



**HAL**  
open science

# Impact of Heat Stress on Transposable Element Expression and Derived Small RNAs in *Drosophila subobscura*

Alejandra Bodelón, Marie Fablet, Daniel Siqueira de Oliveira, Cristina Vieira,  
Maria Pilar García Guerreiro

► **To cite this version:**

Alejandra Bodelón, Marie Fablet, Daniel Siqueira de Oliveira, Cristina Vieira, Maria Pilar García Guerreiro. Impact of Heat Stress on Transposable Element Expression and Derived Small RNAs in *Drosophila subobscura*. *Genome Biology and Evolution*, 2023, 15 (11), pp.evad189. 10.1093/gbe/evad189 . hal-04276658

**HAL Id: hal-04276658**




**<https://hal.science/hal-04276658v1>**

Submitted on 4 Sep 2024

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

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

# Impact of Heat Stress on Transposable Element Expression and Derived Small RNAs in *Drosophila subobscura*

Alejandra Bodelón<sup>1</sup>, Marie Fablet <sup>2,3</sup>, Daniel Siqueira de Oliveira<sup>2,4</sup>, Cristina Vieira <sup>2</sup>, and Maria Pilar García Guerreiro <sup>1,\*</sup>

<sup>1</sup>Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Departament de Genètica i Microbiologia (Edifici C), Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>2</sup>Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon; Université Lyon 1; CNRS; UMR 5558, Villeurbanne, France

<sup>3</sup>Institut universitaire de France, Paris, France

<sup>4</sup>Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (Unesp), São Paulo, Brazil

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

Accepted: October 09, 2023

## Abstract

Global warming is forcing insect populations to move and adapt, triggering adaptive genetic responses. Thermal stress is known to alter gene expression, repressing the transcription of active genes, and inducing 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 *Drosophila subobscura* populations, 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 heat stress, sex-specific changes in gene expression together with a trend toward over-expression, mainly of heat shock response-related genes, were observed. We did not observe large changes of TE expression nor small RNA production due to stress. Only population and sex-specific expression changes of some TE families (mainly retrotransposons), or the amounts of siRNAs and piRNAs, derived from specific TE families were observed, as well as the piRNA production from some piRNA clusters. Changes in small RNA amounts and TE expression could not be clearly correlated, indicating that other factors as 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*.

**Key words:** *Drosophila*, stress, transposable elements, piRNAs, RNA-seq, 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 (TE) 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.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Introduction

Temperature is a well-known stressful factor that can alter gene expression in *Drosophila* (Pauli et al. 1992) and other species (Srikanth et al. 2020). High temperatures have been observed to repress the transcription of several genes and activate that of others, mostly those related to the stress response (Tanguay 1983). When the organisms return to optimal temperatures, their cellular activities are gradually restored at rate depending on the severity of the stress (Pauli et al. 1992). The most important group of genes up-regulated after a heat stress and other stressful factors are those coding heat shock proteins (Hsps) (Sørensen et al. 2003). Hsps are molecular chaperones that modulate the structure and folding of other proteins (Tower 2011). They play important roles in transport, signal pathways, and activation of enzymes and receptors (Parsell and Lindquist 1993). These functions provide the organism a temporary enhanced tolerance to stress (Sørensen et al. 2003). Even though the Hsps have been widely studied (Tanguay 1983), the global effects of high temperatures on the genomes are poorly understood (Sørensen et al. 2005).

Some authors also associated the heat stress to an increase of transposition of some specific transposable element (TE) families in *Drosophila* (Junakovic et al. 1986; Vasilyeva et al. 1999; Kofler et al. 2018), such as *copia* (Strand and McDonald 1985) or *412* (Bubenshchikova et al. 2002), but a lack of mobilization was found in other cases (Arnault et al. 1991, 1997). In *Drosophila* somatic cells, some TEs are silenced by endogenous small interfering RNAs, endo-siRNAs (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008). In the ovarian somatic and germ cells, transposition is mainly controlled by piwi-interacting RNAs (piRNAs) (Brennecke et al. 2007; Li et al. 2009; Malone et al. 2009; Luo and Lu 2017), even though the siRNA pathway is also active (Czech et al. 2008). Most of gonadal piRNAs are produced from specific TE-enriched *loci*, called piRNA clusters, that cover around 3.5% of the *Drosophila melanogaster* genome (Brennecke et al. 2007). These *loci* are transcribed from one or two DNA strands into long piRNA precursors, which once processed give rise to trailing piRNAs (Li et al. 2009; Malone et al. 2009; Ozata et al. 2019). In the germline, trailing piRNAs are amplified in a ping-pong amplification cycle, producing initiator or responder piRNAs (Ozata et al. 2019), which magnifies the TE repression (Brennecke et al. 2007; Gunawardane et al. 2007), a process that has been observed to be more relevant in females than in males (Saint-Leandre et al. 2020). On the other hand, it is also known that the Hsps Hsp70 (Cappucci et al. 2019) and Hsp83 (Specchia et al. 2010) 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 (Funikov et al. 2015), but not others (Cappucci et al. 2019). However, no correlation between changes in piRNA levels and TE transcripts has been observed under heat stress (Funikov et al. 2015; Cappucci et al. 2019).

*Drosophila subobscura* is a species of the *obscura* group of the subgenus *Sophophora* (Collin 1936). 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 (Galludo et al. 2018). This species has a rich inversion polymorphism in most of its six chromosomes (A, J, U, E, O, and dot), with more than 65 identified inversions, most of them located on the O chromosome (Karageorgiou et al. 2019). Because some of them have adaptive roles, showing repeatable spatiotemporal patterns in frequencies related to temperature (Rego et al. 2010), they can be used to monitor global warming (Balanyá et al. 2006). 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 (Calabria et al. 2012). In contrast, comparable basal Hsp70 mRNA levels in populations bearing different O chromosomal arrangements were detected in another study (Puig Giribets et al. 2020). The aim of this work is to study how heat stress affects the whole genome expression of two *D. subobscura* populations, adapted to different temperature regimes, Madeira and Curicó, coming from a native Palearctic region and from a Chilean colonizer region, respectively, and both bearing the same O chromosome arrangement. This study was performed in gonads because germline changes are inherited, the piRNA pathway is active in this tissue (Brennecke et al. 2007; Li et al. 2009; Malone et al. 2009; Luo and Lu 2017), and the effects of heat stress on the gonads 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 toward overexpression, was observed in genes involved in the stress response, such as Hsps. The heat stress changed the expression of only some specific TE families, mostly in population- and tissue-specific patterns. Even though this stress did not globally impact the amounts of TE-derived small RNAs, we detected changes in the siRNA and piRNA

amounts derived from some specific TE families and the piRNA production from some piRNA clusters. Furthermore, our findings demonstrate that small RNA changes have a mild influence on TE activation, suggesting that other factors, such as changes in the epigenome, could also be involved on TE activity.

## Results

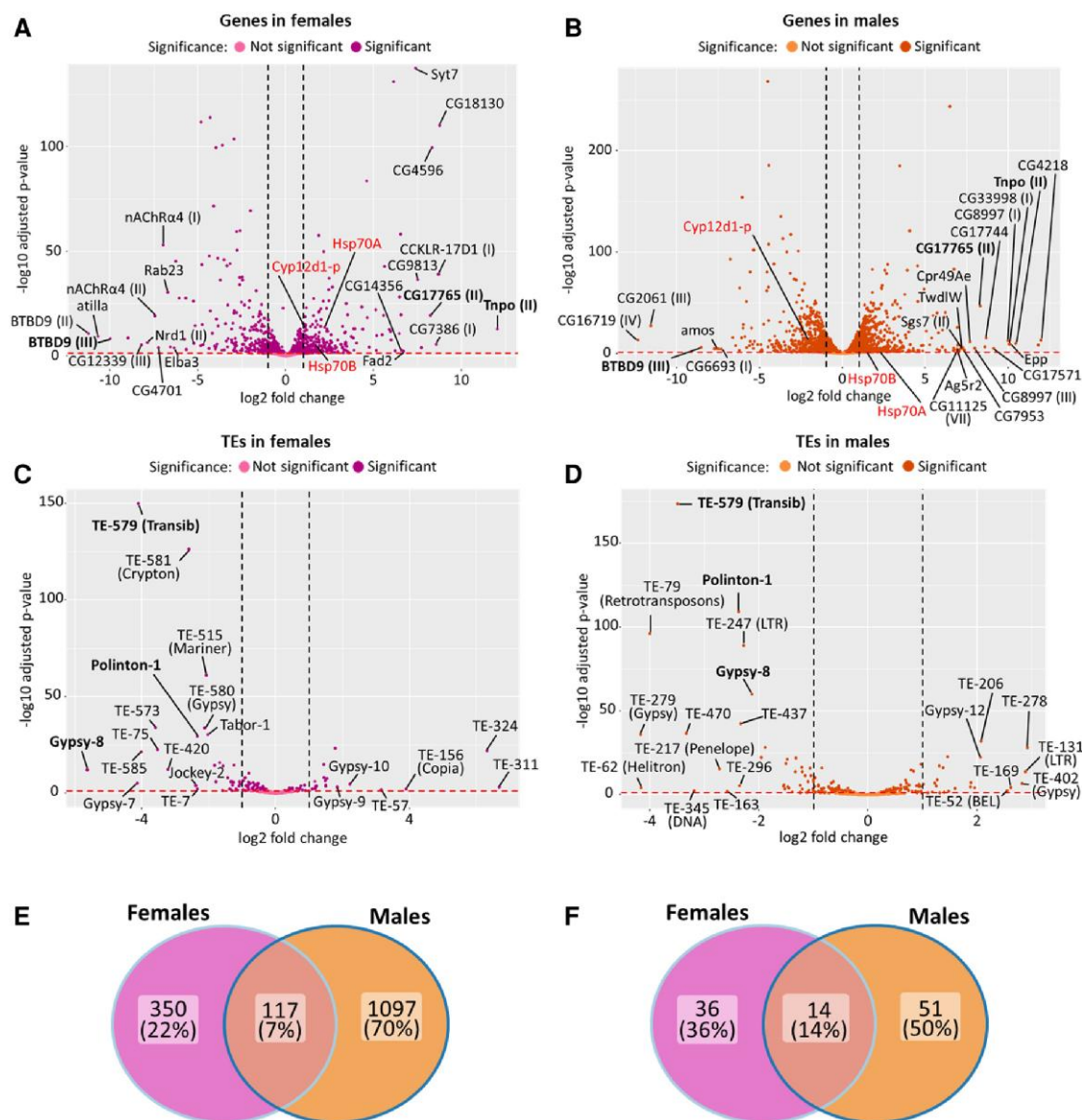
### Differential Expression between Populations in Control Conditions

To compare the expression levels prior to stress (control conditions) between populations adapted to different environments, including different temperature regimes, we studied the gonadal transcriptome of males and females from two strains derived 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 ( $A_2$ ,  $J_1$ ,  $E_{st}$ , and  $O_{3+4}$ ), except on U ( $U_{1+2}$  arrangement in Madeira and  $U_{1+8+2}$  in Curicó) (supplementary fig. S1, Supplementary Material online). All these inversions are frequent in each region (Madrenas et al. 2020; Muñoz 2022). We compared gene, TE, and small RNA expression in testes and ovaries between two populations using RNA-seq under normal and heat stress conditions. We found more genes differentially expressed between populations in testes (9.68%) than in ovaries (4.23%) (supplementary tables S1 and S2, Supplementary Material online; two-proportion Z-test,  $P < 0.001$ ). Overall, the gene expression was more similar between populations (Spearman's rank correlation coefficient  $\rho = 0.974$  and  $0.970$  in females and males, respectively,  $P < 0.001$  in both cases, supplementary table S3, Supplementary Material online) than between sexes ( $\rho = 0.611$  and  $0.568$  in Curicó and Madeira, respectively,  $P < 0.001$  in both cases, supplementary table S3, Supplementary Material online). The Gene Ontology (GO) enrichment analysis on differentially expressed genes between populations showed enrichment in GO terms involved in general processes (cellular, metabolic, and multicellular organismal processes) in all samples, but also in GO terms, sometimes population and/or sex specific, likely related to their environmental adaptation (supplementary fig. S2, Supplementary Material online). For example, the *Cyp12d1-p* gene, encoding a protein in response to the insecticide dichlorodiphenyltrichloroethane (DDT), was more expressed in Curicó males and Madeira females, and both copies of the *Hsp70* gene were more expressed in both sexes of Madeira population (fig. 1A and B). GO terms related to immune system processes were only enriched in Curicó females, but the *Tnpo* gene was more expressed in both sexes of Madeira than in Curicó populations (fig. 1A and B). In the same way, GO terms related to rhythmic processes (e.g.,

*BTBD9* gene) were only enriched in the Curicó population (fig. 1A and B). We also detected an enrichment of growth GO terms when males of both populations were compared (supplementary fig. S2, Supplementary Material online). Moreover, even though an enrichment of reproduction GO terms was not observed, the gene *Ag5r2*, involved in multicellular organism reproduction, was highly expressed in Madeira males (fig. 1B). 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 (fig. 1E).

TE expression differences between populations in gonadal tissues were studied using a list of de novo annotated TEs in the *D. subobscura* reference genome (Karageorgiou et al. 2019), as described in Materials and Methods. We used the TE insertions of the reference genome because divergence of TE consensus sequences versus the reference genome and versus our populations, using four de novo reference transcriptomes (supplementary fig. S3A, Supplementary Material online), was similar. Moreover, a similar distribution of the percentage of identical positions in TE sequences of all reference transcriptomes versus the reference genome (supplementary fig. S3B, Supplementary Material online) was observed. In addition, the correlation of the percentage of identical positions was lower ( $r = 0.680$  in Curicó females vs. Curicó males, supplementary table S4, Supplementary Material online) or similar ( $r = 0.773$  in Madeira females vs. Madeira males, supplementary table S4, Supplementary Material online) between sexes of the same population than in the same sex between populations ( $r = 0.805$  in Curicó females vs. Madeira females and  $r = 0.766$  in Curicó males vs. Madeira males, supplementary table S4, Supplementary Material online), avoiding a population bias in our results.

We found that 12.83% of the reference genome, with an estimated size of around 150 Mb (Karageorgiou et al. 2019), was covered by putative TE families (supplementary fig. S4A, Supplementary Material online): 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 fig. S4B, Supplementary Material online). 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 S5, Supplementary Material online). However, we found a similar percentage of TE families showing differential expression between populations in ovaries (8.09%) and in testes (9.88%) (supplementary tables S5 and S6, Supplementary Material online, two-proportion Z-test,  $P = 0.265$ ). The differentially expressed TEs belonged to different classes: DNA class



**FIG. 1.**—Changes of expression in gonads between populations. (A–D) Differences of expression in gene (A, B) and TE families (C, D) in Madeira versus Curicó populations: females (A–C) and males (B–D). Positive log<sub>2</sub>FC values correspond to genes and TEs more expressed in Madeira. The names of genes showing the 20 highest log<sub>2</sub>FC values, and displaying an ortholog in *D. melanogaster*, are shown in (A) and (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* (Puig Giribets et al. 2020). (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.

(*Polinton*, *mariner*, *Transib*, or *Helitron*) and retrotransposons (*copia*, *gypsy*, *BEL*, *Penelope*, or *jockey* superfamilies) (fig. 1C 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 (fig. 1F). Globally, we found that TE expression is fairly similar between populations ( $\rho = 0.968$  and  $0.974$  in females and males, respectively,  $P < 0.001$ , [supplementary table S7](#), [Supplementary Material](#) online) and between sexes ( $\rho = 0.837$  and  $0.828$  in Curicó and Madeira, respectively,  $P < 0.001$ , [supplementary table S7](#),

[Supplementary Material](#) online), unlike what was observed when the expression of genes was compared (more differences between sexes).

### Sex-Specific Gene Expression Response after a Heat Shock

We then studied how the heat stress affected the transcriptome of ovaries and testes from Madeira and Curicó populations by comparing samples under heat shock versus



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 S8, Supplementary Material online). 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 S9, Supplementary Material online), a higher number of differentially expressed genes were observed in Madeira males compared with Curicó (two-proportion Z-test,  $P=0.004$ , supplementary table S9, Supplementary Material online), suggesting a stronger response of heat stress in the males of this population. Moreover, a trend to overexpression versus underexpression was observed (fig. 2A–D; supplementary table S8, Supplementary Material online), 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 (Sørensen et al. 2005). As expected, the genes showing the highest overexpression in most samples corresponded mainly to the Hsp family (fig. 2A–D), whereas underexpressed genes were involved in other biological functions. Differentially expressed genes after a heat stress were mainly enriched in general GO terms involved in cellular, developmental, metabolic, and multicellular organismal processes, as well as biological regulation (supplementary fig. S5, Supplementary Material online). When we compare the GO terms of overexpressed genes versus underexpressed, we observed a greater representation of genes in 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 underexpressed category, in addition to reproduction in males (supplementary fig. S5, Supplementary Material online). 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).

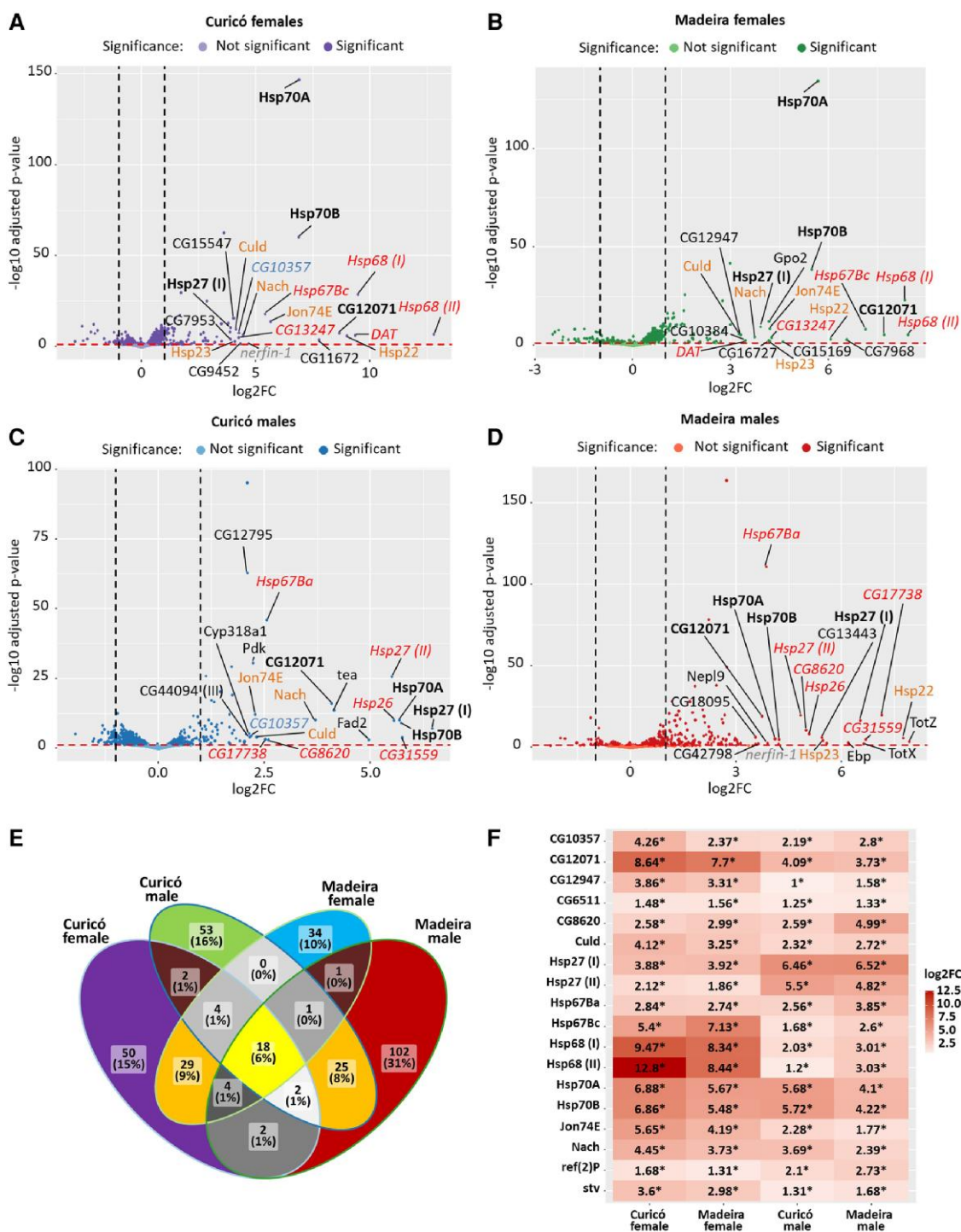
We also looked for common differentially expressed genes after a heat shock (fig. 2E) between populations or sexes. We found a similar percentage of genes with changes in expression unique for each population-sex, except in Madeira males where almost twice as many differentially expressed genes were detected. 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, fig. 2E), than those shared between sexes within a population (two genes in Curicó and one in Madeira, fig. 2E). 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 (fig. 2F). 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 (fig. 2F). In addition, 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–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).

### Heat Shock Effects on Genes of Heat Response and Piwi Pathway

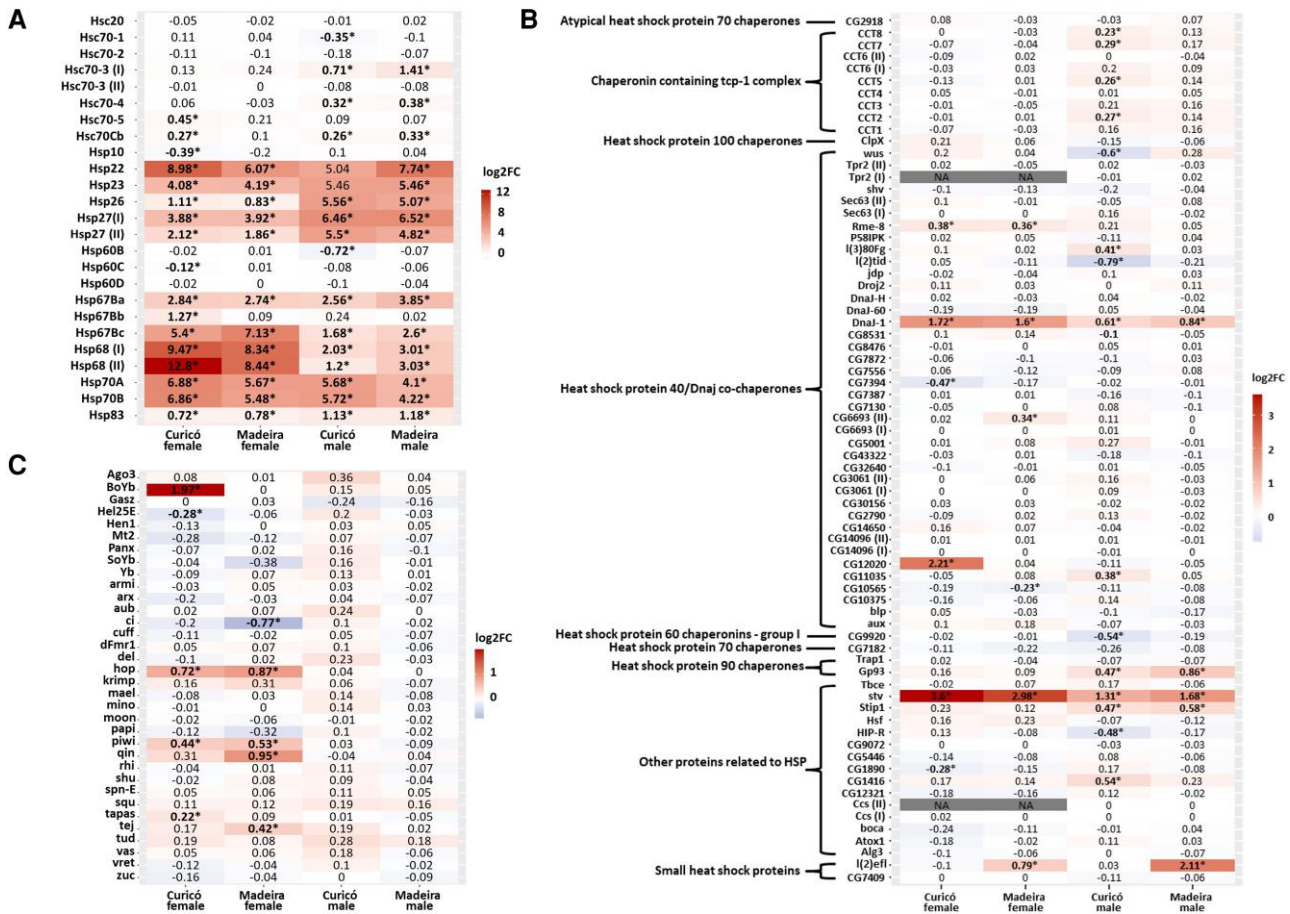
To gain insight into the heat shock response, we studied in depth the heat shock genes encoding for proteins having an important role on stress response. In figure 3A, we observed that most of the heat shock genes were overexpressed after a heat shock in the gonads of all populations and sexes. Some 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* 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 (Dworniczak and Mirault 1987), we found less changes in expression than those observed in the Hsp 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 (fig. 3B). We observed overexpression of *stv* (encodes an *Hsp70* cochaperone during stress recovery) and *DnaJ-1* (encodes a heat shock protein cofactor), and in both cases, the expression tends to be higher in females than in males. In Madeira population, we observed overexpression of *I(2)efl* (a family of small Hsp genes). *Gp93* (encodes an Hsp *Hsp90* family) and *Stip1* (enables *Hsp90* protein binding



**FIG. 2.**—Changes of gene expression in gonads after a heat shock. (A–D) Differential gene expression analysis in heat shock versus controls in Curicó females (A) and males (C) and Madeira females (B) and males (D). Positive log<sub>2</sub>FC values correspond to genes more expressed after a heat shock versus control. Only the names of genes showing the 20 highest log<sub>2</sub>FC values and displaying an ortholog in *D. melanogaster* are shown. *Hsp70* is called as characterized in *D. subobscura* (Puig Giribets et al. 2020). 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 gray 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).

Downloaded from https://academic.oup.com/gbe/article/15/11/evad189/7319546 by guest on 04 September 2024



**FIG. 3.**—Expression changes in gonads of specific gene groups in heat-shocked samples versus controls ( $\log_2FC$ ). (A) Hsp 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 ( $\log_2FC$ ). Values that could not be computed, due to low number of reads, are shown as NA (not applicable).

activity) genes were only overexpressed in testes. In ovaries, we also detected an overexpression of the *Rme-8* gene, which belongs to the Hsp 40/*Dnaj* cochaperones 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 specifically heat shock promoter elements (HSE) and activates Hsp 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 3C, 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  $\log_2FC$  values ( $|\log_2FC| < 1$ ), except *BoYb* which expression is 4-fold higher after a heat shock in the females of Curicó versus control ( $\log_2FC$  of 1.97). It is interesting to highlight that other genes belonging to the Hsp group (*Hsp83*, *Hsc70-4*, and *Hsp70*), which changed their expression, have also an involvement in the piRNA pathway (Cappucci et al. 2019).

### 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 family transcription levels after a heat shock versus 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 S10, Supplementary Material online), and this percentage is similar both between sexes and between populations (two-



proportion Z-test, all  $P \geq 0.447$ , [supplementary table S11, Supplementary Material](#) online). We observed a trend of TE families toward the overexpression (0.46–1.75% overexpressed vs. 0–0.31% underexpressed, [supplementary table S10, Supplementary Material](#) online; [fig. 4A–D](#)). After annotation of the differentially expressed TEs, we noticed that most of them were retrotransposons, mainly from the *gypsy* superfamily ([fig. 4A–D](#)). We also found that most of the TEs with changes in expression were specific of a population or sex ([fig. 4A–E](#)), but some differentially expressed TE families were only shared by sex (one in females and three in males, [fig. 4E](#)). For example, a *gypsy* family was underexpressed in ovaries ([fig. 4A, B, and E](#)), and other three TE families, one *gypsy* and two unknown, were overexpressed in testes ([fig. 4C–E](#)). However, only one overexpressed TE (*gypsy* superfamily) was commonly overexpressed in both populations and sexes (bold name in [fig. 4A–D and 4E](#)) after the heat shock, which was validated using a quantitative real-time polymerase chain reaction (qRT-PCR) (Wilcoxon signed-rank test for  $2^{-\Delta C_t}$ ,  $P \leq 0.046$  in all cases, [supplementary table S12, Supplementary Material](#) online). Finally, when we focused on the 23 differentially expressed TE families of [figure 4F](#), 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 ([fig. 4F](#)).

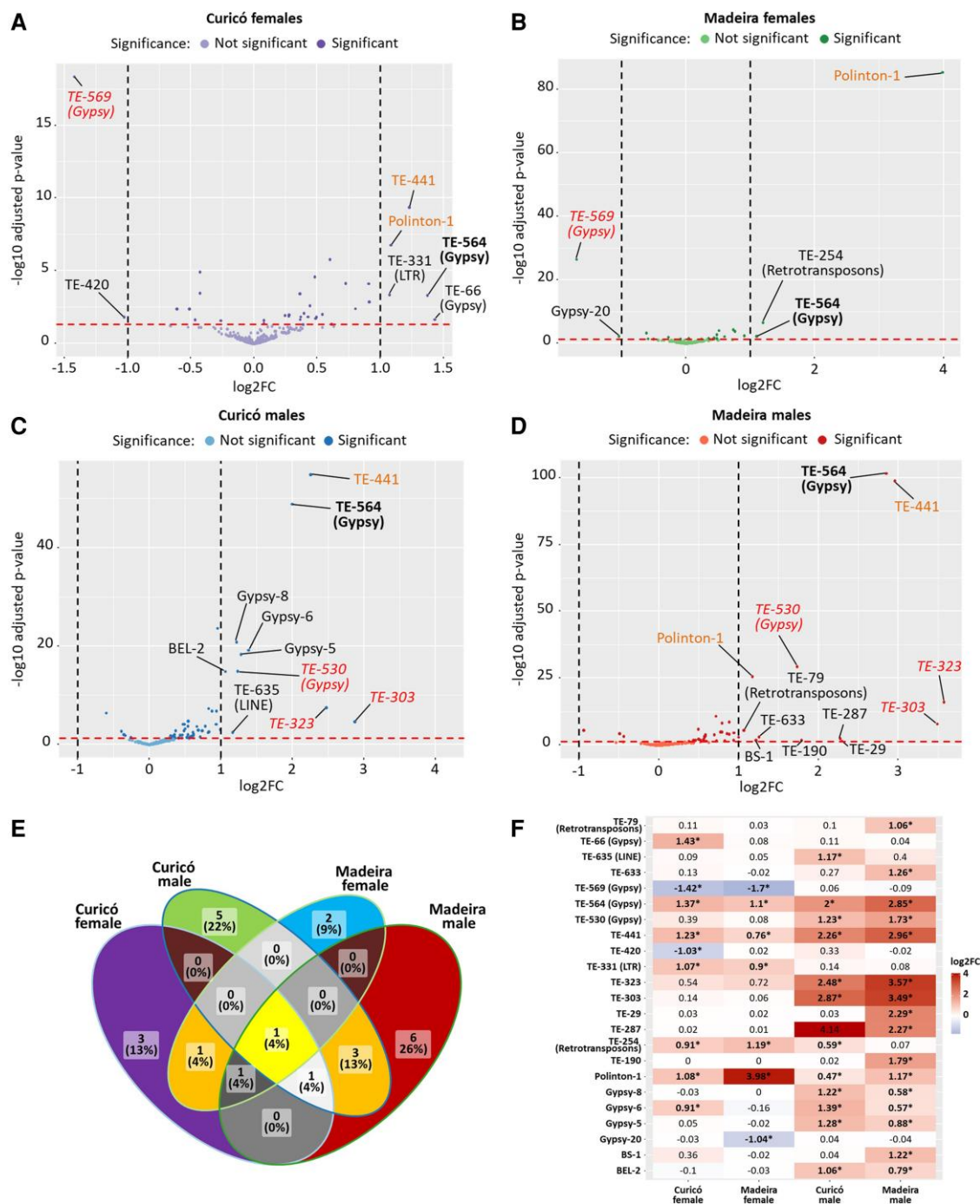
### Impact of the Heat Stress on the Small RNA Amounts

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* (Karageorgiou et al. 2019) to have an overview of their distribution along chromosomes. We used the software proTRAC (Rosenkranz and Zischler 2012), which predicts and analyzes genomic piRNA clusters based on mapped piRNA sequence reads, as explained in Materials 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 S13, Supplementary Material](#) online 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 ([fig. 5A](#)). In addition, 15 of the 20 clusters producing most of the piRNA in each population were the same in Madeira and Curicó (highlighted lines in [supplementary table S13, Supplementary Material](#) online), 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 extra clusters in Madeira and 73 in Curicó ([supplementary tables S14 and S15, Supplementary Material](#) online and piRNA alignment examples of population-specific clusters in [supplementary fig. S6, Supplementary Material](#) online, and data S2 and S3, available at <https://figshare.com/s/ac3d1dc5f9eb0dd3f2dc> and <https://figshare.com/s/80cce38193134e9305b2>, respectively), covering a 0.80% and 0.84% of the reference genome, respectively. We then checked the piRNA 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 piRNA production of the 85 common clusters detected in ovaries ([supplementary table S13, Supplementary Material](#) online), and we found that this stress did not globally modify piRNA production from piRNA clusters in any population and sex (Kruskal–Wallis test, all  $P = 0.942$ , [supplementary table S16, Supplementary Material](#) online). However, heat stress seemed to increase the production of piRNAs from some clusters and to decrease that from others, in some cases showing double or half of the piRNA production ( $|\log_2 FC| > 1$ ) after heat stress ([supplementary fig. S7, Supplementary Material](#) online). The general comparisons of piRNA production per cluster before and after heat stress, considering both kinds of clusters, pointed to an effect of the piRNA cluster production only in Curicó females because in this sample, most of the clusters tend to decrease their piRNA production after stress (Wilcoxon signed-rank test,  $P < 0.001$  in Curicó females, [supplementary table S17, Supplementary Material](#) online). 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  $P < 0.001$ , [supplementary tables S18 and S13, Supplementary Material](#) online). 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 S18, Supplementary Material](#) online).

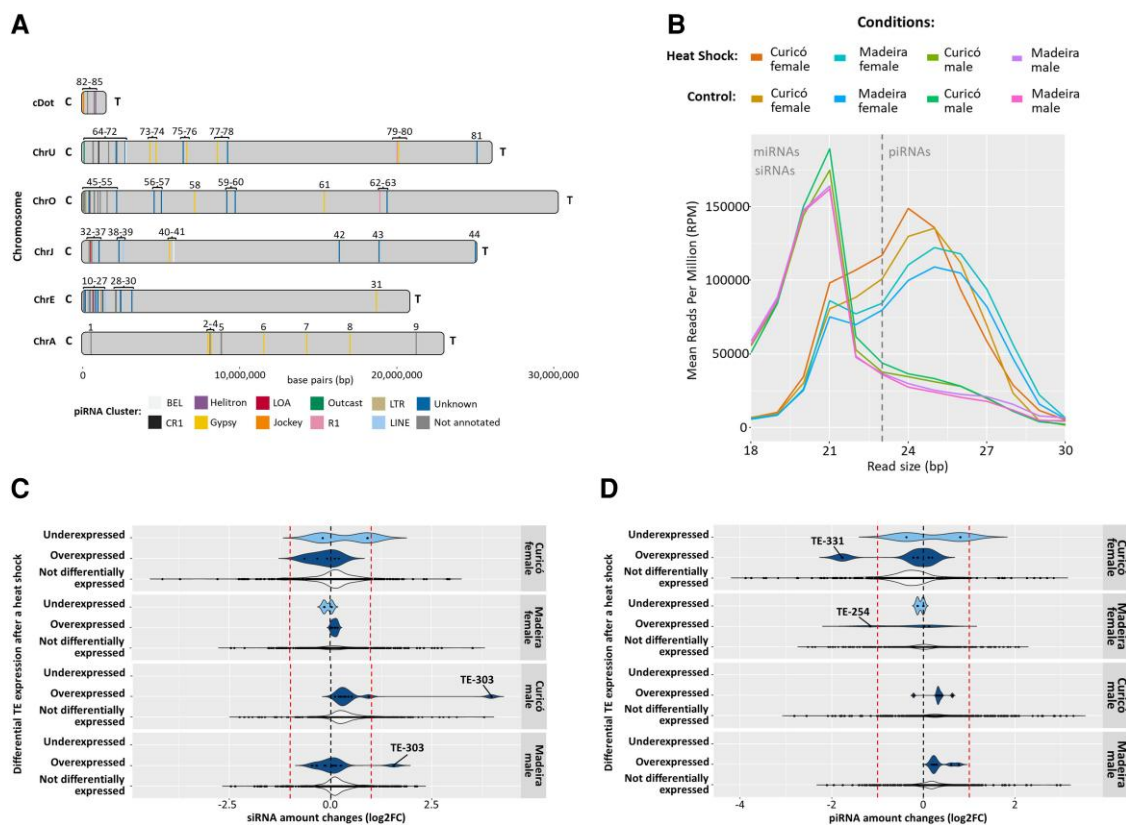
We then studied the global amounts of the small RNAs due to their role in the TE silencing in the *Drosophila* germline. An overview of the quantity of small RNAs, normalized by reads per million (RPM), present in our samples ([fig. 5B](#)) showed higher quantities of piRNAs and lower of miRNAs and siRNAs in females than in the males, as expected due to the higher efficiency of this pathway in ovaries versus testes (Saint-Leandre et al. 2020). 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



**Fig. 4.**—Changes of TE expression in gonads after a heat shock. (A–D) Differential TE expression analysis in heat shock versus control in Curicó females (A) and males (C) and Madeira females (B) and males (D). Positive  $\log_2FC$  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 ( $|\log_2FC| > 1$ ) 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.

$P \geq 0.157$ , supplementary table S19, Supplementary Material online). We then analyzed the impact of the heat stress on the small RNAs, normalized using miRNAs, derived from each TE family. We observed changes in their amounts

after a heat stress (points outside the red-dashed lines in fig. 5C and D). We then checked if TE expression modulation could be explained by the siRNA and piRNA amount changes after the heat stress, and we did not find a strong



**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; T, telomere) and its annotation based in the main TE superfamily/class present. (B) Mean RPM of small RNAs in all samples. (C, D) Differential small RNA expression analyses in heat shock versus control by population and sex. Positive log<sub>2</sub>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<sub>2</sub>FC value considered as significant. The names of the differentially expressed TE families with significant differences in small RNAs are shown.

association between them, except in Curicó males that exhibited a trend to more TE expression and more piRNA amounts, respectively (Fisher's exact test,  $P=0.002$ , [supplementary table S20](#) and [figs. S8 and S9](#), [Supplementary Material](#) online). In addition, changes in small RNAs mainly affected nondifferentially expressed TE families ([fig. 5C 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, which showed an increase of siRNAs after a heat shock. In the same way, *TE-331* and *TE-254* families, overexpressed in the females of Curicó and Madeira populations respectively, exhibited a decrease in their piRNA amounts after a heat stress (names in [fig. 5C and D](#); an example of the siRNA and piRNA alignment with the overexpressed TEs is shown in [supplementary figs. S10 and S11](#), [Supplementary Material](#) online).

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 the ping-pong signal in control

versus 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  $P \geq 0.979$ , [supplementary table S21](#), [Supplementary Material](#) online). We then focused on the specific TE families whose ping-pong signal changed in control versus 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 figs. S12 and S13](#), [Supplementary Material](#) online). In Curicó females, five out of seven TE families only show ping-pong signal after stress ([supplementary fig. S12](#), [Supplementary Material](#) online). Nevertheless, four out of five TE families and all three TE families in Madeira females and Curicó males, respectively, only show ping-pong signal in control conditions ([supplementary figs. S12 and S13](#), [Supplementary Material](#) online). Finally, in Madeira males, this signal is only detected in control conditions in one-half of the TE families (out of four TE families, [supplementary fig. S13](#), [Supplementary Material](#) online). However, the

differences in the ping-pong signal upon heat shock are small and not significant, with some specific exceptions mainly unique by population-sex.

## Discussion

### Adaptive Expression Differences in Two *D. 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 (Juneja et al. 2016), similar (Wurmser et al. 2013) or even lower (Huylmans and Parsch 2014) 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 (Banho et al. 2021). 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 (Juneja et al. 2016); signaling and response to stimulus (Juneja et al. 2016), such as response to chemicals used in pesticides (Catalán et al. 2012; Wurmser et al. 2013) or temperature (Catalán et al. 2012); immune system processes (Huylmans and Parsch 2014; Juneja et al. 2016); growth (Huylmans and Parsch 2014), such as body size (Loeschcke et al. 2000); rhythmic processes, such as the circadian rhythm (Juneja et al. 2016); and reproduction (Wurmser et al. 2013; Juneja et al. 2016). These GOs could be related to the adaptation of *D. subobscura* to new environments during its colonization from the Palearctic region to America (Ayala et al. 1989), where Madeira and Curicó are located, respectively. However, because of the possible effects of the Madeira island colonization from the continent populations (Khadem et al. 1998), 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 (Balanyá et al. 2006). In fact, previous studies in populations bearing different O arrangements showed differences in the Hsp70 protein amounts in control conditions (Calabria et al. 2012), but invariable

mRNA expression was detected in another study (Puig Giribets et al. 2020). Our populations only differ on the U chromosomal arrangement (with both inversions, the  $U_{1+2}$  of Madeira and  $U_{1+8+2}$  of Curicó, related to a “warm” thermal adaptation; Galludo et al. 2018), 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 (Karageorgiou et al. 2019), 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* (Mérel et al. 2021). 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 (Lerat et al. 2017). 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 (García Guerreiro et al. 2008) and of TE insertions, in other species (Vieira et al. 1999; Mérel et al. 2021), has 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 (Catalán et al. 2012; Huylmans and Parsch 2014). 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 (Lebo et al. 2009), including a different efficiency in TE regulation (Chen et al. 2021). These results are also supported by the detection of more similarities in gene expression between populations than between sexes.

### Response to Heat Stress in the Germline

We found from 91 to 155 genes with changes in expression after a heat stress in our species, with a slight trend toward the overexpression. These results contrast with those of previous studies in *D. melanogaster*, where heat shock



resulted in a trend toward underexpression (Sørensen et al. 2005; Landis et al. 2012; Lecheta et al. 2020). 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 Hsps, and protects the individual (Teves and Henikoff 2013). However, these experiments have many differences with ours—in species and in the nature of the heat shock. We especially have to highlight that most of previous heat shock studies were performed in whole flies (Sørensen et al. 2005; Landis et al. 2012; Lecheta et al. 2020), with little information on the real effect in germinal tissues. It is possible that the stress affects the germline in a different way, for example, the oogenesis processes (Gruntenko et al. 2003), the sperm function and viability (Chakir et al. 2002), or even the recovery once the stress disappears. We also cannot rule out the existence of some bias resulting from the analysis of only two biological replicates. Notwithstanding, a similar expression of most temperature-responsive genes in *D. melanogaster* ovaries and whole bodies was suggested in a previous work (Fast et al. 2017). In addition, the detected gene activation 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 Hsps were among the genes showing the highest overexpression after a heat stress, in agreement with what has been previously described in the literature (Sørensen et al. 2005; Landis et al. 2012). We also detected overexpression in all populations and sexes of genes that have been previously reported, such as *stv* (Landis et al. 2012; Lecheta et al. 2020) and *ref(2)P* (Landis et al. 2012; Lecheta et al. 2020). We as well noticed other common genes with changes in expression whose functions were related to the heat response, such as gene expression (Sørensen et al. 2005; Lecheta et al. 2020). 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 (Pauli et al. 1992) 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 (Lecheta et al. 2020), developmental (Lecheta et al. 2020), and immune system processes (Landis et al. 2012; Teets and Hahn 2018); rhythmic processes, such as circadian function (Teets and Hahn 2018); and response to stimulus, such as response to heat or stress (Sørensen et al. 2005; Lecheta et al. 2020). 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* (Jørgensen et al. 2020) related to a shorter recovery time (Rodgers et al. 2007). 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 (Chakir et al. 2002): male sterility in *D. subobscura* can be induced at 25° (Krimbas 1993).

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 (Hsu et al. 2020), the higher heat resistance observed in *D. subobscura* females (Erić et al. 2022), and the sex-specific response after a heat stress observed in other species (Srikanth et al. 2020; Bedulina et al. 2021). It could also be influenced by the differential expression of genes involved in chromatin remodeling in females but not in males. These sex-specific expression differences also affected genes encoding some Hsps. Differences in the induction of a few small Hsps in testes and ovaries were previously described (Michaud et al. 1997), 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 (Dworniczak and Mirault 1987), or other genes related to the heat response. We also could not detect an overexpression of the *hsf* either, contrary to what was observed in a previous study (Lecheta et al. 2020). However, this transcription factor is already present as an inactive monomer without stress and trimerizes during stress to bind to the Hsp promoters (Teves and Henikoff 2013), 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 (Gámez-Visairas et al. 2020) and after a 48-h housing at 29 °C (Fast et al. 2017). The only exception was *BoYb*, essential for the primary piRNA pathway in the germline (Handler et al. 2011), 4-fold more expressed in Curicó females after the heat shock, and likely involved in the piRNA production decrease detected in most piRNA clusters in Curicó ovaries. In addition, the heat shock genes involved in this pathway (*Hsp83*, *Hsc70-4*, and *Hsp70*) (Cappucci et al. 2019) changed their expression, except for *Droj2* (Cappucci et al. 2019), 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 (Strand and McDonald 1985; Funikov et al. 2015; Cappucci et al. 2019) or transposition (Junakovic et al. 1986; Vasilyeva et al. 1999) of some specific TE families were detected in some cases, and absence of transposition induction in the same or other TE families (Arnault et al. 1997, 1991) was 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 (Junakovic et al. 1986), pointed out the importance of the genomic context (Arnault et al. 1997) 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 (Cappucci et al. 2019) 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 (Saint-Leandre et al. 2020). The higher TE activation in males versus 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 (Pimpinelli et al. 1995). 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 (Funikov et al. 2015). 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* (Strand and McDonald 1985; Junakovic et al. 1986; Vasilyeva et al. 1999; Funikov et al. 2015; Cappucci et al. 2019).

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 clusters 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* (Brennecke et al. 2007). However, differences in the detection method, as well as the higher TE percentage of *D. melanogaster* (McCullers and Steiniger 2017), and its lower genome compaction (Adams et al. 2000), in comparison to *D. subobscura* (Karageorgiou et al. 2019), can explain these differences. Nevertheless, the total number of piRNA clusters per population was similar to *D. melanogaster* (142 piRNA clusters) (Brennecke et al. 2007). Consistently with what was previously observed in *D. melanogaster*, most of the piRNA clusters found in this study were dual stranded (Théron et al. 2014; Gebert et al. 2021) and located in the pericentromeric area (Brennecke et al. 2007; Théron et al. 2014), which was also the richest TE insertion region in the reference genome of *D. subobscura* (Karageorgiou et al. 2019). 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 (Brennecke et al. 2007; Gebert et al. 2021). Additionally, we detected that slightly less than half of the piRNA clusters identified in Madeira and Curicó seemed to be unique for each population, which could be explained by the fast piRNA cluster evolution suggested in previous studies (Gebert et al. 2021). piRNA cluster differences within species were also found in other studies but in a lesser extent (20–30%) (Wierzbicki et al. 2023). 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 (Wierzbicki et al. 2023). 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 (Saint-Leandre et al. 2020). We additionally found that the heat stress seemed to have an impact on the expression of the TE piRNA clusters leading to a decrease of piRNA production in most piRNA clusters 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* (Funikov et al. 2015). However, considering that no changes in piRNA levels were detected globally, further explorations are necessary to fully understand the production of piRNAs by piRNA clusters. In fact, a recent study in *D. melanogaster* detected that acute heat shock disrupted the localization of several piRNA pathway proteins and piRNA cluster heterochromatin, leading to a reduction of transcripts in piRNA clusters but not of piRNAs mapping to clusters or TEs (Rice et al. 2023).

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* (Saint-Leandre et al. 2020), 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 RNA studies (Fast et al. 2017), 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 (Funikov et al. 2015), whereas no significant effect in the piRNAs targeting other set of TE families was shown in another (Cappucci et al. 2019). In addition, a transcriptome-based small RNA study found temperature-dependent changes of most transposon-derived piRNAs (Fast et al. 2017). 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 (Funikov et al. 2015) and three heat shock repetitions (Cappucci et al. 2019) or 48-h housing at 29 °C (Fast et al. 2017). 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 (Cernilogar et al. 2011) whereas in plants, an increase of siRNA amounts targeting some genes and a decrease of the amounts targeting other genes has been reported (Ahmed et al. 2021). 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 have 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 (Funikov et al. 2015; Fast et al. 2017). This could be expected by the complexity of the piRNA-mediated TE regulation, which includes both a transcriptional and posttranscriptional silencing, making difficult to detect a direct correlation between these two variables (Senti et al. 2015). 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 (Kelleher and Barbash 2013). 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 (Saint-Leandre et al. 2020), and piRNAs with the ping-pong signature seem

to be even more abundant in *Drosophila* spermatogonia (Emilie et al. 2016). When we studied the effect of the heat shock on the ping-pong signal, we could not find a significant 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 (Funikov et al. 2015). However, other study detected an increase of this signal after 48-h *Drosophila* housing at 29 °C (Fast et al. 2017), 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 concordance with other studies (Funikov et al. 2015; Fast et al. 2017) and highlighted 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 (Bodelón et al. 2022). Specifically, *Hsp83* was described to be involved in epigenetic modification (Tariq et al. 2009), as well as other changes in the epigenome have been observed after a heat stress (Arrigo 1983; Pauli et al. 1992). Nonetheless, other studies have shown that H3K9me3 and H3K27me3 did not significantly change after a heat shock in seven TEs studied in the germline (Cappucci et al. 2019). They propose that the interaction of Hsp70 with the chaperone–Ago3 complex in the germline induces 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 has been described in different Hsps in *Drosophila* (Lerman et al. 2003; Walser et al. 2006). 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, supplementary Material online) showing that these insertions are also present in our species. Nevertheless, these TEs were not activated after a heat stress and 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.

## Materials 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 18 °C for this species (Rego et al. 2010). Twenty generations later, crosses of 25 males × 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-curved recessive marker strain (*ch-cu*). This strain is homokaryotic for the standard arrangement in all chromosomes, except the O chromosome, which is homozygous for the  $O_{3+4}$  arrangement (Karageorgiou et al. 2019).

### Chromosomal Inversions

For the line inversion identification, we first incubated clean and dry slides in a 3xSSC/1xDenhardt's solution for 2 and 30 min in a water bath at 65 °C. Slides were cleaned with distilled water and immersed in ethanol/acetic acid (3:1) at room temperature for 20 min. After air dry, slides were stored at 4 °C until use. The glass coverslips were siliconized with Repel-Silane (Amersham) solution for 10 s and washed in ethanol and distilled water for a few seconds. Third instar larvae salivary glands were dissected in NaCl 0.8%, immersed in acetic acid 45% for 1 min, and fixed in a 1:2:3 solution (lactic acid:water:acetic acid) for 4–8 min.

### 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 min. Then, they were kept at 18 °C for 30 min before gonadal dissection in PBT (1x phosphate-buffered saline, 0.2% Tween 20). Non-heated samples were maintained at the optimal temperature of 18 °C and then 5-day-old individuals dissected. All gonad samples were frozen in liquid nitrogen and stored at –80 °C 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 FASTER SA. Small RNAs of 18–30 nucleotides were purified using polyacrylamide gels, and then the libraries were prepared (two 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 (Bolger et al. 2014), with 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 (Flynn et al. 2020) and EDTA (Ou et al. 2019) with default parameters were used to annotate TE families de novo in the *D. subobscura* reference genome (Karageorgiou et al. 2019), detecting 313 and 542 consensus sequences, respectively. Consensus sequences smaller than 100 bp (37 sequences) were removed, and a MegaBLAST (Morgulis et al. 2008) 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 were 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 (Smit et al. 2013–2015) with the *--norna* and *--nolow* parameters and as a custom library the Repbase (Bao et al. 2015) 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 (Smit et al. 2013–2015) 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” (Bailey-Bechet et al. 2014) 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 (Lerat et al. 2017). To verify that TE insertions in the reference genome were representative of TE insertions in our populations, we produced a de novo transcriptome for all four populations-sexes using Trinity v2.15.1 (Grabherr et al. 2011) with default parameters. The transcriptomes were then masked using RepeatMasker v4.1.2 (Smit et al. 2013–2015), with the same parameters described above and the list of de novo TE consensus as the custom library. Trinity transcripts masked with less than 80% of their length with a TE consensus were filtered out to exclude gene transcripts and real or artificial chimeras. RepeatMasker v4.1.2 (Smit et al. 2013–2015) with the same options was also used to obtain the alignment files (including the `-a` option) of all filtered de novo transcriptomes, and the script `calcDivergenceFromAlign.pl` (included in RepeatMasker) was used to compute TE divergence. Finally, a MegaBLAST (Morgulis et al. 2008) alignment between TE insertions from the reference transcriptome and the reference genome was performed to calculate the percentage of identical positions.

For the first global piRNA cluster annotation, both replicates of the same population were merged, and only female samples (higher piRNA production; Saint-Leandre et al. 2020) and control conditions were used, to avoid possible expression changes due to the heat treatment. The proTRAC (Rosenkranz and Zischler 2012) 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 (Rosenkranz and Zischler 2012). 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` (Zytnicki and Gaspin 2022) for piRNA mapping in the reference genome keeping only the best alignments. `Reallocate.pl` (available at <https://www.smallnagroup.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 (Rosenkranz and Zischler 2012) 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 (Gebert et al. 2021). To compare the annotated clusters in each population, the regions were converted in a bed file using `convert2bed` of BEDOPS v2.4.38 (Neph et al. 2012) and the bedtools v2.29.2 (Quinlan and Hall 2010) `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 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. Differences in piRNA cluster production in heat shock versus control samples were manually computed as  $\log_2FC$ .

### Gene and TE Differential Expression Analyses

RNA-seq sequenced reads were trimmed using Trimmomatic software v0.39 (Bolger et al. 2014), 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 (Karageorgiou et al. 2019) using STAR v2.7.9a (Dobin et al. 2013) with the `--quantMode` to count the reads per gene using `htseq-count` (Putri et al. 2022) and giving the transcript annotation available in the browser. TE expression was then analyzed using the TEcount module of the TETools pipeline (Lerat et al. 2017). First, the RNA-seq data were aligned with the fasta file including all TE insertions in the reference genome of *D. subobscura* using Bowtie2 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, 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 (Lerat et al. 2017). `DESeq2` function from the R Bioconductor package `DESeq2 v1.34.0` (Love et al. 2014) 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` (Love et al. 2014) 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 (Love et al. 2014). 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

(Benjamini et al. 2001) with an false discovery rate  $p$ -value cutoff of 0.05. The  $\log_2FC$  was shrunken using the default and recommended *apecglm* algorithm (Zhu et al. 2019) 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 ( $|\text{shrunken } \log_2FC| > 1$ ), 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 (Alexa and Rahnenfuhrer 2020) (“weight01” algorithm and Fisher’s statistic). GO belonging to each gene were obtained using eggNOG (Huerta-Cepas et al. 2019). Because of the high impact of the  $P$  value adjustment, only the top 25% significant gene ontologies having the lowest  $P$  values were considered as enriched. To simplify the analysis, we manually grouped the enriched GOs by the most general 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 (Goubert) (available at <https://github.com/clemgoub/TE-Aid>) and online BLAST (*nblast* and *blastx*) (Camacho et al. 2009).

Finally, to confirm heat shock and piRNA pathway gene annotation and/or annotate the ones not previously annotated, a reciprocal *tblast* v2.10.1. (Camacho et al. 2009) of the *D. melanogaster*, *Drosophila pseudoobscura*, or *D. subobscura* (if available) proteins, downloaded from UniProt (UniProt Consortium 2021), 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 (Griffiths-Jones et al. 2006), with the *D. subobscura* reference genome using *blat* (Kent 2002). The best alignment (*pslReps*) was then retrieved and converted in bed format (*convert2bed* of BEDOPS v2.4.38; Neph et al. 2012), 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 (Langmead 2010) 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 (Roberts and Pachter 2013). To ensure that miRNA production was not affected by stress, the miRNA aligned reads were normalized by RPM, and their amounts, before and after stress, were compared, finding that the differences were not significant ( $t$ -test, all  $P \geq 0.705$ , [supplementary](#)

[table S22, Supplementary Material](#) online). Small RNA data were then analyzed with the TEcount module of the TETools pipeline (Lerat et al. 2017), and reads of 21 nucleotides (siRNAs) and 23–30 nucleotides (piRNAs) were aligned to our custom TE library using Bowtie v1.3.0 (Langmead 2010) 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 versus control were computed manually calculating  $\log_2FC$ . 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 nucleotides) against these TE family consensus using Bowtie v1.3.0 (Langmead 2010) (-S option), and we checked for the presence of 10-nt overlapping sense–antisense read pairs using the *signature.py* pipeline (Antoniewski 2014). The mean  $Z$ -scores and overlapping pair reads for each overlap between the two replicates of each sample were 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.

### Quantitative Real-Time Polymerase Chain Reaction

qRT-PCR was used to validate the overexpression of the *gypsy* TE-564 in all populations-sexes after a heat stress, detected in the bioinformatic analysis. TE expression was quantified by measuring fluorescence intensity using iQ SYBR Green Supermix (BioRad, Hercules CA, USA) on a CFX96 BioRad Real-Time LightCycler, using TE-specific primers and an annealing temperature of 59 °C. All assays were performed in testes and ovaries in the same conditions as RNA-seq samples, using four biological replicates with three technical replicates for each one. The *rp49* housekeeping gene was used as an endogenous control by its expression stability, following the protocol described in previous *D. subobscura* work (Puig Giribets et al. 2020). Relative expression of TE-564 was then calculated with the comparative CT method (Schmittgen and Livak 2008).

TE-564-specific primers were designed in a conserved region of some expressed TE insertions of this TE family and different from other *gypsy* TE families, to assure family

specificity. The primers used for amplification were as follows: TE564\_F (5'-AAC TCG AAT GCT TGG CGG TA-3') and TE564\_R (5'-ACT TCG CAT TCC AAA ACC GC-3') that amplified a 226-bp TE region of the putative *pol* gene (supplementary fig. S14, Supplementary Material online). For the *rp49* housekeeping gene, the primers RP49F (5'-ACA TCG GTT ATG GCT CCA AC-3') and RP49R (5'-GAT TTC CTT GCG CTT CTT TG-3') were designed from a *D. subobscura* GenBank sequence (accession: AJ228921) and amplified a 212-bp segment from the gene's second exon. Primer efficiencies of TE-564 and *rp49*, calculated on a standard curve slope, using serially diluted templates, were 100.1% in both cases.

### Statistical Tests and Visualization

Five main statistical tests were used in this article, and all were performed using R v4.1.3 (R Core Team 2020). The two-proportion Z-test was used to compare the distributions of significant genes and TE families across comparisons. The Spearman rank correlation was used to compare general gene and TE expression between sexes and populations, and the Pearson correlation coefficient was used to determine intra- and interpopulation divergence correlation. The Wilcoxon signed-rank test was used to compare the cluster production of piRNAs (two and one tailed) and the expression of *gypsy* TE-564 by qRT-PCR before and after the heat stress. The *t*-test was used to compare the normalized aligned count reads in heat shock versus control for small RNAs. The Fisher exact test under independence assumption was computed using a 2 × 2 contingency table to detect associations in TE expression changes in heat shock versus control and the corresponding small RNA amount changes. 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. Finally, the Kruskal–Wallis test was used to compare piRNA cluster production changes in heat shock versus control conditions. All the results were corrected for multiple testing using the Benjamin and Hochberg (Benjamini et al. 2001) method. The plots and visualization of the results were performed using the R package *ggplot2* v3.3.5 (Wickham 2016), and alignment images were obtained using Integrative Genomics Viewer (IGV) v2.16.1 (Robinson et al. 2011).

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

The authors wish to thank Sergio Sánchez Moragues for his contribution in the *hsp83* sequencing and Ariadna Gustems

Garcia and Aina Rotger Bosch for their contribution in the chromosomal preparations. They want to thank as well Lars Ootes for his contribution and advises in the clustering step of the TE pipeline annotation and Miriam Solé Morillas for her contribution to qRT-PCR experiments. We also would like to thank Andrea Betancourt and the anonymous reviewers for their valuable comments on the manuscript. This work was supported by Ministerio de Ciencia e Innovación (Spain) (grant number PID2021-127107NB-I00) and Generalitat de Catalunya (Spain) (grant number 2021 SGR 00526). A.B. was supported by a PIF predoctoral fellowship from the Universitat Autònoma de Barcelona (Spain).

### Data Availability

Raw RNA-seq and small RNA-seq data generated from this article have been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA 1023019; accession numbers from SRR26249410 to SRR26249441. 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>.

### Literature Cited

- Adams MD, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195.
- Ahmed W, et al. 2021. Identification and analysis of small interfering RNAs associated with heat stress in flowering Chinese cabbage using high-throughput sequencing. *Front Genet.* 12:1–9.
- Alexa A, Rahnenfuhrer J. 2020. topGO: enrichment analysis for gene ontology. <https://bioconductor.org/packages/release/bioc/html/topGO.html>.
- Antoniewski C. 2014. Computing siRNA and piRNA overlap signatures. *Methods Mol Biol.* 1173:135–146.
- Arnault C, Heizmann A, Loevenbruck C, Biémont C. 1991. Environmental stresses and mobilization of transposable elements in inbred lines of *Drosophila melanogaster*. *Mutat Res.* 248:51–60.
- Arnault C, Loevenbruck C, Biémont C. 1997. Transposable element mobilization is not induced by heat shocks in *Drosophila melanogaster*. *Naturwissenschaften* 84:410–414.
- Arrigo A-PP. 1983. Acetylation and methylation patterns of core histones are modified after heat or arsenite treatment of *Drosophila* tissue culture cells. *Nucleic Acids Res.* 11:1389–1404.
- Ayala FJ, Serra L, Prevosti A. 1989. A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. *Genome* 31:246–255.
- Bailly-Bechet M, Haudry A, Lerat E. 2014. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. *Mob DNA.* 5:13.
- Balanyá J, Oller JM, Huey RB, Gilchrist GW, Serra L. 2006. Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science* 313:1773–1775.

- Banho CA, Mérel V, Oliveira TYK, Carareto CMA, Vieira C. 2021. Comparative transcriptomics between *Drosophila mojavensis* and *D. arizonae* reveals transgressive gene expression and under-expression of spermatogenesis-related genes in hybrid testes. *Sci Rep.* 11:1–15.
- Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 6:11.
- Bedulina D, et al. 2021. Proteomics reveals sex-specific heat shock response of Baikal amphipod *Eulimnogammarus cyaneus*. *Sci Total Environ.* 763:143008.
- Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 125:279–284.
- Bodelón A, Fablet M, Veber P, Vieira C, García Guerreiro MP. 2022. High stability of the epigenome in *Drosophila* interspecific hybrids. *Genome Biol Evol.* 14:1–20.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Brennecke J, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089–1103.
- Bubenshchikova EV, Antonenko OV, Vasil'eva LA, Ratner VA. 2002. Induction of MGE 412 transposition individually by heat and cold shock in spermatogenesis in *Drosophila* males. *Genetika* 38: 46–55.
- Calabria G, et al. 2012. Hsp70 protein levels and thermotolerance in *Drosophila subobscura*: a reassessment of the thermal co-adaptation hypothesis. *J Evol Biol.* 25:691–700.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Cappucci U, et al. 2019. The Hsp70 chaperone is a major player in stress-induced transposable element activation. *Proc Natl Acad Sci U S A.* 116:17943–17950.
- Catalán A, Hutter S, Parsch J. 2012. Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* 13:654.
- Cernilogar FM, et al. 2011. Chromatin-associated RNAi components contribute to transcriptional regulation in *Drosophila*. *Nature* 480:391–395.
- Chakir M, Chafik A, Moreteau B, Gibert P, David JR. 2002. Male sterility thermal thresholds in *Drosophila*: *D. simulans* appears more cold-adapted than its sibling *D. melanogaster*. *Genetika* 114: 195–205.
- Chen P, et al. 2021. piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in *D. melanogaster* male germline. *Genes Dev.* 35:914–935.
- Collin JE. 1936. Note: *Drosophila subobscura* sp. n. male, female. *J Genet.* 33:60.
- Czech B, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453:798–802.
- Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- Dworniczak B, Mirault ME. 1987. Structure and expression of a human gene coding for a 71 kd heat shock 'cognate' protein. *Nucleic Acids Res.* 15:5181–5197.
- Emilie Q, Amit A, Kai T. 2016. The piRNA pathway is developmentally regulated during spermatogenesis in *Drosophila*. *RNA* 22: 1044–1054.
- Erić K, et al. 2022. Stress resistance traits under different thermal conditions in *Drosophila subobscura* from two altitudes. *Insects* 13: 1–17.
- Fast I, et al. 2017. Temperature-responsive miRNAs in *Drosophila* orchestrate adaptation to different ambient temperatures. *RNA* 23: 1352–1364.
- Flynn JM, et al. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 117: 9451–9457.
- Funikov SY, et al. 2015. The peculiarities of piRNA expression upon heat shock exposure in *Drosophila melanogaster*. *Mob Genet Elements.* 5:72–80.
- Galludo M, et al. 2018. Climatic adaptation of chromosomal inversions in *Drosophila subobscura*. *Genetica* 146:433–441.
- 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 (Basel)* 11:1–15.
- García Guerreiro MP, Chávez-Sandoval BE, Balanyà J, Serra L, Fontdevila A. 2008. Distribution of the transposable elements bilbo and gypsy in original and colonizing populations of *Drosophila subobscura*. *BMC Evol Biol.* 8:234.
- Gebert D, et al. 2021. Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol Cell.* 81:3965–3978.e5.
- Ghildiyal M, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320:1077–1081.
- Goubert C. TE-aid. <https://github.com/clemgoub/TE-Aid>.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29: 644–652.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34:D140–D144.
- Gruntenko NE, Bownes M, Terashima J, Sukhanova MZ, Raushenbach IY. 2003. Heat stress affects oogenesis differently in wild-type *Drosophila virilis* and a mutant with altered juvenile hormone and 20-hydroxyecdysone levels. *Insect Mol Biol.* 12:393–404.
- Gunawardane LS, et al. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315:1587–1590.
- Handler D, et al. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* 30: 3977–3993.
- Hsu SK, et al. 2020. Rapid sex-specific adaptation to high temperature in *Drosophila*. *Elife* 9:1–16.
- Huerta-Cepas J, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47: D309–D314.
- Huylmans AK, Parsch J. 2014. Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. *G3 (Bethesda)* 4:2307–2315.
- Jørgensen LB, Robertson RM, Overgaard J. 2020. Neural dysfunction correlates with heat coma and CTmax in *Drosophila* but does not set the boundaries for heat stress survival. *J Exp Biol.* 223: jeb218750.
- Junakovic N, Di Franco C, Barsanti P, Palumbo G. 1986. Transposition of copia-like nomadic elements can be induced by heat shock. *J Mol Evol.* 24:89–93.
- Juneja P, Quinn A, Jiggins FM. 2016. Latitudinal clines in gene expression and cis-regulatory element variation in *Drosophila melanogaster*. *BMC Genomics* 17:1–11.
- Karageorgiou C, Gámez-Visairas V, Tarrío R, Rodríguez-Trelles F. 2019. 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* 20:223.



- Kawamura Y, et al. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453:793–797.
- Kelleher ES, Barbash DA. 2013. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. *Mol Biol Evol.* 30:1816–1829.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.
- Khadem M, Rozas J, Segarra C, Brehm A, Aguadé M. 1998. Tracing the colonization of Madeira and the Canary Islands by *Drosophila subobscura* through the study of the rp49 gene region. *J Evol Biol.* 11: 439–452.
- Kofler R, Senti K-A, Nolte V, Tobler R, Schlötterer C. 2018. Molecular dissection of a natural transposable element invasion. *Genome Res.* 28:824–835.
- Krimbas CB. 1993. *Drosophila subobscura*: biology, genetics and inversion polymorphism. Hamburg: Verlag Dr. Kovac.
- Landis G, Shen J, Tower J. 2012. Gene expression changes in response to aging compared to heat stress, oxidative stress and ionizing radiation in *Drosophila melanogaster*. *Aging (Albany, NY)* 4: 768–789.
- Langmead B. 2010. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma.* Chapter 11:Unit 11.7.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.
- Lebo MS, Sanders LE, Sun F, Arbeitman MN. 2009. Somatic, germline and sex hierarchy regulated gene expression during *Drosophila* metamorphosis. *BMC Genomics* 10:1–23.
- Lecheta MC, et al. 2020. Integrating GWAS and transcriptomics to identify the molecular underpinnings of thermal stress responses in *Drosophila melanogaster*. *Front Genet.* 11:1–17.
- Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. 2017. TETOOLS facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* 45:e17.
- Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. 2003. Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Mol Biol Evol.* 20: 135–144.
- Li C, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137:509–521.
- Loeschcke V, Bundgaard J, Barker JSF. 2000. Variation in body size and life history traits in *Drosophila aldrichi* and *D. buzzatii* from a latitudinal cline in eastern Australia. *Heredity (Edinb).* 85:423–433.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15:550.
- Luo S, Lu J. 2017. Silencing of transposable elements by piRNAs in *Drosophila*: an evolutionary perspective. *G3 (Bethesda)* 15: 164–176.
- Madrenas R, Balanyà J, Arenas C, Khadem M, Mestres F. 2020. Global warming and chromosomal inversion adaptation in isolated islands: *Drosophila subobscura* populations from Madeira. *Entomol Sci.* 23:74–85.
- Malone CD, et al. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137:522–535.
- McCullers TJ, Steiniger M. 2017. Transposable elements in *Drosophila*. *Mob Genet Elements.* 7:1–18.
- Mérel V, et al. 2021. 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.* 38:4252–4267.
- Michaud S, Marin R, Tanguay RM. 1997. Regulation of heat shock gene induction and expression during *Drosophila* development. *Cell Mol Life Sci.* 53:104–113.
- Morgulis A, et al. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764.
- Muñoz MM. 2022. The Bogert effect, a factor in evolution. *Evolution (NY)* 76:49–66.
- Neph S, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28:1919–1920.
- Ou S, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20:275.
- Ozata DM, et al. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet.* 20:89–108.
- Parsell DA, Lindquist S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu Rev Genet.* 27:437–496.
- Pauli D, Arrigo AP, Tissières A. 1992. Heat shock response in *Drosophila*. *Experientia* 48:623–629.
- Pimpinelli S, et al. 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc Natl Acad Sci U S A.* 92:3804–3808.
- Puig Giribets M, Santos M, García Guerreiro MP. 2020. 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.* 20:17.
- Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. 2022. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* 38:2943–2945.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- R Core Team. 2020. R: A language and environment for statistical computing. <https://www.r-project.org/>.
- Rego C, et al. 2010. Clinal patterns of chromosomal inversion polymorphisms in *Drosophila subobscura* are partly associated with thermal preferences and heat stress resistance. *Evolution* 64:385–397.
- Rice N, Ho S, Weng Z, Theurkauf WE. 2023. Rapid disassembly and Piwi-independent reassembly of *Drosophila* piRNA cluster heterochromatin following acute heat shock. *bioRxiv.* 2023.07.26.549158; doi: <https://doi.org/10.1101/2023.07.26.549158>
- Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods.* 10:71–73.
- Robinson JT, et al. 2011. Integrative genomics viewer. *Nat Biotechnol.* 29:24–26.
- Rodgers CI, et al. 2007. Stress preconditioning of spreading depression in the locust CNS. *PLoS One* 2:e1366.
- Rosenkranz D, Zischler H. 2012. proTRAC—a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* 13:5.
- Saint-Leandre B, Capy P, Hua-Van A, Filée J. 2020. piRNA and transposon dynamics in *Drosophila*: a female story. *Genome Biol Evol.* 12:931–947.
- Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc.* 3:1101–1108.
- Senti K-A, Jurczak D, Sachidanandam R, Brennecke J. 2015. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev.* 29:1747–1762.
- Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Sørensen JG, Kristensen TN, Loeschcke V. 2003. The evolutionary and ecological role of heat shock proteins. *Ecol Lett.* 6:1025–1037.
- Sørensen JG, Nielsen MM, Kruhøffer M, Justesen J, Loeschcke V. 2005. Full genome gene expression analysis of the heat stress response in *Drosophila melanogaster*. *Cell Stress Chaperones.* 10:312–328.
- Specchia V, et al. 2010. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* 463:662–665.

- Srikanth K, et al. 2020. Genome-wide transcriptome and metabolome analyses provide novel insights and suggest a sex-specific response to heat stress in pigs. *Genes (Basel)* 11:540.
- Strand DJ, McDonald JF. 1985. Copia is transcriptionally responsive to environmental stress. *Nucleic Acids Res.* 13:4401–4410.
- Tanguay RM. 1983. Genetic regulation during heat shock and function of heat-shock proteins: a review. *Can J Biochem Cell Biol.* 61: 387–394.
- Tariq M, Nussbaumer U, Chen Y, Beisel C, Paro R. 2009. Trithorax requires Hsp90 for maintenance of active chromatin at sites of gene expression. *Proc Natl Acad Sci U S A.* 106:1157–1162.
- Teets NM, Hahn DA. 2018. Genetic variation in the shape of cold-survival curves in a single fly population suggests potential for selection from climate variability. *J Evol Biol.* 31:543–555.
- Teves SS, Henikoff S. 2013. The heat shock response: a case study of chromatin dynamics in gene regulation. *Biochem Cell Biol.* 91: 42–48.
- Théron E, Dennis C, Brassat E, Vaury C. 2014. Distinct features of the piRNA pathway in somatic and germ cells: from piRNA cluster transcription to piRNA processing and amplification. *Mob DNA.* 5: 1–11.
- Tower J. 2011. Heat shock proteins and *Drosophila* aging. *Exp Gerontol.* 46:355–362.
- UniProt Consortium. 2021. UniProt: the universal protein knowledge-base in 2021. *Nucleic Acids Res.* 49:D480–D489.
- Vasilyeva LA, Bubenshchikova E V, Ratner VA. 1999. Heavy heat shock induced retrotransposon transposition in *Drosophila*. *Genet Res.* 74:111–119.
- Vieira C, Lepetit D, Dumont S, Biémont C. 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol.* 16:1251–1255.
- Walser JC, Chen B, Feder ME. 2006. Heat-shock promoters: targets for evolution by P transposable elements in *Drosophila*. *PLoS Genet.* 2: e165.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. <https://ggplot2.tidyverse.org>.
- Wierzbicki F, Kofler R, Signor S. 2023. Evolutionary dynamics of piRNA clusters in *Drosophila*. *Mol Ecol.* 32:1306–1322.
- Wurmser F, Mary-Huard T, Daudin JJ, Joly D, Montchamp-Moreau C. 2013. Variation of gene expression associated with colonisation of an anthropized environment: comparison between African and European populations of *Drosophila simulans*. *PLoS One* 8: e79750.
- Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 35:2084–2092.
- Zytnicki M, Gaspin C. 2022. Snamapper: an optimal mapping tool for sRNA-Seq reads. *BMC Bioinformatics.* 23:495.

**Associate editor:** Andrea Betancourt