A beginner's guide to manual curation of transposable elements. Mob DNA. 2022;13:7.

106. Goodwin TJD, Poulter RTM. A new group of tyrosine recombinase-encoding retrotransposons. Mol Biol Evol. 2004;21:746–59.

107. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 2007;41:331–68.

108. Spradling AC, Bellen HJ, Hoskins RA. Drosophila P elements preferentially transpose to replication origins.

Proc Natl Acad Sci. 2011;108:15948–53.

109. Hickman AB, Dyda F. DNA transposition at work. Chem Rev. 2016;116:12758–84.

110. Meng Q, Chen K, Ma L, Hu S, Yu J. A systematic identification of Kolobok superfamily transposons in Trichomonas vaginalis and sequence analysis on related transposases. J Genet Genomics. 2011;38:63–70.

111. Bao W, Kapitonov V V, Jurka J. Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. Mob DNA. 2010;1:3.

112. Bao W, Jurka MG, Kapitonov V V, Jurka J. New superfamilies of eukaryotic DNA transposons and their internal divisions. Mol Biol Evol. 2009;26:983–93.

113. Kapitonov V V, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 2005;3:e181.

114. Fattash I, Rooke R, Wong A, Hui C, Luu T, Bhardwaj P, et al. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. Genome. 2013;56:475–86.

115. Kojima KK, Jurka J. Crypton transposons: identification of new diverse families and ancient domestication events. Mob DNA. 2011;2:12.

116. Kapitonov V V., Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 2001;98:8714–9.

117. Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat Commun. 2016;7:10716.

118. Thomas J, Pritham EJ. Helitrons, the eukaryotic Rolling-circle transposable elements. Microbiol Spectr. 2015;3.

119. Haapa-Paananen S, Wahlberg N, Savilahti H. Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. Mol Phylogenet Evol. 2014;78:271–4.

120. Kapitonov V V., Jurka J. Self-synthesizing DNA transposons in eukaryotes. Proc Natl Acad Sci U S A. 2006;103:4540–5.

121. Krupovic M, Bamford DH, Koonin E V. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. Biol Direct. 2014;9:6.

122. Sessegolo C, Burlet N, Haudry A. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. Biol Lett. 2016;12.

123. 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.

124. Karageorgiou C, Gámez-Visairas V, Tarrío R, Rodríguez-Trelles F. Long-read based assembly and synteny analysis of a reference Drosophila subobscura genome reveals signatures of structural evolution driven by inversions recombination-suppression effects. BMC Genomics. 2019;20:223.

125. Rius N, Guillén Y, Delprat A, Kapusta A, Feschotte C, Ruiz A. Exploration of the Drosophila buzzatii transposable element content suggests underestimation of repeats in Drosophila genomes. BMC Genomics. 2016;17:344.

126. Haig D. Transposable elements: Self-seekers of the germline, team-players of the soma. Bioessays. 2016;38:1158–66.

127. Loreto ELS, Pereira CM. Somatizing the transposons action. Mob Genet Elements. 2017;7:1–9.

128. Treiber CD, Waddell S. Resolving the prevalence of somatic transposition in Drosophila. Elife. 2017;6.

129. Zamudio N. Transposable elements in the mammalian germline : a comfortable niche or a deadly trap? Heredity. 2010;105:92–104.

130. Laski FA, Rio DC, Rubin M. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. Cell. 1986;44:7–19.

131. Rubin GM, Kidwell MG, Bingham PM. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell. 1982;29:987–94.

132. Picard G, Bregliano JC, Bucheton A, Lavige JM, Pelisson A, Kidwell MG. Non-mendelian female sterility and hybrid dysgenesis in Drosophila melanogaster. Genet Res. 1978;32:275–87.

133. Craig NL. Target site selection in transposition. Annu Rev Biochem. 1997;66:437–74.

134. Boissinot S. On the base composition of transposable elements. Int J Mol Sci. 2022;23.

135. Lerat E, Capy P, Biémont C. Codon usage by transposable elements and their host genes in five species. J Mol Evol. 2002;54:625–37.

136. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol. 2006;4:e180.

137. Courel M, Clément Y, Bossevain C, Foretek D, Vidal Cruchez O, Yi Z, et al. GC content shapes mRNA storage and decay in human cells. Elife. 2019;8.

138. 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;19:1–12.

139. Wu SC-Y, Meir Y-JJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, et al. piggyBac is a flexible and highly active transposon as compared to Sleeping Beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci. 2006;103:15008–13.

140. Lohe AR, Hartl DL. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol. 1996;13:549–55.

141. Levin HL, Moran J V. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. 2011;12:615–627.

142. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD, O'Carroll D, et al. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019;20:89–108.

143. Wang X, Ramat A, Simonelig M, Liu M-F. Emerging roles and functional mechanisms of PIWI-interacting RNAs. Nat Rev Mol Cell Biol. 2023;24:123–41.

144. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. 2008;320:1077–81.

145. Mirkovic-Hösle M, Förstemann K. Transposon defense by endo-siRNAs, piRNAs and somatic pilRNAs in Drosophila: Contributions of Loqs-PD and R2D2. PLoS One. 2014;9:e84994.

146. Vagin V V, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. Science. 2006;313:320–4.

147. Théron E, Dennis C, Brasset E, Vaury C. Distinct features of the piRNA pathway in somatic and germ cells : from piRNA cluster transcription to piRNA processing and amplification. Mob DNA. 2014;5:1–11.

148. Fagegaltier D, Bougé A-L, Berry B, Poisot E, Sismeiro O, Coppée J-Y, et al. The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. Proc Natl Acad Sci U S A. 2009;106:21258–63.

149. Yin H, Sweeney S, Raha D, Snyder M, Lin H. A High-Resolution Whole-Genome map of key chromatin modifications in the adult Drosophila melanogaster. PLoS Genet. 2011;7.

150. Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. Genome Biol Evol. 2018;10:1185–97.

151. Brennecke J, Aravin AA, 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;128:1089–103.

152. Pélisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J. 1994;13:4401–11.

153. Shpiz S, Ryazansky S, Olovnikov I, Abramov Y, Kalmykova A. Euchromatic transposon insertions trigger production of novel pi- and endo-siRNAs at the target sites in the Drosophila germline. PLoS Genet. 2014;10.

154. Wierzbicki F, Kofler R, Signor S. Evolutionary dynamics of piRNA clusters in Drosophila. Mol Ecol. 2021; December:1–17.

155. Fu Y, Yang Y, Zhang H, Farley G, Wang J, Quarles KA, et al. The genome of the Hi5 germ cell line from Trichoplusia ni, an agricultural pest and novel model for small RNA biology. Elife. 2018;7.

156. Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep. 2014;15:411–8.

157. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell. 2014;157:1364–79.

158. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev. 2014;28:1667–80.

159. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, et al. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell. 2009;138:1137–49.

160. Chen P, Luo Y, Aravin AA. RDC complex executes a dynamic piRNA program during Drosophila spermatogenesis to safeguard male fertility. PLoS Genet. 2021;17:e1009591.

161. Chen P, Kotov AA, Godneeva BK, Bazylev SS, Olenina L V., Aravin AA. piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in D. melanogaster male germline. Genes Dev. 2021;35:914–35.

162. Saint-Leandre B, Capy P, Hua-Van A, Filée J. piRNA and transposon dynamics in Drosophila: a female story. Genome Biol Evol. 2020;12:931–47.

163. Gebert D, Neubert LK, Lloyd C, Gui J, Lehmann R, Teixeira FK, et al. Large Drosophila germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. Mol Cell. 2021;81:3965-3978.e5.

164. Sarot E, Payen-Groschêne G, Bucheton A, Pélisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics. 2004;166:1313–21.

165. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137:522–35.

166. Luo S, Zhang H, Duan Y, Yao X, Clark AG, Lu J. The evolutionary arms race between transposable elements and piRNAs in Drosophila melanogaster. BMC Evol Biol. 2020;20:14.

167. Shevelyov YY. Copies of a Stellate gene variant are located in the X heterochromatin of Drosophila melanogaster and are probably expressed. Genetics. 1992;132:1033–7.

168. Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, et al. Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. RNA. 2007;13:1911–22.

169. Bozzetti MP, Massari S, Finelli P, Meggio F, Pinna LA, Boldyreff B, et al. The Ste locus, a component of the parasitic cry-Ste system of Drosophila melanogaster, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. Proc Natl Acad Sci U S A. 1995;92:6067–71.

170. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature. 2009;461:1296–9.

171. Klein JD, Qu C, Yang X, Fan Y, Tang C, Peng JC. c-Fos repression by Piwi regulates Drosophila ovarian germline formation and tissue morphogenesis. PLoS Genet. 2016;12:e1006281.

172. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science. 2008;322:1387–92.

173. Akkouche A, Mugat B, Barckmann B, Varela-Chavez C, Li B, Raffel R, et al. Piwi is required during Drosophila embryogenesis to license dual-strand piRNA clusters for transposon repression in adult ovaries. Mol Cell. 2017;66:411-419.e4.

174. Fabry MH, Falconio FA, Joud F, Lythgoe EK, Czech B, Hannon GJ. Maternally inherited piRNAs direct transient heterochromatin formation at active transposons during early Drosophila embryogenesis. Elife. 2021;10.

175. Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J. The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. Elife. 2022;11.

176. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157:1353–63.

177. Andersen PR, Tirian L, Vunjak M, Brennecke J. A heterochromatin-dependent transcription machinery drives piRNA expression. Nature. 2017;549:54–9.

178. Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, et al. PiRNA-guided genome defense: From biogenesis to silencing. Annu Rev Genet. 2018;52:131–57.

179. Pane A, Jiang P, Zhao DY, Singh M, Schüpbach T. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J. 2011;30:4601–15.

180. Chen Y-CA, Stuwe E, Luo Y, Ninova M, Le Thomas A, Rozhavskaya E, et al. Cutoff suppresses RNA Polymerase II termination to ensure Expression of piRNA precursors. Mol Cell. 2016;63:97–109.

181. Hur JK, Luo Y, Moon S, Ninova M, Marinov GK, Chung YD, et al. Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev. 2016;30:840–55.

182. Sato K, Siomi MC, Ato BKS, Iomi MCS. The piRNA pathway in Drosophila ovarian germ and somatic cells. Proc Jpn Acad Ser B Phys Biol Sci. 2020;96:32–42.

183. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, et al. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell. 2012;151:871–84.

184. Wang W, Yoshikawa M, Han BW, Izumi N, Tomari Y, Weng Z, et al. The initial uridine of primary piRNAs does not create the tenth adenine that Is the hallmark of secondary piRNAs. Mol Cell. 2014;56:708–16.

185. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 2007;315:1587–90.

186. Huang X, Hu H, Webster A, Zou F, Du J, Patel DJ, et al. Binding of guide piRNA triggers methylation of the unstructured N-terminal region of Aub leading to assembly of the piRNA amplification complex. Nat Commun. 2021;12:4061.

187. Webster A, Li S, Hur JK, Wachsmuth M, Bois JS, Perkins EM, et al. Aub and Ago3 are recruited to nuage through two mechanisms to form a Ping-Pong complex assembled by Krimper. Mol Cell. 2015;59:564–75.

188. Sato K, Iwasaki YW, Shibuya A, Carninci P, Tsuchizawa Y, Ishizu H, et al. Krimper enforces an antisense bias on piRNA pools by binding AGO3 in the Drosophila germline. Mol Cell. 2015;59:553–63.

189. Xiol J, Spinelli P, Laussmann MA, Homolka D, Yang Z, Cora E, et al. RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell. 2014;157:1698–711.

190. Zhang Z, Koppetsch BS, Wang J, Tipping C, Weng Z, Theurkauf WE, et al. Antisense piRNA amplification, but not piRNA production or nuage assembly, requires the Tudor-domain protein Qin. The EMBO journal. 2014;33:536–9.

191. Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, et al. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell. 2011;44:572–84.

192. Cappucci U, Noro F, Casale AM, Fanti L, Berloco M, Alagia AA, et al. The Hsp70 chaperone is a major player in stress-induced transposable element activation. Proc Natl Acad Sci U S A. 2019;116:17943–50.

193. Olivieri D, Senti K-A, Subramanian S, Sachidanandam R, Brennecke J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. Mol Cell. 2012;47:954–69.

194. Feltzin VL, Khaladkar M, Abe M, Parisi M, Hendriks G-J, Kim J, et al. The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in Drosophila. Aging Cell. 2015;14:443–52.

195. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348:817–21.

196. 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;138:1863–73.

197. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348:812–7.

198. Vagin V V, Yu Y, Jankowska A, Luo Y, Wasik KA, Malone CD, et al. Minotaur is critical for primary piRNA biogenesis. RNA. 2013;19:1064–77.

199. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev. 2007;21:1603–8.

200. Pastore B, Hertz HL, Price IF, Tang W. pre-piRNA trimming and 2'-O-methylation protect piRNAs from 3' tailing and degradation in C. elegans. Cell Rep. 2021;36:109640.

201. Ott KM, Nguyen T, Navarro C. The DExH box helicase domain of spindle-E is necessary for retrotransposon silencing and axial patterning during Drosophila oogenesis. G3 (Bethesda). 2014;4:2247–57.

202. Patil VS, Kai T. Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol. 2010;20:724–30.

203. Patil VS, Anand A, Chakrabarti A, Kai T. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC Biol. 2014;12:61.

204. Lin Y, Suyama R, Kawaguchi S, Iki T, Kai T. Tejas functions as a core component of nuage assembly and precursor processing in Drosophila piRNA biogenesis. bioRxiv. 2022;:2022.07.12.499660.

205. Ge DT, Wang W, Tipping C, Gainetdinov I, Weng Z, Zamore PD. The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria. Mol Cell. 2019;74:982-995.e6.

206. Huang H, Li Y, Szulwach KE, Zhang G, Jin P, Chen D. AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. J Cell Biol. 2014;206:217–30.

207. Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, et al. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 2011;30:3977–93.

208. Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, et al. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development. 2011;138:4039–50.

209. Munafò M, Manelli V, Falconio FA, Sawle A, Kneuss E, Eastwood EL, et al. Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. Genes Dev. 2019;33:844–56.

210. Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, et al. PIWI slicing and RNA Elements in precursors instruct directional primary piRNA biogenesis. Cell Rep. 2015;12:418–28.

211. Izumi N, Shoji K, Suzuki Y, Katsuma S, Tomari Y. Zucchini consensus motifs determine the mechanism of pre-piRNA production. Nature. 2020;578:311–6.

212. Stein CB, Genzor P, Mitra S, Elchert AR, Ipsaro JJ, Benner L, et al. Decoding the 5' nucleotide bias of PIWI-interacting RNAs. Nat Commun. 2019;10:828.

213. Zhang Y, Liu W, Li R, Gu J, Wu P, Peng C, et al. Structural insights into the sequence-specific recognition of

Piwi by Drosophila Papi. Proc Natl Acad Sci U S A. 2018;115:3374–9.

214. Yashiro R, Murota Y, Nishida KM, Yamashiro H, Fujii K, Ogai A, et al. Piwi nuclear localization and its regulatory mechanism in Drosophila ovarian somatic cells. Cell Rep. 2018;23:3647–57.

215. Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ. shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA. 2012;18:1446–57.

216. Yamaguchi S, Oe A, Nishida KM, Yamashita K, Kajiya A, Hirano S, et al. Crystal structure of Drosophila Piwi. Nat Commun. 2020;11:858.

217. Darricarrère N, Liu N, Watanabe T, Lin H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc Natl Acad Sci. 2013;110:1297–302.

218. Gainetdinov I, Colpan C, Arif A, Cecchini K, Zamore PD. A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. Mol Cell. 2018;71:775-790.e5.

219. Ilyin AA, Ryazansky SS, Doronin SA, Olenkina OM, Mikhaleva EA, Yakushev EY, et al. Piwi interacts with chromatin at nuclear pores and promiscuously binds nuclear transcripts in Drosophila ovarian somatic cells. Nucleic Acids Res. 2017;45:7666–80.

220. Zhao K, Cheng S, Miao N, Xu P, Lu X, Zhang Y, et al. A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. Nat Cell Biol. 2019;21:1261–72.

221. Dönertas D, Sienski G, Brennecke J. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev. 2013;27:1693–705.

222. Onishi R, Sato K, Murano K, Negishi L, Siomi H, Siomi MC. Piwi suppresses transcription of Brahma-dependent transposons via Maelstrom in ovarian somatic cells. Sci Adv. 2020;6.

223. Batki J, Schnabl J, Wang J, Handler D, Andreev VI, Stieger CE, et al. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. Nat Struct Mol Biol. 2019;26:720–31.

224. Fabry MH, Ciabrelli F, Munafò M, Eastwood EL, Kneuss E, Falciatori I, et al. piRNA-guided co-transcriptional silencing coopts nuclear export factors. Elife. 2019;8.

225. Murano K, Iwasaki YW, Ishizu H, Mashiko A, Shibuya A, Kondo S, et al. Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. EMBO J. 2019;38:e102870.

226. Eastwood EL, Jara KA, Bornelöv S, Munafò M, Frantzis V, Kneuss E, et al. Dimerisation of the PICTS complex via LC8/Cut-up drives co-transcriptional transposon silencing in Drosophila. Elife. 2021;10.

227. Schnabl J, Wang J, Hohmann U, Gehre M, Batki J, Andreev VI, et al. Molecular principles of Piwi-mediated cotranscriptional silencing through the dimeric SFiNX complex. Genes Dev. 2021;35:392–409.

228. Ninova M, Chen Y-CA, Godneeva B, Rogers AK, Luo Y, Fejes Tóth K, et al. Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. Mol Cell. 2020;77:556-570.e6.

229. Osumi K, Sato K, Murano K, Siomi H, Siomi MC. Essential roles of Windei and nuclear monoubiquitination of Eggless/SETDB1 in transposon silencing. EMBO Rep. 2019;20:e48296.

230. Sienski G, Batki J, Senti K-A, Dönertas D, Tirian L, Meixner K, et al. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev. 2015;29:2258–71.

231. Iwasaki YW, Murano K, Ishizu H, Shibuya A, Iyoda Y, Siomi MC, et al. Piwi modulates chromatin accessibility by regulating multiple factors including histone H1 to repress transposons. Mol Cell. 2016;63:408–19.

232. Mugat B, Nicot S, Varela-Chavez C, Jourdan C, Sato K, Basyuk E, et al. The Mi-2 nucleosome remodeler and the Rpd3 histone deacetylase are involved in piRNA-guided heterochromatin formation. Nat Commun. 2020;11:2818.

233. Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, et al. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science. 2015;350:339–42.

234. Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, et al. Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. EMBO J. 1999;18:2659–69.

235. Jones BC, Wood JG, Chang C, Tam AD, Franklin MJ, Siegel ER, et al. A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. Nat Commun. 2016;7:13856.

236. Perrat PN, Dasgupta S, Wang J, Theurkauf W, Weng Z, Rosbash M, et al. Transposition driven genomic heterogeneity in the Drosophila brain. Science. 2013;340.

237. Claycomb JM. Ancient endo-siRNA pathways reveal new tricks. Curr Biol. 2014;24:R703–15.

238. Okamura K, Chung W-J, Ruby JG, Guo H, Bartel DP, Lai EC. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature. 2008;453:803–6.

239. Miyoshi K, Miyoshi T, Hartig JV, Siomi H, Siomi MC. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. Rna. 2010;16:506–15.

240. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, et al. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell. 2010;39:292–9.

241. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell. 2005;123:607–20.

242. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17:1265–72.

243. Marques JT, Kim K, Wu P-H, Alleyne TM, Jafari N, Carthew RW. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat Struct Mol Biol. 2010;17:24–30.

244. van Lopik J, Trapotsi M-A, Hannon GJ, Bornelöv S, Nicholson BC, Lopik J van, et al. Unistrand piRNA clusters are an evolutionary conserved mechanism to suppress endogenous retroviruses across the Drosophila genus. bioRxiv. 2023;:2023.02.27.530199.

245. Desset S, Meignin C, Dastugue B, Vaury C. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. Genetics. 2003;164:501–9.

246. Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A. Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics. 1995;139:697–711.

247. Mével-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics. 2007;175:1615–24.

248. Ishizu H, Iwasaki YW, Hirakata S, Ozaki H, Iwasaki W, Siomi H, et al. Somatic primary piRNA biogenesis driven by cis-acting RNA elements and trans-acting Yb. Cell Rep. 2015;12:429–40.

249. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41:324–37.

250. Dennis C, Brasset E, Sarkar A, Vaury C. Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in Drosophila. Nat Commun. 2016;7:13739.

251. Munafò M, Lawless VR, Passera A, MacMillan S, Bornelöv S, Haussmann IU, et al. Channel nuclear pore complex subunits are required for transposon silencing in Drosophila. Elife. 2021;10.

252. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, et al. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. 2010;24:2493–8.

253. Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. 2010;29:3301–17.

254. Qi H, Watanabe T, Ku H-Y, Liu N, Zhong M, Lin H. The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. J Biol Chem. 2011;286:3789–97.

255. Hirakata S, Ishizu H, Fujita A, Tomoe Y, Siomi MC. Requirements for multivalent Yb body assembly in transposon silencing in Drosophila. EMBO Rep. 2019;20:e47708.

256. Szakmary A, Reedy M, Qi H, Lin H. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. J Cell Biol. 2009;185:613–27.

257. Dennis C, Zanni V, Brasset E, Eymery A, Zhang L, Mteirek R, et al. "Dot COM", a nuclear transit center for the primary piRNA pathway in Drosophila. PLoS One. 2013;8:e72752.

258. Sokolova OA, Ilyin AA, Poltavets AS, Nenasheva V V, Mikhaleva EA, Shevelyov YY, et al. Yb body assembly on the flamenco piRNA precursor transcripts reduces genic piRNA production. Mol Biol Cell. 2019;30:1544–54.

259. Murota Y, Ishizu H, Nakagawa S, Iwasaki YW, Shibata S, Kamatani MK, et al. Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly. Cell Rep. 2014;8:103–13.

260. Dennis C, Brasset E, Vaury C. flam piRNA precursors channel from the nucleus to the cytoplasm in a temporally regulated manner along Drosophila oogenesis. Mob DNA. 2019;10:28.

261. Ishizu H, Kinoshita T, Hirakata S, Komatsuzaki C, Siomi MC. Distinct and collaborative functions of Yb and Armitage in transposon-targeting piRNA biogenesis. Cell Rep. 2019;27:1822-1835.e8.

262. Yamashiro H, Negishi M, Kinoshita T, Ishizu H, Ohtani H, Siomi MC. Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. EMBO Rep. 2020;21:e48769.

263. Thomas CAJ. The genetic organization of chromosomes. Annu Rev Genet. 1971;5:237–56.

264. Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. Transposons, genome size, and evolutionary insights in animals. Cytogenet Genome Res. 2015;147:217–39.

265. Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. Genetica. 2002;115:49–63.

266. Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. Nat Commun. 2010;1:77.

267. Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, et al. Giant lungfish genome elucidates the conquest of land by vertebrates. Nature. 2021;590:284–9.

268. Almojil D, Bourgeois Y, Falis M, Hariyani I, Wilcox J, Boissinot S. The structural, functional and evolutionary impact of Transposable Elements in eukaryotes. Genes. 2021;12.

269. Michael TP. Plant genome size variation: bloating and purging DNA. Brief Funct Genomics. 2014;13:308–17.

270. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A. 2017;114:E1460–9.

271. van de Lagemaat LN, Gagnier L, Medstrand P, Mager DL. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 2005;15:1243–9.

272. Symonová R, Suh A. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. Mob DNA. 2019;10:49.

273. Fullerton SM, Bernardo Carvalho A, Clark AG. Local rates of recombination are positively correlated with GC content in the human genome. Molecular biology and evolution. 2001;18:1139–42.

274. Rizzon C, Marais G, Gouy M, Biémont C. Recombination rate and the distribution of transposable elements in the Drosophila melanogaster genome. Genome Res. 2002;12:400–7.

275. Dubin MJ, Mittelsten Scheid O, Becker C. Transposons: a blessing curse. Curr Opin Plant Biol. 2018;42:23–9.

276. Kazazian HHJ, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature. 1988;332:164–6.

277. Kidd S, Young MW. Transposon-dependent mutant phenotypes at the Notch locus of Drosophila. Nature. 1986;323:89–91.

278. Colonna Romano N, Fanti L. Transposable elements: major players in shaping genomic and evolutionary patterns. Cells. 2022;11.

279. Carr B, Anderson P. Imprecise excision of the Caenorhabditis elegans transposon Tc1 creates functional 5' splice sites. Mol Cell Biol. 1994;14:3426–33.

280. Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, et al. Human retroelements may introduce intragenic polyadenylation signals. Cytogenet Genome Res. 2005;110:365–71.

281. Snyder MP, Kimbrell D, Hunkapiller M, Hill R, Fristrom J, Davidson N. A transposable element that splits the promoter region inactivates a Drosophila cuticle protein gene. Proc Natl Acad Sci U S A. 1982;79:7430–4.

282. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 2014;24:1963–76.

283. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007;8:272–85.

284. Iida S, Morita Y, Choi J-D, Park K-I, Hoshino A. Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. Adv Biophys. 2004;38 Complete:141–59.

285. Sienski G, Dönertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151:964–80.

286. Zattera ML, Bruschi DP. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. Cells. 2022;11.

287. Su LK, Steinbach G, Sawyer JC, Hindi M, Ward PA, Lynch PM. Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. Hum Genet. 2000;106:101–7.

288. Cáceres M, Puig M, Ruiz A. Molecular characterization of two natural hotspots in the Drosophila buzzatii genome induced by transposon insertions. Genome Res. 2001;11:1353–64.

289. Szak ST, Pickeral OK, Landsman D, Boeke JD. Identifying related L1 retrotransposons by analyzing 3' transduced sequences. Genome Biol. 2003;4:R30.

290. Brosius J. Retroposons--seeds of evolution. Science. 1991;251:753.

291. Pasero P, Sjakste N, Blettry C, Got C, Marilley M. Long-range organization and sequence-directed curvature of Xenopus laevis satellite 1 DNA. Nucleic Acids Res. 1993;21:4703–10.

292. Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534:102–5.

293. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet. 2013;45:836–41.

294. Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, et al. Diverse splicing patterns of exonized Alu elements in human tissues. PLoS Genet. 2008;4:e1000225.

295. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 2013;23:169–80.

296. Miller WJ, McDonald JF, Pinsker W. Molecular domestication of mobile elements. Genetica. 1997;100:261–70.

297. Capkova Frydrychova R, Biessmann H, Mason JM. Regulation of telomere length in Drosophila. Cytogenet Genome Res. 2008;122:356–64.

298. Chang C-H, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, et al. Islands of retroelements are major components of Drosophila centromeres. PLoS Biol. 2019;17:e3000241.

299. Emera D, Wagner GP. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. Brief Funct Genomics. 2012;11:267–76.

300. Curcio MJ, Belfort M. The beginning of the end: links between ancient retroelements and modern telomerases. Proc Natl Acad Sci U S A. 2007;104:9107–8.

301. Zhang Y, Cheng TC, Huang G, Lu Q, Surleac MD, Mandell JD, et al. Transposon molecular domestication and the evolution of the RAG recombinase. Nature. 2019;569:79–84.

302. Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, et al. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell. 2016;166:102–14.

303. Dazenière J, Bousios A, Eyre-Walker A. Patterns of selection in the evolution of a transposable element. G3 (Bethesda). 2022;12.

304. Sousa A, Bourgard C, Wahl LM, Gordo I. Rates of transposition in Escherichia coli. Biol Lett. 2013;9:20130838.

305. Harada K, Yukuhiro K, Mukai T. Transposition rates of movable genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1990;87:3248–52.

306. Gerasimova TI, Matjunina L V, Mizrokhi LJ, Georgiev GP. Successive transposition explosions in Drosophila melanogaster and reverse transpositions of mobile dispersed genetic elements. EMBO J. 1985;4:3773–9.

307. Kocíncová D, Singh AK, Beretti J-L, Ren H, Euphrasie D, Liu J, et al. Spontaneous transposition of IS1096 or ISMsm3 leads to glycopeptidolipid overproduction and affects surface properties in Mycobacterium smegmatis. Tuberculosis (Edinb). 2008;88:390–8.

308. Ikeda K, Nakayashiki H, Takagi M, Tosa Y, Mayama S. Heat shock, copper sulfate and oxidative stress activate the retrotransposon MAGGY resident in the plant pathogenic fungus Magnaporthe grisea. Mol Genet Genomics. 2001;266:318–25.

309. Jouan-Dufournel I, Cosset FL, Contamine D, Verdier G, Biémont C. Transposable elements behavior following viral genomic stress in Drosophila melanogaster inbred line. J Mol Evol. 1996;43:19–27.

310. O'Neill RJ, O'Neill MJ, Graves JA. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature. 1998;393:68–72.

311. Smukowski Heil C, Patterson K, Hickey AS-M, Alcantara E, Dunham MJ. Transposable element mobilization in interspecific yeast hybrids. Genome Biol Evol. 2021;13.

312. Ungerer MC, Strakosh SC, Zhen Y. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Current biology : CB. 2006;16:R872-3.

313. Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell. 1984;38:153–63.

314. Kidwell MG, Kidwell JF, Sved JA. Hybrid Dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics. 1977;86:813–33.

315. Yannopoulos G, Stamatis N, Monastirioti M, Hatzopoulos P, Louis C. hobo is responsible for the induction of hybrid dysgenesis by strains of Drosophila melanogaster bearing the male recombination factor 23.5MRF. Cell. 1987;49:487–95.

316. Hill T, Schlötterer C, Betancourt AJ. Hybrid dysgenesis in Drosophila simulans associated with a rapid invasion of the P-Element. PLoS Genet. 2016;12:e1005920.

317. Blumenstiel JP. Whole genome sequencing in Drosophila virilis identifies Polyphemus, a recently activated Tc1-like transposon with a possible role in hybrid dysgenesis. Mob DNA. 2014;5:6.

318. Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER. Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A. 1995;92:8050–4.

319. Tasnim S, Kelleher ES. p53 is required for female germline stem cell maintenance in P-element hybrid dysgenesis. Dev Biol. 2018;434:215–20.

320. Lama J, Srivastav S, Tasnim S, Hubbard D, Kelleher E. Natural tolerance to transposition is associated with Myc-regulation and DNA repair. bioRxiv. 2021. https://doi.org/10.1101/2021.04.30.441852.

321. Serrato-Capuchina A, Wang J, Earley E, Peede D, Isbell K, Matute DR. Paternally inherited P-element copy number affects the magnitude of hybrid Dysgenesis in Drosophila simulans and D. melanogaster. Genome Biol Evol. 2020;12:808–26.

322. Gorbunova V, Seluanov A, Mita P, McKerrow W, Fenyö D, Boeke JD, et al. The role of retrotransposable elements in ageing and age-associated diseases. Nature. 2021;596:43–53.

323. Barton NH, Hewitt GM. Analysis of Hybrid Zones. Annu Rev Ecol Syst. 1985;16:113–48.

324. Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. Annu Rev Genet. 2011;45:331–55.

325. Anderson E, Hubricht L. Hybridization in tradescantia. III. The evidence for introgressive hybridization. Am J Bot. 1938;25:396–402.

326. Rius M, Darling JA. How important is intraspecific genetic admixture to the success of colonising populations? Trends Ecol Evol. 2014;29:233–42.

327. Go AC, Civetta A. Hybrid incompatibilities and transgressive gene expression between two closely related subspecies of Drosophila. Front Genet. 2020;11:599292.

328. Grant PR, Grant BR. Hybridization increases population variation during adaptive radiation. Proc Natl Acad Sci U S A. 2019;116:23216–24.

329. Kotov AA, Adashev VE, Godneeva BK, Ninova M, Shatskikh AS, Bazylev SS, et al. piRNA silencing contributes to interspecies hybrid sterility and reproductive isolation in Drosophila melanogaster. Nucleic Acids Res. 2019;47:4255–71.

330. 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;7:40618.

331. Naveira H, Fontdevila A. The evolutionary history of Drosophila buzzatii. IX. High frequencies of new chromosome rearrangements induced by introgressive hybridization. Chromosoma. 1985;91:87–94.

332. Dobzhansky T. Genetics and the origin of species. New York: Columbia Univ. Press; 1937.

333. E M. Systematic and the Origin of Species. New York: Columbia University Press; 1942.

334. Coyne JA, Orr HA. Speciation. Sunderland Mass: Sinauer Associates; 2004.

335. Vallejo-Marín M, Hiscock SJ. Hybridization and hybrid speciation under global change. New Phytol. 2016;211:1170–87.

336. Owens GL, Huang K, Todesco M, Rieseberg LH. Re-evaluating homoploid reticulate evolution in Helianthus sunflowers. Mol Biol Evol. 2023;40:msad013.

337. Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M. Speciation by hybridization in Heliconius butterflies. Nature. 2006;441:868–71.

338. Schwarz D, Matta BM, Shakir-Botteri NL, McPheron BA. Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature. 2005;436:546–9.

339. Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. Rapid hybrid speciation in Darwin's finches. Science. 2018;359:224–8.

340. Amaral AR, Lovewell G, Coelho MM, Amato G, Rosenbaum HC. Hybrid speciation in a marine mammal: the clymene dolphin (Stenella clymene). PLoS One. 2014;9:e83645.

341. Zhu W, Hu B, Becker C, Doğan ES, Berendzen KW, Weigel D, et al. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. Genome Biol. 2017;18:157.

342. Vrana PB, Fossella JA, Matteson P, del Rio T, O'Neill MJ, Tilghman SM. Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in Peromyscus. Nat Genet. 2000;25:120–4.

343. 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.

344. Michalak P, Malone JH. Testis-derived microRNA profiles of African clawed frogs (Xenopus) and their sterile hybrids. Genomics. 2008;91:158–64.

345. Zhang G, Li J, Zhang J, Liang X, Zhang X, Wang T, et al. Integrated analysis of transcriptomic, miRNA and proteomic changes of a Novel hybrid yellow catfish uncovers key roles for miRNAs in heterosis. Mol Cell Proteomics. 2019;18:1437–53.

346. Li J, Zhou J, Zhang Y, Yang Y, Pu Q, Tao D. New Insights Into the Nature of Interspecific Hybrid Sterility in Rice. Front Plant Sci. 2020;11:555572.

347. Naveira H, Fondevila 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 (Pt 1):57–72.

348. Suda K, Hayashi SR, Tamura K, Takamatsu N, Ito M. Activation of DNA Transposons and Evolution of piRNA Genes Through Interspecific Hybridization in Xenopus Frogs. Front Genet. 2022;13:766424.

349. Wang Y, Zhang Y, Zhang Q, Cui Y, Xiang J, Chen H, et al. Comparative transcriptome analysis of panicle development under heat stress in two rice (Oryza sativa L.) cultivars differing in heat tolerance. PeerJ. 2019;7:e7595.

350. Kelleher ES, Edelman NB, Barbash DA. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 2012;10:e1001428.

351. Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin Y V. Transposition of mobile genetic elements in interspecific hybrids of Drosophila. Chromosoma. 1982;85:375–86.

352. Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, et al. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. Proc Natl Acad Sci U S A. 2009;106:17835–40.

353. Naveira H, Fontdevila A. The evolutionary history of D. buzzatii. XXII. Chromosomal and genic sterility in male hybrids of Drosophila buzzatii and Drosophila koepferae. Heredity. 1991;66 (Pt 2):233–9.

354. García-Franco F, Barandica-Cañon LM, Arandiabarrios J, Galindo-Pérez EJ, Binnqüist Cervantes GS, Martínez García M, et al. Evaluation of Drosophila chromosomal segments proposed by means of simulations of possessing hybrid sterility genes from reproductive isolation. J Genet. 2020;99.

355. Marín I. Genetic architecture of autosome-mediated hybrid male sterility in Drosophila. Genetics. 1996;142:1169–80.

356. Carvajal AR, Gandarela MR, Naveira HF. A three-locus system of interspecific incompatibility underlies male inviability in hybrids between Drosophila buzzatii and D. koepferae. Genetica. 1996;98:1–19.

357. Morán T, Fontdevila A. Genome-wide dissection of hybrid sterility in Drosophila confirms a polygenic threshold architecture. J Hered. 2014;105:381–96.

358. Carvajal A. Developmental disruptions in hybrids between D. buzzatii and D. koepferae. Int J Dev Biol. 1996;Suppl 1:93S-94S.

359. Fontdevila A. Hybrid genome evolution by transposition: an update. J Hered. 2019;110:124–36.

360. Labrador M, Fontdevila A. High transposition rates of Osvaldo, a new Drosophila buzzatii retrotransposon. Mol Gen Genet. 1994;245:661–74.

361. Labrador M, Farré M, Utzet F, Fontdevila A. Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol. 1999;16:931–7.

362. 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.

363. Romero-Soriano V, Burlet N, Vela D, Fontdevila A, Vieira C, García Guerreiro MP. Drosophila females undergo genome expansion after interspecific hybridization. Genome biology and evolution. 2016;8:556–61.

364. 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.

365. 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.

366. Santer B, Taylor KE, Penner J. A search for human influences on the thermal structure of the atmosphere. Nature. 1995;382.

367. Yoganathan D, Rom WN. Medical aspects of global warming. Am J Ind Med. 2001;40:199–210.

368. Mora C, McKenzie T, Gaw IM, Dean JM, von Hammerstein H, Knudson TA, et al. Over half of known human pathogenic diseases can be aggravated by climate change. Nat Clim Chang. 2022;12:869–75.

369. Li K, Pan J, Xiong W, Xie W, Ali T. The impact of 1.5 °C and 2.0 °C global warming on global maize production and trade. Sci Rep. 2022;12:17268.

370. Ali MZ, Carlile G, Giasuddin M. Impact of global climate change on livestock health: Bangladesh perspective. Open Vet J. 2020;10:178–88.

371. Barbarossa V, Bosmans J, Wanders N, King H, Bierkens MFP, Huijbregts MAJ, et al. Threats of global warming to the world's freshwater fishes. Nat Commun. 2021;12:1701.

372. Shockley KR, Ward DE, Chhabra SR, Conners SB, Montero CI, Kelly RM. Heat shock response by the hyperthermophilic archaeon Pyrococcus furiosus. Appl Environ Microbiol. 2003;69:2365–71.

373. Arbona M, De Frutos R. Stress response in Drosophila subobscura. II. Puff activity during anoxia and recovery from anoxia. Biol cell. 1987;60:173–81.

374. Zhong M, Orosz A, Wu C. Direct sensing of heat and oxidation by Drosophila heat shock transcription factor. Mol Cell. 1998;2:101–8.

375. Vass K, Welch WJ, Nowak TSJ. Localization of 70-kDa stress protein induction in gerbil brain after ischemia. Acta Neuropathol. 1988;77:128–35.

376. Bournias-Vardiabasis N, Buzin CH. Developmental effects of chemicals and the heat shock response in Drosophila cells. Teratog Carcinog Mutagen. 1986;6:523–36.

377. Lindquist S. The heat-shock response. Annu Rev Biochem. 1986;55:1151–91.

378. Ritossa F. A new puffing pattern induced by temperature shock and DNP in drosophila. Experientia. 1962;18:571–3.

379. Berendes HD. Factors involved in the expression of gene activity in polytene chromosomes. Chromosoma. 1968;24:418–37.

380. Pascual L, Latorre A, de Frutos R. Stress response in Drosophila subobscura. III. Variability of heat shock puffs. Biol cell. 1987;61:15–21.

381. Tissières A, Mitchell HK, Tracy UM. Protein synthesis in salivary glands of Drosophila melanogaster: relation to chromosome puffs. J Mol Biol. 1974;84:389–98.

382. Ashburner M, Bonner JJ. The induction of gene activity in drosophilia by heat shock. Cell. 1979;17:241–54.

383. Ashburner M. Patterns of puffing activity in the salivary gland chromosomes of Drosophila. V. Responses to environmental treatments. Chromosoma. 1970;31:356–76.

384. Pascual L, de Frutos R. Stress response in Drosophila subobscura. Chromosoma. 1988;97:164–70.

385. Moltó MD, Pascual L, de Frutos R. Puff activity after heat shock in two species of the Drosophila obscura group. Experientia. 1987;43:1225–7.

386. Moltó MD, Pascual L, Martínez-Sebastián MJ, de Frutos R. Genetic analysis of heat shock response in three Drosophila species of the obscura group. Genome. 1992;35:870–80.

387. Westwood JT, Wu C. Activation of Drosophila heat shock factor: conformational change associated with a monomer-to-trimer transition. Mol Cell Biol. 1993;13:3481–6.

388. Marchler G, Wu C. Modulation of Drosophila heat shock transcription factor activity by the molecular chaperone DROJ1. EMBO J. 2001;20:499–509.

389. Zou J, Guo Y, Guettouche T, Smith DF, Voellmy R. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. Cell. 1998;94:471–80.

390. Sawarkar R, Sievers C, Paro R. Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. Cell. 2012;149:807–18.

391. Solomon JM, Rossi JM, Golic K, McGarry T, Lindquist S. Changes in hsp70 alter thermotolerance and heat-shock regulation in Drosophila. New Biol. 1991;3:1106–20.

392. Parker CS, Topol J. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell. 1984;37:273–83.

393. Fujikake N, Nagai Y, Popiel HA, Kano H, Yamaguchi M, Toda T. Alternative splicing regulates the transcriptional activity of Drosophila heat shock transcription factor in response to heat/cold stress. FEBS Lett. 2005;579:3842–8.

394. Orosz A, Wisniewski J, Wu C. Regulation of Drosophila heat shock factor trimerization: global sequence requirements and independence of nuclear localization. Mol Cell Biol. 1996;16:7018–30.

395. Hightower LE. Heat shock, stress proteins, chaperones, and proteotoxicity. Cell. 1991;66:191–7.

396. Amin J, Ananthan J, Voellmy R. Key features of heat shock regulatory elements. Mol Cell Biol. 1988;8:3761–9.

397. Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ, Lis JT. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev. 2016;30:1731–46.

398. Tian S, Haney RA, Feder ME. Phylogeny disambiguates the evolution of heat-shock cis-regulatory elements in Drosophila. PLoS One. 2010;5:e10669.

399. Rasmussen EB, Lis JT. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A. 1993;90:7923–7.

400. Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A, et al. GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet. 2015;11:e1005108.

401. Fritsch M, Wu C. Phosphorylation of Drosophila heat shock transcription factor. Cell Stress Chaperones. 1999;4:102–17.

402. Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, et al. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Mol Cell Biol. 2006;26:250–60.

403. Baler R, Zou J, Voellmy R. Evidence for a role of Hsp70 in the regulation of the heat shock response in mammalian cells. Cell Stress Chaperones. 1996;1:33–9.

404. Feder JH, Rossi JM, Solomon J, Solomon N, Lindquist S. The consequences of expressing hsp70 in Drosophila cells at normal temperatures. Genes Dev. 1992;6:1402–13.

405. Gonsalves SE, Moses AM, Razak Z, Robert F, Westwood JT. Whole-genome analysis reveals that active heat shock factor binding sites are mostly associated with non-heat shock genes in Drosophila melanogaster. PLoS One. 2011;6:e15934.

406. Tutar L, Tutar Y. Heat shock proteins; an overview. Curr Pharm Biotechnol. 2010;11:216–22.

407. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. Mol Cell. 2010;40:253–66.

408. Stollar EJ, Smith DP. Uncovering protein structure. Essays Biochem. 2020;64:649–80.

409. Tang Y-C, Chang H-C, Hayer-Hartl M, Hartl FU, Tang Y-C, Hayer-Hartl M, et al. SnapShot: molecular chaperones, Part I. Cell. 2007;128:412.

410. Craig EA, Ingolia TD, Manseau LJ. Expression of Drosophila heat-shock cognate genes during heat shock and development. Dev Biol. 1983;99:418–26.

411. Shaner L, Morano KA. All in the family: atypical Hsp70 chaperones are conserved modulators of Hsp70 activity. Cell Stress Chaperones. 2007;12:1–8.

412. Qiu X-B, Shao Y-M, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci. 2006;63:2560–70.

413. Rutherford S, Hirate Y, Swalla BJ. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. Crit Rev Biochem Mol Biol. 2007;42:355–72.

414. Gangaraju VK, Yin H, Weiner MM, Wang J, Huang XA, Lin H. Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. Nat Genet. 2011;43:153–8.

415. Tariq M, Nussbaumer U, Chen Y, Beisel C, Paro R. Trithorax requires Hsp90 for maintenance of active chromatin at sites of gene expression. Proc Natl Acad Sci U S A. 2009;106:1157–62.

416. Matsushima Y, Hirofuji Y, Aihara M, Yue S, Uchiumi T, Kaguni LS, et al. Drosophila protease ClpXP specifically degrades DmLRPPRC1 controlling mitochondrial mRNA and translation. Sci Rep. 2017;7:8315.

417. Höhfeld J, Hartl FU. Role of the chaperonin cofactor Hsp10 in protein folding and sorting in yeast mitochondria. J Cell Biol. 1994;126:305–15.

418. Monzo K, Dowd SR, Minden JS, Sisson JC. Proteomic analysis reveals CCT is a target of Fragile X mental retardation protein regulation in Drosophila. Dev Biol. 2010;340:408–18.

419. Kim A-R, Choi K-W. TRiC/CCT chaperonins are essential for organ growth by interacting with insulin/TOR signaling in Drosophila. Oncogene. 2019;38:4739–54.

420. Dabbaghizadeh A, Tanguay RM. Structural and functional properties of proteins interacting with small heat shock proteins. Cell Stress Chaperones. 2020;25:629–37.

421. Morrow G, Heikkila JJ, Tanguay RM. Differences in the chaperone-like activities of the four main small heat shock proteins of Drosophila melanogaster. Cell Stress Chaperones. 2006;11:51–60.

422. Jagla T, Dubińska-Magiera M, Poovathumkadavil P, Daczewska M, Jagla K. Developmental expression and functions of the small heat shock proteins in Drosophila. Int J Mol Sci. 2018;19.

423. Corces V, Holmgren R, Freund R, Morimoto R, Meselson M. Four heat shock proteins of Drosophila melanogaster coded within a 12-kilobase region in chromosome subdivision 67B. Proc Natl Acad Sci U S A. 1980;77:5390–3.

424. Telonis-Scott M, Ali Z, Hangartner S, Sgrò CM. Temporal specific coevolution of Hsp70 and co-chaperone stv expression in Drosophila melanogaster under selection for heat tolerance. J Therm Biol. 2021;102:103110.

425. Huang J, Wang H. Hsp83/Hsp90 physically associates with insulin receptor to promote neural stem cell reactivation. Stem cell reports. 2018;11:883–96.

426. Cuenca JB, Galindo MI, Saura AO, Sorsa V, de Frutos R. Ultrastructure of regions containing homologous loci in polytene chromosomes of Drosophila melanogaster and Drosophila subobscura. Chromosoma. 1998;107:113–26.

427. Calabria G, Dolgova O, Rego C, Castañeda LE, Rezende EL, Balanyà J, et al. Hsp70 protein levels and thermotolerance in Drosophila subobscura: a reassessment of the thermal co-adaptation hypothesis. J Evol Biol. 2012;25:691–700.

428. Puig Giribets M, Santos M, García Guerreiro MP. Basal hsp70 expression levels do not explain adaptive variation of the warm- and cold-climate O(3 + 4 + 7) and O(ST) gene arrangements of Drosophila subobscura. BMC Evol Biol. 2020;20:17.

429. Molto MD, Martinez-Sebastian MJ, De Frutos R. Phylogenetic relationships between Drosophila subobscura, D. guanche and D. madeirensis based on Southern analysis of heat shock genes. Hereditas. 1994;120:217–23.

430. Arbona M, de Frutos R, Tanguay RM. Transcriptional and translational study of the Drosophila subobscura hsp83 gene in normal and heat-shock conditions. Genome. 1993;36:694–700.

431. Becerra M, Lombardía LJ, González-Siso MI, Rodríguez-Belmonte E, Hauser NC, Cerdán ME. Genome-wide analysis of the yeast transcriptome upon heat and cold shock. Comp Funct Genomics. 2003;4:366–75.

432. Sørensen JG, Nielsen MM, Kruhøffer M, Justesen J, Loeschcke V. Full genome gene expression analysis of the heat stress response in Drosophila melanogaster. Cell Stress Chaperones. 2005;10:312–28.

433. Rastrojo A, Corvo L, Lombraña R, Solana JC, Aguado B, Requena JM. Analysis by RNA-seq of transcriptomic changes elicited by heat shock in Leishmania major. Sci Rep. 2019;9:6919.

434. Kim W-J, Lee K, Lee D, Kim H-C, Nam B-H, Jung H, et al. Transcriptome profiling of olive flounder responses under acute and chronic heat stress. Genes Genomics. 2021;43:151–9.

435. Dou J, Cánovas A, Brito LF, Yu Y, Schenkel FS, Wang Y. Comprehensive RNA-Seq profiling reveals temporal and tissue-specific changes in gene expression in Sprague-Dawley rats as response to heat stress challenges. Front Genet. 2021;12:651979.

436. Srikanth K, Park J-E, Ji SY, Kim KH, Lee YK, Kumar H, et al. Genome-wide transcriptome and metabolome analyses provide novel insights and suggest a sex-specific response to heat stress in pigs. Genes. 2020;11.

437. GuhaThakurta D, Palomar L, Stormo GD, Tedesco P, Johnson TE, Walker DW, et al. Identification of a novel cis-regulatory element involved in the heat shock response in Caenorhabditis elegans using microarray gene expression and computational methods. Genome Res. 2002;12:701–12.

438. Klepsatel P, Gáliková M, Xu Y, Kühnlein RP. Thermal stress depletes energy reserves in Drosophila. Sci Rep. 2016;6:33667.

439. Vollmer JH, Sarup P, Kærsgaard CW, Dahlgaard J, Loeschcke V. Heat and cold-induced male sterility in Drosophila buzzatii: genetic variation among populations for the duration of sterility. Heredity. 2004;92:257–62.

440. Beck CD, Rankin CH. Heat shock disrupts long-term memory consolidation in Caenorhabditis elegans. Learn Mem. 1995;2:161–77.

441. Cantet JM, Yu Z, Ríus AG. Heat stress-mediated activation of immune-inflammatory pathways. Antibiot (Basel, Switzerland). 2021;10.

442. Fedyaeva A V, Stepanov A V, Lyubushkina I V, Pobezhimova TP, Rikhvanov EG. Heat shock induces production of reactive oxygen species and increases inner mitochondrial membrane potential in winter wheat cells. Biochemistry (Mosc). 2014;79:1202–10.

443. Yost HJ, Lindquist S. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell. 1986;45:185–93.

444. Liang Z, Anderson SN, Noshay JM, Crisp PA, Enders TA, Springer NM. Genetic and epigenetic variation in transposable element expression responses to abiotic stress in maize. Plant Physiol. 2021;186:420–33.

445. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O, et al. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell. 2010;22:3118–29.

446. Ogasawara H, Obata H, Hata Y, Takahashi S, Gomi K. Crawler, a novel Tc1/mariner-type transposable element in Aspergillus oryzae transposes under stress conditions. Fungal Genet Biol. 2009;46:441–9.

447. Ryan CP, Brownlie JC, Whyard S. Hsp90 and physiological stress are linked to autonomous transposon mobility and heritable genetic change in nematodes. Genome Biol Evol. 2016;8:3794–805.

448. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, et al. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell. 2008;29:499–509.

449. Baev V, Milev I, Naydenov M, Vachev T, Apostolova E, Mehterov N, et al. Insight into small RNA abundance and expression in high- and low-temperature stress response using deep sequencing in Arabidopsis. Plant Physiol Biochem PPB. 2014;84:105–14.

450. He J, Jiang Z, Gao L, You C, Ma X, Wang X, et al. Genome-wide transcript and small RNA profiling reveals transcriptomic responses to heat stress. Plant Physiol. 2019;181:609–29.

451. Graham AM, Barreto FS. Novel microRNAs are associated with population divergence in transcriptional response to thermal stress in an intertidal copepod. Mol Ecol. 2019;28:584–99.

452. Belicard T, Jareosettasin P, Sarkies P. The piRNA pathway responds to environmental signals to establish intergenerational adaptation to stress. BMC Biol. 2018;16:103.

453. Schreiner WP, Pagliuso DC, Garrigues JM, Chen JS, Aalto AP, Pasquinelli AE. Remodeling of the Caenorhabditis elegans non-coding RNA transcriptome by heat shock. Nucleic Acids Res. 2019;47:9829–41.

454. Sermek A, Feliciello I, Ugarković Đ. Distinct regulation of the expression of satellite DNAs in the beetle Tribolium castaneum. Int J Mol Sci. 2020;22.

455. López M-E, Roquis D, Becker C, Denoyes B, Bucher E. DNA methylation dynamics during stress response in woodland strawberry (Fragaria vesca). Hortic Res. 2022;9:uhac174.

456. Migicovsky Z, Yao Y, Kovalchuk I. Transgenerational phenotypic and epigenetic changes in response to heat stress in Arabidopsis thaliana. Plant Signal Behav. 2014;9:e27971.

457. Norouzitallab P, Baruah K, Vandegehuchte M, Van Stappen G, Catania F, Vanden Bussche J, et al. Environmental heat stress induces epigenetic transgenerational inheritance of robustness in parthenogenetic Artemia model. FASEB J Off Publ Fed Am Soc Exp Biol. 2014;28:3552–63.

458. Liu H, Able AJ, Able JA. Small RNA, Transcriptome and Degradome Analysis of the Transgenerational Heat Stress Response Network in Durum Wheat. Int J Mol Sci. 2021;22.

459. Ni JZ, Kalinava N, Chen E, Huang A, Trinh T, Gu SG. A transgenerational role of the germline nuclear RNAi pathway in repressing heat stress-induced transcriptional activation in C. elegans. Epigenetics Chromatin. 2016;9:3.

460. Junakovic N, Di Franco C, Barsanti P, Palumbo G. Transposition of copia-like nomadic elements can be induced by heat shock. J Mol Evol. 1986;24:89–93.

461. Zabanov SA, Vasil'eva LA, Ratner VA. Expression of the quantitative trait radius incompletus in Drosophila and localization of mobile elements MDG1 and copia. Genetika. 1990;26:1144–53.

462. Arnault C, Biémont C. Heat shocks do not mobilize mobile elements in genomes of Drosophila melanogaster inbred lines. J Mol Evol. 1989;28:388–90.

463. Arnault C, Loevenbruck C, Biémont C. Transposable element mobilization is not induced by heat shocks in Drosophila melanogaster. Naturwissenschaften. 1997;84:410–4.

464. Fast I, Hewel C, Wester L, Schumacher J, Gebert D, Zischler H, et al. Temperature-responsive miRNAs in Drosophila orchestrate adaptation to different ambient temperatures. RNA. 2017;23:1352–64.

465. Funikov SY, Ryazansky SS, Kanapin AA, Logacheva MD, Penin AA, Snezhkina A V, et al. Interplay between RNA interference and heat shock response systems in Drosophila melanogaster. Open Biol. 2016;6.

466. Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, et al. Chromatin-associated RNAi components contribute to transcriptional regulation in Drosophila. Nature. 2011;480:391–5.

467. Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, et al. Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature. 2011;480:391–5.

468. Funikov SY, Ryazansky SS, Zelentsova ES, Popenko VI, Leonova OG, Garbuz DG, et al. The peculiarities of piRNA expression upon heat shock exposure in Drosophila melanogaster. Mob Genet Elements. 2015;5:72–80.

469. Arrigo AP. Acetylation and methylation patterns of core histones are modified after heat or arsenite treatment of Drosophila tissue culture cells. Nucleic Acids Res. 1983;11:1389–404.

470. Seong K-H, Li D, Shimizu H, Nakamura R, Ishii S. Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell. 2011;145:1049–61.

471. Michalak P, Noor MAF. Genome-wide patterns of expression in Drosophila pure species and hybrid males. Mol Biol Evol. 2003;20:1070–6.

472. Moehring AJ, Teeter KC, Noor MAF. 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.

473. Vela D, García Guerreiro MP, Fontdevila A. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques. 2011;50:247–50.

474. 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.

475. Malone JH, Michalak P. Gene expression analysis of the ovary of hybrid females of Xenopus laevis and X. muelleri. BMC Evol Biol. 2008;8:82.

476. Mack KL, Nachman MW. Gene Regulation and Speciation. Trends Genet. 2017;33:68–80.

477. 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.

478. 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.

479. Schulze SR, Wallrath LL. Gene regulation by chromatin structure: Paradigms established in Drosophila melanogaster. Annu Rev Entomol. 2007;52:171–92.

480. Ebert A, Lein S, Schotta G, Reuter G. Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosom Res. 2006;14:377–92.

481. He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell. 2010;22:17–33.

482. Satyaki PRV, Cuykendall TN, Wei KHC, Brideau NJ, Kwak H, Aruna S, et al. The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet. 2014;10.

483. García Guerreiro MP. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in Drosophila. Mob Genet Elements. 2014;4:e34394.

484. Gómez GA, Hasson E. Transpecific polymorphisms in an inversion linked esterase locus in Drosophila buzzatii. Mol Biol Evol. 2003;20:410–23.

485. Naveira H, Pla C, Fontdevila A. The evolutionary history of Drosophila buzzatii XI. A new method for cytogenetic localization based on asynapsis of polytene chromosomes in interspecific hybrids of Drosophila. Genetica. 1986;71:199–212.

486. Wurmser F, Ogereau D, Mary-Huard T, Loriod B, Joly D, Montchamp-Moreau C. Population transcriptomics: Insights from Drosophila simulans, Drosophila sechellia and their hybrids. Genetica. 2011;139:465–77.

487. Jurka J, Kapitonov V V., Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7.

488. Boros IM. Histone modification in drosophila. Brief Funct Genomics. 2012;11:319–31.

489. 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.

490. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430:85–8.

491. Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, et al. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics. 2006;172:507–17.

492. 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.

493. 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.

494. Allan CW, Matzkin LM. Genomic analysis of the four ecologically distinct cactus host populations of Drosophila mojavensis. BMC Genomics. 2019;20:732.

495. Coyne JA, Orr HA, Otte D, Endler JA. Speciation and its Consequences. Otte D, Endler JA, Ed. 1989;:180–207.

496. Hollocher H, Wu CI. The genetics of reproductive isolation in the Drosophila simulans clade: X vs. autosomal effects and male vs. female effects. Genetics. 1996;143:1243–55.

497. 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.

498. Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philos Trans R Soc Lond B Biol Sci. 2014;369.

499. Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. Mol Biol Evol. 2014;31:1188–99.

500. Hénault M, Marsit S, Charron G, Landry CR. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. Elife. 2020;9:e60474.

501. Ortíz-Barrientos D, Counterman BA, Noor MAF. Gene expression divergence and the origin of hybrid dysfunctions. Genetica. 2007;129:71–81.

502. De Araujo PG, Rossi M, De Jesus EM, Saccaro NL, Kajihara D, Massa R, et al. Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J. 2005;44:707–17.

503. Zhu A, Greaves IK, Dennis ES, Peacock WJ. Genome-wide analyses of four major histone modifications in Arabidopsis hybrids at the germinating seed stage. BMC Genomics. 2017;18:1–9.

504. He G, Chen B, Wang XXX, Li X, Li J, He H, et al. Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. Genome Biol. 2013;14:1–15.

505. Brown EJ, Bachtrog D. The chromatin landscape of Drosophila: comparisons between species, sexes, and chromosomes. Genome Res. 2014;24:1125–37.

506. modENCODE Consortium, Roy S, Ernst J, Kharchenko P V, Kheradpour P, Negre N, et al. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science. 2011;330:1787–97.

507. Parey E, Crombach A, Alba M. Evolution of the Drosophila melanogaster chromatin landscape and its associated proteins. Genome Biol Evol. 2019;11:660–77.

508. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell. 2010;143:212–24.

509. Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Ekker H, et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. 2020;30:573–88.

510. Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife. 2016;5:1–30.

511. Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, Van Leeuwen F, et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 2004;18:1263–71.

512. Riddle NC, Minoda A, Kharchenko P V., Alekseyenko AA, Schwartz YB, Tolstorukov MY, et al. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res. 2011;21:147–63.

513. Saha P, Sowpati DT, Mishra RK. Epigenomic and genomic landscape of Drosophila melanogaster heterochromatic genes. Genomics. 2019;111:177–85.

514. Moghaddam AMB, Roudier F, Seifert M, Bérard C, Magniette M-LM, Ashtiyani RK, et al. Additive inheritance of histone modifications in Arabidopsis thaliana intra-specific hybrids. Plant J. 2011;67:691–700.

515. Dong X, Reimer J, Göbel U, Engelhorn J, He F, Schoof H, et al. Natural variation of H3K27me3 distribution between two Arabidopsis accessions and its association with flanking transposable elements. Genome Biol. 2012;13:R117.

516. Göbel U, Arce AL, He F, Rico A, Schmitz G, De Meaux J. Robustness of transposable element regulation but no genomic shock observed in interspecific Arabidopsis hybrids. Genome Biol Evol. 2018;10:1403–15.

517. Rebollo R, Horard B, Begeot F, Delattre M, Gilson E, Vieira C. A Snapshot of Histone Modifications within

Transposable Elements in Drosophila Wild Type Strains. PLoS One. 2012;7:1–7.

518. Guio L, Vieira C, González J. Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster. Sci Rep. 2018;8:1–10.

519. Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin S V. Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics. 2009;183:547–61, 1SI-21SI.

520. Lemos B, Araripe LO, Fontanillas P, Hartl DL. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc Natl Acad Sci U S A. 2008;105:14471–6.

521. Goldsborough AS, Kornberg TB. Reduction of transcription by homologue asynapsis in Drosophila imaginal discs. Nature. 1996;381 June:807–10.

522. Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, et al. Mechanistic basis of infertility of mouse intersubspecific hybrids. Proc Natl Acad Sci U S A. 2013;110.

523. Cloutier JM, Mahadevaiah SK, ElInati E, Nussenzweig A, Tóth A, Turner JMA. Histone H2AFX links meiotic chromosome asynapsis to prophase I oocyte loss in mammals. PLoS Genet. 2015;11:1–18.

524. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

525. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

526. 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.

527. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160–5.

528. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

529. Center for Research in Agricultural Genomics, National Center for Genomic Analysis, Genomics Bioinformatics and Evolution Research Group of the Universitat Autònoma de Barcelona, Barcelona Supercomputing Center. Drosophila buzzatii Genome Project.

530. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

531. Tamura K, Stecher G, Kumar S. MEGA11 : Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38:3022–7.

532. Jukes T, Cantor C. Evolution of protein molecules. New York; 1969.

533. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 2013;10:71–3.

534. Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Res. 2017;45:e17.

535. R Core Team. R: A language and environment for statistical computing. 2020.

536. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016.

537. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

538. 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.

539. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35:2084–92.

540. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. 2020.

541. Davidson NM, Hawkins ADK, Oshlack A. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biol. 2017;18:148.

542. Van der Auwera G, O'Connor B. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media; 2020.

543. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.

544. Lindenbaum P. JVarkit: java-based utilities for Bioinformatics. figshare. 2015. https://doi.org/https://doi.org/10.6084/m9.figshare.1425030.

545. Institute B. Picard toolkit. Broad Institute, GitHub repository. 2019;:https://broadinstitute.github.io/picard/.

546. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27:1691–2.

547. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12:656–64.

548. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

549. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–9.

550. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

551. Modolo L, Lerat E. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. BMC Bioinformatics. 2015;16:1–8.

552. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10.

553. Pauli D, Arrigo AP, Tissières A. Heat shock response in Drosophila. Experientia. 1992;48:623–9.

554. Tanguay RM. Genetic regulation during heat shock and function of heat-shock proteins: a review. Can J Biochem cell Biol. 1983;61:387–94.

555. Sørensen JG, Kristensen TN, Loeschcke V. The evolutionary and ecological role of heat shock proteins. Ecol

Lett. 2003;6:1025–37.

556. Tower J. Heat shock proteins and Drosophila aging. Exp Gerontol. 2011;46:355–62.

557. Parsell DA, Lindquist S. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu Rev Genet. 1993;27:437–96.

558. Vasilyeva LA, Bubenshchikova E V, Ratner VA. Heavy heat shock induced retrotransposon transposition in Drosophila. Genet Res. 1999;74:111–9.

559. Kofler R, Senti K-A, Nolte V, Tobler R, Schlötterer C. Molecular dissection of a natural transposable element invasion. Genome Res. 2018;28:824–35.

560. Strand DJ, McDonald JF. Copia is transcriptionally responsive to environmental stress. Nucleic Acids Res. 1985;13:4401–10.

561. Bubenshchikova E V, Antonenko O V, Vasil'eva LA, Ratner VA. Induction of MGE 412 transposition individually by heat and cold shock in spermatogenesis in Drosophila males. Genetika. 2002;38:46–55.

562. Arnault C, Heizmann A, Loevenbruck C, Biémont C. Environmental stresses and mobilization of transposable elements in inbred lines of Drosophila melanogaster. Mutat Res. 1991;248:51–60.

563. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. An endogenous small interfering RNA pathway in Drosophila. Nature. 2008;453:798–802.

564. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature. 2008;453:793–7.

565. Luo S, Lu J. Silencing of transposable elements by piRNAs in Drosophila: an evolutionary perspective. Genomics Proteomics Bioinformatics. 2017;15:164–76.

566. Li C, Vagin V V, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137:509–21.

567. Specchia V, Piacentini L, Tritto P, Fanti L, D'Alessandro R, Palumbo G, et al. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. Nature. 2010;463:662–5.

568. Muñoz MM. The Bogert effect, a factor in evolution. Evolution (N Y). 2022;76:49–66.

569. Dworniczak B, Mirault ME. Structure and expression of a human gene coding for a 71 kd heat shock "cognate" protein. Nucleic Acids Res. 1987;15:5181–97.

570. Rosenkranz D, Zischler H. proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics. 2012;13:5.

571. Juneja P, Quinn A, Jiggins FM. Latitudinal clines in gene expression and cis-regulatory element variation in Drosophila melanogaster. BMC Genomics. 2016;17:1–11.

572. Wurmser F, Mary-Huard T, Daudin JJ, Joly D, Montchamp-Moreau C. Variation of gene expression associated with colonisation of an anthropized environment: Comparison between African and European populations of drosophila simulans. PLoS One. 2013;8.

573. Huylmans AK, Parsch J. Population- and sex-biased gene expression in the excretion organs of Drosophila melanogaster. G3 Genes, Genomes, Genet. 2014;4:2307–15.

574. Banho CA, Mérel V, Oliveira TYK, Carareto CMA, Vieira C. Comparative transcriptomics between Drosophila mojavensis and D. arizonae reveals transgressive gene expression and underexpression of spermatogenesis-related genes in hybrid testes. Sci Rep. 2021;11:1–15.

575. Catalán A, Hutter S, Parsch J. Population and sex differences in Drosophila melanogaster brain gene expression. BMC Genomics. 2012;13.

576. Loeschcke V, Bundgaard J, Barker JSF. Variation in body size and life history traits in Drosophila aldrichi and D. buzzatii from a latitudinal cline in eastern Australia. Heredity. 2000;85:423–33.

577. Khadem M, Rozas J, Segarra C, Brehm A, Aguadé M. Tracing the colonization of Madeira and the Canary Islands by Drosophila subobscura through the study of the rp49 gene region. J Evol Biol. 1998;11:439–52.

578. Mérel V, Gibert P, Buch I, Rodriguez Rada V, Estoup A, Gautier M, et al. The worldwide invasion of Drosophila suzukii Is accompanied by a Large increase of transposable element load and a small number of putatively adaptive insertions. Mol Biol Evol. 2021;38:4252–67.

579. García Guerreiro MP, Chávez-Sandoval BE, Balanyà J, Serra L, Fontdevila A. Distribution of the transposable elements bilbo and gypsy in original and colonizing populations of Drosophila subobscura. BMC Evol Biol. 2008;8.

580. Vieira C, Lepetit D, Dumont S, Biémont C. Wake up of transposable elements following Drosophila simulans worldwide colonization. Mol Biol Evol. 1999;16:1251–5.

581. Lebo MS, Sanders LE, Sun F, Arbeitman MN. Somatic, germline and sex hierarchy regulated gene expression during Drosophila metamorphosis. BMC Genomics. 2009;10:1–23.

582. Landis G, Shen J, Tower J. Gene expression changes in response to aging compared to heat stress, oxidative stress and ionizing radiation in Drosophila melanogaster. Aging (Albany NY). 2012;4:768–89.

583. Lecheta MC, Awde DN, O'Leary TS, Unfried LN, Jacobs NA, Whitlock MH, et al. Integrating GWAS and transcriptomics to identify the molecular underpinnings of thermal stress responses in Drosophila melanogaster. Front Genet. 2020;11 June:1–17.

584. Teves SS, Henikoff S. The heat shock response: A case study of chromatin dynamics in gene regulation. Biochem Cell Biol. 2013;91:42–8.

585. Gruntenko NE, Bownes M, Terashima J, Sukhanova MZ, Raushenbach IY. Heat stress affects oogenesis differently in wild-type Drosophila virilis and a mutant with altered juvenile hormone and 20-hydroxyecdysone levels. Insect Mol Biol. 2003;12:393–404.

586. Chakir M, Chafik A, Moreteau B, Gibert P, David JR. Male sterility thermal thresholds in Drosophila: D. simulans appears more cold-adapted than its sibling D. melanogaster. Genetica. 2002;114:195–205.

587. Teets NM, Hahn DA. Genetic variation in the shape of cold-survival curves in a single fly population suggests potential for selection from climate variability. J Evol Biol. 2018;31:543–55.

588. Jørgensen LB, Robertson RM, Overgaard J. Neural dysfunction correlates with heat coma and CTmax in Drosophila but does not set the boundaries for heat stress survival. J Exp Biol. 2020;223.

589. Rodgers CI, Armstrong GAB, Shoemaker KL, LaBrie JD, Moyes CD, Robertson RM. Stress preconditioning of spreading depression in the locust CNS. PLoS One. 2007;2.

590. Hsu SK, Jakšić AM, Nolte V, Lirakis M, Kofler R, Barghi N, et al. Rapid sex-specific adaptation to high

temperature in drosophila. Elife. 2020;9:1–16.

591. Bedulina D, Drozdova P, Gurkov A, von Bergen M, Stadler PF, Luckenbach T, et al. Proteomics reveals sex-specific heat shock response of Baikal amphipod Eulimnogammarus cyaneus. Sci Total Environ. 2021;763:143008.

592. Michaud S, Marin R, Tanguay RM. Regulation of heat shock gene induction and expression during Drosophila development. Cell Mol Life Sci. 1997;53:104–13.

593. Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, et al. Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc Natl Acad Sci. 1995;92:3804–8.

594. McCullers TJ, Steiniger M. Transposable elements in Drosophila. Mob Genet Elements. 2017;7:1–18.

595. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of Drosophila melanogaster. Science. 2000;287:2185–95.

596. Ahmed W, Xia Y, Li R, Zhang H, Siddique KHMM, Guo P. Identification and analysis of small interfering RNAs associated with heat stress in flowering Chinese cabbage using High-Throughput Sequencing. Front Genet. 2021;12 November:1–9.

597. Senti K-A, Jurczak D, Sachidanandam R, Brennecke J. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev. 2015;29:1747–62.

598. Kelleher ES, Barbash DA. Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. Mol Biol Evol. 2013;30:1816–29.

599. Emilie Q, Amit A, Kai T. The piRNA pathway is developmentally regulated during spermatogenesis in Drosophila. RNA. 2016;22:1044–54.

600. Bodelón A, Fablet M, Veber P, Vieira C, García Guerreiro MP. High stability of the epigenome in Drosophila interspecific hybrids. Genome Biol Evol. 2022;14:1–20.

601. Arrigo A-PP. Acetylation and methylation patterns of core histones are modified after heat or arsenite treatment of Drosophila tissue culture cells. Nucleic Acids Res. 1983;12:8235–51.

602. Walser JC, Chen B, Feder ME. Heat-shock promoters: Targets for evolution by P transposable elements in Drosophila. PLoS Genet. 2006;2:1541–55.

603. Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. Mol Biol Evol. 2003;20:135–44.

604. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117:9451–7.

605. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275.

606. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. Bioinformatics. 2008;24:1757–64.

607. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. :http://www.repeatmasker.org.

608. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.

609. Bailly-Bechet M, Haudry A, Lerat E. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mob DNA. 2014;5:13.

610. Zytnicki M, Gaspin C. srnaMapper: an optimal mapping tool for sRNA-Seq reads. bioRxiv. 2021;:2021.01.12.426326.

611. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, et al. BEDOPS: high-performance genomic feature operations. Bioinformatics. 2012;28:1919–20.

612. 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.

613. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics. 2022;38:2943–5.

614. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.

615. Goubert C. TE-Aid.

616. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34 suppl_1:D140–4.

617. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma. 2010;Chapter 11:Unit 11.7.

618. Antoniewski C. Computing siRNA and piRNA overlap signatures. Methods Mol Biol. 2014;1173:135–46.

619. Rani K, Dutt S, Rana R. Brief review on alkaline phosphatase-an overview. Int J Microbiol Bioinforma. 2012;2:1–4.

620. Diaz GJ. Toxicosis by Plant Alkaloids in Humans and Animals in Colombia. Toxins (Basel). 2015;7:5408–16.

621. Nanni A V, Martinez N, Graze R, Morse A, Newman JRB, Jain V, et al. Sex-biased expression is associated with chromatin state in D. melanogaster and D. simulans. bioRxiv : the preprint server for biology. 2023.

622. Wittkopp PJ. Variable gene expression in eukaryotes: a network perspective. J Exp Biol. 2007;210:1567–75.

623. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40:346–50.

624. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. Science. 2007;317:118–21.

625. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. Nat Genet. 2005;37:544–8.

626. Piccinali R, Aguadé M, Hasson E. Comparative molecular population genetics of the Xdh locus in the cactophilic sibling species Drosophila buzzatii and D. koepferae. Mol Biol Evol. 2004;21:141–52.

627. Benowitz KM, Coleman JM, Allan CW, Matzkin LM. Contributions of cis- and trans-regulatory evolution to transcriptomic divergence across populations in the Drosophila mojavensis larval brain. Genome Biol Evol. 2020;12:1407–18.

628. Ranz JM, Go AC, González PM, Clifton BD, Gomes S, Jaberyzadeh A, et al. Gene expression differentiation in the reproductive tissues of Drosophila willistoni subspecies and their hybrids. Mol Ecol. 2023;n/a n/a.

629. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in Drosophila. Genome Res. 2014;24:797–808.

630. Peng P-C, Khoueiry P, Girardot C, Reddington JP, Garfield DA, Furlong EEM, et al. The Role of Chromatin Accessibility in cis-Regulatory Evolution. Genome Biol Evol. 2019;11:1813–28.

631. Huang Y, Shukla H, Lee YCG. Species-specific chromatin landscape determines how transposable elements shape genome evolution. Elife. 2022;11:e81567.

632. Hauenschild A, Ringrose L, Altmutter C, Paro R, Rehmsmeier M. Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. PLoS Biol. 2008;6:e261.

633. Omholt SW, Plahte E, Oyehaug L, Xiang K. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. Genetics. 2000;155:969–80.

634. Wittkopp PJ, Haerum BK, Clark AG. Parent-of-origin effects on mRNA expression in Drosophila melanogaster not caused by genomic imprinting. Genetics. 2006;173:1817–21.

635. Coolon JD, Stevenson KR, McManus CJ, Graveley BR, Wittkopp PJ. Genomic imprinting absent in Drosophila melanogaster adult females. Cell Rep. 2012;2:69–75.

636. McGirr JA, Martin CH. Ecological divergence in sympatry causes gene misexpression in hybrids. Mol Ecol. 2020;29:2707–21.

637. Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, et al. FlyBase: a guided tour of highlighted features. Genetics. 2022;220:iyac035.

638. Festucci-Buselli RA, Carvalho-Dias AS, de Oliveira-Andrade M, Caixeta-Nunes C, Li H-M, Stuart JJ, et al. Expression of Cyp6g1 and Cyp12d1 in DDT resistant and susceptible strains of Drosophila melanogaster. Insect Mol Biol. 2005;14:69–77.

639. Le Goff G, Hilliou F, Siegfried BD, Boundy S, Wajnberg E, Sofer L, et al. Xenobiotic response in Drosophila melanogaster: sex dependence of P450 and GST gene induction. Insect Biochem Mol Biol. 2006;36:674–82.

640. Freeman A, Pranski E, Miller RD, Radmard S, Bernhard D, Jinnah HA, et al. Sleep fragmentation and motor restlessness in a Drosophila model of Restless Legs Syndrome. Curr Biol. 2012;22:1142–8.

641. Mazzotta GM, Damulewicz M, Cusumano P. Better sleep at night: how light influences sleep in Drosophila. Front Physiol. 2020;11:997.

642. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol. 2008;6:e251.

643. Czech B, Preall JB, McGinn J, Hannon GJ. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell. 2013;50:749–61.

644. Jedlicka P, Mortin MA, Wu C. Multiple functions of Drosophila heat shock transcription factor in vivo. EMBO J. 1997;16:2452–62.

645. Chang TH, Mattei E, Gainetdinov I, Colpan C, Weng Z, Zamore PD. Maelstrom Represses Canonical

Polymerase II Transcription within Bi-directional piRNA Clusters in Drosophila melanogaster. Mol Cell. 2019;73:291-303.e6.

646. Izumi N, Kawaoka S, Yasuhara S, Suzuki Y, Sugano S, Katsuma S, et al. Corrigendum: Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. RNA (New York, N.Y.). 2015;21:1217.

647. Marin I, Labrador M, Fontdevila A. The evolutionary history of Drosophila buzzatii. XXIII. High content of nonsatellite repetitive DNA in D. buzzatii and in its sibling D. koepferae. Genome. 1992;35:967–74.

648. 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.

649. Jaquet V, Wallerich S, Voegeli S, Túrós D, Viloria EC, Becskei A. Determinants of the temperature adaptation of mRNA degradation. Nucleic Acids Res. 2022;50:1092–110.

650. Chen S, Li X. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. BMC Evol Biol. 2007;7:46.

651. 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;47:6842–57.

652. Schneemann H, Munzur AD, Thompson KA, Welch JJ. The diverse effects of phenotypic dominance on hybrid fitness. Evolution (N Y). 2022;76:2846–63.

653. Banho CA, Oliveira DS, Haudry A, Fablet M, Vieira C, Carareto CMA. Transposable element expression and regulation profile in gonads of interspecific hybrids of Drosophila arizonae and Drosophila mojavensis wrigleyi. Cells. 2021;10.

654. Mombach DM, da Fontoura Gomes TMF, Loreto ELS. Stress does not induce a general transcription of transposable elements in Drosophila. Mol Biol Rep. 2022;49:9033–40.

655. Serrato-Capuchina A, Matute DR. The role of transposable elements in speciation. Genes. 2018;9.

656. Anderson DW, Evans BJ. Regulatory evolution of a duplicated heterodimer across species and tissues of allopolyploid clawed Frogs (Xenopus). J Mol Evol. 2009;68:236–47.

7. Acknowledgments

Esta tesis muestra todo el trabajo llevado a cabo durante casi 5 años, cuando empecé a trabajar en el proyecto, y todos los conocimientos y habilidades que tengo la suerte de haber aprendidos durante el camino. Este viaje ha incluido muchas experiencias, con diferentes altibajos, pero en general muy positivas, que me acompañarán durante toda la vida y que me han enriquecido en muchos sentidos. No obstante, este trabajo no habría sido posible sin la participación, la ayuda, el apoyo y el soporte de muchas personas, quienes me han permitido aprender y desarrollarme a nivel laboral, pero también personal, así como me han permitido estar siempre acompañada.

Primero de todo, me gustaría agradecer a mi directora y tutora de la tesis, la Dra. Maria del Pilar Garcia Guerreiro, quien me dio la oportunidad de poder trabajar en este proyecto. Ha sido una gran suerte poder trabajar y aprender tanto de ella como he aprendido durante estos años. Además, le agradezco toda la confianza depositada en mí, así como el esfuerzo, la entrega, el tiempo, el apoyo y la paciencia mostradas, en los buenos y en los malos momentos.

Quiero también dar un enorme agradecimiento a todo el Grupo de Genómica, Bioinformática y Evolución, en el que he tenido la gran suerte de trabajar. Especialmente, quiero agradecer al Dr. Santos, Dr. Fontdevila y Dr. Barbadilla, por su ayuda e interés en mi trabajo. Además, quiero agradecer a todas las personas que han pasado por el grupo, y que han contribuido a este trabajo de una marea u otra. Incluyendo la gente de prácticas, como Sergio, Ariadna y Aina, así como los técnicos Marta y Camila, y otros doctorandos o post-docs, como Hayley y Marta.

Me gustaría agradecer a todo el Departamento de Genética y Microbiología, y especialmente al Departamento de Genética, por toda la ayuda proporcionada para tirar este trabajo adelante. Haciendo algunas menciones de personas, grupos o áreas que han hecho posible este trabajo, quiero agradecer el imprescindible trabajo de los técnicos del departamento, incluyendo Montse, Marta, Miriam, Gloria, y especialmente a Raquel, quien me ha acompañado durante estos 5 años y que además de ser una compañera, también es una amiga. Me gustaría incluir un agradecimiento a toda la administración del Departamento, y especialmente a Elena, Maite, Maria Josep y Dori, su eficiencia me ha permitido avanzar más rápido. Quiero también agradecer a todo el Grupo de Mutagénesis por incluirme en muchas de las experiencias del grupo, especialmente a Alba, y a Gloria por su ayuda.

El día a día del laboratorio no habría sido lo mismo sin los compañeros del Grupo de Mutagénesis, quienes me han acompañado durante estos 5 años. En la primera etapa, quiero agradecer especialmente a Irene y Pepa por las charlas, las comidas y las quedadas, incluyendo los preciosos cuadros que hemos tenido la suerte de pintar durante el doctorado. Durante la segunda etapa, quiero también agradecer a mis compañeros, Aliro, Lourdes, Ali, Gooya, Joan, y

especialmente Javi, por los juegos, las palomitas, la compañía y las visitas a mi laboratorio. También, a los post-docs del grupo, incluyendo a Laura y Alba.

I would also like to thank the people of the laboratory *Laboratoire De Biométrie Et Biologie Évolutive* (LBBE) of Lyon, who included me in their group from the beginning. I would like to thank Cristina to give me the opportunity to have this experience. I also really appreciate all the help and knowledge that her and Marie gave me, not only during my internship, but also during all the years of my PhD. Thank you very much for your contribution in making every paper better than what it was in the beginning. I would also like to thank all the support that the group gave me, including my friends, who made my internship one of the best experiences of the PhD. Worth to mention Diego, Miriam, Camille, Marius, Alicia, Valentina, William, and specially Daniel, who invest a lot of time teaching me.

Además de todas las personas incluidas en el sitio de trabajo, también me gustaría agradecer a las personas que me han apoyado en todo momento fuera de él. Quiero agradecer a mis amigos por el apoyo recibido durante toda la etapa del doctorado. Especialmente, a mis amigos de la carrera de Genética (José, Llorenç, Marc, Miguel y Alejandro) y del Máster de Bioinformática (Maria José, Alex y Joan), con los que he compartido, a la vez, este viaje y he podido disfrutar de muchas celebraciones, fiestas, juegos, viajes y charlas durante el camino. También a mis amigos de AIESEC (Miquel, Laura y Marc) por los *Skypes* y las visitas durante estos años. Además, me gustaría mencionar a Dani, a quien agradezco que siempre ha mostrado un especial interés en todo mi trabajo y ha sido un apoyo durante todo este proceso. Finalmente, me gustaría agradecer a mis mejores amigos Alba y Max, quienes por muy lejos u ocupados que estén, siempre han tenido un hueco para mi durante estos años para apoyarme, escucharme y celebrar mis avances.

Por último, me gustaría agradecerle a toda mi familia el soporte que me han dado todos estos años. Con especial mención a mi abuela, quien siempre ha celebrado todos mis logros en el camino, y mis tías Marista y Miriam, quienes han estado pendientes de mi durante todo el viaje.

Quiero también agradecer a mi familia más cercana. Primero, me gustaría agradecerle a mi hermana Carmen y a su marido Quim, quienes vivieron conmigo durante parte de estos años, la paciencia y el apoyo. En segundo lugar, quiero agradecerles a mis padres, por la confianza, los consejos y el apoyo en los momentos más duros, que han hecho que este viaje sea más fácil. Especialmente, quiero agradecerle a mi madre todas las horas invertidas en mi durante estos años, incluyendo las llamadas, los "*tuppers*", las visitas y especialmente el soporte, que me han ayudado a hacer más llevaderos los momentos más sufridos. Gracias por ser una persona tan excepcional, y por tu amor y apoyo incondicional.

Finally, I would like to acknowledge to Lars, not only for his contribution to this thesis, but also for his enormous support during these almost 5 years. There are simply not words to express my gratitude. Thank you very much for being always by my side, this thesis could definitely not be possible without your help, unconditional support and love. Always.

A todos y a cada uno de ellos, y los que me faltan por nombrar,

Gracias.

8. Annexes

8.1 Supplementary data of "High Stability of the Epigenome in *Drosophila* Interspecific Hybrids"

8.1.1 Supplementary file 1

Table S1: Differences in expression between the parental species D. buzzatii vs D. koepferae

Percentage of differentially expressed genes and TE families between the parental species.

	Total	More expressed in *D. buzzatii* than in *D. koepferae*	More expressed in *D. koepferae* than in *D. buzzatii*	Total of differentially expressed	% of differentially expressed
Genes	**13621**	384	422	806	**5.92**
TE families	**658**	101	94	195	**29.64**

Table S2. Two Proportion Z-Test

Comparison of the number of differentially expressed genes and TE families in hybrids vs each parental species.

	Hybrids in comparison to	Differentially expressed	Total	Proportions	Total proportion	Z-score	Associated p-value:
Genes	*D. buzzatii*	623	13621	0.046	0.043	2.364	**1.81E-02 ***
	D. koepferae	544	13621	0.040			
TE families	*D. buzzatii*	151	658	0.229	0.236	0.520	6.03E-01
	D. koepferae	159	658	0.242			

Significance: *$p < 0.05$

Table S3. ChiSquare

Comparison of the number of differentially expressed (DE) genes in each chromosome in hybrids vs the parental species, and TE families per order.

		Chromosome	Observed DE genes	Expected frequency	Expected DE genes	χ^2	Associated p-value	Adjusted p-value (FDR)
Genes	*D. buzzatii*	Chr 2	120	20.690	128.899	0.614	0.000	**3.80E-05 *****
		Chr 3	113	17.190	107.094	0.326		
		Chr 4	100	18.270	113.822	1.679		
		Chr 5	108	18.600	115.878	0.536		
		Chr 6	12	0.540	3.364	22.168		
		Chr X	102	16.340	101.798	0.000		
		Unknown	68	8.380	52.207	4.777		
		Total	623	-	623.062	**30.100**		
	D. koepferae	Chr 2	100	20.690	112.554	1.400	0.000	**2.40E-05 *****
		Chr 3	105	17.190	93.514	1.411		
		Chr 4	90	18.270	99.389	0.887		
		Chr 5	97	18.600	101.184	0.173		
		Chr 6	12	0.540	2.938	27.957		
		Chr X	88	16.340	88.890	0.009		
		Unknown	52	8.380	45.587	0.902		
		Total	544	-	544.054	**32.739**		
		Order	**Observed DE TE families**	**Expected frequency**	**Expected DE TE families**	χ^2	Associated p-value	Adjusted p-value (FDR)
TE families	*D. buzzatii*	DNA	39	24.320	36.723	0.141	0.165	1.83E-01
		LINE	19	18.690	28.222	3.013		
		LTR	76	46.350	69.989	0.516		
		RC	9	3.650	5.512	2.208		
		unknown	8	6.990	10.555	0.618		
		Total	151	100.000	151.000	**6.497**		
	D. koepferae	DNA	37	24.320	38.669	0.072	0.183	1.83E-01
		LINE	23	18.690	29.717	1.518		
		LTR	81	46.350	73.697	0.724		
		RC	10	3.650	5.804	3.034		
		unknown	8	6.990	11.114	0.873		
		Total	159	100.000	159.000	**6.221**		

* DE: Differentially expressed

Significance: ***$p < 0.001$

		Chromosome (all chromosomes excluding the Chr 6)	Observed DE genes	Expected frequency	Expected DE genes	χ^2	Associated p-value	Adjusted p-value (FDR)
Genes	*D. buzzatii*	Chr 2	120	20.690	128.899	0.614	0.160	3.20E-01
		Chr 3	113	17.190	107.094	0.326		
		Chr 4	100	18.270	113.822	1.679		
		Chr 5	108	18.600	115.878	0.536		
		Chr X	102	16.340	101.798	0.000		
		Unknown	68	8.380	52.207	4.777		
		Total	611	-	619.698	**7.932**		
	D. koepferae	Chr 2	100	20.690	112.554	1.400	0.443	4.43E-01
		Chr 3	105	17.190	93.514	1.411		
		Chr 4	90	18.270	99.389	0.887		
		Chr 5	97	18.600	101.184	0.173		
		Chr X	88	16.340	88.890	0.009		
		Unknown	52	8.380	45.587	0.902		
		Total	532	-	541.117	**4.782**		

Table S4: Linear Model

Contribution of the epigenome to the gene and TE expression, using a Linear Model with the formula: RNA ~ K4 + K9 + K27 + Input, for each species.

		General statistics			Effect of each LM component		
	Species	**Adjusted r-squared**	**p-value**		**Components of LM**	**Size effect**	**p-value**
Genes	*D. buzzatii*	0.624	< 2.2E-16 ***		Intercept	-1.773	< 2E-16 ***
					H3K4me3	1.649	< 2E-16 ***
					H3K9me3	0.496	**1.71E-15 ***
					H3K27me3	-2.466	< 2E-16 ***
					Input	1.209	< 2E-16 ***
	D. koepferae	0.623	< 2.2E-16 ***		Intercept	-1.816	< 2E-16 ***
					H3K4me3	1.595	< 2E-16 ***
					H3K9me3	0.346	**4.60E-09 ***
					H3K27me3	-2.241	< 2E-16 ***
					Input	1.186	< 2E-16 ***
	Hybrids	0.628	< 2.2E-16 ***		Intercept	-1.809	< 2E-16 ***
					H3K4me3	1.746	< 2E-16 ***
					H3K9me3	0.467	**2.29E-14 ***
					H3K27me3	-2.556	< 2E-16 ***
					Input	1.224	< 2E-16 ***
TE families	*D. buzzatii*	0.645	< 2.2E-16 ***		Intercept	-0.255	0.118
					H3K4me3	1.578	< 2E-16 ***
					H3K9me3	-1.045	**9.63E-15 ***
					H3K27me3	0.384	**0.031 *
					Input	-0.211	0.358
	D. koepferae	0.752	< 2.2E-16 ***		Intercept	-0.310	**0.014 *
					H3K4me3	0.961	**2.04E-13 ***
					H3K9me3	-1.149	< 2E-16 ***
					H3K27me3	-0.326	**0.029 **
					Input	1.307	< 2E-16 ***
	Hybrids	0.610	< 2.2E-16 ***		Intercept	-0.247	0.133
					H3K4me3	1.530	< 2E-16 ***
					H3K9me3	-1.156	**5.36E-16 ***
					H3K27me3	0.715	**2.27E-04***
					Input	-0.412	0.089

* Intercept: expected mean value of Y when all X=0

Significance: *p < 0.05; **p < 0.01; ***p < 0.001

Table S5. Two Proportion Z-Test

Comparison of the number of differentially enriched genes and TE families in hybrids *vs* each parental species.

	Chromatin mark	**Hybrids in comparison to**	**Differentially enriched**	**Total**	**Proportions**	**Total proportion**	**Z-score**	**Associated p-value**	**Adjusted p-value (FDR)**
Genes	H3K4me3	*D. buzzatii*	264	13621	0.019	0.024	4.843	< 0.00001	**3.00E-05 ***
		D. koepferae	386	13621	0.028				
	H3K9me3	*D. buzzatii*	191	13621	0.014	0.015	0.808	0.419	0.419
		D. koepferae	207	13621	0.015				
	H3K27me3	*D. buzzatii*	244	13621	0.018	0.021	3.589	0.000	**4.98E-04 ***
		D. koepferae	329	13621	0.024				
TE families	H3K4me3	*D. buzzatii*	26	658	0.040	0.037	0.437	0.662	0.817
		D. koepferae	23	658	0.035				
	H3K9me3	*D. buzzatii*	10	658	0.015	0.014	0.231	0.817	0.817
		D. koepferae	9	658	0.014				
	H3K27me3	*D. buzzatii*	61	658	0.093	0.079	1.839	0.066	0.198
		D. koepferae	43	658	0.065				

Significance: ***p < 0.001

Table S6. Differences in enrichment between the parental species *D. buzzatii vs D. koepferae*

Percentage of differentially enriched genes and TE families between the parental species.

	Total	Chromatin mark	More enriched in *D. buzzatii* than in *D. koepferae*	More enriched in *D. koepferae* than in *D. buzzatii*	Total of differentially enriched	% of differentially enriched
Genes	13621	H3K4me3	222	118	340	**2.50**
		H3K9me3	120	108	228	**1.67**
		H3K27me3	214	154	368	**2.70**
TE families	658	H3K4me3	19	17	36	**5.47**
		H3K9me3	8	5	13	**1.98**
		H3K27me3	36	28	64	**9.73**

Table S7. ChiSquare

Comparison of the number of differentially enriched genes in each chromosome in hybrids vs the parental species, and TE families per order.

	Chromatin mark	Hybrids in comparison to	Chromosome	Observed differentially enriched genes	Expected frequency	Expected differentially enriched genes	χ^2	Associated p-value	Adjusted p-value (FDR)
Genes	H3K4me3	*D. buzzatii*	Chr 2	58	20.690	54.622	0.209	0.085	0.448
			Chr 3	54	17.190	45.382	1.637		
			Chr 4	43	18.270	48.233	0.568		
			Chr 5	35	18.600	49.104	4.051		
			Chr 6	1	0.540	1.426	0.127		
			Chr X	41	16.340	43.138	0.106		
			Unknown	32	8.380	22.123	4.409		
			Total	264	-	264.026	**11.107**		
		D. koepferae	Chr 2	65	20.690	79.863	2.766	0.358	0.448
			Chr 3	79	17.190	66.353	2.410		
			Chr 4	65	18.270	70.522	0.432		
			Chr 5	72	18.600	71.796	0.001		
			Chr 6	2	0.540	2.084	0.003		
			Chr X	71	16.340	63.072	0.996		
			Unknown	32	8.380	32.347	0.004		
			Total	386	-	386.039	**6.613**		
	H3K9me3	*D. buzzatii*	Chr 2	39	20.690	39.518	0.007	0.305	0.448
			Chr 3	37	17.190	32.833	0.529		
			Chr 4	35	18.270	34.896	0.000		
			Chr 5	26	18.600	35.526	2.554		
			Chr 6	0	0.540	1.031	1.031		
			Chr X	31	16.340	31.209	0.001		
			Unknown	23	8.380	16.006	3.056		
			Total	191	-	191.019	**7.179**		
		D. koepferae	Chr 2	41	20.690	42.828	0.078	0.374	0.448
			Chr 3	33	17.190	35.583	0.188		
			Chr 4	33	18.270	37.819	0.614		
			Chr 5	33	18.600	38.502	0.786		
			Chr 6	2	0.540	1.118	0.696		
			Chr X	45	16.340	33.824	3.693		
			Unknown	20	8.380	17.347	0.406		
			Total	207	-	207.021	**6.461**		

				Observed	Expected frequency	Expected	χ²	Associated p-value	Adjusted p-value (FDR)
Genes	H3K27me3	*D. buzzatii*	Chr 2	45	20.690	50.484	0.596	0.566	0.566
			Chr 3	48	17.190	41.944	0.875		
			Chr 4	36	18.270	44.579	1.651		
			Chr 5	46	18.600	45.384	0.008		
			Chr 6	1	0.540	1.318	0.077		
			Chr X	42	16.340	39.870	0.114		
			Unknown	26	8.380	20.447	1.508		
			Total	244	-	244.024	**4.828**		
		D. koepferae	Chr 2	74	20.690	68.070	0.517	0.179	0.448
			Chr 3	59	17.190	56.555	0.106		
			Chr 4	50	18.270	60.108	1.700		
			Chr 5	52	18.600	61.194	1.381		
			Chr 6	0	0.540	1.777	1.777		
			Chr X	57	16.340	53.759	0.195		
			Unknown	37	8.380	27.570	3.225		
			Total	329	-	329.033	**8.901**		

	Chromatin mark	Hybrids in comparison to	Order	Observed differentially enirched TE families	Expected frequency	Expected differentially enirched TE families	χ²	Associated p-value	Adjusted p-value (FDR)
TE families	H3K4me3	*D. buzzatii*	DNA	8	24.316	6.322	0.445	0.594	0.949
			LINE	3	18.693	4.860	0.712		
			LTR	14	46.353	12.052	0.315		
			RC	0	3.647	0.948	0.948		
			unknown	1	6.991	1.818	0.368		
			Total	26	-	26.000	**2.788**		
		D. koepferae	DNA	4	24.316	5.593	0.454	0.951	0.951
			LINE	5	18.693	4.299	0.114		
			LTR	11	46.353	10.661	0.011		
			RC	1	3.647	0.839	0.031		
			unknown	2	6.991	1.608	0.096		
			Total	23	-	23.000	**0.705**		
	H3K9me3	*D. buzzatii*	DNA	2	24.316	2.432	0.077	0.763	0.949
			LINE	3	18.693	1.869	0.684		
			LTR	5	46.353	4.635	0.029		
			RC	0	3.647	0.365	0.365		
			unknown	0	6.991	0.699	0.699		
			Total	10	-	10.000	**1.853**		
		D. koepferae	DNA	3	24.316	2.188	0.301	0.791	0.949
			LINE	1	18.693	1.682	0.277		
			LTR	5	46.353	4.172	0.164		
			RC	0	3.647	0.328	0.328		
			unknown	0	6.991	0.629	0.629		
			Total	9	-	9.000	**1.700**		
	H3K27me3	*D. buzzatii*	DNA	17	24.316	14.833	0.317	0.093	0.279
			LINE	5	18.693	11.403	3.595		
			LTR	30	46.353	28.275	0.105		
			RC	1	3.647	2.225	0.674		
			unknown	8	6.991	4.264	3.272		
			Total	61	-	61.000	**7.964**		
		D. koepferae	DNA	8	24.316	10.456	0.577	0.010	0.063
			LINE	8	18.693	8.038	0.000		
			LTR	17	46.353	19.932	0.431		
			RC	1	3.647	1.568	0.206		
			unknown	9	6.991	3.006	11.951		
			Total	43	-	43.000	**13.166**		

Table S8: Fisher's exact test

Association between gene and TE family expression changes in hybrids in comparison to the parental species and the corresponding changes in chromatin marks.

	Chromatin mark	Hybrids in comparison to	Enrichment	More enriched	Less enriched	Total	p-value	Adjusted p-value (FDR)
Genes	H3K4me3	*D. buzzatii*	**Overexpressed**	2547	2849	**5396**	0.000	**5.35E-04 *****
			Underexpressed	2517	3248	**5765**		
			Total	**5064**	**6097**	**11161**		
		D. koepferae	**Overexpressed**	2324	2974	**5298**	0.000	**4.26E-14 *****
			Underexpressed	3004	2859	**5863**		
			Total	**5328**	**5833**	**11161**		
	H3K9me3	*D. buzzatii*	**Overexpressed**	2715	2678	**5393**	0.636	7.63E-01
			Underexpressed	2873	2887	**5760**		
			Total	**5588**	**5565**	**11153**		
		D. koepferae	**Overexpressed**	2621	2671	**5292**	0.970	9.70E-01
			Underexpressed	2900	2961	**5861**		
			Total	**5521**	**5632**	**11153**		
	H3K27me3	*D. buzzatii*	**Overexpressed**	2721	2674	**5395**	0.596	7.63E-01
			Underexpressed	2878	2886	**5764**		
			Total	**5599**	**5560**	**11159**		
		D. koepferae	**Overexpressed**	2462	2837	**5299**	0.077	1.55E-01
			Underexpressed	2821	3039	**5860**		
			Total	**5283**	**5876**	**11159**		
TE families	H3K4me3	*D. buzzatii*	**Overexpressed**	96	62	**158**	0.000	**8.02E-05 *****
			Underexpressed	129	192	**321**		
			Total	**225**	**254**	**479**		
		D. koepferae	**Overexpressed**	155	25	**180**	0.790	7.90E-01
			Underexpressed	254	45	**299**		
			Total	**409**	**70**	**479**		
	H3K9me3	*D. buzzatii*	**Overexpressed**	85	76	**161**	0.000	**1.37E-04 *****
			Underexpressed	106	210	**316**		
			Total	**191**	**286**	**477**		
		D. koepferae	**Overexpressed**	109	71	**180**	0.000	**2.85E-12 *****
			Underexpressed	80	217	**297**		
			Total	**189**	**288**	**477**		
	H3K27me3	*D. buzzatii*	**Overexpressed**	75	86	**161**	0.002	**2.25E-03 ****
			Underexpressed	103	222	**325**		
			Total	**178**	**308**	**486**		
		D. koepferae	**Overexpressed**	158	28	**186**	0.000	**1.37E-04 *****
			Underexpressed	208	92	**300**		
			Total	**366**	**120**	**486**		

Significance: **p < 0.01;

Table S9: Linear Model

Linear relationship between chromatin marks in TE families.

	General statistics			Effect of each LM component		
	Formula	**Adjusted r-squared**	**p-value**	**Components of LM**	**Size effect**	**p-value**
D. buzzatii	**H3K4me3 ~ H3K9me3**	0.442	**< 2.2E-16 *****	**Intercept**	-0.138	**2.16E-05 *****
				H3K9me3	0.959	**< 2E-16 *****
	H3K4me3 ~ H3K27me3	0.499	**< 2.2E-16 *****	**Intercept**	-0.227	**8.09E-13 *****
				H3K27me3	1.183	**< 2E-16 *****
D. koepferae	**H3K4me3 ~ H3K9me3**	0.091	**1.669E-15 *****	**Intercept**	-1.047	**< 2E-16 *****
				H3K9me3	0.445	**1.67E-15 *****
	H3K4me3 ~ H3K27me3	0.258	**< 2.2E-16 *****	**Intercept**	-0.620	**< 2E-16 *****
				H3K27me3	0.834	**< 2E-16 *****

* Intercept: expected mean value of Y when all X=0   Significance: ***p < 0.001

Table S10. Regulatory divergence between parental species

Table of the numbers of genes included in each regulatory divergence class and their percentage, and Two proportion Z-test to compare the number of genes with *cis-* and *trans*-regulatory divergence.

Regulatory divergence classes	Number of genes	Percentage	Two Proportion Z-Test *cis- vs trans*- regulatory classes	
Trans -regulatory divergence	499	6.828	Z-score	Associated p-value
Cis- regulatory divergence	456	6.240	1.968	4.91E-02 *
Cis- + trans-	265	3.626		
Cis- x *trans-*	38	0.520		
Compensatory	716	9.797		
Ambiguous	1121	15.339		
Conserved	4213	57.649	Significance: *$p < 0.05$	
Total	7308	100.000		

Table S11. Regulatory divergence in differential expressed genes between parental species

Association between *trans-* and *cis*-regulated genes and differential expression between parental species *D. buzzatii* and *D. koepferae.*

	trans- regulation	Remaining genes	Total	p-value	Adjusted p-value (FDR)
Differentially expressed genes between parental species	227	283	510	< 2.2E-16	4.40E-16 ***
Not differentially expressed genes between parental species	272	6526	6798		
Total	499	6809	7308		

Significance: ***$p < 0.001$

	cis- regulation	Remaining genes	Total	p-value	Adjusted p-value (FDR)
Differentially expressed genes between parental species	40	470	510	0.128	0.128
Not differentially expressed genes between parental species	416	6382	6798		
Total	456	6852	7308		

Table S12. Regulatory divergence in differentially expressed genes in hybrids *vs* parental species

Numbers of differentially expressed genes in hybrids *vs D. koepferae* and *D. buzzatii* (divided by expression category) belonging to a regulatory divergence class. Fisher's exact test comparing the number of Comensatory regulated genes in the Deregulated category *vs* the rest of expression categories.

Expression categories	Regulation divergence class	Number of genes	Fisher's exact test of Compensatory vs remaining genes	Adjusted p-value (FDR)
Additive expression	*Cis-* regulatory divergence	0	**Additive expression *vs* deregulated**	**0.039 ***
	Cis- + *trans-*	3		
	Cis- x *trans-*	0		
	Compensatory	0	0.039	
	Trans- regulatory divergence	39		
	Total	42		
***D. buzzatii* -like expression**	*Cis-* regulatory divergence	11	***D. buzzatii-* like expression *vs* deregulated**	**4.42E-03 *****
	Cis- + *trans-*	5		
	Cis- x *trans-*	14		
	Compensatory	3	0.003	
	Trans- regulatory divergence	80		
	Total	113		
***D. koepferae* -like expression**	*Cis-* regulatory divergence	15	***D. koepferae-* like expression *vs* deregulated**	**4.42E-03 *****
	Cis- + *trans-*	7		
	Cis- x *trans-*	11		
	Compensatory	5	0.002	
	Trans- regulatory divergence	89		
	Total	127		
Deregulated	*Cis-* regulatory divergence	4		
	Cis- + *trans-*	2		
	Cis- x *trans-*	0		
	Compensatory	11		
	Trans- regulatory divergence	11	**Significance: *p < 0.05; ***p < 0.001**	
	Total	28		

Table S13: Fisher's exact test

piRNA amount changes and TE family transcript amount changes integration.

Hybrids in comparison to	Expression	More piRNA amounts	Less piRNA amounts	Total	p-value	Adjusted p-value (FDR)
D. buzzatii	**Overexpressed**	108	26	**134**	0.328	0.328
	Underexpressed	240	44	**284**		
	Total	**348**	**70**	**418**		
D. koepferae	**Overexpressed**	39	115	**154**	0.043	0.086
	Underexpressed	45	221	**266**		
	Total	**84**	**336**	**420**		

Hybrids in comparison to	Class of TE	Expression	More piRNA amounts	Less piRNA amounts	Total	p-value	Adjusted p-value (FDR)
D. buzzatii	Class I	Overexpressed	72	17	89	0.732	0.732
		Underexpressed	129	27	156		
		Total	**201**	**44**	245		
	Class II	Overexpressed	29	8	37	0.160	0.564
		Underexpressed	90	11	101		
		Total	**119**	**19**	138		
D. koepferae	Class I	Overexpressed	19	75	94	0.615	0.732
		Underexpressed	27	129	156		
		Total	**46**	**204**	250		
	Class II	Overexpressed	14	38	52	0.282	0.564
		Underexpressed	15	68	83		
		Total	**29**	**106**	135		

8.1.2 Supplementary file 2

Supplementary file 2: Enriched Gene Ontology terms

Significant Gene Ontologies in comparison to at least one parental specie (white p-values). Grey p-values are not considered significant and green rows are significant in comparison to both parental species.

Expression in hybrids	GO ID	GO summary explanation	Fisher's p-value for *D. buzzatii*	Fisher's p-value for *D. koepferae*
Underexpressed genes in hybrids in comparison to the parental species	GO:0001746	Bolwig's organ morphogenesis	0.18776	0.01566
	GO:0002052	positive regulation of neuroblast prolif...	0.00567	0.05197
	GO:0007156	homophilic cell adhesion via plasma memb...	0.00055	5.00E-04
	GO:0007157	heterophilic cell-cell adhesion via plas...	0.18776	0.01566
	GO:0007254	JNK cascade	0.13833	0.00022
	GO:0007275	multicellular organism development	0.00086	0.05831
	GO:0007304	chorion-containing eggshell formation	1.60E-08	1.60E-05
	GO:0007305	vitelline membrane formation involved in...	1.10E-06	4.20E-05
	GO:0007313	maternal specification of dorsal/ventral...	2.50E-05	0.08373
	GO:0007343	egg activation	0.01181	1
	GO:0007414	axonal defasciculation	0.13801	0.00775
	GO:0007503	fat body development	0.00803	0.00239
	GO:0007516	hemocyte development	0.01181	0.16049
	GO:0008258	head involution	0.00194	0.10197
	GO:0008340	determination of adult lifespan	0.00763	0.31033
	GO:0008406	gonad development	0.01205	0.01145
	GO:0008584	male gonad development	0.01181	0.0114
	GO:0009792	embryo development ending in birth or eg...	0.01995	0.01971
	GO:0010025	wax biosynthetic process	0.00124	0.02049
	GO:0010862	positive regulation of pathway-restricte...	0.00491	0.00474
	GO:0016042	lipid catabolic process	0.0339	0.00173
	GO:0016336	establishment or maintenance of polarity...	0.00803	0.00775
	GO:0016339	calcium-dependent cell-cell adhesion via...	0.00341	0.00324
	GO:0030198	extracellular matrix organization	0.00507	0.30357
	GO:0034334	adherens junction maintenance	0.01181	0.16049
	GO:0034446	substrate adhesion-dependent cell spread...	0.01622	0.01566
	GO:0035332	positive regulation of hippo signaling	0.00863	0.06741
	GO:0035336	long-chain fatty-acyl-CoA metabolic proc...	0.00181	0.02585
	GO:0042981	regulation of apoptotic process	0.1136	0.01852
	GO:0044331	cell-cell adhesion mediated by cadherin	0.00181	0.00172
	GO:0045167	asymmetric protein localization involved...	0.34043	0.00671
	GO:0045176	apical protein localization	0.18776	0.01566
	GO:0045571	negative regulation of imaginal disc gro...	0.00041	0.05197
	GO:0046672	positive regulation of compound eye reti...	0.00803	1
	GO:0046843	dorsal appendage formation	0.02008	0.0189
	GO:0055013	cardiac muscle cell development	0.08524	0.00241
	GO:0055085	transmembrane transport	0.00367	0.00893
	GO:0061060	negative regulation of peptidoglycan rec...	0.00567	0.05197
	GO:1900074	negative regulation of neuromuscular syn...	0.11201	0.00474

	GO:0006036	cuticle chitin catabolic process	0.0152	1
	GO:0008039	synaptic target recognition	0.0182	1
	GO:0009609	response to symbiotic bacterium	1	0.01383
	GO:0018990	ecdysis, chitin-based cuticle	0.0137	1
	GO:0019310	inositol catabolic process	0.0152	1
	GO:0019853	L-ascorbic acid biosynthetic process	0.0152	1
	GO:0032499	detection of peptidoglycan	1	0.01383
	GO:0033345	asparagine catabolic process via L-aspar...	1	0.01383
Overexpress	GO:0034435	cholesterol esterification	0.0152	1
ed genes in	GO:0035025	positive regulation of Rho protein signa...	0.0077	1
hybrids in	GO:0035294	determination of wing disc primordium	1	0.01383
comparison	GO:0035694	mitochondrial protein catabolic process	0.0152	0.01383
to the	GO:0035695	mitophagy by induced vacuole formation	0.0152	0.01383
parental	GO:0048150	behavioral response to ether	0.0152	1
species	GO:0055114	oxidation-reduction process	0.0399	0.00062
	GO:0070073	clustering of voltage-gated calcium chan...	0.0152	1
	GO:0070413	trehalose metabolism in response to stre...	0.0152	1
	GO:0070899	mitochondrial tRNA wobble uridine modifi...	0.0152	1
	GO:0071938	vitamin A transport	0.0152	1
	GO:0097118	neuroligin clustering involved in postsy...	0.0152	1
	GO:0140300	serine import into mitochondrion	1	0.01383
	GO:1903006	positive regulation of protein K63-linke...	0.0152	1

8.1.3   Supplementary file 3

Figure S1

Differential gene (**a**) and TE family (**b**) expression analyses in *D. buzzatii vs D. koepferae*. Positive log2FC values correspond to genes (**a**) and TE families (**b**) more expressed in *D. buzzatii*. The genes showing the 20 highest log2FC values and displaying an ortholog in *D. melanogaster* are shown in **a**. The TE superfamilies/orders showing the 10 highest log2F values are shown in **b**. Genes (**a**) and TE families (**b**) common to the examples in the hybrids *vs D. buzzatii* comparison are shown in green, hybrids *vs D. koepferae* are shown in blue and in all three comparisons are in bold.
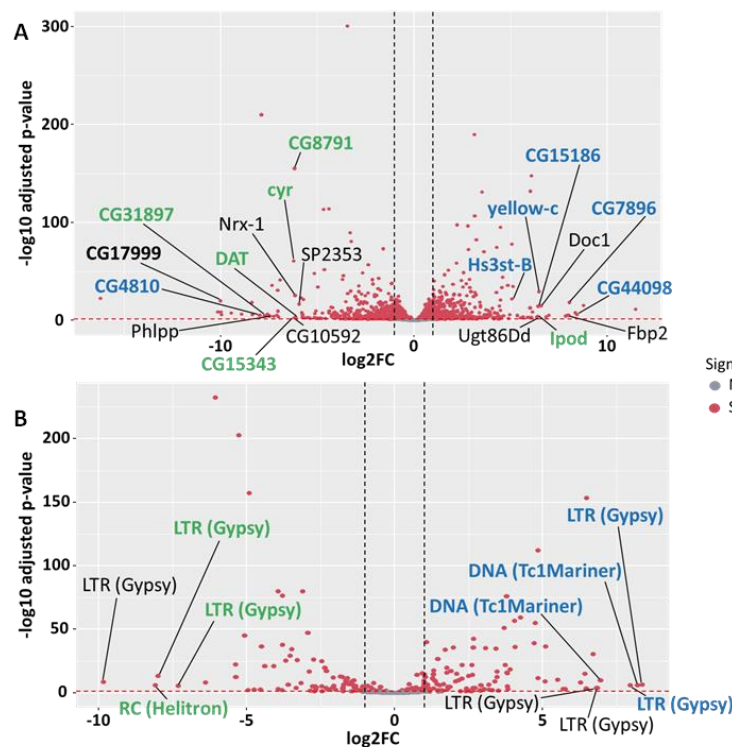
Figure S2

a Violin plots representing the chromosomal distribution of differentially expressed genes in hybrids vs parental species. Genes with an unknown location were categorized as unknown. Points indicate the log2FC of each gene. The percentages of differentially expressed genes per chromosome are framed. The expected (total) gene percentages are at the bottom. Red-dashed lines indicate the log2FC threshold (±1). **b** Differential TE family expression analyses in hybrids vs *D. buzzatii* (green) and *D. koepferae* (blue). Positive log2FC values correspond to TE families more expressed in hybrids. The TE superfamilies/orders showing the 10 highest log2FC values are shown. TE families common to both comparisons are in bold and identified with a subscript.
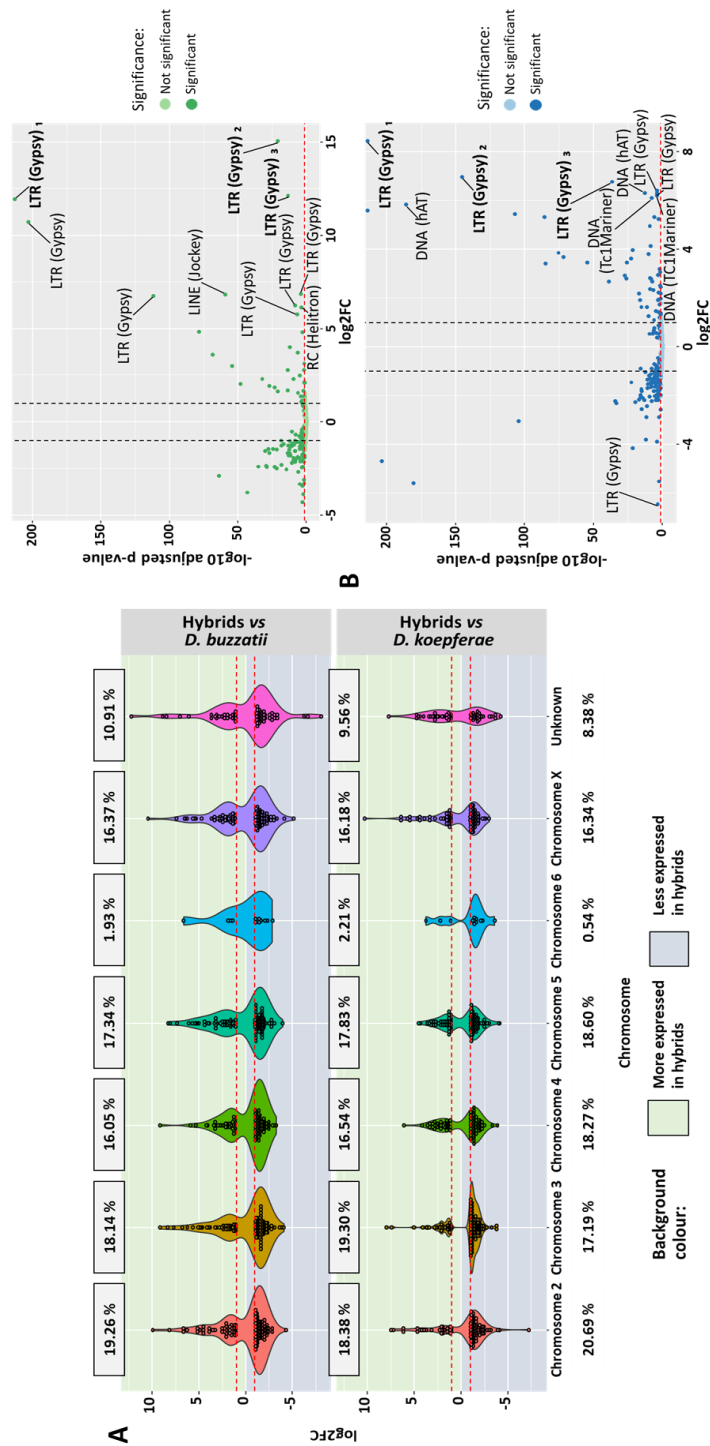
Figure S3

Violin plots representing the chromosomal distribution of differentially H3K4me3, H3K9me3 and H3K27me3 enriched genes in hybrids vs parental species. Points indicate the size effect. The percentages of differentially enriched genes per chromosome are framed. The expected (total) gene percentages per chromosome are at the bottom.
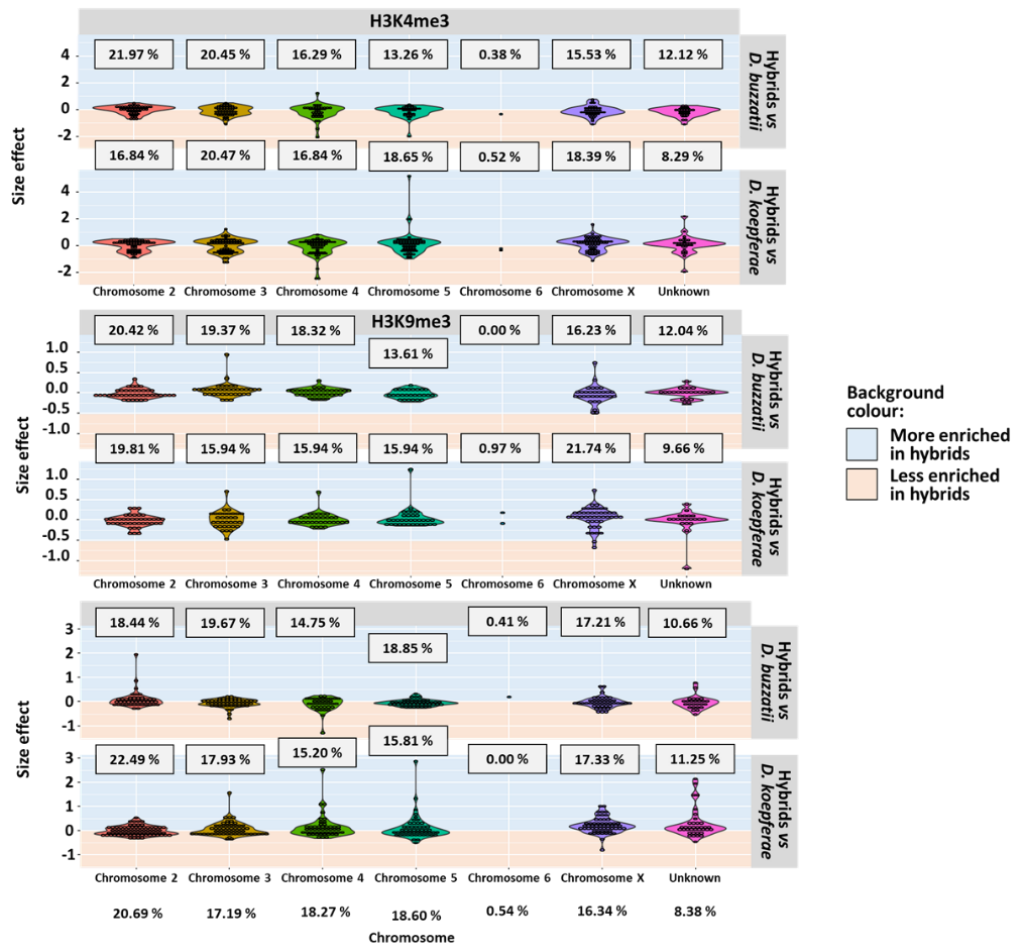
Figure S4

Association between H3K4me3-H3K9me3 and between H3K4me3-H3K27me3 enrichment in hybrids vs *D. buzzatii* and *D. koepferae* respectively. Colours represent TE families differential expression in hybrids vs parental species. Linear model p-values are shown in the corner.
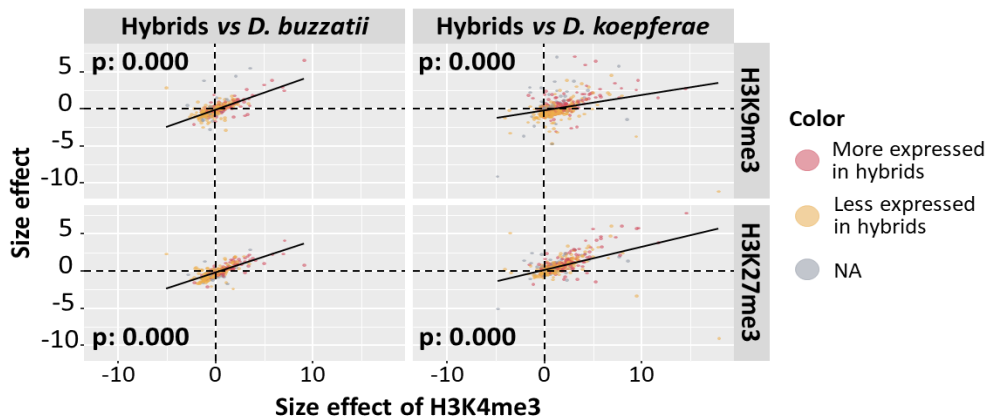
Figure S5

Differential expression (log2FC) and enrichment analyses (size effect) of piRNA Pathway Genes in hybrids vs parental species. Significant values (p < 0.05) are indicated with an asterisk (*). Colours indicate the values of the differences in gene expression (green - orange) and enrichment (purple - red).
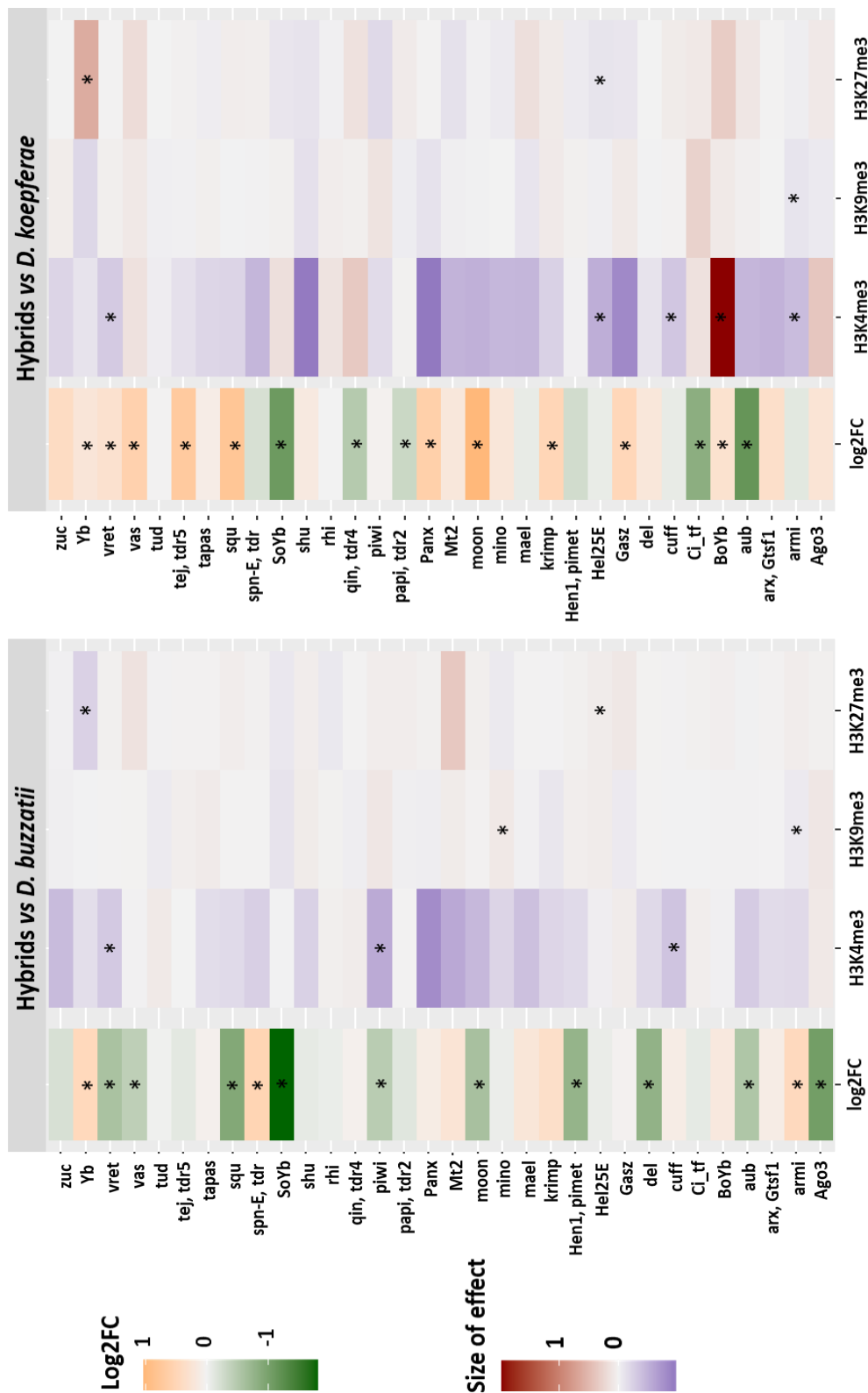
8.1.4 Supplementary file 4

Percentage of alignment of the input with the transcriptome

Percentage of alignment between the transcriptome of *D. buzzatii* and *D. koepferae* and the ChIPseq inputs.

	ChIPseq inputs					
	D. buzzati A	*D. buzzati B*	Hybrids A	Hybrids B	*D. koepferae A*	*D. koepferae B*
D. buzzatii transcriptome	31.43%	33.56%	32.82%	30.81%	36.21%	36.28%
D. koepferae transcriptome	29.27%	31.76%	34.18%	32.15%	40.36%	40.38%

Nucleotide divergence of 40 genes in *D. buzzatii vs D. koepferae*

Number of differences and distance using the Jukes-Cantor model of 40 randomly selected genes.

Gene	Nº of differences	Jukes-Cantor model
TRINITY_DN6741_c0_g1	0	0
TRINITY_DN4937_c5_g1	0	0
TRINITY_DN4394_c2_g1	0	0
TRINITY_DN3936_c3_g1	0	0
TRINITY_DN33366_c0_g1	0	0
TRINITY_DN3272_c5_g2	0	0
TRINITY_DN32714_c0_g1	0	0
TRINITY_DN32055_c0_g1	0	0
TRINITY_DN26642_c0_g1	0	0
TRINITY_DN25744_c0_g1	0	0
TRINITY_DN22061_c0_g1	0	0
TRINITY_DN19973_c0_g1	0	0
TRINITY_DN19240_c0_g1	0	0
TRINITY_DN17550_c0_g3	0	0
TRINITY_DN16569_c2_g1	0	0
TRINITY_DN13618_c0_g1	0	0
TRINITY_DN10041_c0_g1	0	0
TRINITY_DN10014_c0_g1	0	0
TRINITY_DN21508_c0_g1	0	0
TRINITY_DN12521_c0_g1	0	0
TRINITY_DN14225_c0_g3	0	0
TRINITY_DN14427_c0_g1	0	0
TRINITY_DN16326_c0_g1	0	0
TRINITY_DN19343_c5_g1	0	0
TRINITY_DN19880_c1_g1	0	0
TRINITY_DN23959_c0_g1	0	0
TRINITY_DN29_c3_g1	0	0
TRINITY_DN8582_c3_g1	0	0
TRINITY_DN30718_c0_g1	0	0
TRINITY_DN3819_c0_g1	5	0.00182
TRINITY_DN1970_c4_g1	2	0.00255
TRINITY_DN10722_c0_g1	1	0.00346
TRINITY_DN266_c12_g1	1	0.00499
TRINITY_DN3116_c2_g1	5	0.00563
TRINITY_DN515_c0_g1	81	0.00783
TRINITY_DN16064_c0_g1	8	0.0083
TRINITY_DN2626_c0_g1	12	0.00939
TRINITY_DN3536_c0_g1	205	0.0119
TRINITY_DN369_c3_g1	23	0.0241
TRINITY_DN1362_c0_g1	28	0.0265
Average	**9.275**	**0.00266175**

8.2 Supplementary data of "Impact of the heat stress on the Transposable Elements and small RNA expression in *Drosophila subobscura*"

8.2.1 Supplementary Tables

Table S1. Differentially expressed genes between populations in control conditions

Number and percentage of differentially expressed (DE) genes between Madeira and Curicó from the total expressed in control conditions.

Sex	Total expressed	More expressed in Madeira	More expressed in Curicó	Total number of DE	More expressed in Madeira (%)	More expressed in Curicó (%)	DE (%)
Females	11207	250	224	474	2.231	1.999	4.229
Males	12569	623	594	1217	4.957	4.726	9.683

Table S2. Two Proportion Z-Test for genes in control conditions

Comparison of the number of differentially expressed genes between populations in males and females.

Sex	Differentially expressed	Total	Proportions	Total proportion	Z-score	P-value
Females	474	11207	0.042	0.071	-16.330	< 2.2E-16 ***
Males	1217	12569	0.097			

***: p < 0.001

Table S3. Spearman's rank correlation for genes in control conditions

Sperman's rank correlation for genes between sexes and populations using the mean normalized counts of both replicates in control conditions.

Comparison		rho	p-value
Populations	Females	0.974	< 2.2e-16 ***
	Males	0.970	< 2.2e-16 ***
Sexes	Curicó	0.611	< 2.2e-16 ***
	Madeira	0.568	< 2.2e-16 ***

***: p < 0.001

Table S4. Differentially expressed TE families between populations in control conditions

Number and percentage of differentially expressed (DE) TE families between Madeira and Curicó from the total expressed in control conditions.

Sex	Total expressed	More expressed in Madeira	More expressed in Curicó	Total number of DE	More expressed in Madeira (%)	More expressed in Curicó (%)	DE (%)
Females	618	18	32	50	2.913	5.178	8.091
Males	658	23	42	65	3.495	6.383	9.878

Table S5. Two Proportion Z-Test for TEs in control conditions

Comparison of the number of differentially expressed TE families between populations in males and females.

Sex	Differentially expressed	Total	Proportions	Total proportion	Z-score	P-value
Females	50	618	0.081	0.090	-1.115	0.2651
Males	65	658	0.099			

Table S6. Spearman's rank correlation for TEs in control conditions

Sperman's rank correlation for TE families between sexes and populations using the mean normalized counts of both replicates in control conditions.

Comparison		rho	p-value
Populations	Females	0.968	< 2.2e-16 ***
	Males	0.974	< 2.2e-16 ***
Sexes	Curicó	0.837	< 2.2e-16 ***
	Madeira	0.828	< 2.2e-16 ***

***: p < 0.001

Table S7. Differentially expressed genes under heat shock *vs* control

Comparison	Total expressed	Over-expressed	Under-expressed	Total number of DE	Over-expressed (%)	Under-expressed (%)	DE (%)
Curicó females	11664	88	23	111	0.754	0.197	0.952
Madeira females		64	27	91	0.549	0.231	0.780
Curicó males	12722	62	43	105	0.487	0.338	0.825
Madeira males		144	11	155	1.132	0.086	1.218

DE: differentially expressed genes after a heat stress from the total expressed genes

Table S8. Two Proportion Z-Test for genes in heat shock *vs* control

Comparison of the number of differentially expressed genes after a heat shock between Madeira and Curicó populations and between sexes within a population.

Population	Sex	Differentially expressed	Total	Proportions	Total proportion	Z-score	Associated p-value	Adjusted p-value
Madeira	males	155	12722	0.012	0.010	3.117	**0.002**	**0.004 ****
Curicó	males	105	12722	0.008				
Madeira	females	91	11664	0.008	0.009	1.413	0.158	0.210
Curicó	females	111	11664	0.010				
Curicó	females	111	11664	0.010	0.009	1.052	0.293	0.293
Curicó	males	105	12722	0.008				
Madeira	females	91	11664	0.008	0.010	3.421	**0.001**	**0.003 ****
Madeira	males	155	12722	0.012				

**: p < 0.01

Table S9. Differentially expressed TE families under heat shock *vs* control

Comparison	Total expressed	Over-expressed	Under-expressed	Total number of DE	Over-expressed (%)	Under-expressed (%)	DE (%)
Curicó females	646	5	2	7	0.774	0.310	1.084
Madeira females		3	2	5	0.464	0.310	0.774
Curicó males	684	10	0	10	1.462	0.000	1.462
Madeira males		12	0	12	1.754	0.000	1.754

DE: differentially expressed TEs after a heat stress from the total expressed TEs

Table S10. Two Proportion Z-Test for TEs in heat shock *vs* control

Comparison of the number of differentially expressed TEs after a heat shock between Madeira and Curicó populations and between sexes within a population.

Population	Sex	Differentially expressed	Total	Proportions	Total proportion	Z-score	Associated p-value	Adjusted p-value
Madeira	males	12	684	0.018	0.016	0.430	0.667	0.667
Curicó	males	10	684	0.015				
Madeira	females	5	646	0.008	0.009	0.580	0.562	0.667
Curicó	females	7	646	0.011				
Curicó	females	7	646	0.011	0.013	-0.614	0.539	0.667
Curicó	males	10	684	0.015				
Madeira	females	5	646	0.008	0.013	1.591	0.112	0.447
Madeira	males	12	684	0.018				

Table S11. Common piRNA clusters in Madeira and Curicó

Full Table available at: https://figshare.com/s/d33c0f4846f894e4a3be.

piRNA Cluster Name	Chr	Start	End	Direction	Total percentage of the cluster overlapping TEs	piRNA cluster annotation	Curicó Control Females	Curicó HS Females	Madeira Control Females	Madeira HS Females	Curicó Males Control	Curicó Males HS	Madeira Males Control	Madeira Males HS
piRNA cluster no: 1	ChrA	498010	521021	bi:+/-	96.858	Different TE families/classes	860.538	1184.08	956.725	Not detected	700.996	797.433	Not detected	Not detected
piRNA cluster no: 2	ChrA	7967000	7974005	bi:+/-	99.914	GYPSY	109.459	73.397	670.643	712.195	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 3	ChrA	8082054	8096526	bi:-/+	52.418	Different TE families/classes	2625.661	Not detected	1615.71	1227.097	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 4	ChrA	8160047	8196739	mono:+	78.54	GYPSY	41799.147	26292.724	24265.689	21790.914	9679.981	9714.029	6937.704	6657.778
piRNA cluster no: 5	ChrA	8803033	8839945	bi:-/+	38.44	Different TE families/classes	57384.791	67120.886	52360.095	Not detected	11021.794	10987.786	7823.014	8227.082
piRNA cluster no: 6	ChrA	11524094	11540851	mono:+	62.386	GYPSY	3349.895	1974.826	21210.636	21899.152	1969.972	1809.082	10102.614	10287.578
piRNA cluster no: 7	ChrA	14256452	14267733	bi:-/+	73.478	GYPSY	749.585	983.372	1400.257	Not detected	341.718	Not detected	Not detected	Not detected
piRNA cluster no: 8	ChrA	17008457	17018025	mono:-	94.492	GYPSY	7373.912	6193.654	2526.787	2050.731	2156.065	2732.32	Not detected	Not detected
piRNA cluster no: 9	ChrA	21225019	21232904	bi:-/+	17.628	Different TE families/classes	1137.95	Not detected	665.206	Not detected	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 10	ChrE	16024	85025	bi:-/+	83.722	Different TE families/classes	40820.576	36050.046	29364.662	28703.759	15370.947	Not detected	11871.993	Not detected
piRNA cluster no: 11	ChrE	132179	140993	bi:+/-	89.29	CR1	125.037	135.546	122.364	113.881	216.349	217.737	192.573	Not detected
piRNA cluster no: 12	ChrE	148050	176120	bi:+/-	79.672	UNKNOWN	2552.568	Not detected	1901.662	1689.839	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 13	ChrE	333084	352001	bi:+/-	72.136	Different TE families/classes	382.096	Not detected	482.766	434.96	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 14	ChrE	354017	396778	bi:-/+	73.09	LINE	6427.213	Not detected	4435.111	4752.029	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 15	ChrE	410312	514961	mono:+	86.734	Different TE families/classes	30975.959	29458.577	72654.07	73103.463	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 16	ChrE	568107	580008	bi:+/-	79.439	BEL	437.551	368.934	317.175	342.339	Not detected	Not detected	Not detected	604.343
piRNA cluster no: 17	ChrE	604012	610007	bi:+/-	95.48	LINE	66.012	57.7	93.748	103.875	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 18	ChrE	613505	641003	bi:-/+	72.936	Different TE families/classes	1608.607	1881.693	1710.19	1931.184	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 19	ChrE	681194	713852	bi:+/-	85.039	LOA	1057.092	1031.576	3567.312	2505.123	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 20	ChrE	731211	741994	bi:-/+	83.177	LINE	648.88	Not detected	414.478	584.835	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 21	ChrE	781036	806606	mono:+	89.139	Different TE families/classes	2428.975	2465.446	2275.142	2357.164	2341.021	2257.404	Not detected	Not detected
piRNA cluster no: 22	ChrE	821001	833655	bi:-/+	83.073	UNKNOWN	466.995	362.104	346.641	301.991	577.42	Not detected	Not detected	Not detected
piRNA cluster no: 23	ChrE	873248	883003	bi:-/+	83.311	LINE	121.421	112.71	135.993	119.52	370.298	353.546	Not detected	166.938
piRNA cluster no: 24	ChrE	939349	948012	bi:-/+	70.414	UNKNOWN	219.728	161.123	119.181	138.151	Not detected	261.323	Not detected	Not detected
piRNA cluster no: 25	ChrE	1177653	1186737	mono:-	53.336	JOCKEY	9461.469	9732.691	11226.096	Not detected	Not detected	Not detected	Not detected	3481.674
piRNA cluster no: 26	ChrE	1261999	1273368	mono:-	45.343	UNKNOWN	1557.584	1415.952	625.292	837.133	1467.71	1722.372	Not detected	Not detected
piRNA cluster no: 27	ChrE	1420483	1447867	bi:+/-	94.007	LINE	23076.342	26393.963	22721.78	24202.689	10281.907	11927.575	5836.3	6126.02
piRNA cluster no: 28	ChrE	2059037	2137007	bi:+/-	78.138	Different TE families/classes	30620.14	26841.02	24972.834	20561.578	13363.234	14014.534	Not detected	Not detected
piRNA cluster no: 29	ChrE	2396174	2433907	mono:+	57.353	UNKNOWN	11982.973	10548.204	12019.535	10275.379	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 30	ChrE	3108736	3115139	bi:+/-	100	Different TE families/classes	1458.714	1770.387	436.458	375.526	1791.768	1971.235	2241.245	1898.21
piRNA cluster no: 31	ChrE	18690000	18697026	bi:+/-	97.865	GYPSY	84.354	Not detected	94.267	Not detected	2542.654	3097.503	615.003	Not detected
piRNA cluster no: 32	ChrI	379774	398762	bi:+/-	69.354	Different TE families/classes	2361.813	2101.42	2671.107	2691.356	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 33	ChrI	484001	505909	bi:+/-	85.777	LOA	1170.342	1560.866	2402.116	210.656	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 34	ChrI	529714	538396	bi:+/-	89.254	JOCKEY	208.968	125.938	180.028	86.066	Not detected	330.111	Not detected	Not detected
piRNA cluster no: 35	ChrI	555022	563594	bi:-/+	75.665	Different TE families/classes	592.108	Not detected	108.252	Not detected	Not detected	185.51	Not detected	Not detected
piRNA cluster no: 36	ChrI	691206	706918	mono:-	78.831	LINE	7351.174	7384.041	5676.426	7868.464	3760.221	4660.337	Not detected	1644.781
piRNA cluster no: 37	ChrI	1030303	1039966	mono:-	52.675	UNKNOWN	142.159	152.026	3539.529	3143.195	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 38	ChrI	2288744	2314307	bi:+/-	62.915	UNKNOWN	9506.56	8130.416	4539.641	4445.449	25550.286	28227.468	13844.659	11897.019
piRNA cluster no: 39	ChrI	2545076	2562021	bi:+/-	76.925	LINE	1243.451	1223.014	1045.302	1057.524	2143.689	2348.166	Not detected	996.219
piRNA cluster no: 40	ChrI	5501781	5509955	bi:+/-	87.277	GYPSY	139.205	295.318	256.389	670.079	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 41	ChrI	5784222	5791205	bi:+/-	99.928	BEL	2782.221	1959.836	2638.136	3005.635	6233.405	5921.493	12342.281	8993.888
piRNA cluster no: 42	ChrI	16340555	16347650	bi:+/-	77.195	UNKNOWN	5203.21	4591.311	2002.95	3560.065	3904.421	2684.752	3493.891	3349.819
piRNA cluster no: 43	ChrI	18853251	18863026	bi:+/-	25.207	UNKNOWN	122.372	134.297	258.872	223.097	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 44	ChrI	24997035	25018014	bi:-/+	87.378	UNKNOWN	3293.923	2699.371	1312.925	1541.644	Not detected	Not detected	Not detected	Not detected

piRNA Cluster Name	Chr	Start	End	Direction	Total percentage of the cluster overlapping TEs	piRNA cluster annotation	Mapped reads per cluster normalized by Reads Per Million (RPM) - piRNA cluster production							
							Curicó Control Females	Curicó HS Females	Madeira Control Females	Madeira HS Females	Curicó Males Control	Curicó Males HS	Madeira Males Control	Madeira Males HS
piRNA cluster no: 45	ChrO	13	87845	mono:-	91.201	Different TE families/classes	78068.099	81630.727	73721.476	74649.321	39594.073	41493.013	41314.747	40163.272
piRNA cluster no: 46	ChrO	147107	156703	bi:-/+	86.098	Different TE families/classes	1356.054	Not detected	805.657	Not detected	Not detected	Not detected	1179.558	1244.789
piRNA cluster no: 47	ChrO	182078	212937	mono:-	84.413	LTR	14408.013	13712.635	9387.489	11142.041	11015.696	10943.319	5393.07	5591.201
piRNA cluster no: 48	ChrO	373297	394692	bi:+/-	62.692	Different TE families/classes	4321.485	Not detected	2551.216	3268.44	2853.628	Not detected	Not detected	Not detected
piRNA cluster no: 49	ChrO	423003	436816	bi:+/-	63.375	UNKNOWN	1655.01	1293.568	712.156	1285.921	1176.586	1219.971	Not detected	351.432
piRNA cluster no: 50	ChrO	460593	471017	bi:-/+	65.157	Different TE families/classes	208.078	237.409	697.944	555.159	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 51	ChrO	712812	750340	bi:-/+	80.217	Different TE families/classes	5272.789	Not detected	2166.46	2252.556	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 52	ChrO	942082	968016	bi:+/-	75.538	Different TE families/classes	5081.797	4783.748	4314.16	Not detected	2861.359	2941.657	Not detected	1541.063
piRNA cluster no: 53	ChrO	1104163	1117608	bi:-/+	58.907	Different TE families/classes	1852.188	2915.843	3886.404	3759.41	5083.032	4868.209	5398.576	Not detected
piRNA cluster no: 54	ChrO	1532545	1567499	bi:-/+	52.938	Different TE families/classes	29746.344	34562.668	35236.13	36334.645	10817.026	11550.051	8863.575	9367.454
piRNA cluster no: 55	ChrO	2141870	2160570	bi:+/-	73.631	UNKNOWN	3849.285	2629.84	6333.286	6195.177	4098.304	4082.003	5260.029	5391.31
piRNA cluster no: 56	ChrO	4526433	4533166	bi:+/-	81.36	UNKNOWN	3374.292	2797.133	2261.026	2928.539	Not detected	1864.606	2890.004	2775.336
piRNA cluster no: 57	ChrO	4983038	4993587	bi:-/+	29.472	UNKNOWN	3176.763	2619.028	5053.202	5806.491	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 58	ChrO	7115361	7124865	bi:-/+	78.967	GYPSY	1794.951	1054.369	1030.209	1252.589	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 59	ChrO	9157058	9166003	bi:-/+	39.318	UNKNOWN	1802.986	1460.888	758.408	Not detected	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 60	ChrO	9708889	9723489	bi:-/+	29.945	UNKNOWN	3484.316	2519.932	3908.416	6249.864	12539.292	Not detected	10509.28	11341.714
piRNA cluster no: 61	ChrO	15371658	15378789	mono:+	82.541	GYPSY	195.293	150.676	1407.278	1058.063	Not detected	248.505	195.366	220.894
piRNA cluster no: 62	ChrO	18903831	18909209	mono:-	99.833	R1	145.924	Not detected	164.654	93.216	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 63	ChrO	19373350	19381989	mono:+	23.834	UNKNOWN	2055.746	Not detected	732.303	818.236	Not detected	Not detected	2245.848	Not detected
piRNA cluster no: 64	ChrU	65461	77616	bi:-/+	79.877	OUTCAST	468.469	468.615	719.302	710.638	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 65	ChrU	623205	633633	bi:-/+	80.063	LINE	331.153	286.605	329.329	308.617	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 66	ChrU	641131	668015	bi:+/-	88.424	Different TE families/classes	3454.859	2302.718	9029.311	9092.715	3887.375	4584.536	5008.528	5073.701
piRNA cluster no: 67	ChrU	998813	1005007	bi:-/+	61.899	CR1	132.771	Not detected	153.469	138.743	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 68	ChrU	1622868	1633588	bi:+/-	75.979	Different TE families/classes	252.374	238.239	231.592	273.925	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 69	ChrU	2092269	2109794	bi:+/-	46.328	Different TE families/classes	1792.66	2063.947	1395.883	1697.143	800.125	940.953	Not detected	Not detected
piRNA cluster no: 70	ChrU	2136100	2162670	mono:-	48.976	UNKNOWN	7258.219	6924.561	6830.331	7038.263	7593.716	9000.974	9742.439	8178.393
piRNA cluster no: 71	ChrU	2677149	2711021	bi:-/+	35.702	UNKNOWN	4232.014	Not detected	2603.091	3087.982	10235.601	Not detected	Not detected	Not detected
piRNA cluster no: 72	ChrU	2717003	2739969	bi:-/+	44.675	BEL	4094.309	Not detected	3486.416	3258.615	7718.507	Not detected	Not detected	Not detected
piRNA cluster no: 73	ChrU	4256760	4266264	mono:-	96.833	GYPSY	1142.311	1236.692	218.404	Not detected	349.354	Not detected	Not detected	Not detected
piRNA cluster no: 74	ChrU	4646774	4657536	bi:+/-	79.455	GYPSY	469.878	576.331	5308.071	4737.805	275.007	297.696	625.562	415.051
piRNA cluster no: 75	ChrU	6390311	6401639	bi:+/-	26.095	UNKNOWN	1860.747	1439.338	1259.448	1196.823	6768.486	6942.593	4174.228	4658.983
piRNA cluster no: 76	ChrU	6598390	6605521	bi:-/+	99.972	GYPSY	527.027	563.615	353.328	160.068	730.405	1489.126	Not detected	1062.796
piRNA cluster no: 77	ChrU	8542869	8553752	bi:-/+	68.694	GYPSY	861.15	Not detected	2149.83	2727.617	358.518	465.405	1295.273	1120.267
piRNA cluster no: 78	ChrU	9219422	9229365	bi:-/+	71.196	UNKNOWN	751.988	1288.481	5546.649	Not detected	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 79	ChrU	20015418	20020796	mono:-	99.74	R1	2188.522	2439.431	1852.307	2544.695	2917.902	Not detected	692.464	446.582
piRNA cluster no: 80	ChrU	20095266	20104973	bi:-/+	99.279	GYPSY	606.53	501.945	929.339	560.152	Not detected	554.003	285.269	184.684
piRNA cluster no: 81	ChrU	25111585	25124021	mono:+	4.897	UNKNOWN	8298.285	6537.079	9916.916	9129.816	15546.303	15390.741	39886.503	41219.439
piRNA cluster no: 82	cDot	34009	58714	bi:-/+	31.403	JOCKEY	26568.786	25771.858	13078.961	11393.445	16277.333	16989.5	15877.173	14920.602
piRNA cluster no: 83	cDot	282011	295832	bi:+/-	87.613	Different TE families/classes	1896.868	1234.318	1460.575	2134.457	2416.595	1989.601	1242.766	1529.559
piRNA cluster no: 84	cDot	697020	706889	bi:-/+	69.764	Different TE families/classes	178.377	177.914	295.472	Not detected	Not detected	Not detected	Not detected	Not detected
piRNA cluster no: 85	cDot	799466	808962	bi:-/+	55.876	HELITRON	468.041	408.919	382.71	Not detected	702.684	745.386	334.304	Not detected

Table S12. piRNA clusters in Madeira

Full Table available at: https://figshare.com/s/d33c0f4846f894e4a3be.

piRNA Cluster Name	Chr	Start	End	Direction	Mapped reads per cluster normalized by Reads Per Million (RPM) in Madeira Control Females	Total percentage of the cluster overlapping TEs	piRNA cluster annotation
piRNA cluster no: 1	ChrA	2012	15474	bi:+/-	332.435	92.126	OUTCAST
piRNA cluster no: 2	ChrA	371016	390027	bi:+/-	1131.789	99.658	UNKNOWN
piRNA cluster no: 3	ChrA	396000	478026	bi:-/+	6231.917	99.294	Different TE families/classes
piRNA cluster no: 4	ChrA	953097	971021	bi:-/+	2095.892	60.316	UNKNOWN
piRNA cluster no: 5	ChrA	4957001	4965023	bi:+/-	107.49	98.13	GYPSY
piRNA cluster no: 6	ChrA	10282162	10293918	bi:-/+	382.151	58.881	UNKNOWN
piRNA cluster no: 7	ChrA	11603265	11610983	bi:+/-	77.62	83.973	UNKNOWN
piRNA cluster no: 8	ChrA	14898021	14921853	bi:-/+	5177.994	43.358	UNKNOWN
piRNA cluster no: 9	ChrA	16325116	16338291	bi:-/+	599.774	31.575	UNKNOWN
piRNA cluster no: 10	ChrA	22921999	22940838	bi:-/+	3037.267	41.818	UNKNOWN
piRNA cluster no: 11	ChrE	1	9017	mono:-	990.393	52.85	GYPSY
piRNA cluster no: 12	ChrE	199003	207023	bi:+/-	80.408	88.13	UNKNOWN
piRNA cluster no: 13	ChrE	310013	318857	bi:-/+	118.055	79.444	UNKNOWN
piRNA cluster no: 14	ChrE	1458038	1467622	mono:+	462.834	99.176	Different TE families/classes
piRNA cluster no: 15	ChrE	1696145	1707628	bi:+/-	531.365	76.295	GYPSY
piRNA cluster no: 16	ChrE	2663444	2670436	bi:-/+	424.781	57.423	UNKNOWN
piRNA cluster no: 17	ChrE	4109645	4118814	bi:+/-	426.988	33.69	UNKNOWN
piRNA cluster no: 18	ChrE	18113019	18118580	bi:-/+	74.369	97.321	GYPSY
piRNA cluster no: 19	ChrE	20738002	20783144	bi:+/-	28576.854	49.692	UNKNOWN
piRNA cluster no: 20	ChrJ	1999	18021	bi:-/+	454.742	64.998	GYPSY
piRNA cluster no: 21	ChrJ	34137	122996	bi:-/+	34565.236	87.458	Different TE families/classes
piRNA cluster no: 22	ChrJ	199379	215020	mono:+	285.355	83.409	LINE
piRNA cluster no: 23	ChrJ	219006	229026	bi:-/+	126.151	79.721	Different TE families/classes
piRNA cluster no: 24	ChrJ	252255	260026	bi:-/+	91.774	60.069	DNA
piRNA cluster no: 25	ChrJ	314073	330716	mono:+	239.088	69.699	UNKNOWN
piRNA cluster no: 26	ChrJ	423113	477024	bi:+/-	15203.324	76.476	Different TE families/classes
piRNA cluster no: 27	ChrJ	719000	731025	bi:-/+	284.598	75.052	BEL
piRNA cluster no: 28	ChrJ	868058	879022	bi:+/-	202.728	61.811	UNKNOWN
piRNA cluster no: 29	ChrJ	1455016	1474886	bi:-/+	2262.798	81.691	Different TE families/classes
piRNA cluster no: 30	ChrJ	1480002	1491972	bi:+/-	758.177	76.374	UNKNOWN
piRNA cluster no: 31	ChrJ	4296003	4304290	bi:-/+	111.123	99.988	GYPSY
piRNA cluster no: 32	ChrJ	7575519	7585027	bi:-/+	202.648	82.478	GYPSY
piRNA cluster no: 33	ChrJ	7749898	7758023	bi:-/+	128.809	94.24	GYPSY
piRNA cluster no: 34	ChrJ	19105886	19120979	mono:+	7180.359	87.623	Different TE families/classes
piRNA cluster no: 35	ChrO	126999	136311	bi:+/-	1311.333	79.328	Different TE families/classes
piRNA cluster no: 36	ChrO	545192	573953	bi:+/-	4839.669	62.272	Different TE families/classes
piRNA cluster no: 37	ChrO	621255	636961	mono:-	4568.369	76.869	Different TE families/classes
piRNA cluster no: 38	ChrO	15760420	15767521	bi:+/-	426.808	99.972	GYPSY
piRNA cluster no: 39	ChrO	19208999	19215014	bi:-/+	72.207	100	GYPSY
piRNA cluster no: 40	ChrO	25999999	26006024	bi:-/+	67.264	100	GYPSY
piRNA cluster no: 41	ChrO	27070097	27096461	bi:+/-	6950.578	76.119	GYPSY
piRNA cluster no: 42	ChrO	29823521	29833778	bi:+/-	238.378	87.979	GYPSY
piRNA cluster no: 43	ChrO	30227191	30269405	bi:-/+	9133.499	34.73	UNKNOWN
piRNA cluster no: 44	ChrU	3	23007	bi:+/-	3026.655	44.292	GYPSY
piRNA cluster no: 45	ChrU	847999	861476	mono:+	424.731	78.519	Different TE families/classes
piRNA cluster no: 46	ChrU	1211154	1222021	bi:-/+	343.743	76.728	Different TE families/classes
piRNA cluster no: 47	ChrU	1391432	1397808	bi:-/+	119.925	80.772	CR1
piRNA cluster no: 48	ChrU	1491055	1514023	bi:+/-	2078.371	54.794	UNKNOWN
piRNA cluster no: 49	ChrU	1615000	1620768	bi:-/+	63.867	79.057	CR1
piRNA cluster no: 50	ChrU	1871002	1881161	bi:+/-	281.131	72.389	OUTCAST
piRNA cluster no: 51	ChrU	2337286	2373725	bi:+/-	3896.112	84.816	Different TE families/classes
piRNA cluster no: 52	ChrU	2435418	2450735	bi:+/-	510.956	96.122	LINE
piRNA cluster no: 53	ChrU	9720361	9729958	bi:-/+	666.269	79.41	BEL
piRNA cluster no: 54	ChrU	18641000	18646023	bi:-/+	62.092	100	LOA
piRNA cluster no: 55	ChrU	23041532	23049039	bi:+/-	1268.927	99.947	GYPSY
piRNA cluster no: 56	ChrU	25981042	26009020	bi:-/+	8261.036	22.682	UNKNOWN
piRNA cluster no: 57	cDot	0	19022	bi:-/+	1545.939	14.783	JOCKEY
piRNA cluster no: 58	cDot	1207565	1218004	bi:-/+	258.772	82.489	UNKNOWN

Table S13. piRNA clusters in Curicó

Full Table available at: https://figshare.com/s/d33c0f4846f894e4a3be.

piRNA Cluster Name	Chr	Start	End	Direction	Mapped reads per cluster normalized by Reads Per Million (RPM) in Curicó Control Females	Total percentage of the cluster overlapping TEs	piRNA cluster annotation
piRNA cluster no: 1	ChrA	387	17500	bi:+/-	1470.913	77.28	OUTCAST
piRNA cluster no: 2	ChrA	311118	348024	bi:+/-	3069.351	75.066	Different TE families/classes
piRNA cluster no: 3	ChrA	374033	387966	bi:-/+	1066.505	99.677	UNKNOWN
piRNA cluster no: 4	ChrA	405000	454025	mono:+	3298.337	99.105	Different TE families/classes
piRNA cluster no: 5	ChrA	461000	478026	mono:-	298.454	99.742	LOA
piRNA cluster no: 6	ChrA	12711658	12721985	mono:-	278.67	46.693	GYPSY
piRNA cluster no: 7	ChrA	14905049	14921687	mono:+	2285.044	60.194	UNKNOWN
piRNA cluster no: 8	ChrA	16327999	16338004	bi:-/+	166.226	30.165	UNKNOWN
piRNA cluster no: 9	ChrA	21289110	21317024	bi:-/+	818.576	62.438	UNKNOWN
piRNA cluster no: 10	ChrA	22914999	22940839	mono:-	3327.897	57.577	UNKNOWN
piRNA cluster no: 11	ChrE	272115	279955	bi:+/-	67.613	84.936	UNKNOWN
piRNA cluster no: 12	ChrE	2663514	2673630	bi:-/+	1659.201	51.374	UNKNOWN
piRNA cluster no: 13	ChrE	16286704	16293023	bi:+/-	81.605	97.674	GYPSY
piRNA cluster no: 14	ChrJ	35869	92948	bi:-/+	31839.095	88.928	Different TE families/classes
piRNA cluster no: 15	ChrJ	98067	122755	bi:-/+	8583.224	84.515	Different TE families/classes
piRNA cluster no: 16	ChrJ	221228	229027	bi:-/+	100.413	77.946	Different TE families/classes
piRNA cluster no: 17	ChrJ	250104	267008	bi:-/+	478.381	64.559	Different TE families/classes
piRNA cluster no: 18	ChrJ	325051	330714	bi:-/+	84.057	81.565	UNKNOWN
piRNA cluster no: 19	ChrJ	424083	430018	bi:+/-	67.942	95.282	HELITRON
piRNA cluster no: 20	ChrJ	431635	446903	bi:-/+	161.578	53.622	Different TE families/classes
piRNA cluster no: 21	ChrJ	452226	478026	mono:-	6017.375	92.031	Different TE families/classes
piRNA cluster no: 22	ChrJ	677167	685697	mono:-	613.562	67.327	Different TE families/classes
piRNA cluster no: 23	ChrJ	868909	882569	bi:-/+	259.59	69.07	UNKNOWN
piRNA cluster no: 24	ChrJ	1448022	1492713	bi:-/+	7200.681	73.722	UNKNOWN
piRNA cluster no: 25	ChrJ	2702341	2710021	mono:+	107.162	99.909	GYPSY
piRNA cluster no: 26	ChrJ	4294339	4306508	mono:-	516.486	81.765	GYPSY
piRNA cluster no: 27	ChrJ	7575023	7587017	bi:-/+	279.883	77.164	GYPSY
piRNA cluster no: 28	ChrJ	7753003	7758026	mono:+	58.961	100	GYPSY
piRNA cluster no: 29	ChrJ	19102181	19122310	bi:+/-	4029.131	84.306	POLINTON
piRNA cluster no: 30	ChrJ	21664010	21672479	mono:+	144.46	100	GYPSY
piRNA cluster no: 31	ChrO	102024	117906	mono:-	737.435	59.445	UNKNOWN
piRNA cluster no: 32	ChrO	488000	498920	bi:+/-	373.022	66.987	Different TE families/classes
piRNA cluster no: 33	ChrO	537184	573666	bi:+/-	3784.853	57.327	UNKNOWN
piRNA cluster no: 34	ChrO	614129	637776	mono:-	7620.927	65.361	UNKNOWN
piRNA cluster no: 35	ChrO	681099	691332	bi:-/+	203.914	81.941	Different TE families/classes
piRNA cluster no: 36	ChrO	2262070	2272021	bi:-/+	150.419	58.286	HELITRON
piRNA cluster no: 37	ChrO	3607387	3613082	mono:-	299.681	83.09	GYPSY
piRNA cluster no: 38	ChrO	10784334	10789680	mono:+	124.297	100	R1
piRNA cluster no: 39	ChrO	12780221	12797147	bi:-/+	3522.611	15.272	UNKNOWN
piRNA cluster no: 40	ChrO	15758186	15767520	mono:-	917.978	76.077	GYPSY
piRNA cluster no: 41	ChrO	16848857	16855999	bi:+/-	114.726	46.36	UNKNOWN
piRNA cluster no: 42	ChrO	27064086	27098026	bi:+/-	9470.18	77.917	GYPSY
piRNA cluster no: 43	ChrO	29824999	29831012	bi:-/+	77.144	99.983	GYPSY
piRNA cluster no: 44	ChrO	30269170	30283473	bi:-/+	437.951	63.693	UNKNOWN
piRNA cluster no: 45	ChrU	5000	23007	bi:+/-	4050.44	56.584	GYPSY
piRNA cluster no: 46	ChrU	462069	469025	mono:-	72.397	79.14	OUTCAST
piRNA cluster no: 47	ChrU	498686	503957	bi:-/+	2244.903	77.196	CR1
piRNA cluster no: 48	ChrU	720000	726694	mono:+	84.351	91.829	LOA
piRNA cluster no: 49	ChrU	1209046	1223023	mono:+	1141.039	71.489	UNKNOWN
piRNA cluster no: 50	ChrU	1247780	1254993	mono:+	418.301	81.866	LOA
piRNA cluster no: 51	ChrU	1477074	1484007	bi:-/+	94.819	75.682	CR1
piRNA cluster no: 52	ChrU	1491055	1502254	mono:+	2569.533	66.229	UNKNOWN
piRNA cluster no: 53	ChrU	1594896	1604591	bi:-/+	133.737	80.01	Different TE families/classes
piRNA cluster no: 54	ChrU	1824509	1832337	bi:-/+	119.525	77.836	UNKNOWN
piRNA cluster no: 55	ChrU	1869311	1889564	mono:+	847.484	60.189	LINE
piRNA cluster no: 56	ChrU	1992002	1999989	bi:-/+	91.938	95.317	GYPSY
piRNA cluster no: 57	ChrU	2339000	2349008	bi:+/-	119.152	88.419	LOA
piRNA cluster no: 58	ChrU	2355068	2372747	bi:+/-	1433.115	81.232	Different TE families/classes
piRNA cluster no: 59	ChrU	2426002	2450022	bi:+/-	1121.215	84.088	LINE
piRNA cluster no: 60	ChrU	3044050	3053975	mono:+	139.208	100	GYPSY
piRNA cluster no: 61	ChrU	4079001	4085726	bi:+/-	87.832	100	GYPSY
piRNA cluster no: 62	ChrU	6825471	6832387	mono:+	1029.744	65.905	UNKNOWN
piRNA cluster no: 63	ChrU	7281121	7287762	mono:+	889.943	75.636	BEL
piRNA cluster no: 64	ChrU	12794719	12801845	bi:+/-	254.067	100	POLINTON
piRNA cluster no: 65	ChrU	12825999	12832022	mono:-	64.219	100	GYPSY
piRNA cluster no: 66	ChrU	15933390	15938630	mono:-	222.553	99.847	GYPSY
piRNA cluster no: 67	ChrU	21384070	21389277	bi:+/-	131.483	94.603	GYPSY
piRNA cluster no: 68	ChrU	23038086	23049040	mono:+	328.701	68.925	GYPSY
piRNA cluster no: 69	cDot	0	15024	mono:-	9081.816	18.717	JOCKEY
piRNA cluster no: 70	cDot	972143	979428	mono:+	230.624	89.595	UNKNOWN
piRNA cluster no: 71	cDot	1138125	1145022	bi:+/-	92.625	90.88	LOA
piRNA cluster no: 72	cDot	1250301	1269758	mono:-	1086.592	61.762	GYPSY
piRNA cluster no: 73	cDot	1377565	1410011	bi:-/+	14146.621	32.633	UNKNOWN

Table S14. Wilcoxon signed rank test for all piRNA Clusters

Comparison of the general production of piRNA clusters before and after a heat shock, using two-tailed and one-tailed tests.

Sample	p-value (two-tailed)	Adjusted p-value	More expression after a heat shock (one-tailed) - Adjusted p-value	Less expression after a heat shock (one-tailed) - Adjusted p-value
Curicó females	0.000	**8.88E-05 ***	1	**4.44E-05 ***
Madeira females	0.398	0.530	1	0.398
Curicó males	0.314	0.530	0.627	0.845
Madeira males	0.816	0.816	0.816	0.796

***: p < 0.001

Table S15. Wilcoxon signed rank test for piRNA clusters separated by their production

Comparison of the piRNA cluster production in heat stress vs control. Clusters producing more piRNAs after a heat shock are separated from those producing less.

Sample	Hypotesis using one-tailed Wilcoxon test	Number of clusters	p-value (one-tailed)	Adjusted p-value
Curicó females	More expression after a Heat	25	2.98E-08	**4.768E-08 ***
	Less expression after a Heat	60	8.357E-12	**2.229E-11 ***
Madeira females	More expression after a Heat	39	1.819E-12	**7.276E-12 ***
	Less expression after a Heat	46	1.421E-14	**1.137E-13***
Curicó males	More expression after a Heat	31	4.657E-10	**9.314E-10***
	Less expression after a Heat	20	9.537E-07	**1.090E-06***
Madeira males	More expression after a Heat	22	2.384E-07	**3.179e-07 ***
	Less expression after a Heat	18	3.815E-06	**3.815E-06***

***: p < 0.001

Table S16. T-test for small RNA read counts

Comparison of small RNA read counts in heat shock vs control conditions.

	Comparison	Condition and replicate	Total normalized reads	Associated p-value	Adjusted p-value
siRNAs	Curicó females	Heat shock 1	506053.09	0.118	0.157
		Heat shock 2	506588.63		
		Control 1	489707.00		
		Control 2	481957.17		
	Madeira females	Heat shock 1	392677.38	0.461	0.461
		Heat shock 2	370581.91		
		Control 1	371361.80		
		Control 2	367255.14		
	Curicó males	Heat shock 1	868135.14	0.104	0.157
		Heat shock 2	915825.10		
		Control 1	750149.00		
		Control 2	747707.87		
	Madeira males	Heat shock 1	601671.72	0.093	0.157
		Heat shock 2	598152.85		
		Control 1	573924.98		
		Control 2	562274.10		

	Curicó females	Heat shock 1	8233533.67	0.632	0.632
		Heat shock 2	6130385.64		
		Control 1	7579927.00		
		Control 2	8152679.32		
	Madeira females	Heat shock 1	8183582.54	0.547	0.632
		Heat shock 2	8439096.70		
		Control 1	8364581.35		
		Control 2	7654410.71		
piRNAs	Curicó males	Heat shock 1	1987801.78	0.174	0.632
		Heat shock 2	2158443.40		
		Control 1	1533888.00		
		Control 2	1825099.16		
	Madeira males	Heat shock 1	1265373.50	0.540	0.632
		Heat shock 2	975792.53		
		Control 1	1028409.12		
		Control 2	959887.46		

Table S17. Fisher's exact test for TE expression and small RNA changes

Comparison of changes in small RNA amounts and changes and TE expression after heat stress.

	Comparison	More small RNA amounts after a heat shock (HS)	Less small RNA amounts after HS	TE expression	Associated p-value	Adjusted p-value
siRNAs	Curicó females	213	129	Overexpressed after a HS	0.494	0.659
		168	90	Underexpressed after a HS		
	Madeira females	203	112	Overexpressed after a HS	0.862	0.862
		172	92	Underexpressed after a HS		
	Curicó males	429	65	Overexpressed after a HS	0.209	0.418
		111	24	Underexpressed after a HS		
	Madeira males	253	114	Overexpressed after a HS	0.036	0.143
		191	58	Underexpressed after a HS		
piRNAs	Curicó females	79	222	Overexpressed after a HS	0.415	0.748
		51	170	Underexpressed after a HS		
	Madeira females	148	141	Overexpressed after a HS	0.601	0.748
		128	111	Underexpressed after a HS		
	Curicó males	355	73	Overexpressed after a HS	**0.000**	**0.002 *****
		70	35	Underexpressed after a HS		
	Madeira males	215	77	Overexpressed after a HS	0.748	0.748
		141	46	Underexpressed after a HS		

***: $p < 0.001$

Table S18. Chi-square for TE families with Ping-Pong signal

Comparison of TE families with and without Ping-Pong signal before and after heat stress.

| Comparison | Number of TE families | | Conditions | Associated p-value | Adjusted p-value |
	Without Ping-Pong signal	With Ping-Pong signal			
Curicó females	78	72	Control	0.729	0.979
	75	75	Heat Shock		
Madeira females	74	82	Control	0.734	0.979
	77	79	Heat Shock		
Curicó males	78	64	Control	0.720	0.979
	81	61	Heat Shock		
Madeira males	71	64	Control	1.000	1.000
	71	64	Heat Shock		

Table S19. T-test of miRNA read counts

Comparisons of miRNA read counts in heat shocked samples vs controls.

Comparison	Condition and replicate	Normalized reads (RPM)	Associated p-value	Adjusted p-value
Curicó females	Heat shock 1	2587501	0.529	0.705
	Heat shock 2	6188782		
	Control 1	2635360		
	Control 2	2856317		
Madeira females	Heat shock 1	2476361	0.436	0.705
	Heat shock 2	3132178		
	Control 1	2632168		
	Control 2	2185291		
Curicó males	Heat shock 1	1079012	0.259	0.705
	Heat shock 2	787832		
	Control 1	1206311		
	Control 2	1287477		
Madeira males	Heat shock 1	1332911	0.823	0.823
	Heat shock 2	1076545		
	Control 1	1302219		
	Control 2	1182767		

8.2.2 Supplementary Figures and text

Fig. S1.

Chromosomal inversions detected in the Madeira (A) and Curicó (B) populations.
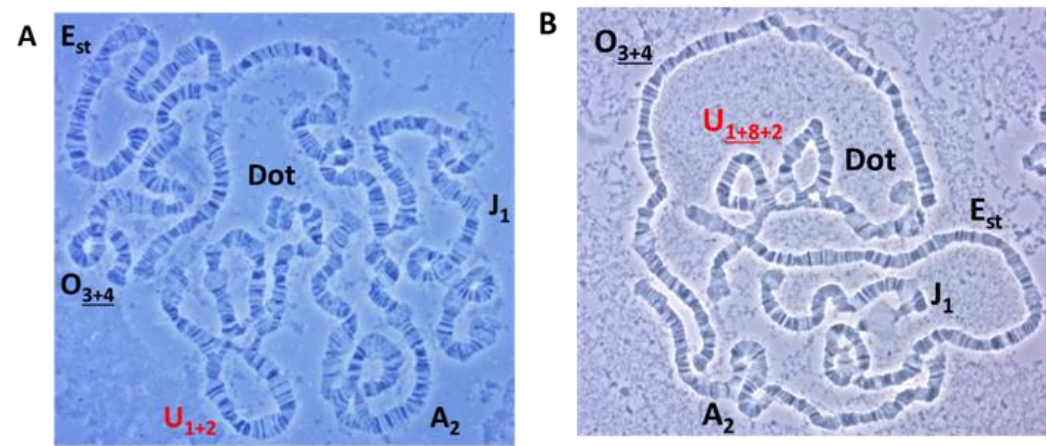
Fig. S2.

Enriched Gene Ontologies (GO) when populations are compared. Pie plots showing the proportion of enriched GO of genes more expressed in each population and sex.
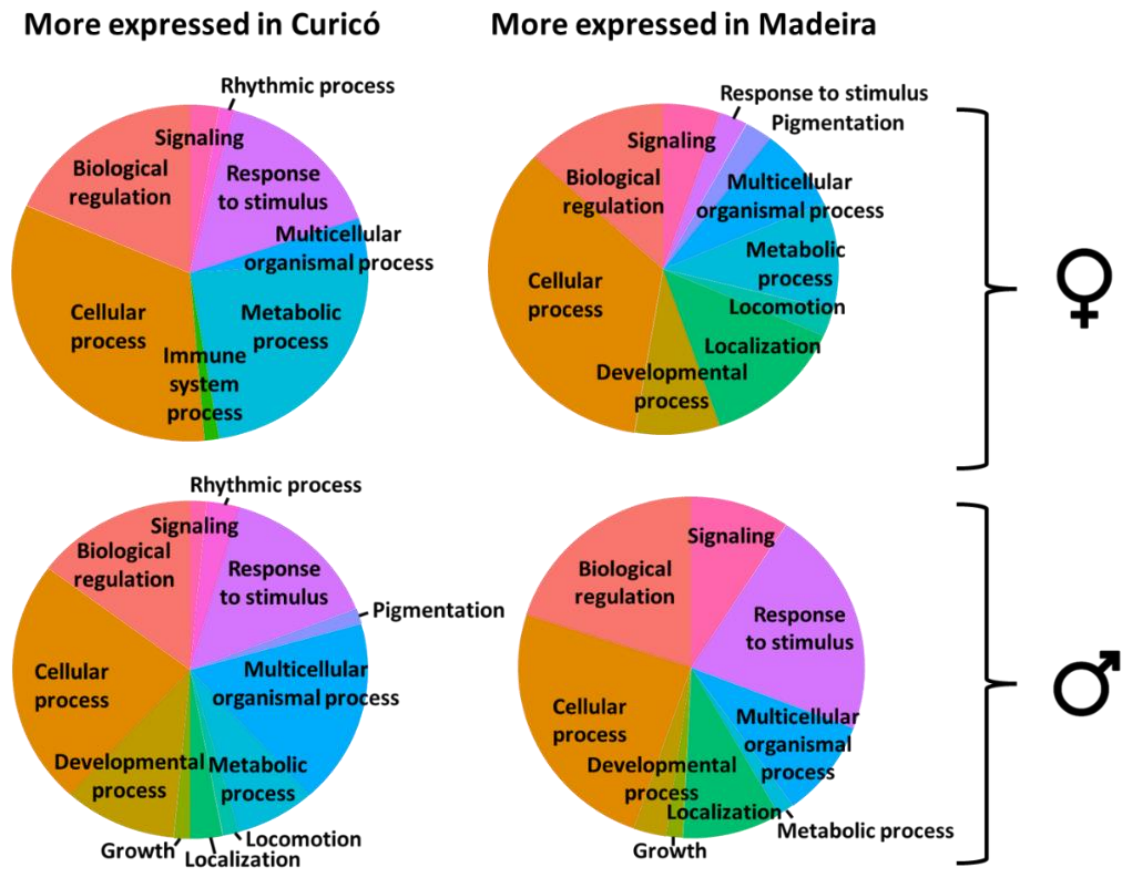
Fig. S3.

Annotation of TE families in the reference genome of *D. subobscura*. (A) Percentage of the genome covered by different TE classes. (B) Number of insertions per TE superfamily, excluding the unknown.
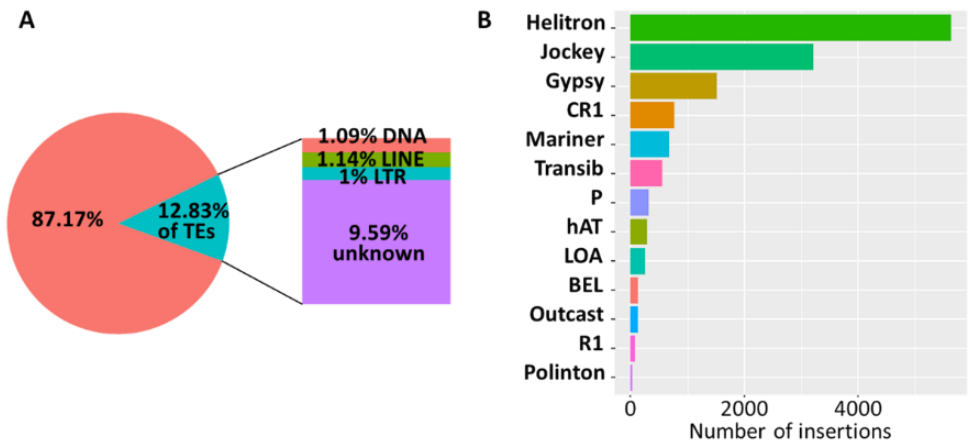
Fig. S4.

Enriched Gene Ontologies (GO) in heat shock vs control conditions. Pie plots showing the proportion of enriched GO of genes over and underexpressed by sex in both populations after a heat stress.
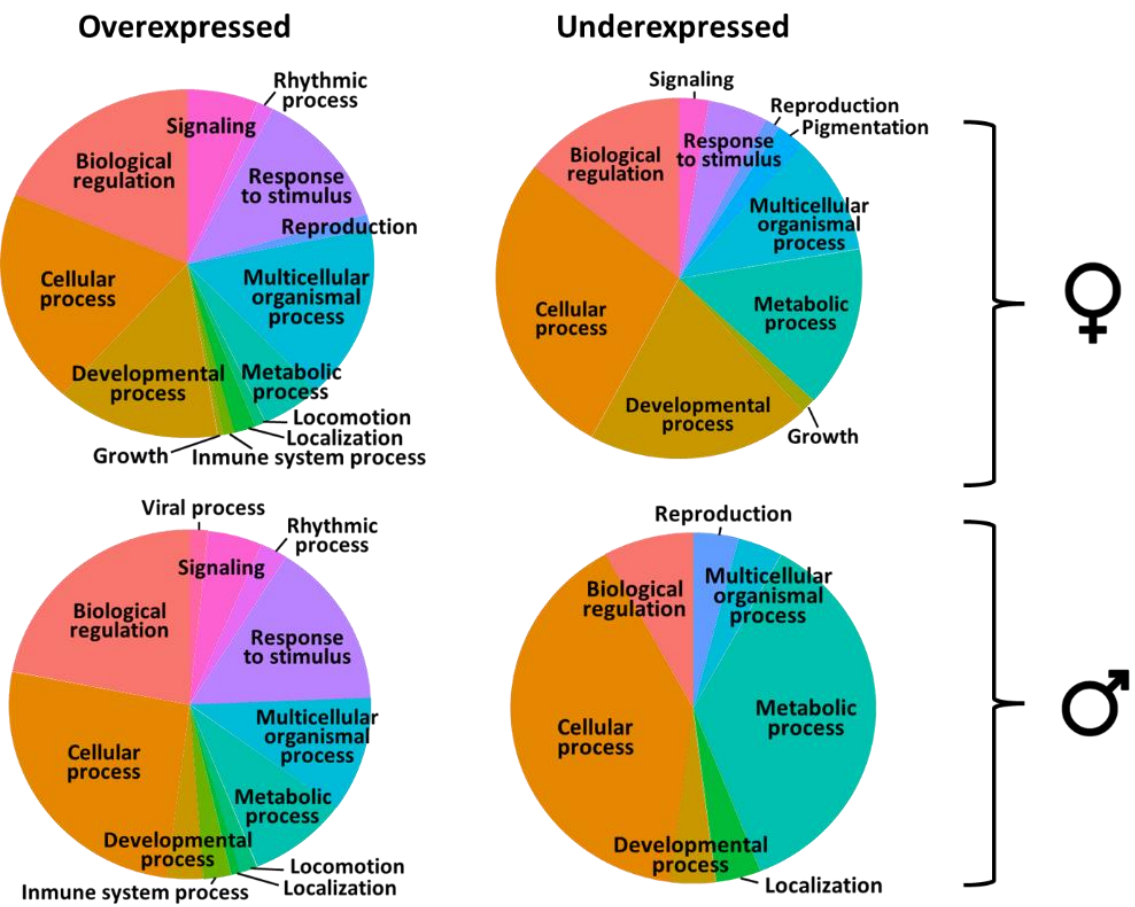
Fig. S5.

Differential expression values of TEs and siRNA amounts in heat shock vs control samples. Fisher's exact test p-values are shown.
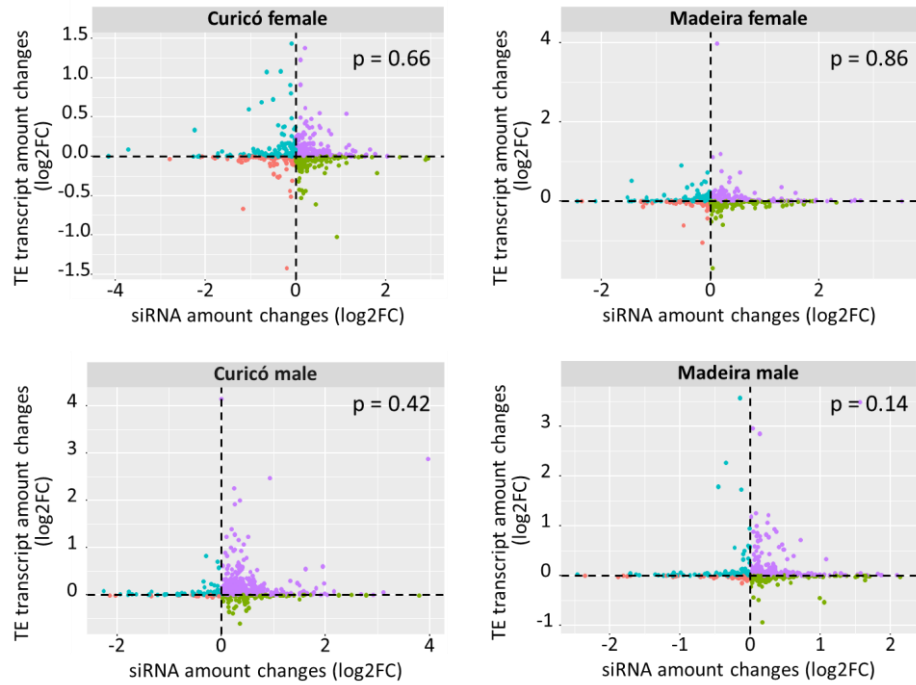
Fig. S6.

Differential expression values of TEs and piRNA amounts in heat shock vs control samples. Fisher's exact test p-values are shown.
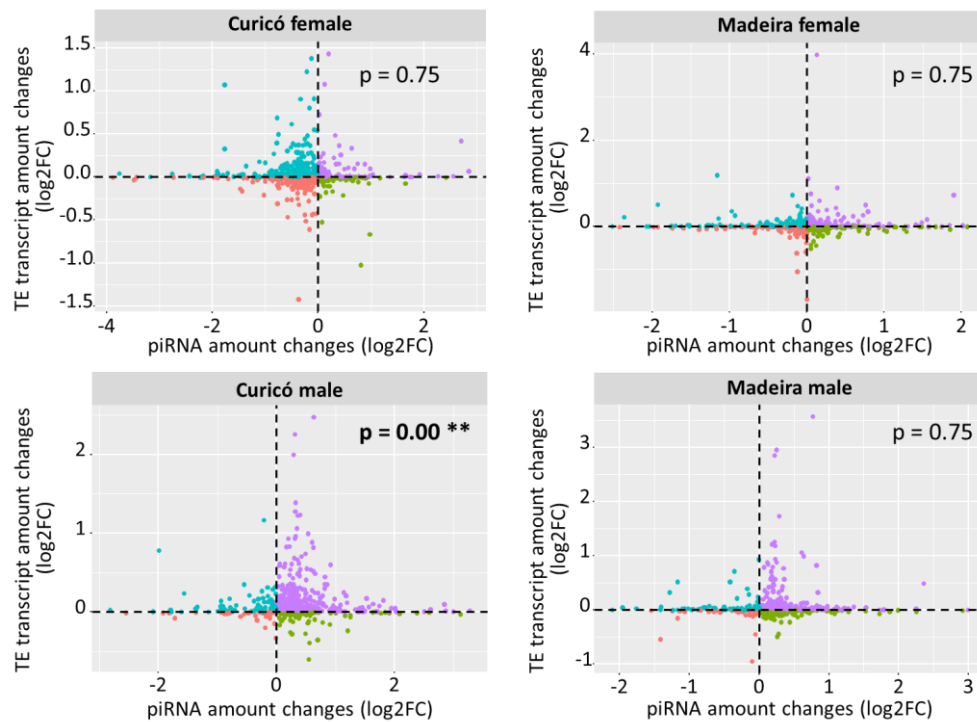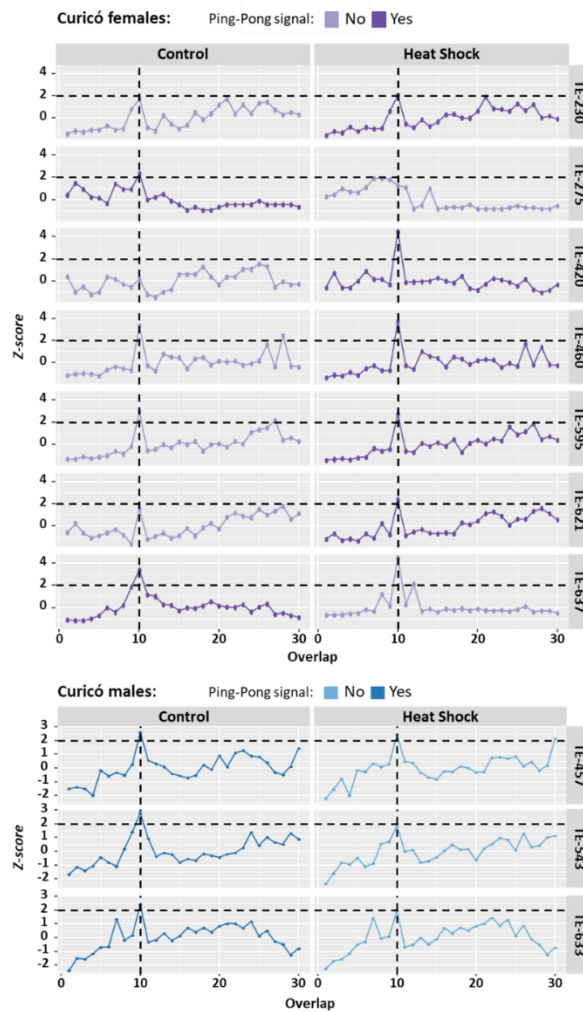
Fig. S7.

Ping-Pong signal z-scores. Ping-Pong signal z-scores for all overlaps of TE families with changes in their Ping-Pong signal in control vs heat shock conditions in the Curicó population: females in purple and males in blue.

Fig. S8.

Ping-Pong signal z-scores. Ping-Pong signal z-scores for all overlaps of TE families with changes in their Ping-Pong signal in control vs heat shock conditions in the Madeira population: females in green and males in red.
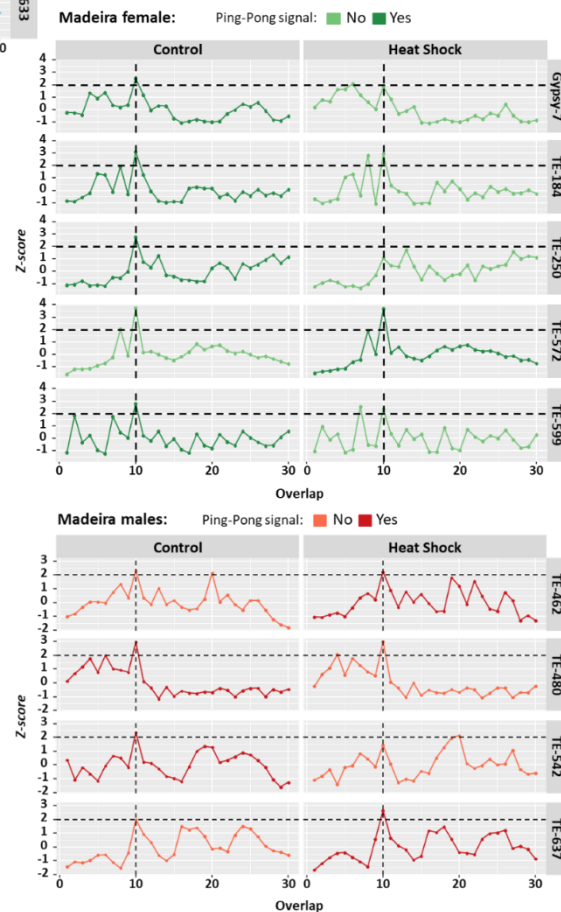
Supplementary text

Text S1. *Hsp83* gene sequence in Madeira and Curicó. Alignment between *Hsp83* gene sequence in Madeira and Curicó and study of the important regions:

- Green: insertion of *TE-622*
- Blue: insertion of *TE-500*
- Pink: insertion of *TE-584* (LTR/Copia)
- Yellow: Heat Shock Elements (HSE), detected using the HSE sequences described in *D. pseudoobscura* and *D. persimilis*, extracted from: [398].
- Grey: TATA box
- Purple: start codon
- Red: Stop codon

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 10         20         30         40         50         60         70         80         90        100
Madeira   TCACAAACAG ATTTTAAGAT ACTTAAGTTG TGTGAAAAAA GAAACCGTTA CCTCTGAAAA AACCTTAAAA AAAATACCAG ATTTTAAAAA CTCTTAAAAA
Curicó    TCACAAACAG ATTTTAAGAT ACTTAAGTTG TGTGAAAAAA GAAACCGTTA CCTCTGAAAA AACCTTAAAA AAAATACCAG ATTTTAAAAA CTCTTAAAAA

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                110        120        130        140        150        160        170        180        190        200
Madeira   AAAGTTCATA AAAAATGTAA TTCCCGAAAT CTATCAGAAA AGGTCGGTAT GGTTCCTTAT ATCCCATTAAT TATCCATTAA TGCGGTGTCA TAATAAATTT
Curicó    AA-GTTCATA AAAAATGTAA TTCCCGAAAT CTATCAGAAA AGGTCGGTAT GGTTCCTTAT ATCCCATTAAT TATCCATTAA TGCGGTGTCA TAATAAATTT

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                210        220        230        240        250        260        270        280        290        300
Madeira   TCATTTTAGA CGGGTTTCAT GTTGTAAGGA AACACAAAAC AGTGTACCCA GAGAGCATGT AACTTGCCCA CTTTGGTCAT ACTGTGTACA AACCGAAATT
Curicó    TCATTTTAGA CGGGTTTCAT GTTGTAAGGA AACACAAAAC AGTGTACCCA GAGAGCATGT AACTTGCCCA CTTTGGTCAT ACTGTGTACA AACCGAAATT

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                310        320        330        340        350        360        370        380        390        400
Madeira   TACGGCACCT TCTTTCGCAG AAAAAATAAG CGAATCTGAG TACTGGGTTT TAAGCTTTTT TTGTGACCTT TCTGGCCTGA ACCTTATTTT GTAGGCATAT
Curicó    TACGGCACCT TCTTTCGCAG AAAAAATAAG CGAATCTGAG TACTGGGTTT TAAGCTTTTT TTGTGACCTT TCTGGCCTGA ACCTTATTTT GTAGGCATAT

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                410        420        430        440        450        460        470        480        490        500
Madeira   CCAATGAAAA AAATCATGAA ATGTGGCCAA TCTTGGGCAC AAGCTATTCA ACAATGGGAT AAAATTATTT TTTAATGGCT TA-------- ----------
Curicó    CCAATGAAAA AAATCATGAA ATGTGGCCAA TCTTGGGCAC AAGTGATTCA ACAATGGGAT AAAACTATTT TTTAATGGCT TAGAGACCTA TCCAGCAGGT

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                510        520        530        540        550        560        570        580        590        600
Madeira   -ATGGTTAAG CGAACATCAA TATCGAGGAA TATGTTTTAT TTCGGTATAC TTTCAAAATG AGATTGGTAT ATTTCTGAGG GTCACACTAG CCTCGGAGCA
Curicó    AATGTTGAAG CGAACATCAA TATCGAGGAA TATGTTTTAT TTCGGTATAC TTTCAAAATG AGATTGGTAT ATTTCTGAGG GTCACACTAG CCTCGGAGCA

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                610        620        630        640        650        660        670        680        690        700
Madeira   ACTCTGTGCG AGCACCGGAA GTAGCAGCCC TGGATTCTCG TAGCCTCTAG AAACGTCTAG AACATTCTAC GCTTGGGGTT GATTTTCTAT AAAAGCAGGC
Curicó    ACTCTGTGCG AGCACCGGAA GTAGCAGCCC TGGATTCTCG TAGCCTCTAG AAACGTCTAG AACATTCTAC GCTTGGGGTT GATTTTCTAT AAAAGCAGGC

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                710        720        730        740        750        760        770        780        790        800
Madeira   TCGAGTAAAA TTTTCGTTCA GTGAAAAAAG TGCGTTGCGT AGCGCAGTGT AAAAGTGAAT TTATTCTACA CAAATCGAAG TGAAAATACA TATAATTGTA
Curicó    TCGAGTAAAA TTTTCGTTCA GTGAAAAAAG TGCGTTGCGG AGCGCAGTGT AAAAGTGAAT TTATTCTACA CAAATCGAAG TGAAAATACA TATAATTGTA

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                810        820        830        840        850        860        870        880        890        900
Madeira   TCTCTTTGCT GTGAAATAAA ACACATTCAA GGTAAGCTGT TGTAATGAGA GTGCATTTAC TTAACAAAAC GTAGAGATGT GCGGTGTGCA GTGCTCGCGG
Curicó    TCTCTTTGCT GTGAAATAAA ACACATTCAA GGTAAGCTGT TGTAATGAGA GTGCATTTAC TTAACAAAAC GTAGAGATGT GCGGTGTGCA GTGCTCGCGG

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                910        920        930        940        950        960        970        980        990       1000
Madeira   AGCGTAGGCG GCTGAAAAGT TCTTGACCCA AATGCAGAAA ATCAATAGAA TGTTTGAGAT TCCTTCCAAG ATCTTATAAT CGGTAAACGC CTGTTTTCCG
Curicó    AGCGTAGGCG GCTGAAAAGT TCTTGACCCA AATGCAGAAA ATCAATAGAA TGTTTGAGAT TCCTTGCAAG ATCTTATAAT CGGTAAACGC CTGTTTTCCG

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
Madeira   CGAAATTATA ACCACAAGCG GTAAAAAAGT AACAAAGTCT CGCACACATG CAAACATACA CACGGACACT GAAAAATTAT GCCGGTAGTA ATTTGTGCTC
Curicó    CGAAATTATA ACCACAAGCG GTAAAAAAGT AACAAAGTCT CGCACACATG CAAACATACA CACGGACACT GAAAAATTAT GCCGGTAGTA AATTGTGCTC

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
Madeira   TGTGCGAATT TTCTAGAAGA CCGGTTTTCG GACATTCAAC CATGCCCAAC CAATTCTGTT TGACAATTTG CACAACTTAA TTTTCTTGGC ACACGTGCCT
Curicó    TGCGCGAATT TTCTAGAAGA CCGGTTTTTG GACATTCAAC CATGCCCAAC CAATTCTGTT TGACCATTTG CACAACTTAA TTTTCTTGGC ACACGTGCCT
```

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1210       1220       1230       1240       1250       1260       1270       1280       1290      1300
Madeira   TTTTT-CTCA TTTAAATTCA GTCTTTTGCA CAAGGCATAC ACACACATGC AGCTCATACA TACAGTCATG GATGGCCACA ATCGAGAAAA TTCGAAAAAC
Curicó    TTTTTTCTCA TTTAAATTCA GTCTTTTGCA CAAGGCATAC ACACACATGC AGCTCATACA TACAGGCATG GATGGCCACA ATCGAGAAAA TTCGAAAAAC


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1310       1320       1330       1340       1350       1360       1370       1380       1390      1400
Madeira   GTGAATCAAA CTTGGGGCAT GAATCGAATA TCTAAGAGGT GTCATCGGAG AGGCTTAGTT TCCTTCTGGT TCTAGAAGTT TCCATTGGCA TTCTGGGATT
Curicó    GTGAATCAAA CTTGGGGCAT GAATCGAATA TCTAAGAGGT GTCATCGGAG AGGCTTAGTT TCCTTCTGGT TCTAGAAGTT TCCATTGGCA TTCTGGGATT


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1410       1420       1430       1440       1450       1460       1470       1480       1490      1500
Madeira   CCTAATGCAC ATGTACTCGA CCACAGCCAC ACATGAATGT GCCATACAAA ATGTACAGAA TACCTTTTTC GTCAAGGAGA GTTTTTCTTC CAGAAAGTTT
Curicó    CCTAATGCAC ATGTACTCGA CCACAGCCAC ACATGAATGT GCCATGCAAA ATGTACAGAA TACCTTTTTC GTCAAGGAGA GTTTTTCTTC CAGAAAGTTT


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1510       1520       1530       1540       1550       1560       1570       1580       1590      1600
Madeira   CAATTAGAGT GGTTTTTCGG GGATGAGGAC AACCCCATGA GTCTAGAAAC TTCCCATACA ATGGGTACAT CCCTATGCGT GCGTGTGTAT GCGTGTTCGC
Curicó    CAATTAGAGT GGTTTTTCGG GGATGAGGAC AACCCCATGA GTCTAGAAAC TTCCCATACA ATGGGTACAT CCCTATGCGT GCGTGTGTAT GCGTGTTCGC


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1610       1620       1630       1640       1650       1660       1670       1680       1690      1700
Madeira   CAGCTGTAAT GTATGTGTTA GTGTGCGAAA GAGAACAAAG AATGAGGCGT CTGCCATTTT GAAGTAAAAA CATTTTTGCA TTTGGTGGAA ACACTTTTTC
Curicó    CAGCTGTAAT GTATGTGTTA GTGTGCGAAA GAGAACAAAG AATGAGGCGT CTGCCATTTT GAAGTAAAAA CATTTTTGCA TTTGGTGGAA ACACTTTTTC


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1710       1720       1730       1740       1750       1760       1770       1780       1790      1800
Madeira   TCGAAAATAG TCAGCTGCAT AAGCAATGAA ATGTGACTAA TTCTGGTATG GTATTTTCTT GCTCTTCCAG ATGCCCGAAG AAGCTGAGAC TTTCGCATTC
Curicó    TCGAAAATAG TCAGCTGCAT AAGCAATGAA ATGTGACTAA TTCTGGTATG GTATTTTCTT GCTCTTCCAG ATGCCCGAAG AAGCTGAGAC TTTCGCATTC


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1810       1820       1830       1840       1850       1860       1870       1880       1890      1900
Madeira   CAGGCTGAGA TTGCTCAGCT TATGTCGTTG ATCATCAACA CATTCTATTC GAACAAGGAG ATCTTCTTGC GTGAATTGAT TTCGAACGCA TCTGATGCCC
Curicó    CAGGCTGAGA TTGCTCAGCT TATGTCGTTG ATCATCAACA CATTCTATTC GAACAAGGAG ATCTTCTTGC GTGAATTGAT TTCGAACGCA TCTGATGCCC


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1910       1920       1930       1940       1950       1960       1970       1980       1990      2000
Madeira   TCGACAAGAT CCGCTATGAG TCGCTGACGG ATCCCAGCAA GCTCGACTCG GGAAAGGAGC TGTACATCAA GCTGATTCCC AACAAGACGG CTGGTACTCT
Curicó    TCGACAAGAT CCGCTATGAG TCGCTGACGG ATCCCAGCAA GCTCGACTCG GGAAAGGAGC TGTACATCAA GCTGATTCCC AACAAGACGG CTGGTACTCT


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2010       2020       2030       2040       2050       2060       2070       2080       2090      2100
Madeira   GACCATCATT GATACCGGTA TTGGCATGAC CAAGTCCGAC TTGGTTAACA ACTTGGGAAC CATCGCCAAG TCCGGCACCA AGGCTTTCAT GGAGGCATTG
Curicó    GACCATCATT GATACCGGTA TTGGCATGAC CAAGTCCGAC TTGGTTAACA ACTTGGGAAC CATCGCCAAG TCCGGCACCA AGGCTTTCAT GGAGGCATTG


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2110       2120       2130       2140       2150       2160       2170       2180       2190      2200
Madeira   CAGGCTGGTG CTGACATTTC CATGATTGGC CAATTCGGCG TGGGCTTCTA CTCGGCCTAC CTGATTGCCG ACCGTGTTAC TGTCACCTCG AAGAACAACG
Curicó    CAGGCTGGTG CTGACATTTC CATGATTGGC CAATTCGGCG TGGGCTTCTA CTCGGCCTAC CTGATTGCCG ACCGTGTTAC TGTCACATCG AAGAACAACG


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2210       2220       2230       2240       2250       2260       2270       2280       2290      2300
Madeira   ATGATGAGCA GTACGTCTGG GAGTCGTCCG CCGGCGGCTC GTTCACCGTG AGGGCTGACA ACTCAGAGCC CCTAGGTCGC GGCACAAAGA TCGTGCTCTA
Curicó    ATGATGAGCA GTACGTCTGG GAGTCGTCCG CCGGCGGCTC GTTCACCGTG AGGGCCGACA ACTCAGAGCC CCTAGGTCGC GGCACAAAGA TCGTGCTCTA


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2310       2320       2330       2340       2350       2360       2370       2380       2390      2400
Madeira   CATCAAGGAG GATCAGACCG ATTACCTTGA GGAGAGCAAG ATTAAGGAAA TCGTTAACAA GCACTCCCAG TTCATTGGGT ATCCCATCAA ACTGCTGGTG
Curicó    CATCAAGGAG GATCAGACCG ATTACCTTGA GGAGAGCAAG ATTAAGGAAA TCGTTAACAA GCACTCCCAG TTCATTGGGT ATCCCATCAA ACTGCTGGTG


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2410       2420       2430       2440       2450       2460       2470       2480       2490      2500
Madeira   GAGAAGGAGC GCGAGAAGGA GGTCAGCGAC GATGAGGCTG ATGATGAGAA GAAGGATGAG GAGGTCAAGA AGGACATGGA CACCGATGAG CCCAAGATCG
Curicó    GAGAAGGAGC GCGAGAAGGA GGTCAGCGAC GATGAGGCTG ATGATGAGAA GAAGGATGAG GAGGTCAAGA AGGACATGGA CACCGATGAG CCCAAGATCG


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2510       2520       2530       2540       2550       2560       2570       2580       2590      2600
Madeira   AAGATGTCGG CGAGGATGAG GATGCCGACA AGAAGGACAA GGATGGCAAG AAGAAGAAGA CCATCAAGGA GAAGTACACT GAAGACGAGG AGCTGAACAA
Curicó    AAGATGTCGG CGAGGATGAG GATGCCGACA AGAAGGACAA GGATGGCAAG AAGAAGAAGA CCATCAAGGA GAAGTACACT GAAGACGAGG AGCTGAACAA


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2610       2620       2630       2640       2650       2660       2670       2680       2690      2700
Madeira   GACCAAGCCA ATCTGGACCC GCAACCCCGA TGATATCTCC CAGGAGGAGT ACGGCGAGTT CTACAAGTCC CTGACCAACG ACTGGGAGGA TCATCTGTGT
Curicó    GACCAAGCCA ATCTGGACCC GCAACCCCGA TGATATCTCC CAGGAGGAGT ACGGCGAGTT CTACAAGTCC CTGACCAACG ACTGGGAGGA TCATCTGTGT


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2710       2720       2730       2740       2750       2760       2770       2780       2790      2800
Madeira   GTCAAGCACT TCTCGGTCGA GGGTCAGTTG GAGTTCCGCG CACTCCTCTT CATTCCCCGT CGCACCCCCT TCGATCTTTT TGAGAACCAG AAGAAGCGCA
Curicó    GTCAAGCACT TCTCGGTCGA GGGTCAGTTG GAGTTCCGCG CACTCCTCTT CATTCCCCGT CGCACCCCCT TCGATCTTTT TGAGAACCAG AAGAAGCGCA


          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2810       2820       2830       2840       2850       2860       2870       2880       2890      2900
Madeira   ACAACATCAA GCTGTACGTG CGCCGCGTGT TCATCATGGA CAACTGCGAG GATCTCATTC CCGAGTACTT GAACTTTATC AAGGGAGTGG TCGACTCTGA
Curicó    ACAACATCAA GCTGTACGTG CGCCGCGTGT TCATCATGGA CAACTGCGAG GATCTCATTC CCGAGTACTT GAACTTTATC AAGGGAGTGG TCGACTCTGA
```
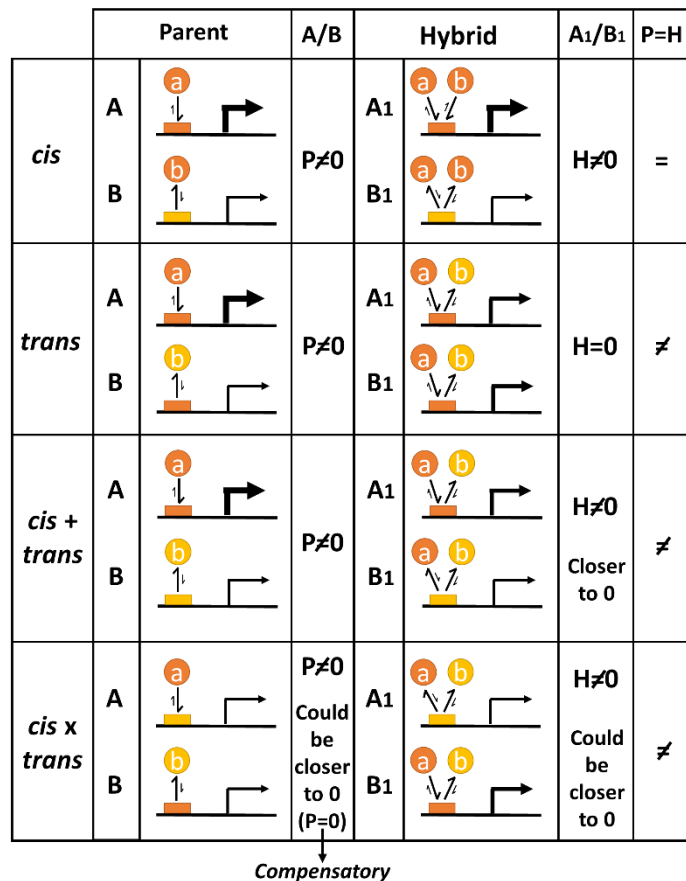
200

```
              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
Madeira       GGATCTGCCT CTGAACATCT CCCGTGAGAT GTTGCAGCAG AACAAAGTTC TGAAGGTGAT CCGCAAGAAT CTGGTGAAGA AGACCATGGA GCTGATCGAG
Curicó        GGATCTGCCC CTGAACATCT CCCGTGAGAT GTTGCAGCAG AACAAAGTTC TGAAGGTGAT CCGCAAGAAT CTGGTGAAGA AGACCATGGA GCTGATCGAG

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3010       3020       3030       3040       3050       3060       3070       3080       3090       3100
Madeira       GAGCTCACCG AGGACAAGGA GAACTACAAA AAGTTCTACG ACCAGTTCAG CAAGAACTTG AAACTGGGTG TCCACGAGGA CAGCAACAAC CGCGCCAAGC
Curicó        GAGCTCACCG AGGACAAGGA GAACTACAAA AAGTTCTACG ACCAGTTCAG CAAGAACTTG AAACTGGGTG TCCACGAGGA CAGCAACAAC CGCGCCAAGC

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
Madeira       TCGCCGATTT TCTGCGCTTC CACACTTCGG CCTCTGGTGA CGATTTCTGC TCCCTGTCGG ACTATGTGTC CCGCATGAAG GAGAACCAGA AGCACGTCTA
Curicó        TCGCCGATTT TCTGCGCTTC CACACTTCGG CCTCTGGTGA CGATTTCTGC TCCCTGTCGG ACTATGTGTC CCGCATGAAG GAGAACCAGA AGCACGTCTA

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3210       3220       3230       3240       3250       3260       3270       3280       3290       3300
Madeira       CTTCATCACT GGCGAGTCCA AGGACCAGGT CAGCAACTCT GCCTTCGTTG AGCGTGTCAA GGCCCGCGGC TTTGAGGTTG TCTACATGAC CGAGCCCATT
Curicó        CTTCATCACT GGCGAGTCCA AGGACCAGGT CAGCAACTCT GCCTTCGTTG AGCGTGTCAA GGCCCGCGGC TTTGAGGTTG TCTACATGAC CGAGCCCATT

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3310       3320       3330       3340       3350       3360       3370       3380       3390       3400
Madeira       GATGAGTATG TCATCCAGCA CTTGAAGGAG TACAAGGGAA AGCAGCTGGT GTCCGTCACC AAGGAGGGTC TGGAGCTGCC AGAGGATGAT GCCGAGAAGA
Curicó        GATGAGTATG TCATCCAGCA CTTGAAGGAG TACAAGGGAA AGCAGCTGGT GTCCGTCACC AAGGAGGGTC TGGAGCTGCC AGAGGATGAT GCCGAGAAGA

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3410       3420       3430       3440       3450       3460       3470       3480       3490       3500
Madeira       AGAAGCGTGA GGAAGATAAG GCCAAGTTCG AGGGTCTCTG CAAGCTGATG AAGTCCATCC TGGACAGCAA AGTTGAGAAG GTGGTCGTCT CGAACCGTCT
Curicó        AGAAGCGTGA GGAAGATAAG GCCAAGTTCG AGGGTCTCTG CAAGCTGATG AAGTCCATCC TGGACAGCAA AGTTGAGAAG GTGGTCGTCT CGAACCGTCT

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3510       3520       3530       3540       3550       3560       3570       3580       3590       3600
Madeira       CGTGGATTCG CCCTGCTGCA TCGTCACCTC GCAGTTCGGC TGGTCCGCCA ACATGGAGCG TATCATGAAG GCCCAGGCTC TGCGCGACAC AGCCACCATG
Curicó        CGTGGATTCG CCCTGCTGCA TCGTCACCTC GCAGTTCGGC TGGTCCGCCA ACATGGAGCG TATCATGAAG GCCCAGGCTC TGCGCGACAC AGCCACCATG

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3610       3620       3630       3640       3650       3660       3670       3680       3690       3700
Madeira       GGCTACATGG CCGGCAAGAA GCAGTTGGAA ATCAATCCCG ACCATCCCAT TGTCGAGGCC CTGCGCCAGA AGGCCGATGC CGACAAGAAC GACAAGGCTG
Curicó        GGCTACATGG CCGGCAAGAA GCAGTTGGAA ATCAATCCCG ACCATCCCAT TGTCGAGGCC CTGCGCCAGA AGGCCGATGC CGACAAGAAC GACAAGGCTG

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3710       3720       3730       3740       3750       3760       3770       3780       3790       3800
Madeira       TGAAGGATCT GGTCATCCTT CTGTTTGAGA CCTCGCTGCT GTCTTCCGGA TTCTCTCTGG ACAGCCCTCA GGTTCATGCC AGCCGCATTT ACCGTATGAT
Curicó        TGAAGGATCT GGTCATCCTT CTGTTTGAGA CCTCGCTGCT GTCTTCCGGA TTCTCTCTGG ACAGCCCTCA GGTTCATGCC AGCCGCATTT ACCGTATGAT

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3810       3820       3830       3840       3850       3860       3870       3880       3890       3900
Madeira       CAAGCTGGGT CTTGGCATTG ATGAGGATGA GCCCATGACC ACCGAGGACG CCCACAGTGG TGGTGATGCC CCCAATCTGG TCGAGGACAC CGAGGATGCT
Curicó        CAAGCTGGGT CTTGGCATTG ATGAGGATGA GCCCATGACC ACCGAGGACG CCCACAGTGG TGGTGATGCC CCCAATCTGG TCGAGGACAC CGAGGATGCT

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 3910       3920       3930       3940       3950       3960       3970       3980       3990       4000
Madeira       TCCCATATGG AGGAGGTCGA CTAAACCATT CGCTTACTCA AACAAAACAC ATTCACACAC ACACAAGCAC ACAACACACA ---------- -----CAAAA
Curicó        TCCCATATGG AGGAGGTCGA CTAAACCATT CGCTTACTCA AACAAAACAC ATTCACACAC ACACAAGCAC ACAACACACA AGCACACAAC ACACACAAAA

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 4010       4020       4030       4040       4050       4060       4070       4080       4090       4100
Madeira       TTGTCTAAGT TTCGACTGCT TATAGCTTAT GTTGTTTACA AAGGAGCTAA CTTTGAATTT TCATTAACAT ATTTGTCGCG TTATATGCGG CAATTATACG
Curicó        TTGTCTAAGT TTCGACTGCT TATAGCTTAT GTT---TACA AAGGAGCTAA CTTTGAATTT TCATTAACAT ATTTGTCGCG TTATATGCGG CAATTATACG

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 4110       4120       4130       4140       4150       4160       4170       4180       4190       4200
Madeira       CTTATCCATC CCACATTCAC TAGCGTTCAG AATAGAATCC TCCAGGTTCC CAAAACTAAC TTCAGAAGCA GTTGTCGTTT TAAGAATTTC TAATTTAGAG
Curicó        CTTATCCATC CCACATTCAC TAGCGTTCAG AATAGAATCC TCCAGGTTCC CAAAACTAAC TTCAGAAGCA GTTGTCGTTT TAAGAATTTC TAATTTAGAG

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 4210       4220       4230       4240       4250       4260       4270       4280       4290       4300
Madeira       TCCACATAAT TTATGTAAAA ATAATCTAAA GACTACACGC TCTAGTTTTT AGAATTGAAA TATATATCTT AAGGTTTGAT GACCAGATCG ATCGATGATA
Curicó        TCCACATAAT TTATGTAAAA ATAATCTAAA GACTACACGC TCTAGTTTTT AGAATTGAAA TATATATCTT AAGGTTTGAT GACCAGATCG ATCGATGATA

              ....|....| ....|....| ....|....| ....|....| ....|....| ....
                 4310       4320       4330       4340       4350
Madeira       AAACCAAATA AACAAATTGT GTTTTAAAAT ACATATAAAT GAGGATATAT TTAT
Curicó        AAACCAAATA AACAAATTGT GTTTTAAAAT ACATATAAAT GAGGATATAT TTAT
```

Alignment available at: https://figshare.com/s/26f5a4f78b928bf7a441.

8.3 Regulatory divergence

	Parent		A/B	Hybrid		A1/B1	P=H
cis	A		P≠0	A1		H≠0	=
	B			B1			
trans	A		P≠0	A1		H=0	≠
	B			B1			
cis + trans	A		P≠0	A1		H≠0 Closer to 0	≠
	B			B1			
cis x trans	A		P≠0 Could be closer to 0 (P=0)	A1		H≠0 Could be closer to 0	≠
	B			B1			

Compensatory

Figure adapted from [489] **and** [656]**.**

8.4 Weather in Madeira and Curicó

	Curicó	Madeira	
Latitud	34°58'58" S	32° 22' 17.9976" N	
Longitud	71°14'21" O	16° 16' 29.9928" W	
Average anual temperature	14.7	16.6	°C
Average coldest month temperature	8.0	13.0	°C
Average warmest month temperature	21.9	21.1	°C
Lowest minimum peak temperature	3.7	11.6	°C
Highest maximum peak temperature	28.5	23.8	°C
Annual average rainfall	832.0	571.4	mm
Driest months average precipitations	4.0	8.2	mm
Rainly months average precipitations	199.0	89.4	mm
Humidity	66.7	74.4	%
Average lowest humid month humidity	52.9	72.5	%
Average highest humid month humidity	78.6	76.3	%
Average sun hours	9.3	5.8	h
Average sun hours in the month with less light	6.6	4.5	h
Average sun hours in the month with more light	12.3	7.6	h

Data extracted from: https://en.climate-data.org/.

8.5 High Stability of the Epigenome in Drosophila Interspecific Hybrids – Published version

High Stability of the Epigenome in *Drosophila* Interspecific Hybrids

Alejandra Bodelón[1], Marie Fablet[2,3], Philippe Veber[2], Cristina Vieira[2], and Maria Pilar García Guerreiro[1,*]

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

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

[3]Institut universitaire de France, France

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

Accepted: 2 February 2022

Abstract

Interspecific hybridization is often seen as a genomic stress that may lead to new gene expression patterns and deregulation of transposable elements (TEs). The understanding of expression changes in hybrids compared with parental species is essential to disentangle their putative role in speciation processes. However, to date we ignore the detailed mechanisms involved in genomic deregulation in hybrids. We studied the ovarian transcriptome and epigenome of the *Drosophila buzzatii* and *Drosophila koepferae* species together with their F_1 hybrid females. We found a trend toward underexpression of genes and TE families in hybrids. The epigenome in hybrids was highly similar to the parental epigenomes and showed intermediate histone enrichments between parental species in most cases. Differential gene expression in hybrids was often associated only with changes in H3K4me3 enrichments, whereas differential TE family expression in hybrids may be associated with changes in H3K4me3, H3K9me3, or H3K27me3 enrichments. We identified specific genes and TE families, which their differential expression in comparison with the parental species was explained by their differential chromatin mark combination enrichment. Finally, *cis–trans* compensatory regulation could also contribute in some way to the hybrid deregulation. This work provides the first study of histone content in *Drosophila* interspecific hybrids and their effect on gene and TE expression deregulation.

Key words: epigenome, *Drosophila*, interspecific hybrids, deregulation, transposable elements, histone methylation.

Significance

The genomic stress caused by hybridization between different species is a cause of gene and transposable element (TE) deregulation. Histone modifications were often associated to genomic deregulation, but little is known to date about their role in *Drosophila* hybrid anomalies, and their elucidation is essential to unravel the role of these phenomena in speciation. Using transcriptomic and epigenomic analyses, we found that even though the epigenome of the parental species is highly conserved in hybrids, some changes detected in histone marks were associated to gene and TE deregulation. This work provides the first study of histone modification in *Drosophila* hybrids and its putative impact on gene a TE expression.

Introduction

Interspecific hybridizations are a source of phenotypic variation (Soto et al. 2008; Zhu, Hu, et al. 2017), sterility (Naveira and Fontdevila 1991; Michalak and Noor 2003; Moehring et al. 2007), and genomic instability (Naveira and Fontdevila 1991; Vela et al. 2014, 2011) in hybrid offspring. These phenomena are induced by the fusion of two different parental genomes and epigenomes, which act as a genomic stressful

factor in the hybrid genome (García Guerreiro 2012). This integration leads to deregulation of gene expression in hybrids, which has been described in *Drosophila* (Michalak and Noor 2003; Ranz et al. 2004; Moehring et al. 2007; Romero-Soriano et al. 2017; Gámez-Visairas et al. 2020) and other organisms (Malone and Michalak 2008; Zhu, Hu, et al. 2017). For example, the misexpression of genes involved in spermatogenesis in *Drosophila* male hybrids (Michalak and Noor 2003; Moehring et al. 2007) has been related with their sterility. Deregulation of gene expression was also observed in hybrid females of *Drosophila buzzatii* and *Drosophila koepferae* (Romero-Soriano et al. 2017; Gámez-Visairas et al. 2020) and in other *Drosophila* hybrids (Ranz et al. 2004). Gene expression in *Drosophila* is controlled at different levels, including epigenetic modifications (Yin et al. 2011) and *cis*- and *trans*-regulatory elements, among others. Divergence between parental regulatory elements may play an important role in gene expression deregulation in hybrids (Mack and Nachman 2017). For example, differences in the *cis* and *trans* parental regulatory elements were related with misexpression in *Drosophila* hybrids (Landry et al. 2005).

In addition to gene expression disruption, a growing number of evidence in *Drosophila* suggest that TEs could also play an important role in the hybrid anomalies. For example, increases of transposition rates and/or misexpression were observed in interspecific hybrids between *D. buzzatii* and *D. koepferae* (Labrador et al. 1999; Vela et al. 2014; García Guerreiro 2015; Romero-Soriano and Garcia Guerreiro 2016) and other *Drosophila* species (Kelleher et al. 2012; Lopez-Maestre et al. 2017). In *Drosophila*, germline transposition is controlled at two levels: transcriptional silencing, by heterochromatinization processes, and post-transcriptional by piwi-interacting RNAs (piRNAs) involved in transposon transcript degradation. In this sense, the causes of TE release in hybrids have been the matter of recent research focused on the study of disruption of the germline mechanisms involved in TE regulation, basically those affecting piRNA production (Kelleher et al. 2012; Lopez-Maestre et al. 2017; Romero-Soriano et al. 2017). Recent discoveries pointed to a dysfunction of piRNA pathway in *Drosophila melanogaster–Drosophila simulans* hybrids (Kelleher et al. 2012), which would have a deficient global piRNA production, whereas a similar or higher piRNA production was observed in *D. buzzatii-D. koeferae* ovaries (Romero-Soriano et al. 2017). Moreover, in the later study a high proportion of the TEs overexpressed in hybrids did not have associated piRNAs in parents or hybrids, which suggests a more complex deregulation network, at least, in these hybrids.

Less is known about the behavior of epigenomes in interspecific *Drosophila* hybrids, which is of fundamental interest in evolutionary biology. Histone modification is a post-translational chemical modification of histone proteins. In *Drosophila*, the histone marks H3K4me3 are associated to activation of gene expression, whereas H3K9me3 and H3K27me3 are associated to gene repression (Yin et al. 2011). H3K4me3 marks are abundant in euchromatin, and H3K9me3 and H3K27me3 take part of constitutive and facultative heterochromatin, respectively (Ebert et al. 2006; Schulze and Wallrath 2007). Moreover, it was also shown that TEs are enriched in the H3K9me3 and H3K27me3 heterochromatic marks, suggesting a putative colocalization of these silencing marks in these sequences (Yin et al. 2011). In plants, histone modification occurring during interspecific hybridization has been associated to gene expression variation (He et al. 2010; Zhu, Hu, et al. 2017). However, in *Drosophila*, little is known about the role of histone modification on gene expression and their consequences on hybrid incompatibility and TE deregulation. Studies on *D. melanogaster–D. simulans* hybrids suggest that adaptive divergence of heterochromatin proteins is an important force driving the evolution of genes involved in hybrid incompatibility (Satyaki et al. 2014). Our previous work on *D. buzzatii–D. koepferae* hybrids suggested the existence of interacting phenomena, including incompatibilities of the piRNA pathway, due to a functional divergence between parental species, as one of the causes responsible of TE mobilization (Romero-Soriano et al. 2017; review Fontdevila 2019). However, we cannot disregard other putative mechanisms in these species, including histone modifications causing changes in TE expression.

To unravel the causes of this hybrid incompatibility, we investigated the expression profiles and the patterns of three histone marks: H3K4me3, H3K9me3, and H3K27me3, using RNA-sequencing (RNAseq) and chromatin immunoprecipitation and deep sequencing (ChIPseq) in the genome of the parental species (*D. buzzatii* and *D. koepferae*) and their hybrids. *Drosophila buzzatii* and *D. koepferae* are two cactophilic sibling species, belonging to *repleta* group (García Guerreiro 2014), that diverged approximately 4–5 Ma (Gómez and Hasson 2003; Oliveira et al. 2012; Romero-Soriano et al. 2017). Crosses between *D. buzzatii* males and *D. koepferae* females result in F_1 sterile males and fertile females (Marin et al. 1993), even though a few cases of partial fertility with atrophy in one of the ovaries is also observed (Marin and Fontdevila 1998). Hybrid females can be backcrossed with *D. buzzatii* males (Naveira et al. 1986; Marin and Fontdevila 1998). In this system, we can only obtain hybrids from crosses between *D. koepferae* females and *D. buzzatii* males, the reciprocal cross does not produce offspring (Marin et al. 1993).

We found that both genes and TE families detected as differentially expressed in hybrid ovaries in comparison with *D. buzzatii* and *D. koepferae*, tended to be mostly underexpressed. In contrast, we found a high conservation of the parental chromatin mark patterns in hybrid genes and TEs, showing intermediate levels between the parental species. Nevertheless, we could associate some changes in gene and TE family expression in hybrids with their corresponding histone modifications versus parental species.

Results

Deregulation of Gene and TE Expression in Hybrids

We analyzed and compared the ovarian transcript amounts of *D. buzzatii* and *D. koepferae* and their F_1 female hybrids. We found that, out of a total of 13,621 protein coding genes and 658 TE families, 5.92% and 29.64%, respectively, were differentially expressed between the parental species (supplementary file 1: table S1, Supplementary Material online). Similarly, hybrids also showed a lower percentage of differentially expressed genes, in comparison with the parental species (4.57% vs. *D. buzzatii* and 3.99% vs. *D. koepferae*, fig. 1A), than TE families (22.95% vs. *D. buzzatii* and 24.16% vs. *D. koepferae*, fig. 1B). Gene expression in hybrids was more similar to the maternal species *D. koepferae* than to the paternal species *D. buzzatii* (Z-test, $P = 0.018$, supplementary file 1: table S2, Supplementary Material online), but the number of differentially expressed TE families in hybrids was similar in comparison with both parental species (Z-test, $P = 0.603$, supplementary file 1: table S2, Supplementary Material online).

Globally, we could classify the differentially expressed genes and TEs in hybrids in three categories (fig. 1C): 1) *D. koepferae*- or *D. buzzatii*-like expression, which corresponded to the majority of differentially expressed genes (fig. 1A) and TE families (fig. 1B). 2) Overexpressed or underexpressed in hybrids in comparison with both parental species, which were considered as deregulated in hybrids (fig. 1C). We observed a deregulation trend toward underexpression in hybrids compared with both parental species: 145 genes and 33 TE families underexpressed versus 29 genes and 12 TE families overexpressed (fig. 1A and B). 3) Additive expression which included differentially expressed genes and TE families in hybrids with an intermediate expression between both parental species. Only a small number of genes (29 and 40 genes) were included in this category (fig. 1A). Regarding the TEs, the additive expressed TE families were mostly underexpressed in comparison with *D. buzzatii*, and overexpressed in comparison with *D. koepferae* (fig. 1B). Both results were consistent with the differences between the parental species, where the mean TE family expression was higher in *D. buzzatii* than in *D. koepferae* (mean logarithmic 2-fold change [log2FC] = 0.27 or 1.21-fold increase in *D. buzzatii* vs. *D. koepferae*), and opposite for the genes (mean log2FC= −0.23 or 0.85-fold decrease in *D. buzzatii* vs. *D. koepferae*; supplementary file 1: table S1, Supplementary Material online).

We found that, even though there was a general bias toward underexpression of genes (fig. 1D) and TE families (fig. 1E) in hybrids, those detected as overexpressed exhibit the highest differences of expression compared with parental species. In fact, the values of mean log2FC were 2.53–3.01 (FC = 5.78–8.06) for genes and 3.33–3.56 (FC = 10.06–11.79) for TE families overexpressed in hybrids versus *D. koepferae* and *D. buzzatii*, respectively, whereas they decreased to −1.62–1.72 (FC = 0.33–0.30) for genes and to −1.85–1.74 (FC = 0.28–0.30) for TE families underexpressed in hybrids versus *D. koepferae* and *D. buzzatii*, respectively. Regarding the genes, most of the overexpressed shown in figure 1D, were involved in metabolism (small molecules or protein), development, signaling and stimulus response, reproduction and transportation, and cell organization (supplementary file 2, Supplementary Material online).

An interesting example was the *no hitter gene* (nht), highlighted in bold in figure 1D, which was overexpressed in hybrids compared with both parental species. This gene is involved in spermatogenesis and regulation of gene expression and may be related to female hybrid sterility (Ranz et al. 2004). Moreover, as expected, most of the genes with higher differences between hybrids and the parental species (gene names labeled in fig. 1D) were at the same time detected as the most different between the parental species (green and blue gene names in supplementary file 3: fig. S1a, Supplementary Material online).

Regarding underexpressed genes, the Gene Ontology (GO) analysis revealed a shared enrichment for 11 terms in hybrids versus both parental species (green GO in supplementary file 2, Supplementary Material online) and were mainly related to developmental processes, cell adhesion and reproduction (i.e., gonad development). Furthermore, the fact that most of them had functions related with reproduction, reinforces the idea of a putative role in hybrid fertility loss (Moehring et al. 2007). In the case of the overexpressed genes in hybrids, only two enriched GO terms were shared with both parental species (green GO in supplementary file 2, Supplementary Material online), and were linked to metabolic and cellular process. Additional GO terms, related to the same or other processes (such as response to stimulus and biological regulation, among others) were enriched in the differentially expressed genes, but only in comparison with one of the parental species (supplementary file 2, Supplementary Material online).

To detect a putative location bias of the differentially expressed genes in hybrids in comparison with the parental species, previously reported in some *Drosophila* hybrids (Moehring et al. 2007; Wurmser et al. 2011), we studied their distribution per chromosome. We found that the number of differentially expressed genes across chromosomes in hybrids compared with both parental species (supplementary file 3: fig. S2a, Supplementary Material online), was different from the random expectation (chi-square test, $P < 0.001$, supplementary file 1: table S3, Supplementary Material online). This result is due to a higher number of differentially expressed genes on the dot chromosome 6 (1.93% vs. *D. buzzatii* and 2.21% vs. *D. koepferae*) than the expected (0.54%). When this chromosome was excluded from the test, the differences become non-significant (chi-square test, *D. buzzatii*:
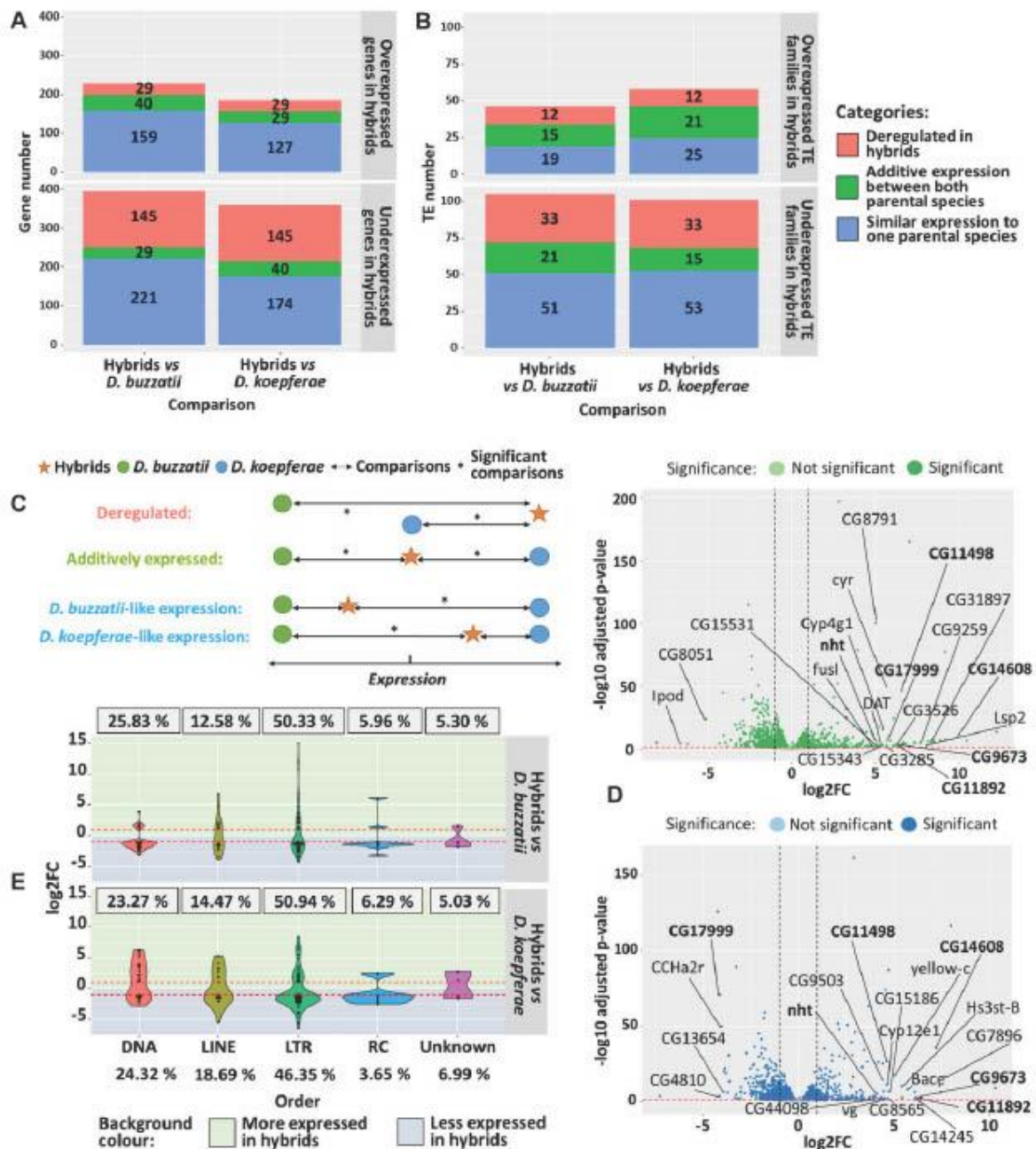
Fig. 1.—Differential expression analyses of genes and TEs: (A) Number of differentially expressed genes and (B) TE families in hybrids versus parental species of the total of 13,621 genes and 658 TE families. Colors indicate gene expression categories. (C) Expression categories in hybrids versus parental species. (D) Differential gene expression analyses in hybrids versus *D. buzzatii* (green) and *D. koepferae* (blue). Positive log2FC values correspond to genes more expressed in hybrids. The genes showing the 20 highest log2FC values and displaying an ortholog in *D. melanogaster* are shown. Genes common to both comparisons are in bold. (E) Violin plots representing the distribution of differentially expressed orders of TEs in hybrids versus parental species. Points indicate the log2FC of each TE family and for each comparison. The percentages of differentially expressed TE families per order are framed. The expected (total) TE family percentages are at the bottom. Red-dashed lines indicate the log2FC threshold for significancy (±1).

$P = 0.320$ and *D. koepferae*: $P = 0.443$, supplementary file 1: table S3, Supplementary Material online).

Concerning the TEs, we studied the distribution and expression of the differentially expressed TE families, per order, in hybrids in comparison with the parental species. We used the Repbase classification (Jurka et al. 2005) where TEs were divided in LTR, LINE, and DNA orders. Due to their particular replication mechanism, we considered the RC/Helitron as a group apart from the DNA order. As shown in figure 1*E*, the number of differentially expressed TE families by order was similar to the random expectation considering their relative proportions in the genome (chi-square test, $P = 0.183$ for *D. buzzatii* and *D. koepferae*, supplementary file 1: table S3, Supplementary Material online). The most extreme underexpression values were observed in LINE and LTR families in hybrids versus *D. koepferae*. In contrast, these families together with RC were the most overexpressed in hybrids versus *D. buzzatii*. Last, the LTR order was the one where some families showed extreme overexpression and underexpression values in hybrids versus both parental species. When we went deeper into the superfamilies of elements belonging to each order (supplementary file 3: fig. S2*b*, Supplementary Material online), highly expressed elements in hybrids included mainly Gypsy (LTR), Helitron (RC), and Jockey (LINE) elements in comparison with *D. buzzatii* and, Gypsy (LTR), hAT (DNA), and TC1Mariner (DNA) elements compared with *D. koepferae*. Some of these TEs were detected as the most differentially expressed between the parental species (green and blue TE names in supplementary file 3: fig. S1*b*, Supplementary Material online). Consistently with what was found in previous works (Romero-Soriano et al. 2017), the most highly overexpressed TEs in hybrids versus both parental species belonged to Gypsy subfamily (supplementary file 3: fig. S2*b*, Supplementary Material online).

Epigenetic Mark Landscapes Are Conserved and Contribute to Gene and TE Expression

In *Drosophila*, the constitutive heterochromatin is enriched in H3K9me3 epigenetic mark, whereas H3K27me3 and H3K4me3 are associated to facultative heterochromatin and euchromatin, respectively (Boros 2012). We investigated the landscape of epigenetic marks in the parental species, unknown to date, and in their hybrids. Analyses of the distribution of histone marks in genes and their surrounding regions (fig. 2*A* and *B*) showed that H3K4me3 euchromatic mark is enriched around the start codon (SC) and throughout the coding sequence in hybrids and their parental species, which is similar to the well-described *D. melanogaster* species. H3K9me3 and H3K27me3 epigenetic marks were mainly depleted in the gene body in all species (fig. 2*A* and *B*).

We then studied the contribution of these histone marks to the gene and TE family expression in the parental species *D. buzzatii*, *D. koepferae*, and their hybrids, using the linear

model (RNA $\sim$ K4 + K9 + K27 + Input) described in Material and Methods. We found that 62% (r^2 adjusted) of gene expression variation was explained by the additive linear relationship with log-transformed histone mark enrichments, with a *P*-value lower than 2.2×10^{-16} in all species (supplementary file 1: table S4, Supplementary Material online). In the case of TE expression, the adjusted r^2 reached values from 61% in hybrids, to 75% in *D. koepferae* (supplementary file 1: table S4, Supplementary Material online).

Globally, we detected that the size effect, representing the contribution of each chromatin mark to the gene expression according to our model, was similar in hybrids and parental species (fig. 2*C*), being H3K27me3 the chromatin mark with the greatest contribution to gene expression. H3K4me3 was strongly positively associated with gene expression showing coefficient values from 1.59 in *D. koepferae* to 1.75 in hybrids, whereas H3K27me3 was strongly negatively associated: from −2.24 in *D. koepferae*, to −2.56 in hybrids (supplementary file 1: table S4, Supplementary Material online). H3K9me3 was the one with the least positive contribution to gene expression (from 0.35 in *D. koepferae* to 0.5 in *D. buzzatii*).

Regarding TEs, the contribution of H3K4me3 and H3K9me3 to the expression was similar in hybrids and parental species (fig. 2*D*). H3K4me3 was positively associated with TE expression, with coefficient values from 0.96 in *D. koepferae* to 1.58 in *D. buzzatii* (supplementary file 1: table S4, Supplementary Material online), whereas H3K9me3 was negatively associated (−1.04 in *D. buzzatii* to −1.16 in hybrids), as expected. Contrary to the pattern observed in genes, H3K27me3 was the chromatin mark with the lowest contribution to TE expression and its contribution depended on the species: positively associated in *D. buzzatii* and hybrids (0.38 and 0.71 respectively) and negatively in *D. kopferae* (−0.33).

Maintenance of the Parental Enrichment of Epigenetic Marks in the Hybrid Genome

We computed read counts corresponding to enrichments in H3K4me3, H3K9me3, and H3K27me3, in the gene bodies and TE families from hybrids and their parental species. We found that only a small percentage of all genes (1.40–2.83%) were differentially enriched in hybrids compared with any parental species and chromatin mark (fig. 3*A*). A lower number of differentially H3K4me3 and H3K27me3 enriched genes was detected in hybrids when they were compared with *D. buzzattii* than to *D. koepferae* (Z-test, $P < 0.001$, supplementary file 1: table S5, Supplementary Material online). Regarding TE families, the hybrid versus parents comparisons revealed very similar patterns for H3K9me3 enrichments (1.52–1.37% of differentially enriched TE families), whereas they were the most contrasted for H3K27me3 enrichments (9.27–6.53% differentially enriched TE families, fig. 3*B*). However, the number of differentially enriched TEs in hybrids
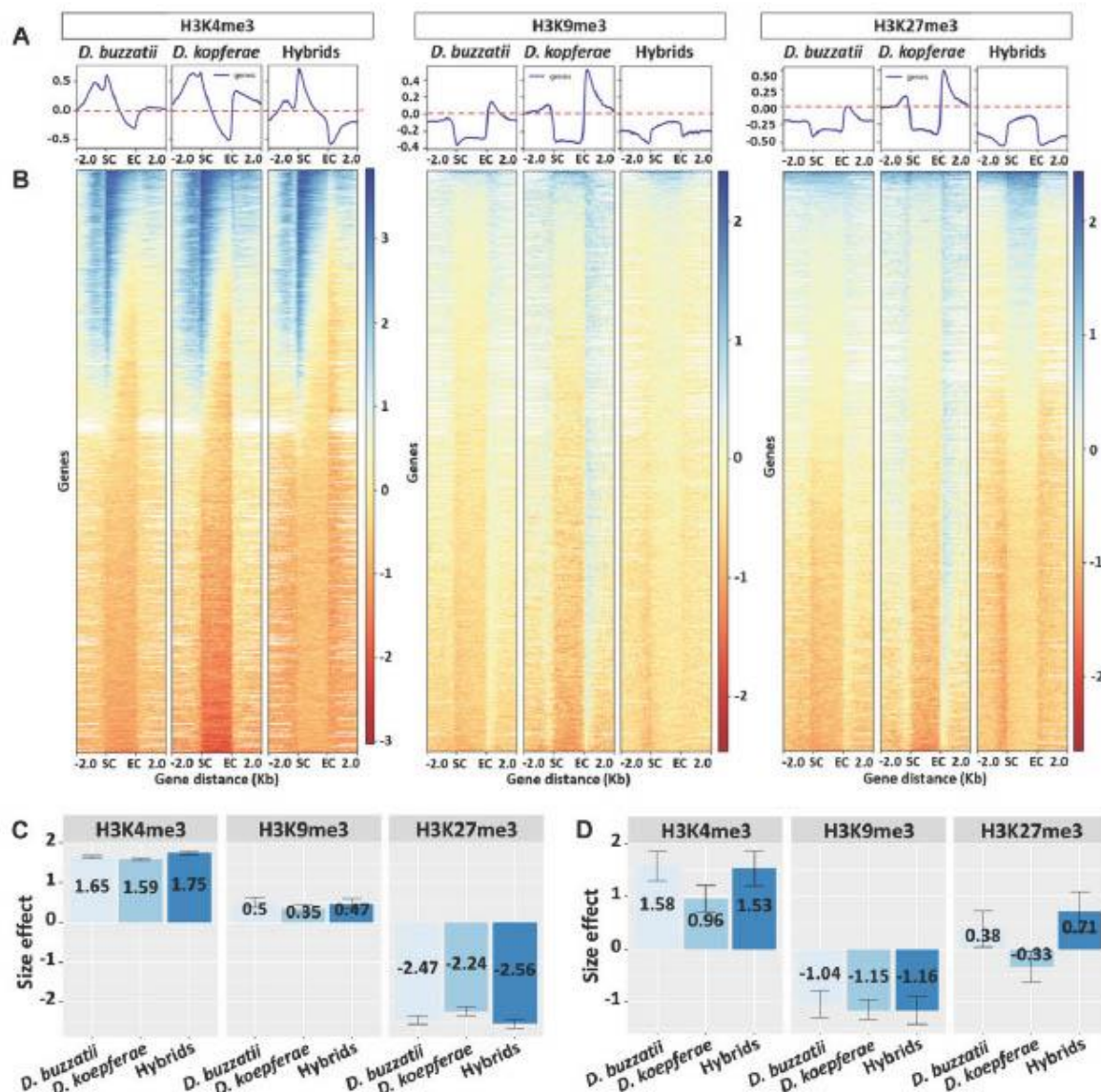
207

Fig. 2.—Global epigenetic patterns in *D. buzzatii*, *D.koepferae*, and hybrids and their contribution to expression: (*A*) Average levels of the histone marks H3K4me3, H3K9me3, and H3K27me3 over all annotated gene sequences (exons and introns) and the ±2 kb surrounding regions. (*B*) Heatmaps showing the density scores of each histone mark in *D. buzzatii*, *D. koepferae*, and their hybrids. White regions in the heatmaps indicate missing data. SC: start codon, EC: end codon. (*C, D*) Mean linear effects of H3K4me3, H3K9me3, and H3K27me3 enrichments on RNA read counts (log transformed) in (*C*) genes and (*D*) TE families. Colors indicate the species. The 95% confidence intervals are indicated.

was similar to one or another parental species (Z-test, $P = 0.817$ H3K4me3 and H3K9me3; $P = 0.198$ H3K27me3; supplementary file 1: table S5, Supplementary Material online). Moreover, H3K9me3 was the chromatin mark with more similar patterns between hybrids and the parental species for both genes and TE families, which was consistent with the similar patterns of this chromatin mark when the parental species were compared (supplementary file 1: table S6,

Supplementary Material online). On the contrary, in TE families H3K27me3 was the mark with the largest differential enrichment in hybrids versus the parental species, showing the largest contrast between the parental species (supplementary file 1: table S6, Supplementary Material online).

In general, most of the differentially enriched genes and TEs in hybrids have *D. buzzatii* or *D. koepferae*-like chromatin mark levels. However, we also observe a high number of
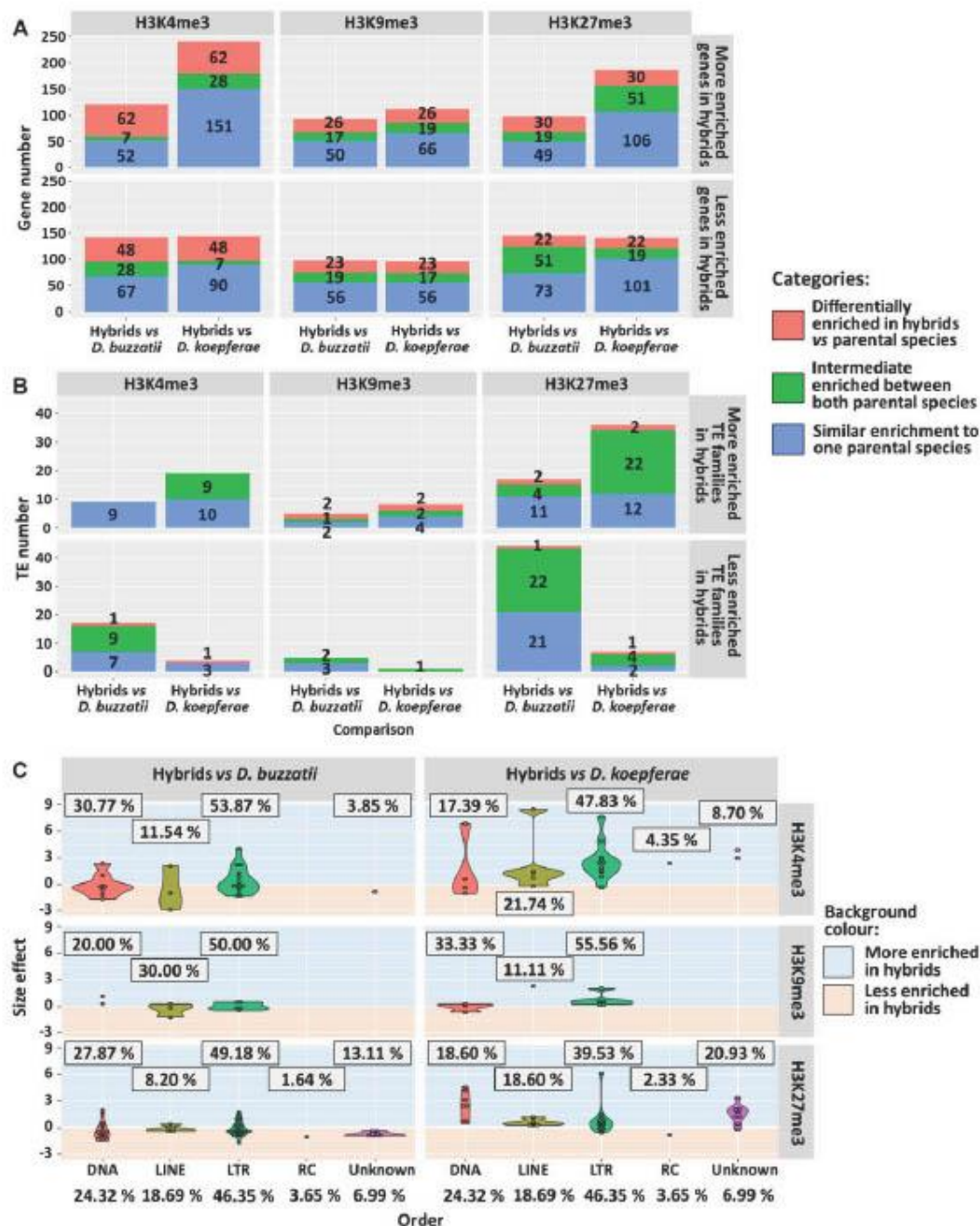
Fig. 3.—Comparison of epigenetic marks in hybrids and parental species: (A) Number of differentially enriched genes and (B) TEs of the total of 13,621 genes and 658 TE families in H3K4me3, H3K9me3, or H3K27me3 chromatin marks in hybrids versus the parental species *D. buzzatii* and *D. koepferae*, respectively. Colors indicate the different categories of enrichment. (C) Violin plots representing the distribution of differentially H3K4me3, H3K9me3, and H3K27me3 enriched TE orders in hybrids versus parental species. Points indicate the size effect. The percentages of differentially enriched TE families per order are framed. The expected (total) TE family percentages are at the bottom.

209

genes displaying histone mark enrichments outside of the range of parental values, especially for H3K4me3—48 less enriched and 62 more enriched in hybrids versus both parental species (fig. 3A). In the same way, a high number of TE families show intermediate patterns of enrichment between the parental species (fig. 3B).

We studied the distribution across chromosomes of the differentially enriched genes for each epigenetic mark in hybrids in comparison with each parental species and we did not find significant differences in any case (chi-square test, $P > 0.05$ in all cases, supplementary file 1: table S7, Supplementary Material online). Globally, we observed that the extreme changes in chromatin mark levels were toward an increase in hybrids in comparison with the parental species (supplementary file 3: fig. S3, Supplementary Material online). For example, some extreme changes of the euchromatic mark H3K4me3 were observed in chromosome 5 in comparison with *D. koepferae*, and of the euchromatic mark H3K27me3 in comparison with *D. koepferae* (chromosomes 4 and 5) and *D. buzzatii* (chromosome 2).

When we studied the orders of the differentially enriched TE families in hybrids versus parental species, we did not find differences (chi-square test, $P > 0.05$ in all cases, supplementary file 1: table S7, Supplementary Material online) for any chromatin mark (fig. 3C). If we focus on the specific enrichment of epigenetic marks, three TEs belonging to DNA, LINE, and LTR families showed the most extreme values of H3K4me3 in hybrids versus *D. koepferae* (ranging from 6 to 9). However, RC and unknown TEs showed values of this chromatin mark similar to both parental species, being only a small percentage differentially enriched. We also found a high increment of H3K27me3 in one LTR element in comparison with *D. koepferae* ($\sim$6), whereas the changes were less extreme in comparison with *D. buzzatii*. As reported for gene enrichment, only small differences of H3K9me3 amounts were observed in hybrids versus parental species.

Small Changes in the Hybrid Epigenome Affect Their Gene and TE Expression

We analyzed the association between gene and TE expression changes in hybrids in comparison with the parental species and the corresponding changes in chromatin marks (fig. 4A–L, and supplementary file 1: table S8, Supplementary Material online) using the whole set of genes and TE families (significant and nonsignificant). Regarding the genes, we found that changes in the enrichment of the heterochromatic marks H3K9me3 and H3K27me3 were not associated with expression changes in hybrids in comparison with the parental species (Fisher's exact test, *D. buzzatii*: $P = 0.763$ both chromatin marks; *D. koepferae*: $P = 0.970$ and $P = 0.155$, respectively, supplementary file 1: table S8, Supplementary Material online). However, changes in H3K4me3 in hybrids seemed to be associated with gene expression changes (Fisher's exact test,

$P \leq 0.001$: comparison with both parental species, supplementary file 1: table S8, Supplementary Material online). Indeed, genes that were underexpressed in hybrids compared with *D. buzzatii* also showed a reduced H3K4me3 enrichment. However, and unexpectedly, genes that were underexpressed in hybrids compared with *D. koepferae* also displayed an increased H3K4me3 enrichment. This apparent inconsistency could be explained by the differences in enrichment of this epigenetic mark between parental species: the genes enriched in H3K4me3 in hybrids versus *D. koepferae* often corresponded to those more enriched in *D. buzzatii* than in *D. kopferae* (blue color, fig. 4A), the opposite was also observed.

TE families, as genes, showed an intermediate inheritance of the chromatin marks in hybrids: TEs more enriched in a chromatin mark in hybrids in comparison with one parental species tended to be less enriched when they were compared with the other parental species (fig. 4G, I and K). In addition, we found an association between TE expression changes in hybrids in comparison with the parental species and the corresponding changes in the epigenome. For example, an impoverishment of the heterochromatic mark H3K9me3 was found in many underexpressed TEs in hybrids compared with both parental species (Fisher's exact test, $P < 0.001$, supplementary file 1: table S8, Supplementary Material online). In the same way, a high TE number showing a decrease of H3K4me3 or H3K27me3 chromatin marks were underexpresed in hybrids in comparison with *D. buzzatii* (Fisher's exact test, H3K4me3: $P < 0.001$ and H3K27me3: $P = 0.002$, supplementary file 1: table S8, Supplementary Material online). These results were opposite to those found in hybrids compared with *D. koepferae*, where an increase of H3K27me3 was observed in underexpressed elements ($P < 0.001$, supplementary file 1: table S8, Supplementary Material online). To determine if there was an association between chromatin marks in TEs, we checked the correlation between the euchromatic mark H3K4me3 and the other two, in hybrids in comparison with the parental species (supplementary file 3: fig. S4, Supplementary Material online). We found that the increase of H3K4me3 in hybrids was associated with an increase of H3K9me3 and H3K27me3 (*Linear model*, $P \leq 1.669 \times 10^{-15}$ in all cases, supplementary file 1: table S9, Supplementary Material online) and the opposite, with some TEs highly expressed and extremely enriched in H3K4me3 in hybrids (red dots in the upper part right of supplementary file 3: fig. S4, Supplementary Material online). When we focus on the genes considered as differentially expressed and differentially enriched in hybrids, we observe that the epigenetic mark associated with the most extreme expression changes in hybrids (absolute log2FC values ranging from 4 to 6) in comparison with both parental species was H3K4me3 (fig. 4M). Changes in the remaining chromatin marks were related with less extreme expression changes (absolute log2FC values from 1 to 3). Regarding the TEs showing
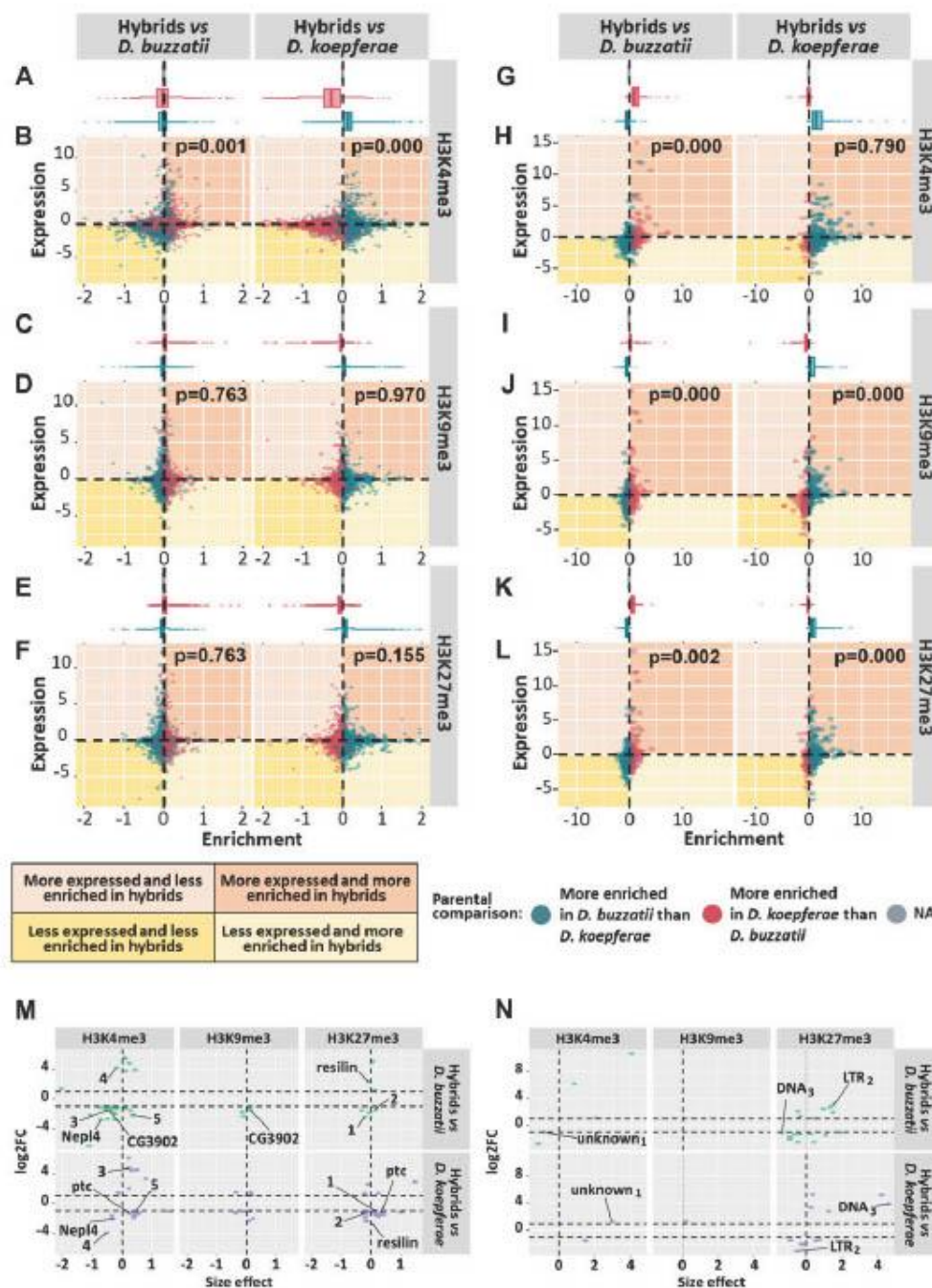
Fig. 4.—Association of expression changes with the corresponding chromatin mark changes: (A–L) Differential expression (log2FC) and differential enrichment (size effect) of H3K4me3, H3K9me3, and H3K27me3 in all genes (B, D, F) and TE families (H, J, L) (significant and nonsignificant) in hybrids versus *D. buzzatii* and *D. koepferae*, respectively. Top plots represent the distribution of the differential enrichment between parental species in genes (A, C, E) and TE families (G, I, K). Colors represent the results of the differential enrichment between parental species, and values that cannot be computed are included in the NA category. Fisher's exact test *P*-values are shown in the right corner. Gene outliers are not included in (A–F). (M, N) Log2FC and size effect representation of significant differentially expressed and enriched genes (M) and TEs (N) in hybrids versus the parental species. Significant genes (M) in at least two comparisons are named using their *D. melanogaster* ortholog or, otherwise, a number. The TE families (N) common to at least two comparisons are marked, indicating their category and a subscript.

211

significant differences in enrichment and expression in hybrids versus parental species (fig. 4N), we observed two TE families whose increase in H3K4me3 was associated to an extreme overexpression in hybrids in comparison with *D. buzzatii* (log2FC from 6 to 11). Additionally, increases in the number of H3K27me3 marks in hybrids in comparison with *D. kopferae* were likely related to extreme overexpression in hybrids (log2FC from 3 to 6). Chromatin mark changes in the remaining TEs were related with less extreme expression changes (absolute log2FC values from 1 to 3).

If we focus on the expression and chromatin enrichment of specific genes (fig. 4M) in hybrids, we observe some examples whose expression is explained by the combination of different chromatin marks. For example, the *CG3902* gene, involved in oxidation–reduction processes, was underexpressed in hybrids in comparison with *D. buzzatii* and less enriched in H3K4me3 but more in H3K9me3. The gene *ptc*, involved in different processes including gonad development, was underexpressed in hybrids in comparison with *D. koepferae* and more enriched in H3K4me3 and H3K27me3. In other cases, the expression changes in genes (fig. 4M) and TEs (fig. 4N) could be explained by the content of a unique epigenetic mark when hybrids are compared with one or another parental species. For example, the gene *Nepl4* (fig. 4M), involved in protein processing, had a low H3K4me3 content in hybrids versus both parental species and was underexpressed in hybrids. On the other hand, the gene named *3* (fig. 4M) and the unknown$_1$ TE (fig. 4N) were detected as underexpressed and less enriched in H3K4me3 in hybrids in comparison with *D. buzzatii*, but overexpressed and more enriched in this chromatin mark in comparison with *D. koepferae*. Finally, LTR$_2$ and DNA$_3$ were enriched in H3K27me3 in comparison with one parental species, but the opposite in comparison with the other (fig. 4N).

Additional Mechanisms Influencing Hybrid Deregulation

The divergence in *cis*- and *trans*- regulatory elements was proposed as one of the causes of expression differences between *Drosophila* species, as well as of gene deregulation in their hybrids (Wittkopp et al. 2004; Landry et al. 2005; McManus et al. 2010). To detect regulatory divergence between *D. buzzatii* and *D. koepferae* and its putative effects on hybrid expression, we compared the relative allele expression in hybrids (H), the differential expression between parental species (P), and the ratio between these two metrics (T). We categorized the genes in different classes, as in (McManus et al. 2010) and described in Material and Methods. As shown in figure 5A, most of the genes (57.65%), categorized as conserved, do not have regulatory divergence between parental species. However, we found slightly more genes showing evidence of *trans*-regulatory divergence (6.83%) than *cis*-regulatory divergence (6.24%) (Z-test, $P = 0.049$, supplementary file 1: table S10, Supplementary Material online). We also studied how the regulatory divergence influences expression differences between parental species. We found more differentially expressed genes between parental species with *trans*-regulatory divergence (Fisher's exact test, $P < 0.01$, supplementary file 1: table S11, Supplementary Material online) than expected, whereas no differences to what was expected were observed in genes with *cis*-regulatory divergence (Fisher's exact test, $P = 0.128$ for *cis*-regulatory, supplementary file 1: table S11, Supplementary Material online). Finally, we studied how the regulatory divergence observed in parents influenced the inheritance of gene expression in hybrids. We analyzed genes with regulatory divergence between *D. koepferae* and *D. buzzatii*, which were differentially expressed in hybrids versus parental species, using the expression categories described previously (fig. 1C). We found a higher number of differentially expressed genes with *trans*-regulatory divergence than with other regulatory classes in all expression categories, except those of genes deregulated in hybrids. In this latter category, we also observed a significantly higher number of genes with compensatory regulation (*cis*- and *trans*-regulatory differences compensate each other) than in the other expression categories (additive or *D. buzzatii/koepferae*-like) (Fisher's exact test, additive expression: $P = 0.039$, *D. buzzatii* and *koepferae*-like expression: $P < 0.001$, supplementary file 1: table S12, Supplementary Material online).

To gain insight into other mechanisms affecting TE deregulation in hybrids, we examined the expression of piRNA pathway genes considering their role in germline development, epigenetic regulation, and TE silencing. We observed that *aubergine* (aub) and *Sister of yellow body* (SoYb) were deregulated toward underexpression in hybrids in comparison with both parental species (supplementary file 3: fig. S5, Supplementary Material online). The chromatin mark levels were similar to the parental species, with only a few genes differentially enriched, mostly in comparison with only one parental species. Globally, differentially enriched genes decreased H3K4me3 levels in hybrids in comparison with the parental species. *Yellow body* (Yb), which was detected as overexpressed in hybrids versus both parental species, was intermediate enriched in H3K27me3 in hybrids versus both parental species, whereas *vreteno* (vret), which had an additive expression in hybrids versus parental species, was less enriched in H3K4me3 in hybrids versus both parental species. *Brother of Yellow body* (BoYb) and *armitage* (armi), which were detected as overexpressed in hybrids in comparison with one parental species, were more enriched in H3K4me3 and less enriched in H3K9me3, respectively, in comparison with the same parental species. However, most differentially enriched genes were not detected as differentially expressed (supplementary file 3: fig. S5, Supplementary Material online).

We next studied the association of piRNA amounts and TE expression in hybrids and parental species. We observed that, on average, parental species showed differences in piRNA
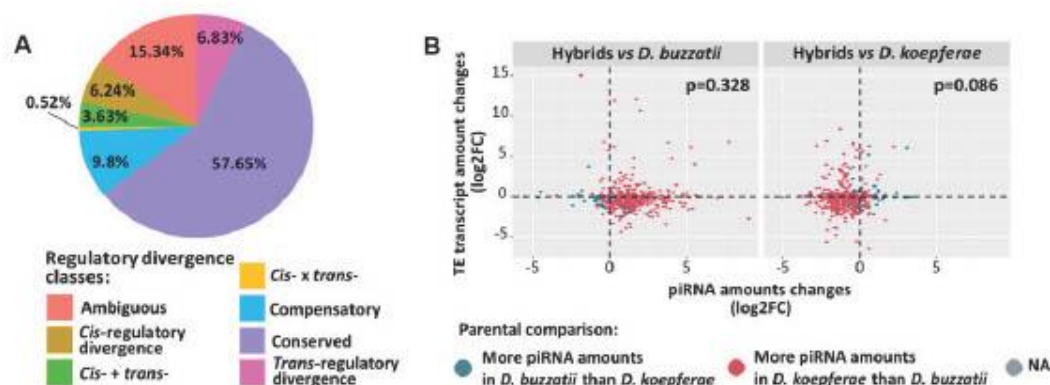
Fig. 5.—Additional factors affecting gene and TE expression, respectively: (A) *cis*- and *trans*- regulatory divergence between parental species. The pie chart represents the percentage of total genes categorized by regulatory divergence class. (B) Differential expression (log2FC) values of TEs and piRNA amounts (log2FC) in hybrids versus *D. buzzatii* and *D. koepferae*, respectively. Colors represent the results of the piRNA amount changes between parental species. Values that cannot be computed are included in the NA category. Fisher's exact test *P*-values are shown in the right corner.

amounts: TE families were associated with more piRNAs in *D. koepferae* than in *D. buzzatii* (mean log2FC of −2.14 in *D. buzzatii* vs. *D. koepferae* comparison, fig. 5B), as observed in our previous work (Romero-Soriano et al. 2017). Hybrids exhibited an additive pattern of piRNA amounts between parental species: less piRNA amounts than *D. koepferae* and more than *D. buzzatii*. Additionally, as shown in figure 5B, the differences in piRNA amounts were not associated to the changes in TE transcript amounts in hybrids in comparison with the parental species (Fisher's exact test, *P* = 0.328 for *D. buzzatii* and *P* = 0.086 for *D. koepferae*, supplementary file 1: table S13, Supplementary Material online). There were also no differences when TE classes were analyzed separately (Fisher's exact test, *P* = 0.732 for Class I and *P* = 0.564 for Class II, supplementary file 1: table S13, Supplementary Material online).

Discussion

Bias to Underexpression of Genes and TEs in Hybrid Females

The analysis of the total ovarian transcriptome of parental species showed that 5.92% and 29.64% of the protein-coding genes and TE families, respectively, were differentially expressed between *D. buzzatii* and *D. koepferae* parental species. These differences are similar to those obtained in other studies with other species of the *repleta* group (Lopez-Maestre et al. 2017). When the transcriptome of F₁ hybrids and their parental species were compared, around ~4% of genes and ~23% of TEs from hybrids were differentially expressed compared with any parental species. A greater deregulation of TEs than of genes was already observed in a previous work with *D. melanogaster*–*D. simulans* artificial hybrids (Kelleher et al. 2012), but the percentages were different: 0.7% of genes were deregulated and 12% of TE

families were overexpressed in comparison with both parental species (no data provided about underexpression). In contrast, ~78% of differentially expressed genes of *D. melanogaster*–*D. simulans* female hybrid versus their parental species were found in other studies (Ranz et al. 2004). These discrepancies could be due to the different methodological approaches used along with the different genetic background of the *Drosophila* stocks.

On the other hand, in our hybrid females most of the differentially expressed genes tended to be underexpressed, as observed in previous studies in plants (Wang et al. 2006) and in *Drosophila* females (Ranz et al. 2004). The underexpression seems also to be the rule in *Drosophila* hybrid males between species of the *melanogaster* group (Michalak and Noor 2003; Moehring et al. 2007), which have been associated to male sterility. In our case, even if most F₁ hybrid females are fertile, some are partially fertile (Marin and Fontdevila 1998), which could explain that GO terms of our underexpressed genes are associated to developmental processes, cell adhesion, and reproduction. Instead, only a few genes were extremely overexpressed in hybrids, for example the *nht*, which is involved in spermatogenesis processes. The overexpression of male-specific genes in female hybrids is not new, and has been attributed to a failure in the mechanisms controlling the expression of these genes in females (Ranz et al. 2004).

If we focus on the resemblance of gene expression in hybrids versus parental species, we found a bias toward genes that resemble more to one of the parental species, being the number of genes having an additive expression between both parental species low, as observed in previous studies with *D. melanogaster*–*D. simulans* hybrids (Ranz et al. 2004). Moreover, the number of genes in hybrids sharing similar transcript amounts with *D. koepferae* was higher than with *D. buzzatii*. Maternal effects were pointed out in *Arabidopsis lyrata* (Videvall et al. 2016), *Xenopus* (Malone and Michalak

213

2008), and in coral (Chan et al. 2021) intraspecific hybrids. Because only one direction of cross can be performed in our case, it is difficult to attribute these results unequivocally to maternal effects.

When the distribution of the derepressed genes in F_1 hybrid females along chromosomes was considered, we did not find an overrepresentation of deregulated genes in any chromosome, except in chromosome 6 (dot). This bias could be explained if this chromosome was different between our parental species, which is reinforced by the highest rate of molecular evolution found in this dot chromosome in the closely related species *D. mojavensis* (Allan and Matzkin 2019). No bias was found in a previous *Drosophila* study (Michalak and Noor 2003). However, an X-chromosome bias of differentially expressed genes was found in other studies performed in *Drosophila* hybrid males (Moehring et al. 2007; Wurmser et al. 2011). The faster evolution of X-linked genes (Coyne and Orr 1989) proposed to explain the differential expression between X and autosomal genes has, however, been questioned by other authors (Hollocher and Wu 1996) who consider that the homozygous autosome effects in reproductive isolation, approach those of X chromosome.

We showed that 6.84% of TE families were completely deregulated in F_1 hybrid females, in comparison with both parental species, and had a trend toward underexpression. These results, obtained using a different approach (normalization of TE counts using gene counts) and updated analysis tools, contrast with our previous results where the number of TE families upregulated slightly exceeded that of downregulated (Romero-Soriano et al. 2017). However, a few TE families belonging to the Gypsy superfamily showed values of expression very high compared with parental species, concordantly with previous results in these species where an increase of transposition and expression of the *Osvaldo* element was shown (Labrador et al. 1999; Vela et al. 2014; García Guerreiro 2015). Results on TE expression in hybrids reported in the literature are very heterogeneous, finding cases of overexpression of specific elements by RT-qPCR (Carnelossi et al. 2014; García Guerreiro 2015) or by RNAseq (Kelleher et al. 2012). However, examples of underexpression in hybrids, affecting most (Renaut et al. 2014) or a high percentage of TEs were also reported (Dion-Côté et al. 2014), but usually results of underexpression, if they exist, are poorly discussed. Finally, no evidence of TE reactivation was found in natural hybrid lineages of *Saccharomyces*, suggesting that other factors like population structure and hybrid genotype are major determinants of TE content (Hénault et al. 2020). The finding of underexpressed TE families in our hybrids highlights that regulation of some TEs exits in a way. Because cases of overexpressed TE families were also observed, we suggest that deregulation processes could be closely related to the TE family and the genetic background of species involved in the hybridization processes.

The Epigenomic Landscapes of Parental Species and Hybrids Are Similar to Other *Drosophila* Species

Gene and TE deregulation in *D. buzzatii* and *koepferae* hybrids were previously described in the literature (Soto et al. 2008; Vela et al. 2014; García Guerreiro 2015; Romero-Soriano and Garcia Guerreiro 2016; Romero-Soriano et al. 2017; Gámez-Visairas et al. 2020), as well as in other hybrids of *Drosophila* (Moehring et al. 2007; Ortíz-Barrientos et al. 2007; Satyaki et al. 2014) and of other organisms (De Araujo et al. 2005; He et al. 2013; Zhu, Greaves, et al. 2017). To get insight about these genomic deregulation mechanisms, we studied the epigenomic landscape of a euchromatic mark (H3K4me3) and two heterochromatic marks (H3K9me3 and H3K27me3) in *D. buzzatii*, *D. koepferae*, and their F_1 hybrids, and we found a high similarity across these species. Our results do not globally differ from those of the *D. melanogaster* species epigenome described in the literature and are consistent with the reported high conservation of the active chromatin epigenome landscapes across *Drosophila* species (Brown and Bachtrog 2014). As reported in other studies (modENCODE Consortium et al. 2011; Yin et al. 2011; Boros 2012; Parey et al. 2019), we found the H3K4me3 active chromatin mark located at the 5′ ends of actively transcribed genes and a depletion of H3K9me3 and H3K27me3 in the gene body. H3K9me3 was previously described to be enriched in promoters but depleted in 5′ transcribed regions of active genes (Yin et al. 2011; Boros 2012), whereas the H3K27me3 was underrepresented in the gene body (Filion et al. 2010; Yin et al. 2011; Boros 2012). However, both were reported to be enriched in transposons and repetitive sequences in *Drosophila* (Yin et al. 2011; Parey et al. 2019) and other organisms (Walter et al. 2016; Montgomery et al. 2020).

We next investigated whether the genomic expression was explained by the epigenetic marks, finding that globally the gene expression was positively correlated with the active mark and was negatively correlated with the H3K27me3 repressive mark, concordantly with previous studies in *Drosophila* (Yin et al. 2011) and plants (He et al. 2010). Correlations between the epigenomic landscape and gene expression in *Drosophila* have been previously described in other studies (Schübeler et al. 2004; Yin et al. 2011). According to our model, 62% of gene expression was explained by epigenetic marks, being H3K27me3 the one that most influence gene expression, followed by H3K4me3 as observed in other *Drosophila* studies (Yin et al. 2011). H3K4me3 was positively associated with gene expression, consistent with its enrichment in actively transcribed genes found in other works (modENCODE Consortium et al. 2011; Yin et al. 2011; Boros 2012; Parey et al. 2019). H3K27me3 was found to be negatively correlated to mRNA levels, consistent with its reported function of binding target for Polycomb repressive complex 1 (PRC1) (Riddle et al. 2011; Yin et al. 2011; Boros 2012; Parey et al.

2019). Surprisingly, the epigenetic mark H3K9me3, usually associated to heterochromatin regions (Yin et al. 2011), was positively associated with the gene expression, even if its contribution is very low. Increases of this epigenetic mark, along with other euchromatic ones, have been described in intronic and 3′ end regions of some active heterochromatic genes in some stages of *D. melanogaster* (Riddle et al. 2011; Saha et al. 2019).

The epigenetic marks found on TEs explained more than 60% of their expression, being H3K9me3 and H3K4me3 the most influential, with a negative and positive association to TE expression, respectively. H3K9me3 together with H3K27me3, are known to be abundant in TEs and are associated to their silencing (Walter et al. 2016; Parey et al. 2019) and H3K4me3 to their activation in *Drosophila* (Riddle et al. 2011; Yin et al. 2011) and other organisms (Walter et al. 2016). The different association of H3K27me3 with some species (positively correlated with TE expression in *D. buzzatii* and hybrids) could be explained by the different chromatin marks associated to each TE copy inside the same family.

Hybrids Exhibited Limited Changes in Histone Marks

We studied differences in the enrichment of three chromatin marks (H3K4me3, H3K9me3, and H3K27me3) in genes and TEs between the parental species *D. buzzatii*, *D. koepferae*, and their hybrids, constituting a pioneer study in the *Drosophila* genus. We found that most of the genomic regions in hybrids show similar histone modification patterns compared with the parental species. This is consistent with the maintenance of the parental histone modification patterns, as well as their additive inheritance in intraspecific (Moghaddam et al. 2011; Dong et al. 2012; Zhu, Greaves, et al. 2017) and interspecific (Zhu, Hu, et al. 2017) plant hybrids described in the literature. However, we also observe that 1.40–2.83% of genes and 1.37–9.27% of TE families are differentially enriched in hybrids compared with any parental species. H3K4me3 seems to be related with most gene expression changes, consistent with the findings in rice and maize hybrids (He et al. 2010, 2013) where gene expression changes were correlated to the enrichment of this epigenetic mark.

In the case of TEs, TE family expression changes in hybrids in comparison with *D. buzzatii* were related to their corresponding enrichment changes in H3K4me3, H3K9me3, and H3K27me3, whereas they were related only to H3K9me3 and H3K27me3 enrichment changes when hybrids were compared with *D. koepferae*. These results are in disagreement with previous results reported in interspecific *Arabidopsis* hybrids, where changes in H3K9me2 and H3K27me3 do not coincide with the TEs having their expression changed (Göbel et al. 2018). These differences could be explained by the small number of differentially expressed TEs found in their hybrids and the use of a different organism and methodology.

Because in our study, we cannot distinguish individual TE copies, they were analyzed at a family level (658 TE families), meaning that changes in expression or chromatin mark amounts are the result of the addition of the different copies per family. Indeed, intraspecific and interspecific variations in histone marks were observed in *Drosophila* TE copies (Rebollo et al. 2012), indicating that specific epigenetic modifications in TE individual copies in our hybrids could go unnoticed. Further inspection showed that increases of H3K4me3 were associated with increases of H3K9me3 and H3K27me3 in hybrids. Changes in the epigenetic status in individuals submitted to other stress were already observed in *Bari-Jheh* TE: a H3K9me3 dominant pattern turned to increases in H3K4me3, H3K9me3, and H3K27me3 enrichments (Guio et al. 2018).

Other Mechanisms May Have a Role in Hybrid Genome Deregulation

Even though in hybrids we found a relationship between epigenetic status and expression changes in genes and TEs, the histone modification patterns did not account for the whole genome deregulation. To get more insight into this aspect, we studied the regulatory divergence between *D. buzzatii* and *D. koepferae*, and its contribution to gene expression differences between them and their hybrids. In general, we found that most genes did not have regulatory divergence between parental species, which is consistent with the small percentage (5.92%) of differentially expressed genes found between *D. buzzatii* and *D. koepferae*. We also detected more *trans*-regulatory than *cis*-regulatory divergence between these species, as well as a higher number of differentially expressed *trans*-regulated genes. These results are in concordance with the high *trans*-regulatory divergence reported between *D. melanogaster* and *D. sechellia* species (McManus et al. 2010) and the association of this class with gene differential expression between parental species, but opposite to other studies in *D. melanogaster* and *D. simulans* (Wittkopp et al. 2004; Graze et al. 2009). When we studied how the regulatory divergence affects the gene inheritance in hybrids, we found that most of the differentially expressed genes in hybrids versus parental species had *trans*-regulatory divergence. These results are in contrast to what was reported in the additive expressed genes, where *cis*-regulatory changes were more frequent than *trans*-regulatory changes (Lemos et al. 2008). Finally, we also found a high number of deregulated genes in hybrids with *cis–trans* compensatory evolution, which was also previously reported in *D. melanogaster* and *D. simulans* hybrids and considered an important cause of hybrid deregulation (Landry et al. 2005). The differences reported, both between previous works and our own, could be due to use of different *Drosophila* species, different methodologies and the use, in this work, of ovarian tissues versus whole adults.

215

We as well examined the expression of piRNA pathway genes for their role in germline TE silencing and we found that *aub* and *SoYb* were both underexpressed, whereas Yb was overexpressed in hybrids versus both parental species. Results of expression of these genes showed the same general trend as observed in a previous study with these species using a different analysis approach of RNAseq data (Romero-Soriano et al. 2017). Nevertheless, overexpression of three additional piwi pathway genes in the same hybrids was detected by RT-qPCR in our previous work (Gámez-Visairas et al. 2020), suggesting that the activation of these genes could be a primary response to the hybridization stress. Discrepancies found between RT-qPCR and RNAseq results highlight the different sensibilities of these two techniques. In addition, we found that the chromatin mark levels of the piRNA pathway genes was similar to the parental species, with only differences in a few genes, which is consistent to the general trend observed in other genes. In addition, we found that the piRNA amount changes were not associated with TE expression changes in hybrids, because we found an intermediate inheritance of piRNAs between the parental species in the hybrid, being *D. koepferae* the parental species with the highest amount, as observed in our previous work (Romero-Soriano et al. 2017).

We suggest that changes in transcript amounts in hybrids are either the result of the enrichment/impoverishment of a specific mark or the disproportion in the active/repressive mark content, together with other mechanisms such as *cis-trans* compensatory regulation. Three histone marks, considered as relevant to the expression in *Drosophila*, were considered in this work, but we cannot discard the effect of other epigenetic marks like H3K9ac highly positively correlated to gene expression (Yin et al. 2011). Other factor that could affect gene and TE deregulation is asynapsis, frequently observed in our hybrids and reported in other *repleta* hybrids (Naveira and Fontdevila 1991). Asynapsis, is known for influencing the *trans* regulation of *Ultrabithorax* gene alleles in *Drosophila* (Goldsborough and Kornberg 1996), for contributing to mouse intersubspecific hybrids infertility (Bhattacharyya et al. 2013) and to the female meiotic losses in mammals (Cloutier et al. 2015).

Conclusions

The hybridization between *D. buzzatii* and *D. koepferae* species promotes a higher deregulation in TE families than in genes, both toward underexpression. The epigenome of the parental species is in general highly preserved in the hybrids, but some changes of the parental chromatin landscape are also observed in hybrids and are associated with their new gene and TE family expression patterns. Finally, *cis–trans* compensatory regulation could also be involved in expression deregulation of some genes. The present study has contributed to a better understanding of the mechanism affecting genomic deregulation in hybrids. Nevertheless, we cannot discard additional mechanisms, resulting from the incompatibility of the two different paternal genomes in the hybrids, which could interact forming a complex gene network and contribute to the deregulation patterns observed. This and the fact that this study is limited to hybrid females, makes that additional studies are necessary to go deeper into a better knowledge of the regulatory mechanisms and the factors involved in hybrid male sterility.

Materials and Methods

Drosophila Stocks and Crosses

We used *D. buzzatii* Bu 28 and *D. koepferae* Ko2 inbred strains described in our previous works (Vela et al. 2014; García Guerreiro 2015; Romero-Soriano et al. 2016; Romero-Soriano and Garcia Guerreiro 2016). Both strains were maintained by brother–sister mating for at least a decade and then by mass culturing. We performed ten different interspecific crosses of ten *D. buzzatii* males to ten *D. koepferae* females, due to the scarce offspring obtained in interspecific crosses. All stocks and crosses were reared at 25 °C in a standard *Drosophila* medium supplemented with yeast.

Chromatin Preparation

Ovaries of 5-day-old *Drosophila* females were dissected in PBT (1× phosphate-buffered saline and 0.2% Tween 20). We performed chromatin extraction from two biological replicates of 50 ovaries per sample of parental species and 70 ovaries for hybrids. Samples were resuspended in Buffer A1 (HEPES 15 mM, sodium butyrate 10 mM, KCl 60 mM, Triton x100 0.5%, NaCl 15 mM) plus 1.8% formaldehyde, homogenized with a Dounce tissue grinder (15 times) and incubated for 10 min at room temperature. The crosslink was then stopped with glycine to a final concentration of 125 mM. Samples were subsequently incubated 3 min, kept on ice and washed three times with Buffer A1. We then added 0.2 ml of lysis buffer (HEPES 15 mM, EDTA 1 mM, EGTA 0.5 mM, sodium butyrate 10 mM, SDS 0.5%, sodium deoxycholate 0.1%, *N*-lauroylsarcosine 0.5%, Triton x100 1%, and NaCl 140 mM) and incubated 1–2 h at 4 °C. After the lysis process, we sonicated the samples using Biorruptor pico sonication device from Diagenode: 32 cycles of 30 s ON/30 s OFF, for parental species, and 35 cycles of 30 s ON/30 s OFF for hybrids. The sheared crosslinked chromatin was recovered from the pellet after a spin step at 10,000 g 4 min at 4 °C.

To check DNA size, samples were previously de-crosslinked with NaCl 5M, boiled 15 min and treated with 1 µl of 10 mg/ml RNAse A (37 °C, 30 min). They were purified with phenol-chloroform, precipitated in absolute ethanol, washed in 70% ethanol, resuspended in 20 µl of DEPC and run in a 1.5%-agarose gel.

For the immunoprecipitation step, Magna ChIPTM G chromatin immunoprecipitation Kit (Millipore) was used together with antibodies against H3K4me3 (Abcam; ab8580), H3K9me3 (Abcam; ab8898), and H3K27me3 (Abcam; ab6002). We separated 20 µl of chromatin for input and the remaining 180 µl were distributed in three aliquots where 1 µl of each antibody plus 20 µl of magnetic beads and dilution buffer up to 530 µl were added. Samples were incubated overnight at 4 °C in an agitation wheel. After the beads were washed with low salt buffer, high salt buffer, LiCl buffer, and TE buffer. Separation of chromatin from the beads was performed using 0.5 ml of CHIP elution buffer. One microliter of Proteinase K was then added to each sample and incubated at 62 °C for 2 h in a shaker at 300 rpm. Samples were then purified with the columns provided by the kit and stored at −20 °C. ChIP enrichment was quantified by real-time PCR of a well-known genomic region enriched for the different histone marks studied: *rp49* for H3K4me3, *kkv* for H3K9me3, and *light* for H3K27me3. The following gene-specific primers were used: rp49-forward: 5′-GTCGTCGCTTCAAGGGCCAA T-3′, rp49-reverse: 5′-ATGGGCGATCTCACCGCAGTA-3′, kkv-forward: 5′-TAATCCAGCCACGCCCATTT-3′, kkv-reverse: 5′-CCCAACGTTTGCATTGCTGA-3′, light-forward: 5′-CGAGTACAAAATGAATAGCTCCG-3′, light-reverse: 5′-GCGGTTCTCCTCAATGAT-3′.

Chromatin Sequencing

Duplicate Truseq ChIP libraries, corresponding to two biological replicates per sample, were performed by Macrogen Inc., Seoul, Korea. Sequencing was carried out using an Illumina Hiseq4000. We obtained 22–34 millions of paired-end reads for each sample, resulting in a total of 659.9 millions of paired-end reads.

ChIPseq Visualization Patterns

ChIPseq sequenced reads were trimmed using Trimmomatic software v0.39 (Bolger et al. 2014) and aligned to the *D. buzzatii* genome downloaded from the *Drosophila buzzatii* Genome Project web page (http://dbuz.uab.cat, last accessed January 7, 2015) using Bowtie2 v2.3.5.1 (Langmead and Salzberg 2012). For the alignment, the default parameters of the --very-sensitive-local mode with two extra-modifications to increase the sensibility (--local -D 20 -R 3 -L 20 -N 1 -p 12 --gbar 3 --mp 5,1 --rdg 4,2 --rfg 4,2) were used to reach the highest percentage of aligned reads with both parental species and their F₁ hybrids. Reads with map quality score lower than 30 and unmapped reads were filtered using Samtools v1.10 (Li et al. 2009) and excluded from further analysis. Deeptools v3.3.2 (Ramírez et al. 2016) was used to visualize the enrichment of each chromatin mark around the SC and the end codon (EC). First, *bamCompare* was used to normalize the ChIPseq samples by depth using the RPKM method and by the input (control). Finally, the read density

values were computed using *computeMatrix* and visualized with *plotHeatmap*.

Gene Alignments

RNAseq reads from Romero-Soriano et al. (2017) were treated the same way as ChIPseq reads to ensure that results were comparable. Gene sequences (only body region: from the SC to the EC) from *D. buzzatii* were obtained with *getfasta* of Bedtools v2.29.2 (Quinlan and Hall 2010) using the genome sequence and its gene annotations (Guillén et al. 2014). Gene sequences were masked using RepeatMasker v4.1.1. (http://www.repeatmasker.org) and the Repeat Masker from the *D. buzzatii* browser (http://dbuz.uab.cat, last accessed January 7, 2015). A total of 37 genes were completely masked and excluded from further analysis, and a total of 13,621 protein coding genes were included in the reference sequences.

To ensure that there was no bias due to the use of only the *D. buzzatii* reference genome, a de novo transcriptome for each parental species was created using Trinity v2.9 (Grabherr et al. 2011) and the corresponding RNAseq. *Drosophila buzzatii* and *D. koepferae* transcriptomes were then aligned against all the ChIPseq inputs (supplementary file 4: Alignment_inputChIPSeq, Supplementary Material online). We also randomly selected 40 genes amongst Trinity outputs, and computed nucleotide divergence between parental species using the Mega software (Tamura et al. 2021) and the Jukes-Cantor model (Jukes and Cantor 1969) (supplementary file 4: Nucleotide_Divergence, Supplementary Material online). Both, the similar alignment percentages and the low average divergence, indicate that bias of using *D. buzzatii* as reference genome, if any, is marginal. Nevertheless, we decided to be conservative and use only the protein coding genes. Trimmed RNAseq and ChIPseq reads were aligned to the *D. buzzatii* masked gene sequences using Bowtie2 v2.3.5.1 (Langmead and Salzberg 2012) and the parameters explained above. eXpress v1.5.1 (Roberts and Pachter 2013) with the default options and then an additional online EM round (to increase the accuracy) was used to quantify read counts for each gene. All isoforms of a gene were considered together. We used rounded effective counts for the following steps.

TE Alignment

TE RNAseq (Romero-Soriano et al. 2017) and ChIPseq read counts were analyzed using the TEcount module of the TEtools pipeline (Lerat et al. 2017) (available at https://github-com/l-modolo/TEtools). First, the manually constructed TE library containing consensus TEs from both *D. buzzatii* and *D. koepferae*, described in a previous work (Romero-Soriano et al. 2017), was aligned with the RNAseq and ChIPseq reads, using Bowtie 2 v2.2.4 (Langmead and Salzberg 2012) with the most sensitive option and keeping a single alignment for

reads mapping to multiple positions (–very-sensitive). Read counts were computed per TE family, adding all reads mapped on copies from the same family. Count tables corresponding to genes and TEs were concatenated and were used for the differential expression and enrichment analyses. Genes counts were used to normalize TE counts, following the guidelines of TEtools pipeline (Lerat et al. 2017).

Differential Expression and Enrichment Analyses

The statistical analyses were performed using R v4.0 (R Core Team 2020). The visualization of the results was performed using the R package ggplot2 v3.3.2 (Wickham 2016). The DESeq2 function from the R Bioconductor package DESeq2 v1.28.1 (Love et al. 2014) was used to normalize read counts, using the default method, and to model the read counts using a negative binomial distribution.

For RNAseq, the DESeq2 package was used to identify differentially expressed genes and TE families while performing a Wald test (Love et al. 2014). The P-values were adjusted for multiple testing using the procedure of Benjamin and Hochberg (Benjamini et al. 2001) with an FDR cutoff of 0.05, and were obtained using the results function from the DESeq2 package. The log2FC was shrunken using the default and recommended apeglm algorithm (Zhu et al. 2019) of the lfcShrink function. Genes with an adjusted P-value lower than 0.05 and at least double of expression in one species above the other (absolute shrunken log2FC > 1) were considered as differentially expressed. GO term enrichment analyses of biological processes were performed for the underexpressed and overexpressed significant genes using the topGO R package v.2.42.0. (Alexa and Rahnenfuhrer 2020) ("weight01" algorithm and Fisher's statistic) and the ortholog genes in D. melanogaster obtained from the D. buzzatii Genome Project web page (http://dbuz.uab.cat, last accessed January 7, 2015). The obtained Fisher's exact test P-values were not adjusted but, from the total gene ontologies with P< 0.05, only the top $\frac{1}{3}$ with the lowest P-values were considered as enriched.

Regarding the ChIPseq data, a handmade script was used to perform the differential enrichment analysis between hybrids and both parental species. Briefly, the "Regularized log" transformed counts obtained using DESeq2 for each chromatin mark were considered and analyzed using the following linear model for each gene: RldKm ~ RldKi + species (RldKm: log-transformed counts for the considered histone mark, RldKi: log-transformed counts for the input). We considered contrasts between pairs of species modalities (D. buzzatii, D. koepferae, hybrid). Because of the high number of tests and the lack of extreme P-values, the Benjamin and Hochberg (Benjamini et al. 2001) adjustment of the P-values had a high effect in our results. For this reason, the P-values were not adjusted and, from the genes and TE families with P< 0.05, the rounded $\frac{1}{3}$ of the top genes and

TE families with the lowest P-values were considered as differentially enriched.

The significant genes were assigned to their respective chromosomes following the scaffold assignation to chromosomes obtained from a previous work (Guillén et al. 2014) to detect potential chromosome biases of differentially expressed genes. Additionally, the order of the significant TE families was analyzed to detect possible order biases in our significant TEs.

Within each genome, we quantified the contribution of histone mark enrichments on transcript amounts using the following linear model on log-transformed read counts for each gene and TE family: RNA ~ K4 + K9 + K27 + Input. In addition, we tested whether changes in RNAseq counts in hybrids were associated with changes in ChIPseq counts using Fisher's exact test on 2 × 2 matrices: genes and TE families were classified as displaying a positive or negative log2FC of RNAseq counts in hybrids versus parents and as displaying an increase or decrease in histone mark enrichment in hybrids versus parents.

Allele-Specific Expression Analysis

We created a de novo transcriptome with the RNAseq of both parental species, together with the hybrids using Trinity v2.9 (Grabherr et al. 2011). A SuperTranscript (Davidson et al. 2017) was then created as a general reference transcriptome. The GATK pipeline for variant calling (Van der Auwera and O'Connor 2020) was used to detect differences between the general reference transcriptome and each parental species. The VCF files were filtered following the GATK guidelines, including a coverage depth of at least 10, and the exclusion of variants only present in one replicate, which were considered assembly errors. These variants were replaced in the general reference transcriptome using the FastaAlternateReferenceMaker GATK tool (Van der Auwera and O'Connor 2020) to create a reference transcriptome for each parental species.

HISAT2 (Kim et al. 2015) with the "no-softclip" option was used to align the RNAseq from hybrids to both parental reference transcriptome. Then, CompareBams (Lindenbaum 2015) of Jvarkit was used to compare the alignments and FilterSamReads from Picard (Broad Institute 2018) to filter out the reads aligning in different position when the data were aligned to each parental species. Samtools v1.10 (Li et al. 2009) was used to remove multimapped and unmapped reads and BamTools v.2.4.0 (Barnett et al. 2011) to keep only reads that align without mismatches. Bedtools multicov (Quinlan and Hall 2010) and a manually updated GTF were used to count reads aligning to each gene.

The reference transcriptomes were annotated using BLAT v.35x1 (Kent 2002) with the parameters –minIdentity = 80 and –maxIntron = 75,000 and the gene sequences of D. buzzatii, keeping the best match with an overlap of at least

50%. Subsequently, statistical analyses were performed using R v4.0 (R Core Team 2020) and DESeq2. The *DESeq2* function was used to normalize read counts, using the default method. First, the *results* function was used to compute the log2FC and the adjusted *P*-values using the Benjamin and Hochberg (Benjamini et al. 2001) method. A FDR cutoff of 0.05 was used to identify differentially expressed genes 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 analyzed for *trans*-regulatory effects (T) by comparing the P and H ratios with the same *results* function. Genes were then categorized in the following groups as reported in a previous work (McManus et al. 2010):

- *Cis*- only: Significant differential expression in P and H, but no significant T.
- *Trans*- only: Significant differential expression in P, and T, but not H.
- *Cis*- + *trans*-: Significant differential expression in P, H, and T. *Cis*- and *trans*- regulatory differences favor expression of the same allele.
- *Cis*- × *trans*-: Significant differential expression in P, H, and T. *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 with no clear biological interpretation.

Piwi Pathway Genes and piRNA Amount Study

Additionally, reciprocal tblast v2.10.1. (Camacho et al. 2009) of the *D. melanogaster* piRNA pathway proteins from UniProt (The UniProt Consortium 2021) and the reference gene sequence list were performed. A total of 31 proteins were detected and associated with *D. buzzatii* genes. The *Argonaute 3* gene (*ago3*), not included in the reference genes, was manually included in the list from our previous results (Gámez-Visairas et al. 2020).

Small RNAseq raw reads from our previous work (Romero-Soriano et al. 2017) were used to study the piRNA regulatory data in the TE expression study. Using *PRINSEQ lite* (Schmieder and Edwards 2011), we isolated 23–30 nt-long reads and considered them as piRNAs. For normalization purposes, we also isolated 20–23 nt-long reads and searched for microRNA sequences: low-quality reads were removed using UrQt (Modolo and Lerat 2015), then trimmed 20–23 nt reads were aligned to the masked genes using Bowtie V.1.3.0

(Langmead et al. 2009) and keeping a single alignment for reads mapping to multiple positions. These reads were considered to correspond to microRNAs and counts were computed using eXpress v1.5.1 (Roberts and Pachter 2013). TE counts among piRNAs were computed using the TEcount module of TEtools (Lerat et al. 2017). piRNA counts were then normalized so that the sum of microRNA counts is the constant across samples.

Regarding TE sequences, RNAseq and CHIPseq data were integrated following the same approaches and processes as for genes. To get insight into the chromatin mark combination, we study the association between the euchromatic mark H3K4me3 and the heterochromatic marks (H3K9me3 and H3K27me3) in hybrids versus each parental species.

Statistical Tests

Three main statistical tests were used in the article and were performed using the R v4.0 (R Core Team 2020) program:

- *Two proportion Z-test* was used to compare distributions of significant sequences (genes and TE families) across comparisons and *cis*- and *trans*-regulatory classes.
- *Chi-square test* under equal assumption was used to detect chromosome-biases and TE-category-biases.
- *Fisher's exact test* under independence assumption was computed using a 2×2 contingency table to detect associations in: gene and TE expression (log2FC) and chromatin mark enrichment (size effect); TE euchromatic mark (H3Kme3) and heterochromatic marks (H3K9me3 and H3K27me3); *trans*- and *cis*-regulatory divergence and differences of expression between parental species; compensatory and remaining classes and gene deregulation in hybrids, and TE expression and piRNA quantities, in hybrids versus each parental species independently.

The results were corrected for multiple testing using the Benjamin and Hochberg (Benjamini et al. 2001) method.

Supplementary Material

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

Acknowledgments

Author Contributions

Data Availability

The data are included in the article and in its Supplementary Material online. Raw ChIPseq data generated from this article have been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA796032: accession numbers from SRR17535990 to SRR17536013.

Literature Cited

Alexa A, Rahnenfuhrer J. 2020. topGO: enrichment analysis for gene ontology. R package version 2.42.0. Available from: https://bioconductor.org/packages/release/bioc/html/topGO.html

Allan CW, Matzkin LM. 2019. Genomic analysis of the four ecologically distinct cactus host populations of Drosophila mojavensis. BMC Genomics. 20(1):732.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27(12):1691–1692.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 125(1–2):279–284.

Bhattacharyya T, et al. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. Proc Natl Acad Sci USA. 110:E468–E477.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Boros IM. 2012. Histone modification in drosophila. Brief Funct Genomics. 11(4):319–331.

Broad Institute. 2018. Picard toolkit. Broad Institute, GitHub Repos. Available from: https://broadinstitute.github.io/picard/. Accessed February 21, 2018.

Brown EJ, Bachtrog D. 2014. The chromatin landscape of Drosophila: comparisons between species, sexes, and chromosomes. Genome Res. 24(7):1125–1137.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1):1–9.

Carnelossi EAG, et al. 2014. Specific activation of an I-like element in Drosophila interspecific hybrids. Genome Biol Evol. 6(7):1806–1817.

Center for Research in Agricultural Genomics, National Center for Genomic Analysis, Genomics Bioinformatics and Evolution Research Group of the Universitat Autònoma de Barcelona, Barcelona Supercomputing Center. Drosophila buzzatii Genome Project. Available from: http://dbuz.uab.cat (accessed November 26, 2020).

Chan WY, Chung J, Peplow LM, Hoffmann AA, van Oppen MJH. 2021. Maternal effects in gene expression of interspecific coral hybrids. Mol Ecol. 30(2):517–527.

Cloutier JM, et al. 2015. Histone H2AFX links meiotic chromosome asynapsis to prophase I oocyte loss in mammals. PLoS Genet. 11(10):e1005462.

Coyne JA, Orr HA. 1989. Two rules of speciation. In: Otte D, Endler JA, editors. Speciation and its consequences. Sunderland (MA): Sinauer Associates. p. 180–207.

Davidson NM, Hawkins ADK, Oshlack A. 2017. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biol. 18(1):148.

De Araujo PG, et al. 2005. Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J. 44(5):707–717.

Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. Mol Biol Evol. 31(5):1188–1199.

Dong X, et al. 2012. Natural variation of H3K27me3 distribution between two Arabidopsis accessions and its association with flanking transposable elements. Genome Biol. 13(12):R117.

Ebert A, Lein S, Schotta G, Reuter G. 2006. Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res. 14(4):377–392.

Filion GJ, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143(2):212–224.

Fontdevila A. 2019. Hybrid genome evolution by transposition: an update. J Hered. 110(1):124–136.

Gámez-Visairas V, Romero-Soriano V, Martí-Carreras J, Segarra-Carrillo E, García Guerreiro MP. 2020. Drosophila interspecific hybridization causes a deregulation of the piRNA pathway genes. Genes 11(2):215.

García Guerreiro MP. 2015. Changes of Osvaldo expression patterns in germline of male hybrids between the species Drosophila buzzatii and Drosophila koepferae. Mol Genet Genomics. 290(4):1471–1483.

García Guerreiro MP. 2012. What makes transposable elements move in the Drosophila genome. Heredity 108(5):461–468.

García Guerreiro MP. 2014. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in Drosophila. Mob Genet Elements. 4:e34394.

Göbel U, et al. 2018. Robustness of transposable element regulation but no genomic shock observed in interspecific Arabidopsis hybrids. Genome Biol Evol. 10(6):1403–1415.

Goldsborough AS, Kornberg TB. 1996. Reduction of transcription by homologue asynapsis in Drosophila imaginal discs. Nature 381(6585):807–810.

Gómez GA, Hasson E. 2003. Transpecific polymorphisms in an inversion linked esterase locus in Drosophila buzzatii. Mol Biol Evol. 20(3):410–423.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV. 2009. Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics 183(2):547–561.

Guillén Y, et al. 2014. Genomics of ecological adaptation in cactophilic Drosophila. Genome Biol Evol. 7(1):349–366.

Guio L, Vieira C, González J. 2018. Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster. Sci Rep. 8:1–10.

He G, et al. 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell. 22(1):17–33.

He G, et al. 2013. Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. Genome Biol. 14(6):R57.

Hénault M, Marsit S, Charron G, Landry CR. 2020. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. *eLife* 9:e60474.

Hollocher H, Wu CI. 1996. The genetics of reproductive isolation in the *Drosophila simulans* clade: x vs. autosomal effects and male vs. female effects. *Genetics* 143(3):1243–1255.

Jukes T, Cantor C. 1969. *Evolution of protein molecules.* New York: Academic press.

Jurka J, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110(1–4):462–467.

Kelleher ES, Edelman NB, Barbash DA. 2012. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol.* 10(11):e1001428.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. *Genome Res.* 12(4):656–664.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 12(4):357–360.

Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. *Mol Biol Evol.* 16(7):931–937.

Landry CR, et al. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. *Genetics* 171(4):1813–1822.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):r25.

Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci USA.* 105(38):14471–14476.

Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. 2017. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* 45(4):e17.

Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Lindenbaum P. 2015. JVarkit: java-based utilities for Bioinformatics. figshare. doi: https://doi.org/10.6084/m9.figshare.1425030.

Lopez-Maestre H, et al. 2017. Identification of misexpressed genetic elements in hybrids between Drosophila-related species. *Sci Rep.* 7:40618–40613.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550–521.

Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends Genet.* 33(1):68–80.

Malone JH, Michalak P. 2008. Gene expression analysis of the ovary of hybrid females of *Xenopus laevis* and *X. muelleri. BMC Evol Biol.* 8(1):82.

Marin I, Fontdevila A. 1998. Stable *Drosophila buzzatii–Drosophila koepferae* hybrids. *J Hered.* 89(4):336–339.

Marin I, Ruiz A, Pla C, Fontdevila A. 1993. Reproductive relationships among ten species of the Drosophila repleta group from South America and the West Indies. *Evolution* 47(5):1616.

McManus CJ, et al. 2010. Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res.* 20(6):816–825.

Michalak P, Noor MAF. 2003. Genome-wide patterns of expression in Drosophila pure species and hybrid males. *Mol Biol Evol.* 20(7):1070–1076.

modENCODE Consortium, et al. 2011. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330:1787–1797.

Modolo L, Lerat E. 2015. UrQt: an efficient software for the unsupervised quality trimming of NGS data. *BMC Bioinformatics.* 16(1):1–8.

Moehring AJ, Teeter KC, Noor MAF. 2007. Genome-wide patterns of expression in Drosophila pure species and hybrid males. I. Examination of multiple-species hybridizations, platforms, and life cycle stages. *Mol Biol Evol.* 24(1):137–145.

Moghaddam AMB, et al. 2011. Additive inheritance of histone modifications in *Arabidopsis thaliana* intra-specific hybrids. *Plant J.* 67(4):691–700.

Montgomery SA, et al. 2020. Chromatin organization in early land plants reveals an ancestral association between H3K27me3. *Curr Biol.* 30(4):573–588.

Naveira H, Fontdevila A. 1991. The evolutionary history of *Drosophila buzzatii*. XXI. Cumulative action of multiple sterility factors on spermatogenesis in hybrids of *D. buzzatii* and *D. koepferae. Heredity* 67(1):57–72.

Naveira H, Pla C, Fontdevila A. 1986. The evolutionary history of *Drosophila buzzatii* XI. A new method for cytogenetic localization based on asynapsis of polytene chromosomes in interspecific hybrids of Drosophila. *Genetica* 71(3):199–212.

Oliveira DCSG, et al. 2012. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Mol Phylogenet Evol.* 64(3):533–544.

Ortíz-Barrientos D, Counterman BA, Noor MAF. 2007. Gene expression divergence and the origin of hybrid dysfunctions. *Genetica* 129(1):71–81.

Parey E, Crombach A, Alba M. 2019. Evolution of the *Drosophila melanogaster* chromatin landscape and its associated proteins. *Genome Biol Evol.* 11(3):660–677.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

R Core Team. 2020. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.r-project.org/.

Ramírez F, et al. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44(W1):W160–W165.

Ranz JM, Namgyal K, Gibson G, Hartl DL. 2004. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans. Genome Res.* 14(3):373–379.

Rebollo R, et al. 2012. A snapshot of histone modifications within transposable elements in Drosophila wild type strains. *PLoS One.* 7(9):e44253.

Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Philos Trans R Soc Lond B Biol Sci.* 369 (1648):20130345.

Riddle NC, et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* 21(2):147–163.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods.* 10(1):71–73.

Romero-Soriano V, et al. 2016. Drosophila females undergo genome expansion after interspecific hybridization. *Genome Biol Evol.* 8(3):556–561.

Romero-Soriano V, et al. 2017. Transposable element misregulation is linked to the divergence between parental piRNA pathways in Drosophila hybrids. *Genome Biol Evol.* 9(6):1450–1470.

Romero-Soriano V, Garcia Guerreiro MP. 2016. Expression of the retrotransposon helena reveals a complex pattern of TE deregulation in Drosophila hybrids. *PLoS One.* 11(1):e0147903.

Saha P, Sowpati DT, Mishra RK. 2019. Epigenomic and genomic landscape of *Drosophila melanogaster* heterochromatic genes. *Genomics* 111(2):177–185.

Satyaki PRV, et al. 2014. The *Hmr* and *Lhr* hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet.* 10(3):e1004240.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.

Schübeler D, et al. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* 18(11):1263–1271.

Schulze SR, Wallrath LL. 2007. Gene regulation by chromatin structure: paradigms established in *Drosophila melanogaster*. *Annu Rev Entomol.* 52:171–192.

Soto EM, Soto IM, Carreira VP, Fanara JJ, Hasson E. 2008. Host-related life history traits in interspecific hybrids of cactophilic Drosophila. *Entomol Exp Appl.* 126:18–27.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 38(7): 3022–3027.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49:D480–D489.

Van der Auwera G, O'Connor B. 2020. *GATK, and WDL in terra genomics in the cloud: using Docker.* 1st ed. Sebastopol (CA): O'Reilly Media, Inc.

Vela D, Fontdevila A, Vieira C, García Guerreiro MP. 2014. A genome-wide survey of genetic instability by transposition in Drosophila hybrids. *PLoS One* 9(2):e88992.

Vela D, García Guerreiro MP, Fontdevila A. 2011. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. *Biotechniques* 50(4): 247–250.

Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. 2016. Strong maternal effects on gene expression in *Arabidopsis lyrata* hybrids. *Mol Biol Evol.* 33(4):984–994.

Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* 5:1–30.

Wang J, et al. 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. *Genetics* 172(1):507–517.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis.* New York: Springer-Verlag. Available from: https://ggplot2.tidyverse.org.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430(6995):85–88.

Wurmser F, et al. 2011. Population transcriptomics: insights from *Drosophila simulans, Drosophila sechellia* and their hybrids. *Genetica* 139(4):465–477.

Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2011. A high-resolution whole-genome map of key chromatin modifications in the adult *Drosophila melanogaster*. *PLoS Genet.* 7(12):e1002380.

Zhu A, Greaves IK, Dennis ES, Peacock WJ. 2017. Genome-wide analyses of four major histone modifications in Arabidopsis hybrids at the germinating seed stage. *BMC Genomics.* 18(1):1–9.

Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 35(12):2084–2092.

Zhu W, Hu B, et al. 2017. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. *Genome Biol.* 18(1):1–16.

Associate editor: Andrea Betancourt