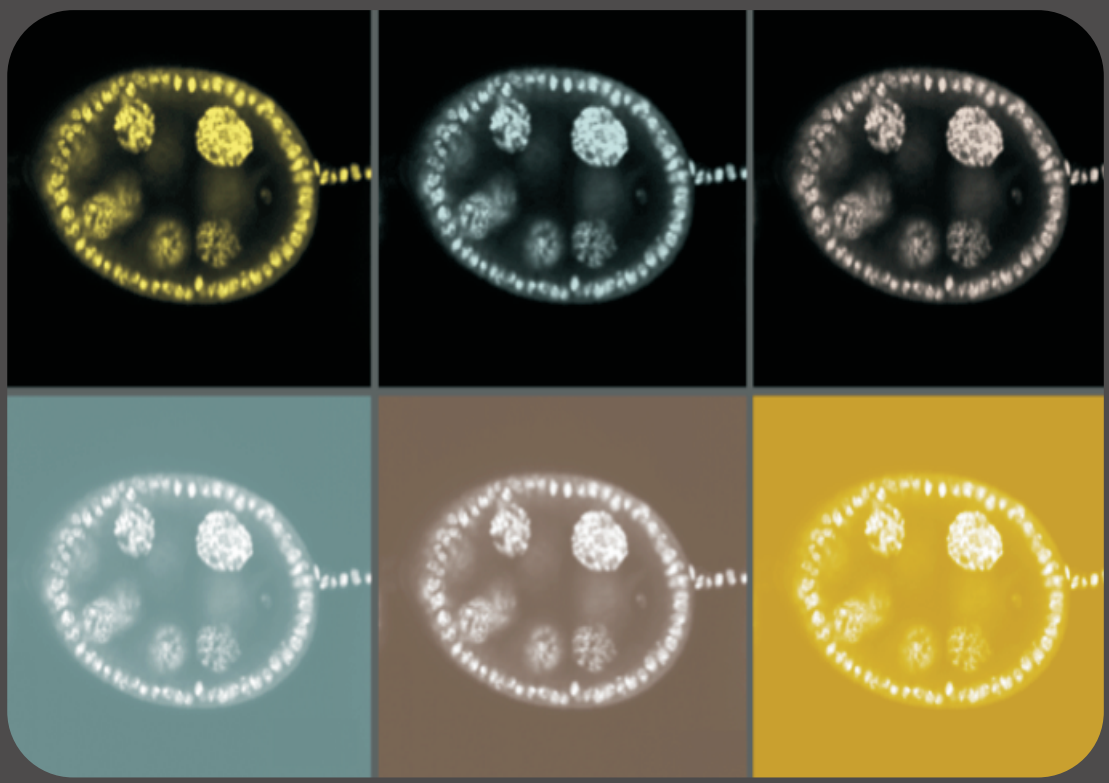
Effects of environmental and genetic factors on transposable element activity

Marta MARIALVA



Dissertation presented to obtain the Ph.D degree in Evolutionary Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
November, 2016

UNIVERSIDADE NOVA DE LISBOA

Effects of environmental and genetic factors on transposable element activity

Marta MARIALVA

Dissertation presented to obtain the Ph.D degree in Evolutionary Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras, November, 2016

Aos meus pais…

Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, INOVAÇÃO E DO ENSINO SUPERIOR

INSTITUTO GULBENKIAN DE CIÊNCIA

The research study described in this thesis was financially supported by Fundação para a Ciência e a Tecnologia (FCT) fellowship SFRH/BD/51882/2012. This thesis was printed with support from Instituto Gulbenkian de Ciência.

ROAD TO EXTINCTION
“carry the weight or crush under it”
Until we are no more (Breathe)

Moonspell

Acknowledgements

There is no successful work that results from the performance of isolated individuals, and it was the personal- and work-related help I had that drove me forward. There are therefore a lot of people that I would like to show my appreciation to.

First, I would like to acknowledge the people that directly influenced me during the PhD and that contributed with work for this thesis. This includes my supervisor, Patrícia Beldade, for accepting me in her lab and for the (almost blind) trust in the work I was doing. She was brave enough to allow me to work on my scientific interests, and to develop this project that went way beyond the focus of the lab. Thank you for letting me grow from my mistakes and for providing the freedom I needed to mature both scientific- and personally. The skills I acquired in your lab are definitely going to influence me for the rest of my life.

During this process I was lucky enough to have Ana Eugénio by my side. Her enthusiasm and hard work always fascinated me, and she greatly contributed with results for this thesis. Ana, your work was really appreciated and this thesis is also yours.

I also received the help of two very talented students (Joana Carvalho and André Alves) that were able to produce good quality data in such a short period of time.

One chapter of this thesis is the result of a collaboration established with Banu Önder laboratory. It was a great pleasure to work with her. Our interactions made me believe (again) that it is still possible to do science in a collaborative, rather than competitive, manner.

I will always be in debt to Oscar Ruiz. He read almost every chapter of this thesis when I needed it the most. Thank you for your critical advises and your words of encouragement. Thank you for being “EXACTLY what friends are for”.

This thesis would not exist if not for the extraordinary environment that exists at Instituto Gulbenkian de Ciência (IGC), for its laboratories without boundaries and the amazing facilities. Many things in this work are the result of the direct interaction with people at IGC, and I feel immensely grateful to those that spend their time to help me. In particular, I would like to thank to the bioinformatics (Renato Alves and Daniel Sobral), microscopy (Ânia Ferreira and Nuno Martins) and fly (Liliana Vieira, Carina Monteiro and Sandra Crisóstomo) facilities. I would also like to acknowledge Nelson Martins and David Duneau for all the help with statistical analysis, and to Alisson Gontijo, Christen Mirth, Élio Sucena and Luis Teixeira laboratories for the important discussions and suggestions regarding this project. Also, from Patrícia Beldade's lab I would like to thank to Maria Adelina Jerónimo, Elvira Lafuente and Carolina da Silva for all the fun during the (long) hours spent at the bench. For understanding what-it-is-like, and for the great discussions we had outside the walls of the lab. Elvira also read the chapter II of this thesis. Thank you for your sincere words, thank you. To Susana Ramos for help with summaries of this thesis and to Catarina Carmo for your comments on the introduction. To my thesis committee, Lounés Chikhi and Miguel Godinho, for the meaningful advises when I needed them the most (even if they were not always followed).

I would like to show my appreciation to Élio Sucena, for teaching me how to think and for introducing me to evo-devo. *"A teacher affects eternity"*, you definitely affected mine.

This thesis would not be the same if not for Luis Teixeira and former members of his laboratory (specially to Gil Ferreira and Sara Esteves). I learned everything I know about *Drosophila* and *Wolbachia* with you. Thank you for all the friendship, and the great time we had while working together. Your critical mind and interest in science motivated me to pursue other questions in biology and significantly affected the way I looked at my data (later) during my PhD. It was a great pleasure to work with you all.

I also thank to Thiago Carvalho – former program director – for selecting me to enter the program only when I was ready for it.

I would like to thank to PIBS 2012, especially to the three musketeers that followed more closely this four-year journey. To Sandra Tavares: for your practical mind and skepticism. Thank you for reading my mind when I didn't want to talk. To Inês Pais: for teaching me that people are worth it and that at the end of each story there will be a rainbow (sometimes). To Sara Esteves: for reminding me of whom I am. You inspire me, and your words will always be with me in every choice I make.

To my "*tatlim*"… My love… Thank you for bringing music into my life. Thank you for being by my side in the good moments and for transforming bad moments into good ones. For all the patience and understanding (specially when I was in front of the computer and I did not listen to anyone around me). Thank you for all the unconditional and "*Untouchable*" love. Thank you for making life worth living.

Aos meus pais… Por me ensinarem a ter os pés no chão e a acreditar que o céu é o limite. Por todo o apoio e carinho constante. Por ignorarem o mau feitio depois de longas horas de trabalho. Por isto e tudo o resto: sem vós, seria nada…

Marta MARIALVA
October, 2016

Table of contents

Summary 13

Resumo 15

Chapter I 17

Chapter II 59

Chapter III 107

Chapter IV 141

Chapter V 177

Chapter VI 199

Summary

Heritable phenotypic variation is the raw material for evolution to occur through natural selection. It is a pervasive property of living organisms and, in order to understand evolutionary processes, we must study the factors that mediate the production of heritable phenotypic, and therefore genotypic variation.

Transposable elements (TEs) are mobile DNA sequences that constitute a large proportion of most eukaryotic genomes and that can drive the production of adaptive genetic variation. There are many studies on transposon mobility, but these generally include only few environmental conditions, TEs and genetic backgrounds. As a consequence, we still know very little about the impact of genetic and environmental factors that can affect transposon dynamics in natural populations. Here, I proposed to overcome those limitations by analysing the impact of a multitude of factors on transposon mobility using *Drosophila melanogaster* as a model organism.

To identify loci carrying natural allelic variants affecting TE copy-number, we performed a genome-wide association study, using a mapping panel of *D. melanogaster* wild-derived genotypes (**Chapter 2**). We then selected four *D. melanogaster* stocks infected with *Wolbachia* – a maternally transmitted α-Proteobacteria, which renders the host resistant to infection by RNA viruses – to test the effect of symbiotic bacteria on the expression of transposons that resemble retroviruses. We further asked whether TE activity in the presence of the endosymbiont was influenced by temperature and host genetic background, thus investigating genetic-by-environment-by-environment (GxExE) interactions (**Chapter 3**). This analysis was extended to other abiotic factors; notably, exposure to oxidative stress and heavy metals in the diet, and their interaction with the presence of *Wolbachia* (ExE interactions) (**Chapter 4**).

Even though the piRNA pathway constitutes the hallmark of organisms' protection against TE activity in the germline, acting through the production of small RNAs that interact with PIWI family proteins (piRNAs), we were only able to find few piRNA pathway loci associated with transposon copy-number (**Chapter 2**). Additionally, this pathway was also found to be robust to

environmental perturbations (**Chapter 4**). Based on the analysis of the associated loci and on the TE responses to perturbation, we propose that changes in transcription factor activity could be at the basis of genotype- and environmental-specific expression of transposons (**Chapter 2, 3 and 4**).

To further assess the impact of *Wolbachia* in the context of TE invasion of natural populations, and given the known correlation between a recent replacement of *Wolbachia* strains and the spread of the TE *P-element*, we characterized a panel of 86 natural genotypes according to the presence of *Wolbachia* and susceptibility to *P-element* invasion (**Chapter 5**). This analysis suggested that the presence of the endosymbiont might have contributed to promote *P-element* spread in natural populations through beneficial and genotype-specific effects on fecundity and/or percentage of eclosion.

In summary, this thesis provides novel insight about the complexity of interactions established between TEs, environmental conditions and host genotypes. We also show that increased TE expression in ovaries can potentially lead to novel insertions in the next generation and that regulation at additional steps of the transposon "*life cycle*" can add an extra layer of complexity in the ability to produce genetic variation.

Resumo

Toda a variação fenotípica que é herdada encontra-se codificada no genoma e constitui uma propriedade universal de todos os sistemas biológicos. Essa variação pode ser filtrada e/ou induzida por determinadas condições ambientais. Sendo crucial para evolução, o seu estudo é determinante para compreender os processos evolutivos.

Transposões são sequências de DNA com propriedades móveis que compõem uma importante proporção dos genomas de eucariotas, e que intervêm na produção de variação genética adaptativa. Embora existam estudos relevantes que descrevem a mobilidade destes elementos, os factores genéticos e ambientais que ativam transposões são ainda amplamente desconhecidos. Esta tese tem como principal objectivo compreender a dinâmica de transposição em diversos genótipos de *Drosophila melanogaster*, e como esta é afetada pela exposição a factores bióticos e abióticos.

Em primeiro lugar, testámos a associação entre o número de cópias de transposões e a variação alélica que segrega em populações naturais de *D. melanogaster*, que existem num painel de linhas estabelecidas no laboratório (**Capítulo 2**). Destas linhas, foram selecionadas para posterior análise algumas que continham infecção com *Wolbachia* – uma α-Proteobacteria transmitida maternalmente e que confere ao hospedeiro resistência a infecções virais – permitindo assim o estudo da interação entre o endossimbionte, a temperatura e o genótipo (AxAxG) do hospedeiro na modulação da expressão de transposões que partilham semelhanças com retrovírus (**Capítulo 3**). Paralelamente, e de forma a enriquecer o estudo com outros factores abióticos, foi também incluída na análise a interação entre a presença de *Wolbachia* e a resposta a stresse oxidativo ou a dietas ricas em metais pesados (AxA) (**Capítulo 4**).

Embora o processo de produção de piRNAs (pequenos RNAs que reconhecem sequências de transposões e que interagem com proteínas PIWI) constitua o principal modo de repressão de elementos móveis na linha germinal de eucariotas, observámos que poucos alelos em genes relacionados com este mecanismo se encontram associados ao número de cópias de transposões

(**Capítulo 2**). A análise de possíveis efeitos ambientais também revelou que este mecanismo é robusto a perturbações bióticas e abióticas (**Capítulo 4**). Tendo em conta a falta de ligação entre o principal mecanismo de repressão de transposões e a resposta destes elementos a factores genéticos e ambientais, propomos que alterações na atividade de factores de transcrição possam estar na base da resposta dos transposões observada, sendo esta específica para os factores testados (**Capítulo 2, 3** e **4**).

Para estudar o impacto do endossimbionte *Wolbachia* durante o processo de invasão de transposões em populações naturais, tivemos em conta a correlação estabelecida entre a substituição de genótipos de *Wolbachia* e a invasão do transposão *P* em populações de *D. melanogaster* (**Capítulo 5**). A nossa análise de 86 genótipos naturais, avaliando a presença de *Wolbachia* e a sua susceptibilidade à invasão de elementos *P,* sugere que a presença do endossimbionte pode ter promovido a invasão deste elemento, ao conferir efeitos benéficos na fecundidade e percentagem de eclosão, efeitos que são específicos para cada genótipo.

Em suma, esta tese revelou interações complexas, estabelecidas entre transposões, condições ambientais e genótipos, onde cada elemento móvel responde de forma específica a cada factor testado. Os nossos resultados sugerem também que um aumento da expressão de transposões nas gónadas poderá conduzir a novas inserções, que serão transmitidas à geração seguinte, e realçam as alterações ambientais como um factor importante na produção de variação genética hereditária mediada por transposição.

Chapter I

General Introduction

Environmental conditions shape evolution by acting both as an agent of natural selection, involved in the sorting of genetic variation, and as a factor influencing the production of novel genetic variation, through effects on recombination and mutation rates. During the past few years, much attention has been devoted towards the exploration of the genetic basis of adaptive evolution[1,2], and important insights have emerged from experimental evolution studies that attempt to link three key components of evolution: genotype, phenotype and fitness[3–7]. However, most of those studies typically use rather simple environments, with changes in few variables, that fall short of properly representing many of the ecological interactions that mediate adaptation in wild populations[8]. Also, few studies succeed at connecting variation in phenotypic traits with variation in allele frequencies as environmental conditions change in nature[1,9,10]. New technologies provide tools that will allow better exploration of the genetics of adaptation to changing environments by allowing the fast genetic mapping of phenotypic traits, and genome-wide analyses of populations evolving in the wild[8,11–15]. Those approaches can help to properly establish the link between genotype, phenotype and fitness, and to understand how and why allele frequencies change in nature.

The evolutionary success of a population depends on available genetic variability that is initially produced through mutagenesis and rearranged by recombination. Intraspecific variation in mutation and recombination rates can also be subjected to natural selection as populations adapt to a new environment[16,17]. In addition, there is phenotypic plasticity of those traits, as inferred by the environmental ability to influence the frequency of mutation and/or recombination in same genotype[18–26]. Environmental factors such as temperature and nutritional intake are important players in the establishment of novel genetic variation[18–20,24–26].

1.1 Response to stressful environments

Environmental heterogeneity is a pervasive property of all ecological niches, and conditions are called stressful when they result in reduced individual fitness[27]. Suboptimal conditions caused by either biotic or abiotic factors can be grouped into chronic, mild or acute based on the severity and/or duration of the stress[28]. Most studies about the effects of dynamic environments focus on acute stresses that require immediate and precise cellular responses.[28–30].

In order to circumvent stress and re-establish homeostasis, organisms need to sense stress and induce proper signal transduction pathways that activate effector processes[28]. Signalling pathways tend to be more conserved than sensing and effector mechanisms, and generally respond to a variety of environmental factors[28,31]. For example, mitogen-activated protein kinase (MAPK) family elements are activated by several stimuli such as gamma rays, heat shock, infection, ultraviolet light, osmotic and oxidative stress[28,31–37]. How specificity is obtained when the same pathway responds to different stresses may rely on 1) the differential activation, in space and time, of the MAPK proteins, and 2) the simultaneous induction of signal transduction pathways that are exclusive to each environment and may create a stimulus-specific network of responsive elements[28].

1.1.1 Mechanisms of response to temperature

In the wild, populations are exposed to temporal and spatial fluctuations in temperature that may affect organismal performance and fitness[38–41]. Individuals evolved different mechanisms to sense thermal conditions and trigger effector responses upon exposure to sub-optimal temperatures[42–45]. The relative contribution of different sensing mechanisms depends on the species, previous exposure and/or the amplitude of thermal perturbation[42,46–48].

Heat shock proteins (Hsps) are the hallmark of temperature-mediated cellular processes and promote proper protein folding in response extreme insults such as heat, oxidative stress, infection and presence of heavy metals[49]. In response

to thermal perturbation, Hsps can undergo conformational changes in a thermosensor domain that are necessary to activate their chaperone activity[42]. In addition, Hsps can also indirectly detect changes in temperature via their high binding affinity to heat-induced denatured proteins[45]. The recruitment of Hsps to temperature-mediated protein aggregates results in the activation of heat shock factors (HSFs) that conversely elicit the transcription of *Hsp* genes (**Figure 1.1**)[50–52].

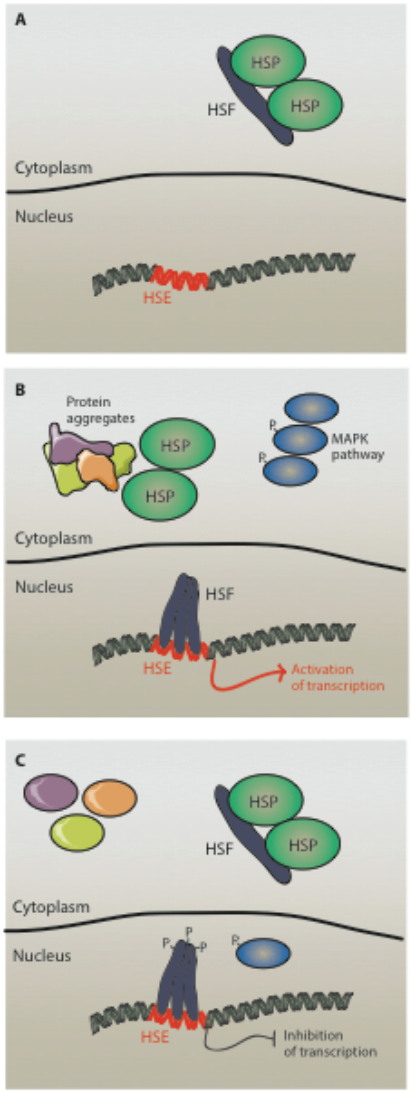


Figure 1.1. Effector processes mediating response to temperature stress. In the absence of thermal perturbations, heat shock factor (HSF) is stabilized in the cytosol through the interaction with heat shock proteins (Hsps)[49] **(A)**. Upon temperature shock, several proteins acquire non-native conformations that ultimately result in the formation of aggregates. As Hsps are recruited to mediate proper protein folding, HSF is released and activated by forming homotrimers[53–57]. HSF trimers travel to the nucleus and induce transcription of temperature responsive genes, including *Hsps*, that contain heat shock *cis*-regulatory elements (HSE)[58]. At the same time, the mitogen-activated protein kinase (MAPK) pathway is induced[59] **(B)**. This mediates HSF phosphorylation and alleviates its influence in gene expression. Additionally, as cytosolic proteins acquire proper configuration, Hsps are released and contribute to re-repressing HSF **(C)**.

Aside from the ability of Hsp to specifically sense thermal fluctuations, almost all biomolecules undergo conformational changes when temperature oscillates and can hence provide competent mediators of direct or indirect responses[28,47]. Also, protein binding affinities and kinetics of enzymatic reactions may be affected by thermal conditions, possibly without changes in protein conformation[60,61]. Effects in DNA topology, RNA secondary structures, and/or in the activity of transcriptional repressors, sensor kinases and transient receptor potential (TRP) cation channels can thus enhance cellular responses to sudden changes in temperature[46,47,62,63]. ThermoTRPs are evolutionarily conserved proteins with temperature-dependent conductances that can be expressed in many cell types including neurons[46]. The activation of TRP channels is crucial to induce thermotaxis and hence mediate behavioural avoidance of harmful temperatures[43,64]. However, and especially when individuals cannot avoid prolonged exposure to stress, the transcriptional control of stress-related genes is fundamental for thermotolerance. To mediate those responses, cells activate inducible translational regulators such as the evolutionary conserved HSFs (**Figure 1.1**)[49]. The tight temporal restriction of HSF-mediated transcription is crucial to ensure cell growth and homeostasis[28].

1.1.2 Mechanisms of response to oxidative stress

Reactive oxygen species (ROS) can be produced by normal intracellular metabolism and play a physiological role in cellular proliferation. Increasing ROS above homeostatic levels, driven by environmental perturbations, can be detrimental and may ultimately result in cellular damage. Antioxidant defences are hence necessary to resist stress. During this process, damaged molecules are repaired or replaced and ROS levels return to normal values. It is the balance between production and destruction of ROS that reflects the degree of oxidative stress[65]. Infection, temperature perturbations and exposure to heavy metals are among the environmental factors known to induce oxidative stress through increase in ROS levels that lead to DNA damage and problems in protein conformation (**Figure 1.2** and **Figure 1.3**)[65–67].

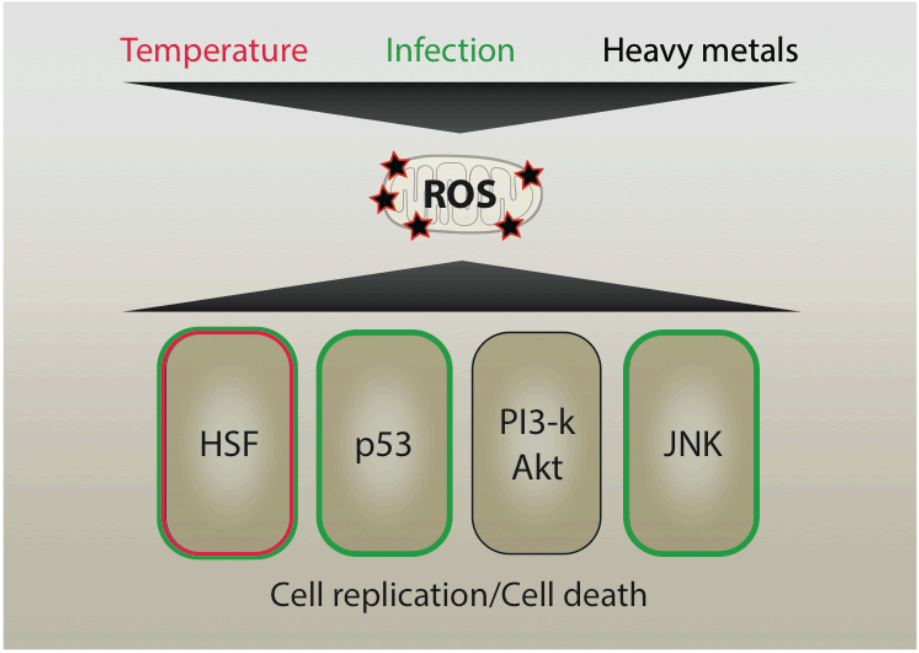


Figure 1.2. Cross-talk between response to temperature stress, infection, and exposure to heavy metals. The production of reactive oxygen species (ROS) in the mitochondria is enhanced after exposure to several environmental perturbations such as temperature stress, infection and presence of heavy metals[65–67]. Upon oxidative stress, cells induce diverse effector processes including the activation of heat shock factor (HSF), p53, c-Jun N-terminal kinase (JNK) cascades, as well as serine/threonine kinase Akt in response to phosphoinositide 3-kinase (PI3-k/Akt)[68]. The severity and type of stress result in different levels of relative activation of each pathway and dictate the cellular ability to replicate or induce programmed cell death[68,69]. In particular, HSF is triggered by both pathogens and thermal perturbation whereas p53 and JNK are induced only after infection[33,70,71].

ROS appear to act as messengers capable of eliciting cellular responses to environmental perturbation[72]. As a consequence of enhanced ROS production, molecules such as DNA, proteins and lipids suffer reversible or irreversible damage that depends on the degree of oxidative stress[73]. Those molecules that are directly affected by ROS can then directly activate signalling pathways that circumvent oxidant injury via pro-survival or apoptotic responses (**Figure 1.3**)[69].

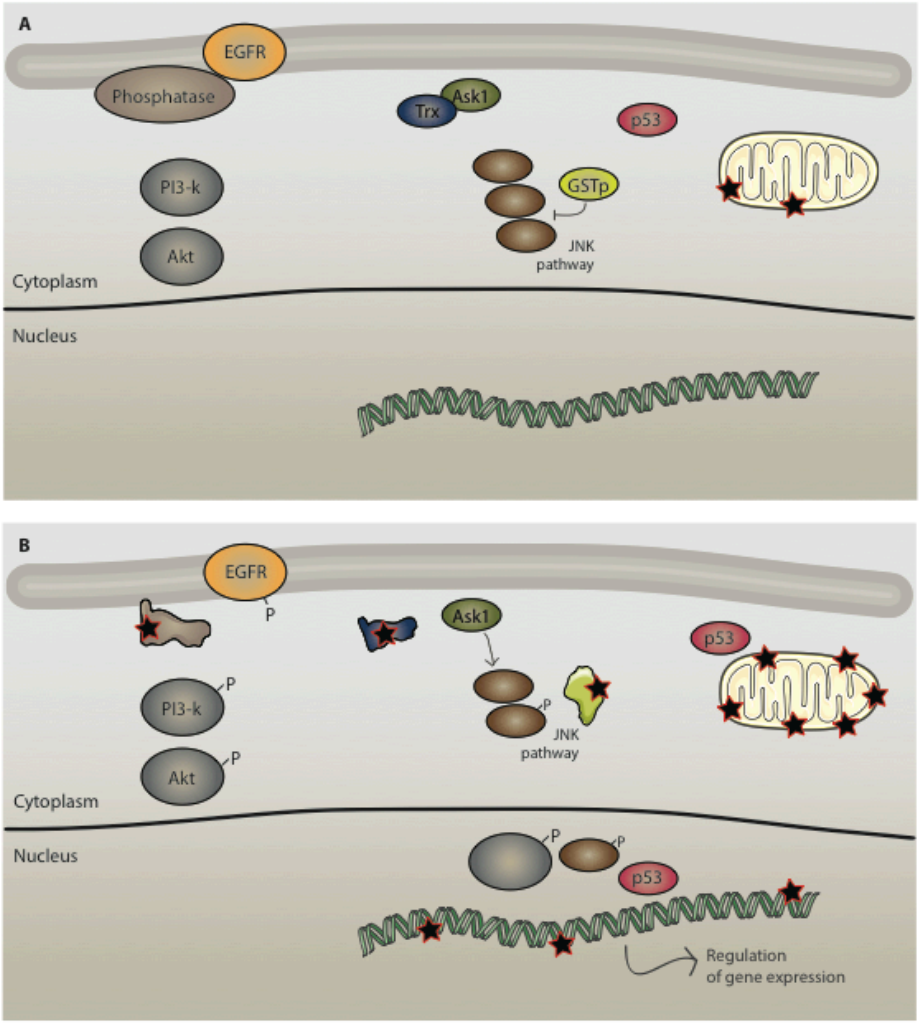


Figure 1.3. Effector processes mediating response to oxidative stress. In normal conditions, thioredoxin (Trx) interacts with signal-regulating kinase (Ask1) inhibiting the c-Jun N-terminal kinase (JNK) cascade[74]. In parallel, glutathione S-transferase Pi (GSTp) protein synergistically represses the JNK pathway[75]. Active phosphatases guarantee that growth factor receptor (EGFR) is kept in an inactive state and do not signal to the phosphoinositide 3-kinase (PI3-k/Akt) pathway[76,77] **(A)**. When cells are exposed to environmental perturbations (e.g. temperature stress, infection or presence of heavy metals), levels of reactive oxygen species (ROS) increase and directly trigger DNA damage and problems in protein conformation. As a consequence of ROS-mediated genomic double strand breaks, p53 is activated; it migrates to the nucleus where it induces gene expression and to the mitochondria where more ROS are produced[78,79]. In

addition, as phosphatases acquire non-native conformation, EGFR becomes active and mediates PI3-k/Akt phosphorylation[76,77]. Also, Trx and GSTp inactivation results in the induction of JNK cascade[74,75]. The balance between these stress-activated pathways dictates the ultimate outcome in gene expression (induction of pro-survival or pro-apoptotic genes) necessary to circumvent stress[68]. Increase in the expression of ROS scavengers and Hsps may contribute to the establishment of antioxidant defences and help cells to evade injury. Alternatively, severely damaged cells may undergo apoptosis and be removed from the multicellular host to preserve the organism[69] **(B)**.

1.1.3 Symbiotic interactions

Symbiosis was initially defined by Heinrich Anton de Bary as "*the living together of two dissimilar organisms, usually in intimate association"*, in reference to the relationship between fungi and algae in lichen formation[80]. Traditionally, the definition includes a spectrum of interactions that range from beneficial to harmful[81]. Among those, obligate heritable symbionts comprise the most extreme cases, as microorganisms depend on the host to live or complete its life cycle, and will be the focus of this essay. Examples include interaction between phylogenetically diverse microorganisms and hosts such as those established between several species of bacteria and arthropodes[82,83], nematodes[84,85], molluscs[86,87] or porifera[88,89]. Here, we focus on *Wolbachia*, a widespread intracellular bacteria, and provide some insights about endosymbiont-host interactions.

Wolbachia are obligatory intracellular and maternally transmitted α-proteobacteria. To increase chances of transmission, these Gram-negative bacteria often manipulates several aspects of host reproduction that depends on bacteria strain and host species. Those manipulations favor females, the host type that can transmit *Wolbachia*, and include promoting parthenogenesis, cytoplasmic incompatibility, male killing, and feminization[82]. Although the mechanisms whereby this is achieved are still not completely understood, we know that many rely on *Wolbachia*'s ability to disrupt cell cycle[82,90,91]. In addition, *Wolbachia*'s usage of the host cells' spindle apparatus, as well as dynein and

kinesin motors, promotes efficient transmission to the germline[92–95] and, only recently, have been shown *Wolbachia*'s ability to promote germline stem cell self-renewal[96].

To fight infection, invertebrates rely exclusively on innate immune responses[33]. Peptidoglycan recognition proteins (PGRPs), including PGRP-LC and PGRP-LE, are expressed in several tissues of *D. melanogaster* (FlyAtlas[97]) and mediate the activation of antimicrobial peptides in the presence of Gram-negative bacteria[98–102]. Although PGRP-LC is a membrane-bound protein that exclusively recognizes bacteria outside the cell, PGRP-LE acts as both extra- and intracellular receptor[103]. PGRP-LE can therefore recognize Gram-negative bacteria inside the cell and trigger immune defence against endosymbionts[33,103] which, on the other hand, have also evolved defensive mechanisms aimed at evading recognition[104,105]. The presence of an extra membrane around *Wolbachia* suggests that their proficiency to persist inside invertebrate cells relies on their ability to disable targeting while using host endoplasmic reticulum membrane as camouflage[106,107].

1.2 Transposable elements (TEs)

Pioneer work of Barbara McClintock regarding genomic instability in maize granted her the Nobel Prize in Physiology or Medicine (1983) for the discovery of motile DNA sequences now known as transposable elements (TEs)[108,109]. The detection of changes in kernel colour patterning associated with TE mobilization suggested those elements as important contributors for the regulation of gene expression. Such a view was later forgotten by fashionable theories that considered TEs as selfish genetic elements with no phenotypic or evolutionary function[110,111]. Only recently, were those two perspectives brought together along with new integrative ideas that consider the relationship between mobile genetic elements and their host genome as ranging from parasitic to mutualistic[112]. Along with those premises, and although most insertions, like any mutation, will have deleterious effects, TE insertions were also accepted as important source of adaptive genetic variation[112–114].

1.2.1 A unified classification system for eukaryotic TEs

Recent breakthroughs in sequencing and software technologies have provided significant advances in the detection of TE sequences[115–117]. Accordingly, extensive sequence diversity and a substantial proportion of mobile genetic elements has been found in almost all genomes analysed so far[117–120], making up for 45% of the human genome, for example. The diversity of TEs found within and across species pose problems in TE annotation and raised a need for a unified classification system of mobile elements[121]. Thomas Wicker *et al.* provided a hierarchical grouping system based on the mechanism of transposition, sequence similarity and structural relationships[121]. TEs are hence divided into two main classes depending on the mechanism of jumping (**Figure 1.4**): 1) retrotransposons move via an RNA intermediate and two copies are formed at the end of each replication cycle (*copy-and-paste* mechanism), and 2) DNA transposons are excised from one genome site and integrated into another and, therefore, number of copies is usually maintained after transposition (*cut-and-paste* mechanism)[121,122]. Among retrotransposons and DNA elements, we will focus on autonomous TEs – elements that encode all proteins necessary for transposition – rather than non-autonomous transposons that lack jumping capacity when alone.

Retrotransposons are similar to retroviruses in that they reverse transcribe an RNA intermediate that is later integrated in another genomic region. According to their sequence properties, retrotransposons are defined as elements with or without long terminal repeats, LTRs and non-LTRs respectively. LTRs are repetitive sequences ranging from a few hundred base pairs to more than 5 kb that appear flanking some retrotransposons[121]. They are usually in the same orientation (**Figure 1.4B**) and start with 5'-TG-3' and end with 5'-CA-3'[121,123]. Genetic variation in these flanking sequences can be used to distinguish different TE families[124,125].

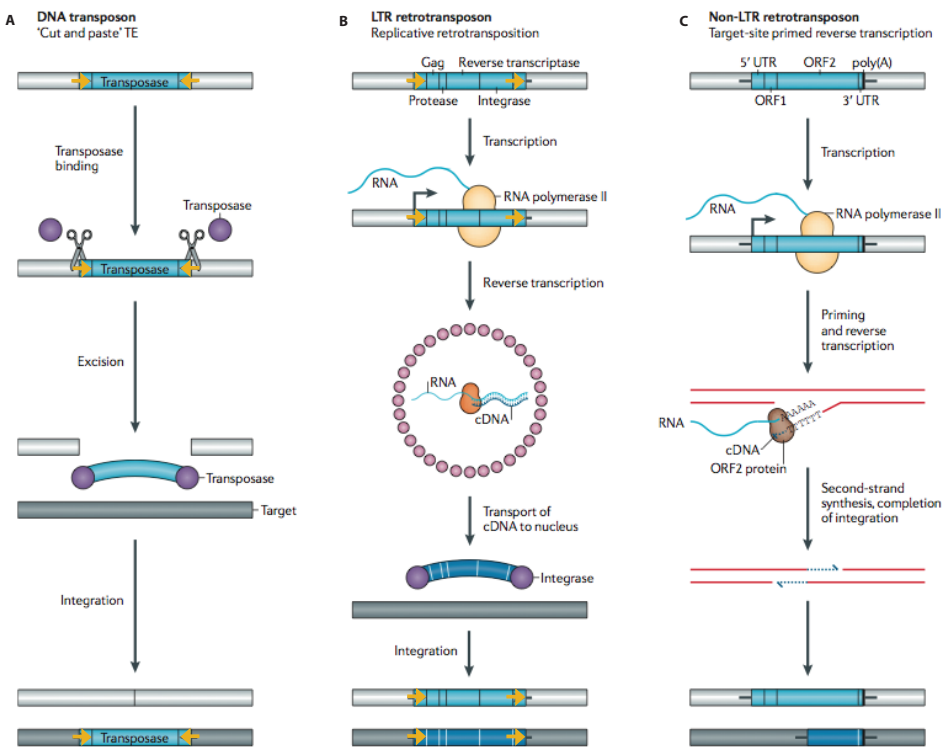


Figure 1.4. Mechanisms of TE jumping. DNA elements **(A)** move via a *cut-and-paste* mechanism mainly mediated by transposase activity. Each DNA element encodes own transposase that specifically recognizes terminal inverted repeats at each TE end (yellow arrows), mediating DNA cleavage and transposon excision. Transposase also catalyzes TE integration in a new genomic area. LTR **(B)** and non-LTR **(C)** elements are retrotransposons and move via an RNA intermediate. LTR retrotransposons are flanked by long terminal repeats (yellow arrows) and code for: protease, viral particle coat (Gag), reverse transcriptase and integrase proteins. Host transcriptional machinery is used to express LTR elements and a viral-like particle is formed in the cytoplasm with Gag isolating TE-expressed proteins and mRNA. Reverse transcription occurs inside those particles. After, cDNA is transported to the nucleus where integrase catalyzes transposon insertion **(B)**. Non-LTR elements code for two enzymes: endonuclease (ORF1 protein) and reverse transcriptase (ORF2 protein). cDNA production occurs at the nucleus while reverse transcriptase uses mRNA as template and DNA overhang (produced by endonuclease activity) as primer **(C)**. Aside detailed information for the reverse transcription process, little is known about other steps of non-LTR transposition[122,126]. This figure was adapted from Levin and Moran (2011)[122].

DNA transposons are ancient TEs found in prokaryotes and almost all eukaryotes that transpose without producing an RNA intermediate[127]. Those are mostly flanked by terminal inverted repeats (TIRs) (**Figure 1.4A**) that can vary in length and sequence, and be used to classify DNA elements[128,129]. In plants, the first bases of TIRs are composed by highly conserved CACTA motifs that are replaced by CCC sequences in animals and fungi[130].

In short, DNA elements that code for transposase mobilize by a *cut-and-paste* mechanism and can only increase in copy number when transposing during DNA replication or by sister chromatid recombination[131]. Alternatively, they can also exploit the gap repair machinery upon excision while the old insertion is restored using flanking homologous regions as primers for DNA synthesis and the sister chromatid containing TE insertion as template[132].

1.2.2 TE activity upon environmental perturbations

Our knowledge about what triggers TE transposition is still rather incomplete. Barbara McClintock proposed that, mobile elements would function as genomic architects by becoming active in response to environmental stresses[133]. She believed that jumping of transposons could elicit the expression of genes necessary to circumvent stress. Only recently have discoveries uncovering a link between environmental factors and TE activity[134–139] confirmed her view. This includes data demonstrating transcriptional induction of plant *Tnt1* and *Tto1* LTR elements in response to several biotic (inoculation with viral, bacteria and fungal pathogens) and abiotic (methyl jasmonate, CuCl2 and salicylic acid) conditions[140–143]. Also, the presence of *cis*-acting elements within the 5' sequence of LTRs was proven to mediate TE activation during environmental change[144]. Conversely, *Tnt1* insertions found in several tobacco plants were shown to promote chloride-resistant phenotypes[123,145].

It is important to note that the life cycle of retrotransposons includes other steps than transcription such as translation, reverse transcription and integration, and that environmental factors can affect any of those levels to influence transposition. For example, change in the phosphorylation state of *Ty5* integrase

that is mediated by nutrient deprivation affects TE insertion preference without influencing TE expression[146,147]. When integrase is phosphorylated, *Ty5* preferentially inserts into heterochromatin minimizing the chances to damage coding sequences. After exposure to nutrient deprivation, integrase is dephosphorylated and *Ty5* is inserted into areas of active gene expression. A mechanism whereby environmental perturbation can regulate TE activity or integration could have been selected for if TE deployment is sometimes, but not always, advantageous. The study of TE-response to environmental fluctuations in other model organisms is needed to understand the conservation and/or generality of such mechanisms.

There have been several attempts to demonstrate induction of mobile elements associated with stressful conditions in *D. melanogaster*[148]. These include circumstantial evidence for TE response to heat[149–151], ethanol, radiation, and viral infection[152,153] though results were strongly dependent on the transposon and genetic background analyzed. While there is evidence for the importance of environmental factors in the regulation of TE activity, empirical research on other host organisms and transposon types is needed to elucidate the occurrence and importance of such strategies.

1.2.3 TEs as source of genetic variation

The famous colour phenotypes of corn kernels first described by Barbara McClintock result from TE instability and insertion in essential pigmentation pathway genes[154]. In that study, transposition occurred during kernel development and therefore induced non-heritable mutations[123,154]. In fact, only TE activity in the germline will contribute to heritable genetic variation, which is what is relevant for evolution. The transposon-mediated mutation rate can range from 10^{-3} and 10^{-5} per element per generation, depending on the element, whereas the classical nucleotide substitution rate fluctuates around 10^{-8} and 10^{-9} per nucleotide per generation[155]. Upon transposition, TEs can influence expression of neighbouring genes, both at the transcriptional and post-transcriptional levels (**Figure 1.5**). Interestingly, several mobile elements adding

new regulatory regions are scattered along the genome and can play an important function in the establishment of genetic networks[156,157]. The preference of DNA elements for *Hsp* promoters in *Drosophila*[158] and the considerable impact of LTR retrotransposons in the transcriptional network of human p53[159] provide good examples of this.

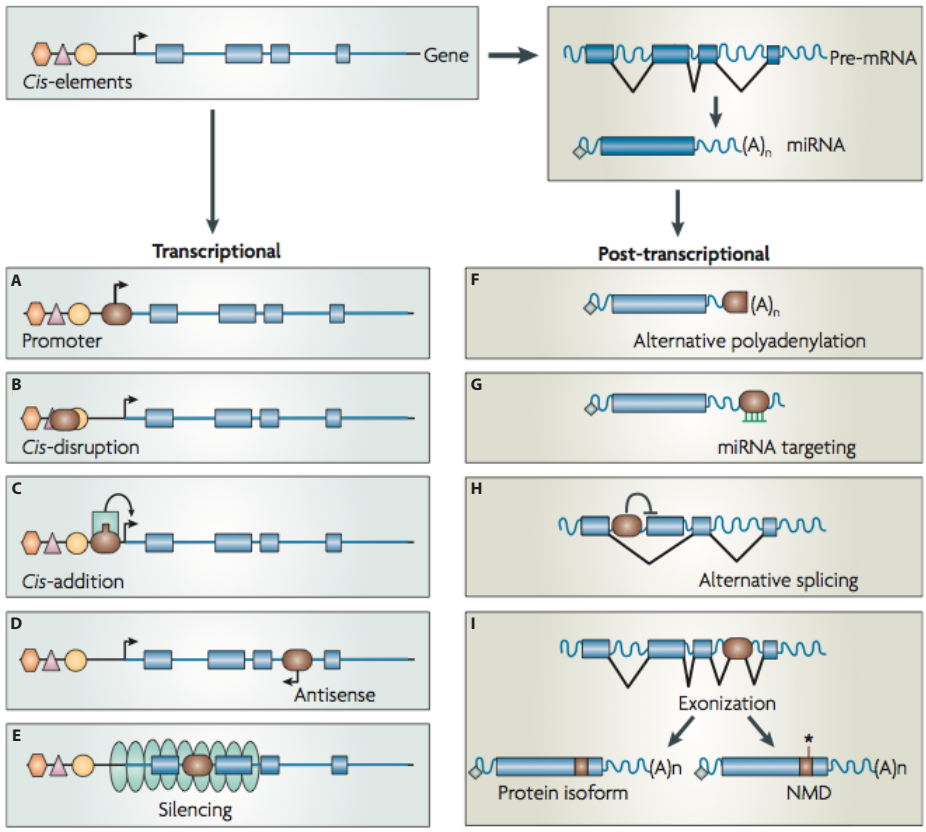


Figure 1.5. Transcriptional and post-transcriptional modifications mediated by TE insertions. Gene expression and/or function can be affected by novel TE insertions. Those can be mediated by the induction of transcriptional **(A-E)** or post-transcriptional modifications **(F-I)**. Given that transposons contain their own promoters and *cis*-regulatory regions, TE insertions upstream of a gene can introduce an alternative transcription start site **(A)**, disrupt expression when inserted within gene regulatory elements **(B)** or add new regulatory elements **(C)**[160–166]. In addition, TE integration can also promote anti-sense transcription **(D)**[167] or recruit chromatin-remodeling proteins that eventually silence gene and TE activity **(E)**[168]. Conversely, TE integration may mediate post-transcriptional

modifications by introducing alternative polyadenylation signals **(F)**[169,170], novel miRNA target sites **(G)**[171] or alternative splicing events **(H)**[172,173]. TE sequences can also be incorporated as an alternative exon (process called exonization) that can result in the translation of a new protein isoform, or mediate mRNA degradation by the nonsense-mediated decay (NMD) if new stop codons are inserted (asterisk) **(I)**[163,174]. This figure was modified from Feschotte (2008)[156].

The contribution of TE activity to the production of genetic variation is a general property of living organisms. In fact, active transposons have been found in the genome assemblies of marine plankton[175] as well as in several plant[176–178] and fish[179,180] species. Additionally, more detailed studies on the selective value of TE insertions segregating in *D. melanogaster* natural populations revealed that many are adaptive with putative function in processes such as pesticide resistance[181,182], oxidative stress[183], amino acid metabolism, immune response to viral infection and eye development[184]. Moreover, experimental evolution in bacteria has shown changes in TE activity and selection of novel insertions during adaptation to gut environment[6]. In summary, the onset of evidence for the importance of mobile elements in adaptation distinguishes those sequences as engineers of transcriptional networks and putative mediators of diversification.

1.2.4 TEs and novelty

The bulk of studies in evolutionary biology entail research about adaptation, speciation and the origin of novelties – lineage specific traits with adaptive value[185]. This last research program advanced the emergence of a new discipline (evolutionary developmental biology) that focuses on the genetic basis of intra- and inter-specific morphological variation[186]. During the past years, it has become clear that TEs are at the foundation of several novel traits and hence demand further attention. Among those, researchers gathered evidence for the evolution of conspicuous male fin colour patterns that are involved in mating of cichlids[187], the origin of mammal placenta[188] and the diversification of fruit colour and shape[189]; all phenotypes deriving from transposon insertions that resulted in *cis*-regulatory changes of neighbouring genes. Aside these examples

of how particular TE insertions affected gene expression or mediated new genomic functions, there is also documentation of various *"domestication events"* that produced new protein-coding genes from transposases[156]. In particular, the V(D)J recombination reaction necessary for adaptive immunity relies on transposase-derived RAG1 that emerged in jawed vertebrates[190]. Other examples also found in metazoans deal with the evolution of highly conserved PAX proteins that function in eye development and cephalization[191] as well as centromere-associated protein B (CENP-B) that is related with cell-cycle[192]. Additionally, a striking scenario of convergence between distant lineages describes the putative recurrent appearance of transposase-derived proteins that are involved in far-red light signalling in yeast, insects and angiosperms[156,193,194]. All in all, it is clear that TEs are critical mediators of diversification and innovation.

1.3 Mechanisms to repress TE activity

Despite the importance of mobile genetic elements in creating adaptive variation and contributing to morphological diversity, they can threaten genomic integrity and reduce fitness when active in the germline[195]. As a result, metazoan species have evolved mechanisms to promote specific TE recognition and guide silencing through the use of small RNAs[196,197]. Small RNAs can be grouped into three main classes that differ in their biogenesis process and biological/biochemical function. Small interfering RNAs (siRNAs) and micro RNAs (miRNAs) are ubiquitously expressed and derive from double stranded RNAs processed by the RNA-cleaving enzyme Dicer, whereas the gonadal specific PIWI-interacting RNAs (piRNAs) are produced through Dicer-independent mechanisms and derive from both single and double stranded RNAs[198,199]. Most information available about piRNA activity comes from studies in *D. melanogaster* that is currently used as reference for TE-silencing systems in other species[197]. The main function of piRNAs is to mediate homology-dependent transposon silencing while interacting with proteins that belong to the PIWI family – Piwi, Aubergine (Aub) and Argonaute 3 (AGO3)[200–202]. piRNAs

belong to a group of sequences that are particularly diverse and transcribed from intergenic inactive repetitive elements that collectively map to genomic clusters. In *D. melanogaster* there are 569 described piRNA clusters (http://pirnabank.ibab.ac.in/), and only some are exclusively expressed in the germline of males and females[203–205]. Conversely, other clusters specifically repress TE activity in the somatic tissues of ovaries, whereas some repress both germline and somatic expression of mobile elements[203].

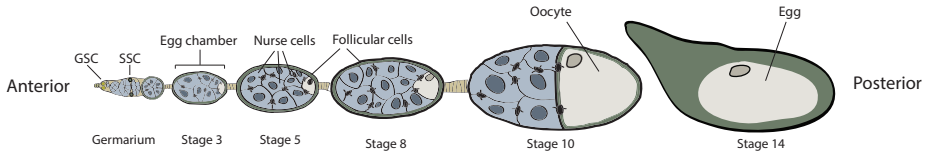


Figure 1.6. Progressive oocyte development in *D. melanogaster* ovarioles. Oogenesis initiates at the germarium that is localized at the anterior tip of ovarioles. There, germline stem cells (GSC) divide to produce oocyte and nurse cells; somatic stem cells (SSC) give rise to follicular cells. When the egg chamber exits the germarium, a layer of follicular cells is protecting the oocyte and nurse cells that maintain a common cytoplasm.

1.3.1 piRNA pathway in the ovary of *D. melanogaster*

Drosophila ovaries are composed of multiple ovarioles each made up of a series of egg chambers with anterior-posterior progressive development (**Figure 1.6**). The somatic support cells establish an intricate relationship with the germline by feeding developmental signals and nutrients into the oocyte and nurse cells syncytium[206]. This close interaction between somatic (follicular cells) and germline (oocyte and nurse cells) tissue exposes the oocyte genome to threats that can derive from both cell types. In fact, several transposons are able to express in the germline where movement can directly drive heritable transmission[195,207–209], but retrotransposons expressed in follicular cells can also form retroviral particles that efficiently invade the oocyte[210–213]. To minimize TE

access to the female gametocyte, the piRNA pathway functions as a genome surveillance system for both cell types. Interestingly, piRNA biogenesis differs between somatic and germline tissue leading to distinct silencing processes used to repress TEs in different cell types[203].

In **follicular cells** over 90% of the piRNA pool is antisense to transposons that are specifically active in somatic gonadal cells and belong to the gypsy family[214]. Most are produced from two X-chromosomal clusters – the flamenco locus and cluster 20A – and derive from long single-stranded transcripts that are processed into smaller sequences[214,215]. Interestingly, the production of those genomic guardians is independent of the expression of active transposons and hence constitutes the primary piRNA biogenesis[214]. The biogenesis step takes place in the cytoplasm, where the Piwi protein (only element of the PIWI-family expressed in follicular cells) is loaded with piRNAs[216]. Piwi-piRNA complexes then migrate to the nucleus to mediate transcriptional repression of mobile elements[217]. The genomic place of TE insertions is recognized by complementarity with piRNA sequences, and Piwi directs the accumulation of heterochromatic marks at those genomic places through direct interaction with heterochromatin protein 1a that is a central player in heterochromatic gene silencing (**Figure 1.7**)[168].

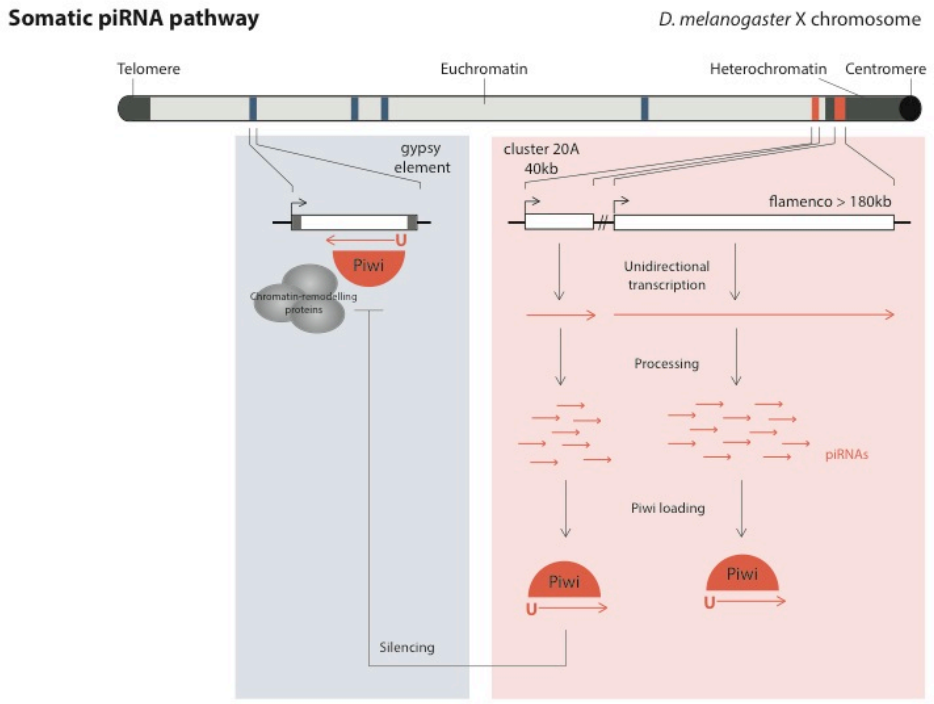


Figure 1.7. Primary piRNA biogenesis in follicular cells. piRNA production in ovarian somatic tissue consists of primary biogenesis resultant from processing of two main clusters (cluster 20A and flamenco, red box). Those are unidirectionally transcribed and processed into smaller sequences (piRNAs). Piwi loaded with piRNAs migrates to the nucleus to promote TE repression by recruiting chromatin-remodeling proteins. A putative TE insertion in the X-chromosome is highlighted in the blue box.

In **germline cells** the architecture of piRNA biogenesis is complex and involves the production of both sense and antisense small RNAs. Those originate from several piRNA clusters that are scattered along the genome and comprehend an extensive collection of fragmented TE sequences[215]. Clusters that are expressed in the somatic and germline tissue of ovaries (such as the previously described cluster 20A) are unidirectionally transcribed; germline-specific clusters (such as the well described cluster 42AB) are bidirectionally transcribed and produce double-stranded transcripts that are fully cleaved until piRNAs[203,214]. Those are then loaded into all three PIWI family proteins: Piwi and Aub bind to antisense

and AGO3 to sense piRNAs. In addition to the primary biogenesis, piRNAs undergo a loop of amplification that is mostly driven by Aub and AGO3. Those two proteins are enriched around the nucleus where Aub-piRNA complexes target and direct TE transcripts for cleavage. As a consequence, sense piRNAs are produced and incorporated by AGO3, that has the potential to specifically trigger piRNA cluster processing and increase the abundance of antisense piRNAs loaded into Aub[215]. In parallel, Piwi-piRNA complexes maintain function to promote chromatin-remodelling changes at the loci where transponsons are inserted. Interestingly, maternal transmission of antisense small RNAs associated with both Piwi and Aub works as a catalyst to produce piRNAs crucial to repress TEs in the next generation[207]. The nature of Aub and AGO3 engagement to produce piRNAs in the germline named this amplification loop as ping-pong cycle[215]. This mechanism represents the hallmark of the germline piRNA pathway as it is conserved in a spectrum of species that range from sponges to mammals (**Figure 1.8**)[218–221].

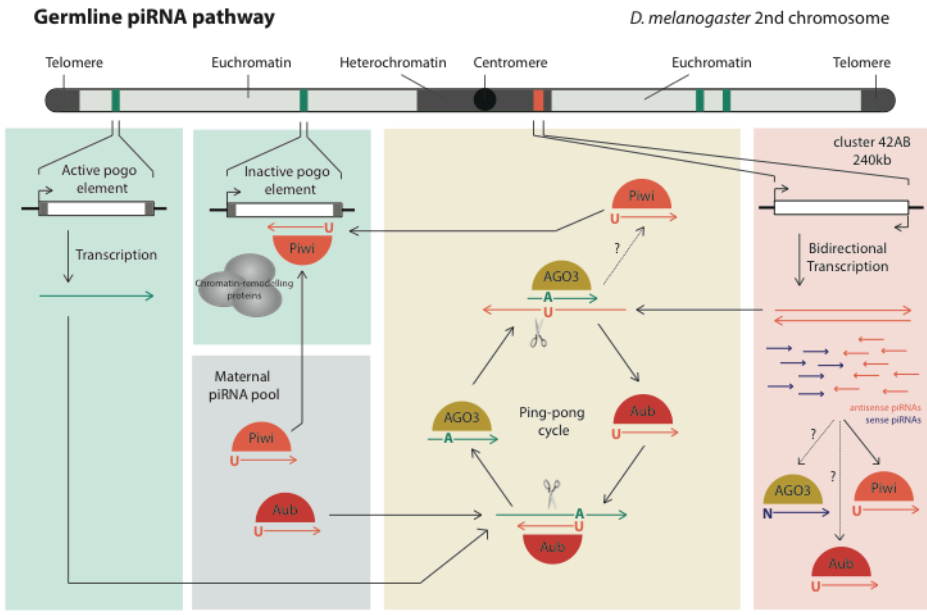


Figure 1.8. Production of piRNAs in the germline. Several piRNA clusters, including cluster 42AB, are bidirectionally expressed in the germline (red box). As a consequence, both sense and antisense piRNAs are produced and loaded into all PIWI-family proteins.

Piwi and Aub specifically select piRNAs antisense to transposon sequences that are enriched with 5' uridine (represented by U). On the other hand, AGO3 binds to sense piRNAs. In addition to the primary biogenesis that occurs in the cytoplasm, accumulation of AGO3 and Aub around the nucleus mediates piRNA production and amplification by the ping-pong cycle (yellow box). During this process, transposon mRNAs derived from active TE insertions (green box on the left) are incorporated and cleaved trough recognition by Aub-piRNA complexes. The resulting sense piRNAs that are enriched in adenine (represented by A) 10 nucleotides downstream of the 5' end are loaded into AGO3 that mediates detection and processing of cluster RNAs. As a consequence, more antisense piRNAs are created and loaded into Aub and Piwi. The ping-pong cycle elicits TE repression by cleavage of their mRNA sequences and by creating more Piwi-piRNA complexes that move to the nucleus to specifically recognize transposon insertions and mediate repression by recruiting chromatin-remodeling proteins (green box on the right). Interestingly, maternal deposition in the oocyte of antisense piRNAs bound to Piwi and Aub is necessary to start the process of TE repression later in the ovary of the progeny (grey box).

1.3.2 piRNA pathway and environment

In order to be fully functional Piwi needs to interact with Hsp90 – a protein induced in stressful conditions[222,223]. Hsp90 has been called a *"capacitor for morphological evolution"* due to its ability to mask standing genetic variation that adds to phenotypic diversification[224–227]. Interestingly, environmental perturbations that include changes in temperature produce phenotypic effects that resemble those observed when Hsp90 is impaired[224–226]. Given the ubiquitous induction of heat shock responses in stressful conditions, it has been long suggested that Hsp90 functions as a regulator of the penetrance of adaptive phenotypic variation in different environmental conditions. More recently, this protein was linked with production of *de novo* variation through transposon-induced mutagenesis[228]. Given that Piwi needs to interact with Hsp90 to be properly phosphorylated[222] and loaded with piRNA percursors[223], non-functional Hsp90 can compromise piRNA-mediated TE repression and lead to increased transposition rates. In this manner, we propose that environmental perturbations affecting Hsp90 activity will eventually restrict Piwi function and drive the production of genetic variation.

1.4 When TEs invade: hybrid dysgenesis

The mechanisms that protect germcells from transposition ensure reproductive vigour. The piRNA pathway is committed to repress TE activity, as transposition events into clusters enable the production of specific piRNAs that mediate TE recognition guided by the primary biogenesis. This system is also prepared to mitigate immediate threats driven by sudden TE overexpression through the incorporation of TE transcripts into the ping-pong cycle[229]. However, when a host first encounters a new transposon, the piRNA pathway is not able to restrain activity which can culminate in severe DNA damage[195]. This parasitic behaviour of transposons is alleviated as new insertions occurring at piRNA clusters allow the production of piRNAs against the invading element[195]. This form of acquired *"immunity"* is then maternally transmitted to the next generation and prevents future bursts of activity of the corresponding TEs.

Sometimes, when the genomes of mating partners differ in TE composition and piRNA libraries, there can be a burst of TE activity in the germline of the hybrid progeny[207]. If mothers are unable to produce piRNAs against particular TEs provided in the father genome, the hybrid females have reduced fertility relative to those obtained from the reciprocal cross. The direction of the cross dictates the amplitude of transposon activity observed in the hybrids as it specifies the maternal transmission, or not, of piRNAs necessary to recognize and inhibit parasitic transposons. This phenomenon is known as hybrid dysgenesis and was initially defined by problems in the fertility of hybrids[230]. In *D. melanogaster*, dysgenesis syndrome is well characterized for both I-element and P-element TEs[207], and the severity of this phenomenon is dependent on environmental factors such as temperature[231,232] (**Figure 1.9**).

Generally speaking, transposons colonize the landscape of all genomes constituting important components for gene network rewiring and phenotypic diversification. Moreover, descriptions of parasitic events and bursts in transposition that culminate in hybrid sterility suggests that mobile elements can also accelerate the process of speciation[199]. For example, when a geographically isolated population that was colonized by a new TE attempts to

interbreed with another population, it would fail to produce fertile offspring due to over-activation of transposons in their germline. Consequently, TE incompatibilities between different populations can present a valid mechanism to produce the reproductive isolation which is considered as one of the first steps in driving speciation[233].

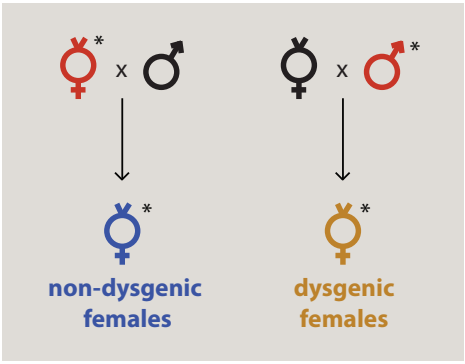


Figure 1.9. Hybrid dysgenesis. When a female containing insertions of a given element (red) mates with males that lack insertions for that (black), F1 females are fertile and able to restrain TE activity in the germline (non-dysgenic females, blue). On the contrary, when a female that does not contain insertions (black) mates with males containing copies of that element (red) the corresponding F1 females are unable to repress activity in the gonads and are sterile (dark yellow). The asterisk (*) represents presence of transposon insertions.

1.5 Study system

Here, we propose to look at the effect of environmental and genetic factors on TE dynamics in the germline, including genotype, environment and genotype-environment interactions. We studied whether and how different environmental perturbations such as temperature perturbation, oxidative stress and presence of heavy metals affect TE activity in different *D. melanogaster* genotypes. We took advantage of the fully sequenced panel of 147 wild-derived fruit fly lines – Drosophila Genetic Reference Panel (DGRP) – that includes information about single nucleotide polymorphisms and TE insertions (number, identity and place of insertion)[15]. We hypothesize that environmental perturbation can divert Hsp90 from piRNA function and mediate increased jumping, possibly in a genotype-dependent way. Temperature challenges, presence of heavy metals and oxidative stress are all abiotic factors able to induce heat-shock responses[49], and therefore possible mediators of TE activity through changes in Hsp90-Piwi interaction. As a read out for TE dynamics we used TE expression in ovaries.

Aside from the abiotic factors mentioned above, we also included a biotic challenge and proposed to study the effect of environment-environment interactions in TE activity. In particular, and given that retrotransposons resemble retroviruses and that *Wolbachia* is able to promote resistance to infection with RNA but not with DNA viruses, we aimed at studying the influence of this endosymbiont in TE response to abiotic stresses. We further suggest that *Wolbachia* may mitigate the activity of retrotransposons, similar to its function in reducing viral infection.

P-element is a DNA transposon that recently invaded *D. melanogaster* populations in the wild. This process of invasion was also associated with hybrid dysgenesis syndrome and with the replacement of *Wolbachia* strains which indicates potential involvement of this bacteria in promoting *P-element* spread within fruit fly populations[234]. In this context, we tested whether the presence of endosymbiont alleviates hybrid dysgenesis severity (another read out of TE dynamics).

In general, this essay proposes to give insights into the relevance of environmental biotic and abiotic factors in the making of new genetic variation using transposon expression as a proxy for activity in *D. melanogaster* ovaries. However, and given that TE activity can be regulated at different steps of its life cycle, the association between higher TE expression and production of novel insertions may not be straightforward. In order to test the generality of this assumption, we further characterized expression of some mobile elements in several DGRP lines containing different number of novel insertions. We also took advantage of those wild-derived fully sequenced genotypes to study the alleles that may influence TE activity in natural populations.

1.6 Overview of the thesis

The aim of this thesis was to explore different aspects (genotypic and environmental) affecting TE dynamics in *D. melanogaster* ovaries. For more details and overview of the thesis see **Figure 1.10**.

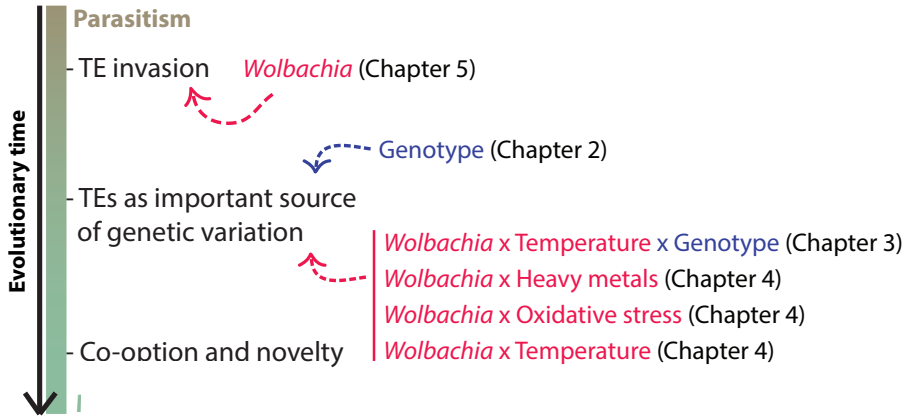


Figure 1.10. Overview of the thesis. TEs can establish intricate relationships with their hosts that may resemble those observed in symbiotic interactions and range from parasitism at one extreme to mutualism at the other[112]. We studied the effect of genetic (blue) and environmental (red) factors when mobile elements are active and expressed at low concentrations. We evaluated the effect of biotic (*Wolbachia*) and abiotic (temperature, heavy metals and oxidative stress) factors in TE activity (Chapter 3 and 4). The effect of genetic background was also further analyzed (Chapter 2 and 3). We also tested the importance of *Wolbachia* in the context of hybrid dysgenesis syndrome and uncontrolled TE activity (Chapter 5).

References

1. Barret, R. D. H., Rogers, S. M. & Schluter, D. Natural Selection on a Major Armor Gene in Threespine Stickleback. *Science (80-.).* **1092,** 255–257 (2008).

2. Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. On the origin and spread of an adaptive allele in deer mice. *Science (80-.).* **325,** 1095–8 (2009).

3. Martins, N. E. *et al.* Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 5938–43 (2014).

4. MacLean, R. C., Hall, A. R., Perron, G. G. & Buckling, A. The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nat. Rev. Genet.* **11,** 405–414 (2010).

5. Barroso-Batista, J., Demengeot, J. & Gordo, I. Adaptive immunity increases the pace and predictability of evolutionary change in

commensal gut bacteria. *Nat. Commun.* **6,** 8945 (2015).

6. Barroso-Batista, J. *et al.* The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.* **10,** (2014).

7. Tobler, R. *et al.* Massive Habitat-Specific Genomic Response in D. melanogaster Populations during Experimental Evolution in Hot and Cold Environments. *Mol. Biol. Evol.* **31,** 364–375 (2014).

8. Barrett, R. D. H. & Hoekstra, H. E. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12,** 767–780 (2011).

9. Colosimo, P. F. *et al.* Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science (80-.).* **307,** 1928–1933 (2005).

10. Marchinko, K. B. Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution (N. Y).* **63,** 127–138 (2009).

11. Hudson, M. E. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.* **8,** 3–17 (2008).

12. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).

13. Stapley, J. *et al.* Adaptation genomics: The next generation. *Trends Ecol. Evol.* **25,** 705–712 (2010).

14. Storz, J. F. & Wheat, C. W. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution (N. Y).* **64,** 2489–2509 (2010).

15. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482,** 173–178 (2012).

16. Denamur, E. & Matic, I. Evolution of mutation rates in bacteria. *Mol. Microbiol.* **60,** 820–827 (2006).

17. Aggarwal, D. D. *et al.* Experimental evolution of recombination and crossover interference in Drosophila caused by directional selection for stress-related traits. *BMC Biol.* **13,** 101 (2015).

18. Abdullah, M. F. & Borts, R. H. Meiotic recombination frequencies are affected by nutritional states in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 14524–14529 (2001).

19. Jackson, S., Nielsen, D. M. & Singh, N. D. Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila. *BMC Evol. Biol.* **15,** 175 (2015).

20. Stern, C. An effect of temperature and age on crossing over in the first

chromosome of Drosophila melanogaster. *Genetics* 530–532 (1926).

21. Henderson, S. A. & Edwards, R. G. Chiasma frequency and maternal age in mammals. *Nature* **218,** 22–28 (1968).

22. Priest, N. K., Roach, D. A. & Galloway, L. F. Mating-induced recombination in fruit flies. *Evolution (N. Y).* **61,** 160–167 (2007).

23. Hussin, J., Roy-Gagnon, M. H., Gendron, R., Andelfinger, G. & Awadalla, P. Age-dependent recombination rates in human pedigrees. *PLoS Genet.* **7,** (2011).

24. Neel, J. V. A Relation between Larval Nutrition and the Frequency of Crossing over in the Third Chromosome of Drosophila Melanogaster. *Genetics* **26,** 506–516 (1941).

25. Bjedov, I. *et al.* Stress-induced mutagenesis in bacteria. *Sicence* **42,** 1404–1409 (2003).

26. Muller, H. J. The Measurement of Gene Mutation Rate IN Drosophila, Its High Variability, and Its Dependence Upon Temperature. *Genetics* 279–357 (1928).

27. Armbruster, P. & Reed, D. H. Inbreeding depression in benign and stressful environments. *Heredity (Edinb).* **95,** 235–242 (2005).

28. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Genet.* (2011). doi:10.1038/nrg3055

29. Azad, P., Zhou, D., Russo, E. & Haddad, G. G. Distinct mechanisms underlying tolerance to intermittent and constant hypoxia in Drosophila melanogaster. *PLoS One* **4,** e5371 (2009).

30. De Angelis, M. & Gobbetti, M. Environmental stress responses in Lactobacillus: A review. *Proteomics* **4,** 106–122 (2004).

31. Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* **9,** 537–549 (2009).

32. Kamada, Y., Jung, U. S., Piotrowski, R. & Levin, D. E. The Protein-Kinase C-Activated Map Kinase Pathway of Saccharomyces-Cerevisiae Mediates a Novel Aspect of the Heat-Shock Response. *Genes Dev.* **9,** 1559–1571 (1995).

33. Lemaitre, B. & Hoffmann, J. The Host Defense of Drosophila melanogaster. *Annu. Rev. Immunol.* **25,** 697–743 (2007).

34. Mittler, R. Oxidative stress, antioxidants and stress tolerance. **7,** 405–410 (2002).

35. Inoue, H. *et al.* The C . elegans p38 MAPK pathway regulates nuclear localization of the transcription factor SKN-1 in oxidative stress response. 2278–2283 (2005). doi:10.1101/gad.1324805.2278

36. Kovtun, Y., Chiu, W. L., Tena, G. & Sheen, J. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2940–2945 (2000).

37. de Nadal, E., Alepuz, P. M. & Posas, F. Dealing with osmostress through MAP kinase activation. *EMBO Rep.* **3,** 735–740 (2002).

38. Thomas, M. B. & Blanford, S. Thermal biology in insect-parasite interactions. *Trends Ecol. Evol.* **18,** 344–350 (2003).

39. Ragland, G. J. & Kingsolver, J. G. The effect of fluctuating temperatures on ectotherm life-history traits: Comparisons among geographic populations of Wyeomyia smithii. *Evol. Ecol. Res.* **10,** 29–44 (2008).

40. Lazzaro, B. P., Flores, H. a., Lorigan, J. G. & Yourth, C. P. Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in Drosophila melanogaster. *PLoS Pathog.* **4,** 1–9 (2008).

41. Colinet, H., Sinclair, B. J., Vernon, P. & Renault, D. Insects in Fluctuating Thermal Environments. *Annu. Rev. Entomol.* **60,** 123–140 (2015).

42. Franzmann, T. M., Menhorn, P., Walter, S. & Buchner, J. Activation of the Chaperone Hsp26 Is Controlled by the Rearrangement of Its Thermosensor Domain. *Mol. Cell* **29,** 207–216 (2008).

43. Hamada, F. N. *et al.* An internal thermal sensor controlling temperature preference in Drosophila. *Nature* **454,** 217–220 (2008).

44. McClung, C. R. & Davis, S. J. Ambient thermometers in plants: From physiological outputs towards mechanisms of thermal sensing. *Curr. Biol.* **20,** 1086–1092 (2010).

45. Richter, K., Haslbeck, M. & Buchner, J. The Heat Shock Response: Life on the Verge of Death. *Mol. Cell* **40,** 253–266 (2010).

46. Dhaka, A., Viswanath, V. & Patapoutian, A. Trp Ion Channels and Temperature Sensation. *Annu. Rev. Neurosci.* **29,** 135–161 (2006).

47. Klinkert, B. & Narberhaus, F. Microbial thermosensors. *Cell. Mol. Life Sci.* **66,** 2661–2676 (2009).

48. Feder, M. E. & Hofmann, G. E. Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and Ecological Physiology. *Annu. Rev. Physiol* **61,** 243–282 (1999).

49. Akerfelt, M., Morimoto, R. I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* **11,** 545–555 (2010).

50. Gasch, a P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11,** 4241–4257

(2000).

51. Causton, H. *et al.* Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12,** 323–337 (2001).

52. Shamovsky, I. & Nudler, E. New insights into the mechanism of heat shock response activation. *Cell. Mol. Life Sci.* **65,** 855–861 (2008).

53. Shi, Y., Mosser, D. D. & Morimoto, R. I. Molecular chaperones as HSF1-specific transcriptional repressors. *Genes Dev.* **12,** 654–666 (1998).

54. Baler, R., Welch, W. J. & Voellmy, R. Heat shock gene regulation by nascent polypeptides and denatured proteins: hsp70 as a potential autoregulatory factor. *J. Cell Biol.* **117,** 1151–1159 (1992).

55. Abravaya, K. & Biology, C. The human heat shock protein hsp70 interacts with HSF , the transcription factor that regulates heat shock gene expression iii ; ii ! i ! iiiii ii. 1153–1164 (1992). doi:10.1101/gad.6.7.1153

56. Goodson, M. L. & Sarge, K. D. Heat-inducible DNA binding of purified heat shock transcription factor. *J. Biol. Chem.* **270,** 2447–2450 (1995).

57. Larson, J. S., Schuetz, T. J. & Kingston, R. E. In vitro activation of purified human heat shock factor by heat. *Biochemistry* **34,** 1902–1911 (1995).

58. Parker, C. S. & Topol, J. A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. *Cell* **36,** 357–369 (1984).

59. Chu, B., Soncin, F., Price, B. D., Stevenson, M. A. & Calderwood, S. K. Sequential Phosphorylation by Mitogen-activated Protein Kinase and Glycogen Synthase Kinase 3Represses Transcriptional Activation by Heat Shock Factor-1. *J. Biol. Chem.* **271,** 30847–30857 (1996).

60. Raison, J. K. The influence of temperature-induced phase changes on the kinetics of respiratory and other membrane-associated enzyme systems. *J. Bioenerg.* **4,** 285–309 (1973).

61. Phillips, R. S. Temperature modulation of the stereochemistry of enzymatic catalysis: Prospects for exploitation. *Trends Biotechnol.* **14,** 13–16 (1996).

62. Brauchi, S., Orio, P. & Latorre, R. Clues to understanding cold sensation: thermodynamics and electrophysiological analysis of the cold receptor TRPM8. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15494–15499 (2004).

63. Voets, T. *et al.* The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* **430,** 748–754 (2004).

64. Rosenzweig, M. *et al.* The Drosophila ortholog of vertebrate TRPA1

regulates thermotaxis. *Genes Dev.* **19,** 419–424 (2005).

65. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408,** 239–247 (2000).

66. Buchon, N., Silverman, N. & Cherry, S. Immunity in Drosophila melanogaster - from microbial recognition to whole-organism physiology. *Nat. Rev. Immunol.* **14,** 796–810 (2014).

67. Betteridge, D. J. What Is Oxidative Stress? *Metabolism* **49,** 3–8 (2000).

68. Finkel, T. Oxidant signals and oxidative stress. *Curr. Opin. Cell Biol.* **15,** 247–254 (2003).

69. Allen, R. G. & Tresini, M. Oxidative stress and gene regulation. *Free Radic. Biol. Med.* **28,** 463–499 (2000).

70. Zhong, M., Orosz, a & Wu, C. Direct sensing of heat and oxidation by Drosophila heat shock transcription factor. *Mol. Cell* **2,** 101–108 (1998).

71. Singh, V. & Aballay, A. Heat Shock and Genetic Activation of HSF-1 Enhance Immunity to Bacteria. **5,** 2443–2446 (2006).

72. Burdon, R. H. Superoxid and hydrogen peroxide in relation to mammalian cell proliferation. **18,** 775–794 (1995).

73. Beckman, K. B. & Ames, B. N. The free radical theory of aging matures. *Physiol. Rev.* **78,** 547–581 (1998).

74. Saitoh, M. *et al.* Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J.* **17,** 2596–2606 (1998).

75. Adler, V. *et al.* Regulation of JNK signaling by GSTp. *EMBO J.* **18,** 1321–34 (1999).

76. Lee, S. R., Kwon, K. S., Kim, S. R. & Rhee, S. G. Reversible inactivation of proteintyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *J. Biol. Chem.* **273,** 15366–15372 (1998).

77. Wang, X., McCullough, K. D., Franke, T. F. & Holbrook, N. J. Epidermal growth factor receptor-dependent Akt activation by oxidative stress enhances cell survival. *J. Biol. Chem.* **275,** 14624–14631 (2000).

78. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389,** 300–305 (1997).

79. Johnson, T. M., Yu, Z. X., Ferrans, V. J., Lowenstein, R. a & Finkel, T. Reactive oxygen species are downstream mediators of p53-dependent apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 11848–11852 (1996).

80. de Bary, H. A. *Die Erscheinung der Symbiose*. (1879).

81. Relman, D. A. 'Til death do us part': coming to terms with symbiotic relationships. *Nat. Rev. Microbiol.* **6,** 721–724 (2008).

82. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6,** 741–751 (2008).

83. Baumann, P., Moran, N. A. & Baumann, L. Genetics of Aphid Endosymbionts. *Bioscience* **47,** 12–20 (1997).

84. Taylor, M. J. & Hoerauf, A. Wolbachia bacteria of filarial nematodes. *Parasitol. Today* **15,** 437–442 (1999).

85. Foster, J. *et al.* The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3,** 0599–0614 (2005).

86. Craig, S. C. & Stephen, J. G. Transovarial inheritance of endosymniotic Bacteria in clams onhabiting deep-sea hydrothermal vents and colds-seeps. *Proc Natl Acad Sci U S A* **90,** 5695–5699 (1993).

87. Stewart, F. J. & Cavanaugh, C. M. Bacterial endosymbioses in Solemya (Mollusca: Bivalvia) - Model systems for studies of symbiont-host adaptation. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **90,** 343–360 (2006).

88. Usher, K. M., Kuo, J., Fromont, J. & Sutton, D. C. Vertical transmission of cyanobacterial symbionts in the marine sponge Chondrilla australiensis (Demospongiae). *Hydrobiologia* **461,** 15–23 (2001).

89. Sharp, K. H., Eam, B., John Faulkner, D. & Haygood, M. G. Vertical transmission of diverse microbes in the tropical sponge Corticium sp. *Appl. Environ. Microbiol.* **73,** 622–629 (2007).

90. Tram, U. & Sullivan, W. Role of Delayed Nuclear Envelope Breakdown and Mitosis in Wolbachia-Induced Cytoplasmic Incompatibility. *Science (80-.).* **296,** 1124–1126 (2002).

91. Stouthamer, R. & Kazmer, D. J. Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in Trichogramma wasps. *Heredity (Edinb).* **73,** 317–327 (1994).

92. Kose, H. & Karr, T. L. Organization of *Wolbachia pipientis* in the *Drosophila* fertilized egg and embryo revealed by an anti-*Wolbachia* monoclonal antibody. *Mech. Dev.* **51,** 275–288 (1995).

93. Ferree, P. M. *et al.* Wolbachia utilizes host microtubules and Dynein for anterior localization in the Drosophila Oocyte. *PLoS Pathog.* **1,** 0111–0124 (2005).

94. Serbus, L. R. & Sullivan, W. A cellular basis for Wolbachia recruitment to the host germline. *PLoS Pathog.* **3,** 1930–1937 (2007).

95. Sheehan, K. B., Martin, M., Lesser, C. F., Isberg, R. R. & Newton, L. G. Identification and Characterization of a Candidate Wolbachia pipientis Type IV Effector That Interacts with the Actin Cytoskeleton. *MBio* **7,** 1–12

(2016).

96. Ote, M., Ueyama, M. & Yamamoto, D. Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in Drosophila Sex-lethal Article Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in Drosophila Sex-lethal Mutants. *Curr. Biol.* **26,** 1–10 (2016).

97. Chintapalli, V. R., Wang, J. & Dow, J. a T. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.* **39,** 715–20 (2007).

98. Gottar, M. *et al.* The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* **416,** 640–644 (2002).

99. Choe, K. M., Werner, T., Stoven, S., Hultmark, D. & Anderson, K. V. Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in Drosophila. *Science (80-.).* **296,** 359–362 (2002).

100. Rämet, M., Manfruelli, P., Pearson, A., Mathey-Prevot, B. & Ezekowitz, R. A. B. Functional genomic analysis of phagocytosis and identification of a Drosophila receptor for E. coli. *Nature* **416,** 644–648 (2002).

101. Kaneko, T. *et al.* Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the Drosophila IMD pathway. *Immunity* **20,** 637–649 (2004).

102. Takehana, A. *et al.* Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in Drosophila immunity. *Embo J* **23,** 4690–4700 (2004).

103. Kaneko, T. *et al.* PGRP-LC and PGRP-LE have essential yet distinct functions in the drosophila immune response to monomeric DAP-type peptidoglycan. *Nat. Immunol.* **7,** 715–723 (2006).

104. Brumell, J. H. & Scidmore, M. a. Manipulation of rab GTPase function by intracellular bacterial pathogens. *Microbiol. Mol. Biol. Rev.* **71,** 636–652 (2007).

105. Roy, C. R. Exploitation of the endoplasmic reticulum by bacterial pathogens. *Trends Microbiol.* **10,** 418–424 (2002).

106. Serbus, L. R., Casper-Lindley, C., Landmann, F. & Sullivan, W. The genetics and cell biology of Wolbachia-host interactions. *Annu. Rev. Genet.* **42,** 683–707 (2008).

107. Voronin, D., Dudkina, N. V & Kiseleva, E. V. A new form of symbiotic bacteria Wolbachia found in the endoplasmic reticulum of early embryos of Drosophila melanogaster. *Dokl Biol Sci* **396,** 227–229 (2004).

108. McClintock, B. Induction of Instability at Selected Loci in Maize. *Genetics* **38,** 579–599 (1953).

109. McClintock, B. The origin and behaviour of mutable loci in maize. *Genetics* **36,** 344–355 (1950).

110. Orgel, L. E. & Crick, F. H. C. Selfish DNA: the ultimate parasite. *Nature* **284,** 604–607 (1980).

111. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284,** 601–603 (1980).

112. Kidwell, M. G. & Lisch, D. R. Perspective: transposable elements, parasitic dna, and genome evolution. *Evolution (N. Y).* **55,** 1–24 (2001).

113. González, J., Macpherson, J. M. & Petrov, D. a. A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila melanogaster. *Mol. Biol. Evol.* **26,** 1949–1961 (2009).

114. González, J. & Petrov, D. A. The adaptive role of transposable elements in the Drosophila genome. *Gene* **448,** 124–33 (2009).

115. Kofler, R., Betancourt, A. J. & Schlötterer, C. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. *PLoS Genet.* **8,** (2012).

116. Fiston-Lavier, A.-S., Barrón, M. G., Petrov, D. A. & González, J. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Res.* **43,** e22 (2014).

117. Kaminker, J. *et al.* The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol.* **3,** 1–20 (2002).

118. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436,** 793–800 (2005).

119. Daboussi, M.-J. & Capy, P. Transposable elements in filamentous fungi. *Annu. Rev. Microbiol.* **57,** 275–299 (2003).

120. Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

121. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

122. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12,** 615–627 (2011).

123. Feschotte, C., Jiang, N. & Wessler, S. R. Plant Transposable Elements: Where Genetics Meets Genomics. *Nat. Rev. Genet.* **3,** 329–341 (2002).

124. Carr, M., Bensasson, D. & Bergman, C. M. Evolutionary Genomics of Transposable Elements in Saccharomyces cerevisiae. **7,** (2012).

125. Kim, J. M. *et al.* Transposable Elements and Genome Organization : A Comprehensive Survey of Retrotransposons Revealed by the Complete Saccharomyces cerevisiae Genome ? Sequence Transposable Elements and Genome Organization : A Comprehensive Survey of Retrotransposons Revea. 464–478 (1998). doi:10.1101/gr.8.5.464

126. Kazazian Jr., H. H. Mobile Elements: Drivers of Genome Evolution. *Science (80-.).* **303,** 1626–1632 (2004).

127. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41,** 331–68 (2007).

128. Kempken, F. & Windhofer, F. The hAT family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110,** 1–9 (2001).

129. Feschotte, C. & Pritham, E. J. Non-mammalian c-integrases are encoded by giant transposable. *BMC Evol. Biol.* **4,** 22–32 (2004).

130. DeMarco, R., Venancio, T. M. & Verjovski-Almeida, S. SmTRC1, a novel Schistosoma mansoni DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily. *BMC Evol. Biol.* **6,** 89 (2006).

131. Greenblatt, I. M. & Brink, A. Twin mutations in medium variegated pericarp maize. *Genetics* **47,** 489–501 (1961).

132. Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.* **14,** 1613–25 (1994).

133. *Nobel Lectures: Physiology or Medicine 1981-1990*. (World Scientific Publishing Co. Pte. Ltd., 1993).

134. Ikeda, K., Nakayashiki, H., Takagi, M., Tosa, Y. & Mayama, S. Heat shock, copper sulfate and oxidative stress activate the retrotransposon MAGGY resident in the plant pathogenic fungus Magnaporthe grisea. *Mol. Genet. Genomics* **266,** 318–325 (2001).

135. Kimura, Y. *et al.* OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* **42,** 1345–1354 (2001).

136. Salazar, M., González, E., Casaretto, J. A., Casacuberta, J. M. & Ruiz-Lara, S. The promoter of the TLC1.1 retrotransposon from Solanum chilense is activated by multiple stress-related signaling molecules. *Plant Cell Rep.* **26,** 1861–1868 (2007).

137. Servant, G., Pennetier, C. & Lesage, P. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon

under severe adenine deficiency. *Mol. Cell. Biol.* **28,** 5543–5554 (2008).

138. Woodrow, P., Pontecorvo, G., Ciarmiello, L. F., Fuggi, A. & Carillo, P. Ttd1a promoter is involved in DNA-protein binding by salt and light stresses. *Mol. Biol. Rep.* **38,** 3787–3794 (2011).

139. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **22,** 1503–1517 (2013).

140. Pouteau, S., Grandbastien, M.-A. & Boccara, M. Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *Plant J.* **5,** 535–542 (1994).

141. Grandbastien, M. A. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3,** 181–187 (1998).

142. Takeda, S., Sugimoto, K., Otsuki, H. & Hirochika, H. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* **18,** 383–393 (1999).

143. Beguiristain, T., Grandbastien, M.-A., Puigdomènech, P. & Casacuberta, J. M. Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants. *Plant Physiol.* **127,** 212–221 (2001).

144. Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* **33,** 479–532 (1999).

145. Grandbastien, M. a, Spielmann, A. & Caboche, M. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337,** 376–380 (1989).

146. Dai, J., Xie, W., Brady, T. L., Gao, J. & Voytas, D. F. Phosphorylation Regulates Integration of the Yeast Ty5 Retrotransposon into Heterochromatin. *Mol. Cell* **27,** 289–299 (2007).

147. Ebina, H. & Levin, H. L. Stress Management: How Cells Take Control of Their Transposons. *Mol. Cell* **27,** 180–181 (2007).

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

149. Junakovic, N., Franco, C. Di, Barsanti, P. & Palumbo, G. Transposition of Copia-like Nomadic Elements Can Be Induced by Heat Shock. *J. Mol. Evol.* **24,** 89–93 (1986).

150. Ratner, V. A., Zabanov, S. A., Kolesnikova, O. V & Vasilyeva, L. A. Induction of the mobile genetic element Dm-412 transpositions in the Drosophila genome by heat shock treatment. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5650–4 (1992).

151. Arnault, C., Loevenbruck, C. & Biémont, C. Transposable element mobilization is not induced by heat shocks in Drosophila melanogaster. *Naturwissenschaften* **84,** 410–4 (1997).

152. Jouan-Dufournel, I., Cosset, F., Contamine, D., Verdier, G. & Biémont, C. Transposable Elements Behavior Following Viral Genomie Stress in Drosophila melanogaster Inbred Line. *J. Mol. Evol.* **43,** 19–27 (1996).

153. Nabirochkin, S. D. *et al.* Oncoviral DNAs induce transposition of endogenous mobile elements in the genome of Drosophila melanogaster. *Mutat. Res.* **403,** 127–36 (1998).

154. Wessler, S. R. Phenotypic Diversity Mediated by the Maize Transposable Elements Ac and Spm. *Science (80-.).* **242,** 399–405 (1988).

155. Biémont, C. & Vieira, C. Genetics: junk DNA as an evolutionary force. *Nature* **443,** 521–524 (2006).

156. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9,** 397–405 (2008).

157. Cowley, M. & Oakey, R. J. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. *PLoS Genet.* **9,** (2013).

158. Walser, J.-C., Chen, B. & Feder, M. E. Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. *PLoS Genet.* **2,** e165 (2006).

159. Wang *et al.* Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18613–18618 (2007).

160. Mariño-Ramírez, L., Lewis, K. C., Landsman, D. & Jordan, I. K. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* **110,** 333–341 (2005).

161. Van De Lagemaat, L. N., Landry, J. R., Mager, D. L. & Medstrand, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **19,** 530–536 (2003).

162. Mikkelsen, T. S. *et al.* Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447,** 167–77 (2007).

163. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441,** 87–90 (2006).

164. Santangelo, A. M. *et al.* Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. *PLoS Genet.* **3,** 1813–1826 (2007).

165. Britten, R. J. Cases of ancient mobile element DNA insertions that now

affect gene regulation. *Mol. Phylogenet. Evol.* **5,** 13–17 (1996).

166. Lowe, C. B., Bejerano, G. & Haussler, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8005–8010 (2007).

167. Conley, A. B., Miller, W. J. & Jordan, I. K. Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet.* **24,** 53–56 (2008).

168. Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev.* **21,** 2300–2311 (2007).

169. Goodchild, N. L., Wilkinson, D. A. & Mager, D. L. A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. *Gene* **121,** 287–294 (1992).

170. Mager, D. L., Hunter, D. G., Schertzer, M. & Freeman, J. D. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics* **59,** 255–263 (1999).

171. Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **13,** 1097–1101 (2006).

172. Varagona, M. J., Purugganan, M. & Wessler, S. R. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* **4,** 811–20 (1992).

173. Davis, M. B., Dietz, J., Standiford, D. M. & Emerson, C. P. Transposable element insertions respecify alternative exon splicing in three Drosophila myosin heavy chain mutants. *Genetics* **150,** 1105–1114 (1998).

174. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21,** 708–718 (2007).

175. Lescot, M. *et al.* Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J.* 1–13 (2015). doi:10.1038/ismej.2015.192

176. Iorizzo, M. *et al.* De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* **12,** 389 (2011).

177. Costa, G. G. L. *et al.* Transcriptome analysis of the oil-rich seed of the bioenergy crop Jatropha curcas L. *BMC Genomics* **11,** 462 (2010).

178. Vicient, C. M. Transcriptional activity of transposable elements in maize. *BMC Genomics* **11,** 601 (2010).

179. Böhne, A., Sengstag, T. & Salzburger, W. Comparative transcriptomics in East African cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biol. Evol.* **6,** 2567–2585 (2014).

180. Nandi, S. *et al.* Repeat structure of the catfish genome: A genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (Ictalurus punctatus). *Genetica* **131,** 81–90 (2007).

181. Daborn, P. J. *et al.* A Single P450 Allele Associated with Insecticide Resistance in Drosophila. *Science (80-.).* **2253,** 2253–2256 (2013).

182. Aminetzach, Y. T., Macpherson, J. M. & Petrov, D. A. Pesticide Resistance via Transposition-Mediated Adaptive Gene Truncation in Drosophila. *Science (80-.).* **309,** 764–767 (2005).

183. Guio, L., Barrón, M. G. & González, J. The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila. *Mol. Ecol.* **23,** 2020–2030 (2014).

184. González, J., Lenkov, K., Lipatov, M., Macpherson, J. M. & Petrov, D. a. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol.* **6,** 2109–2129 (2008).

185. Wagner, G. P. & Lynch, V. J. Evolutionary novelties. *Curr. Biol.* **20,** 48–52 (2010).

186. Stern, D. L. Evolutionary Developmental Biology and the Problem of Variation. *Evolution (N. Y).* **54,** 1079 (2000).

187. Santos, M. E. *et al.* The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* **5,** 5149 (2014).

188. Emera, D. & Wagner, G. P. Transposable element recruitments in the mammalian placenta: Impacts and mechanisms. *Brief. Funct. Genomics* **11,** 267–276 (2012).

189. Lisch, D. How important are transposons for plant evolution? *Nat Rev Genet* **14,** 49–61 (2012).

190. Kapitonov, V. V. & Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3,** 0998–1011 (2005).

191. Breitling, R. & Gerber, J. K. Origin of the paired domain. *Dev. Genes Evol.* **210,** 644–650 (2000).

192. Casola, C., Hucks, D. & Feschotte, C. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol. Biol. Evol.* **25,** 29–41 (2008).

193. Babu, M. M., Iyer, L., Balaji, S. & Aravind, L. The natural history of the

WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* **34,** 6505–6520 (2006).

194. Lin, R. *et al.* Transposase-Derived Transcription in Arabidopsis. *Science (80-.).* **318,** 1302–1305 (2007).

195. Khurana, J. S. *et al.* Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell* **147,** 1551–1563 (2011).

196. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. a. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12,** 246–258 (2011).

197. Juliano, C., Wang, J. & Lin, H. Uniting Germline and Stem Cells: the Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu. Rev. Genet.* 1–26 (2011). doi:10.1146/annurev-genet-110410-132541.Uniting

198. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10,** 94–108 (2009).

199. Malone, C. D. & Hannon, G. J. Small RNAs as Guardians of the Genome. *Cell* **136,** 656–668 (2009).

200. Vagin, V. V *et al.* The RNA Interference Proteins and Vasa Locus are Involved in the Silencing of Retrotransposons in the Female Germline of Drosophila melanogaster. *RNA Biol.* **1,** 54–58 (2004).

201. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. **33,** 2052–2059 (2005).

202. Li, C. *et al.* Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137,** 509–521 (2009).

203. Malone, C. D. *et al.* Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137,** 522–535 (2009).

204. Nagao, A. *et al.* Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. *RNA* **16,** 2503–2515 (2010).

205. Aravin, A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr. Biol.* **11,** 1017–1027 (2001).

206. Spradling, A. C. *Developmental genetics of oogenesis. In 'Development of Drosophila melanogaster'.* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993).

207. Brennecke, J. *et al.* An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (80-.).* **1387,** 1387–1392 (2008).

208. Chambeyron, S. *et al.* piRNA-mediated nuclear accumulation of

retrotransposon transcripts in the Drosophila female germline. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 14964–14969 (2008).

209. Shpiz, S., Kwon, D., Rozovsky, Y. & Kalmykova, A. rasiRNA pathway controls antisense expression of Drosophila telomeric retrotransposons in the nucleus. *Nucleic Acids Res.* **37,** 268–278 (2009).

210. Chalvet, F. *et al.* Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J.* **18,** 2659–2669 (1999).

211. Leblanc, P. *et al.* Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. *J. Virol.* **74,** 10658–69 (2000).

212. Pélisson, a *et al.* Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J.* **13,** 4401–4411 (1994).

213. Akkouche, A. *et al.* Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* **14,** 458–64 (2013).

214. Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res.* 1776–1785 (2009). doi:10.1101/gr.094896.109.In

215. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128,** 1089–1103 (2007).

216. Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461,** 1296–1299 (2009).

217. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell* **151,** 964–980 (2012).

218. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (80-.).* **316,** 744–747 (2007).

219. Kawaoka, S. *et al.* The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* **15,** 1258–64 (2009).

220. Houwing, S., Berezikov, E. & Ketting, R. F. Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J.* **27,** 2702–2711 (2008).

221. Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E. & Blower, M. D. Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. *EMBO J.* **28,** 2945–2958 (2009).

222. Gangaraju, V. K. *et al.* Drosophila Piwi functions in Hsp90-mediated

suppression of phenotypic variation. *Nat. Genet.* **43,** 153–158 (2011).

223. Izumi, N. *et al.* Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. *RNA J.* **19,** 896–901 (2013).

224. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **53,** 1689–1699 (1998).

225. Queitsch, C., Sangster, T. a & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417,** 618–24 (2002).

226. Jarosz, D. F. & Lindquist, S. Hsp90 and Environmental Stress Transform the Adaptive Value of Natural Genetic Variation. *Science (80-.).* **330,** 1820–1825 (2010).

227. Rohner, N. *et al.* Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish. *Science (80-.).* **342,** 1372–1375 (2013).

228. Specchia, V. *et al.* Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463,** 662–665 (2010).

229. Haase, A. D. A Small RNA-Based Immune System Defends Germ Cells against Mobile Genetic Elements. *Stem Cells Int.* 1–9 (2015).

230. Kidwell, M., Kidwell, J. & Sved, J. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traitsincluding mutation, sterility and male recombination. **85,** 1973–1975 (1976).

231. Schaefer, R. E., Kidwell, Z. M. G. & Fausto-sterling, A. Hybrid dysgenesis in Drosophila melanogaster: morphological and cytological studies of ovarian dysgenesis. *Public Health* 1141–1152 (1979).

232. Dramard, X., Heidmann, T. & Jensen, S. Natural epigenetic protection against the I-factor, a Drosophila LINE retrotransposon, by remnants of ancestral invasions. *PLoS One* **2,** e304 (2007).

233. Rose, M. & Doolittle, F. Molecular biological mechanisms of speciation. *Science (80-.).* **220,** 157–162 (1983).

234. Riegler, M., Sidhu, M., Miller, W. J. & O'Neill, S. L. Evidence for a global Wolbachia replacement in Drosophila melanogaster. *Curr. Biol.* **15,** 1428–1433 (2005).

Chapter II

Survey of transposable element copy-number in natural *Drosophila melanogaster* genotypes provides new insights into the mechanisms for *blood* and *gypsy5* activity

2.1 Abstract

Transposable elements (TEs) are mobile DNA sequences that constitute a large fraction of eukaryotic genomes, and mediate the production of genetic and phenotypic variation[1–5]. Because transposition threatens genomic integrity and typically has deleterious effects, organisms evolved a protective mechanism to specifically restrain TE activity in the germline[6–12], the piRNA pathway. Despite much work on this pathway, the genetic basis of natural variation in TE activity remain unknown[5]. To study the loci that contain variation affecting TE activity in natural populations we performed a genome-wide association study using as phenotype the number of *blood, gypsy5, mdg1, roo, I-element, jockey, H-element* and *pogo* insertions that are more likely to result from recent TE activity[13]. We observed little overlap between alleles associated with the activity of different transposons and very few affected piRNA pathway genes. Our analysis on *blood* activity led us to propose that the transcription factor Sin3A might have played an important role in affecting the expression of this TE. Conversely, the burst of *gypsy5* activity observed in a single genotype was related to alleles affecting several genes functioning in follicle cell development and that included the transcription factor Pointed. This study identified an unanticipated set of regulators and highlighted the importance of genome-wide association studies to address questions regarding the genetic factors that influence TE activity in the wild.

2.2 Introduction

Transposable elements (TEs) are mobile DNA sequences that were first discovered by Barbara McClintock's revolutionary work on the effect of genomic instability in the context of color pigmentation in maize kernels[14]. They are abundant, diverse and active components of most eukaryotic genomes, and while transposition events are often deleterious, TEs can contribute to genetic variation that can be adaptive and relevant for evolution[1,3,4,15–18]. Recent studies have also implicated TEs as mediators of phenotypic diversification contributing

for the evolution of novel traits such as placenta[19], fish spots[20] and color pigmentation in British peppered moths[21].

TEs are divided into two main categories that group mobile elements according to their sequence properties, which reflect how transposition occurs. Retrotransposons resemble retroviruses and move via an RNA intermediate that adds new copies to the profile of already existing insertions; DNA elements are excised and integrated into new genomic places thus the number of copies is maintained at each cycle of movement[22,23]. The *copy-and-paste* and *cut-and-paste* mechanisms of transposition described above include a scope of distinct steps such as transcription, translation and integration, upon which hosts can act to control TE-specific activity[24–26]. Moreover, the discovery of a conserved pathway – the piRNA pathway – that produces small RNAs able to target and reduce expression of mobile elements (retrotransposons and DNA elements), converged efforts to the study of TE transcriptional regulation[6–9].

The piRNA pathway functions as a genome surveillance system that specifically minimizes TE access to female and male gametocytes[6,27]. In *D. melanogaster* ovaries, the biogenesis of PIWI-interacting RNAs (piRNAs) differs between somatic (follicular epithelium that establish an intricate relationship with germline) and germline (oocyte and nurse cell syncytium) tissue where the silencing process is specialized to repress TEs exclusive to each cell type[28]. Although transposons that are active in the germline tissue have direct access to the oocyte, TEs expressed in follicular cells need to form viral-like particles in order to be able to invade the female gametocyte[29,30].

Despite the extensive information regarding repressive mechanisms, the genetic factors that modulate TE activity in natural populations remain elusive[5]. Here, we aimed to identify loci in host genomes that carry allelic variation contributing to changes in TE activity and that leads to differences in the number of recent TE insertions. We benefited from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) that includes various fully sequenced genotypes containing alleles that segregated in a natural population[13]. The *in silico* analysis of the sequence of 146 of those isogenic lines provided information on TE insertions,

including the position of insertions, the element inserted, and whether or not each insertion was shared with the *D. melanogaster* reference genome (from another population of flies). TE insertions were classified as *novel* when they were not shared with reference genome; these were also typically not shared between different DGRP genotypes and are, thus, more likely to result from recent TE activity. We performed a genome-wide association study (GWAS) using the number of *novel* insertions of different transposons (*blood*, *gypsy5*, *mdg1*, *roo*, *I-element*, *jockey*, *H-element* and *pogo*) as phenotype. To do the work, we first validated *in silico* predictions of TE composition of naturally-derived sequenced genomes and then used those predictions as a phenotype in the GWAS study. Although we had hypothesized to find allelic variants for piRNA genes, we observed little overlap between loci contributing to differences in the number of *novel* insertions and those encoding genes of the piRNA pathway. Analysis of the GWAS results suggest that activity of the TE *blood*, a retrotransposon expressed in germline and somatic tissue of ovaries, is regulated by the transcription factor Sin3A, and that a burst in the activity of TE *gypsy5*, a retrotransposon specifically expressed in follicular cells, may be related to allelic variation for genes important for the development of follicular epithelia.

2.3 Materials and Methods

TE identity for novel insertions in DGRP genotypes

The DGRP is composed by nearly 200 wild-derived genotypes, 146 of which have available information about TE insertions that were characterized as *shared* or *novel* according to whether they are present or absent, respectively, in the reference genome[13]. For the *novel* insertions, breakpoint estimates (position of insertion) and TE identity were assessed based on sequence available for left and right sides of the insertion, as described in ref 13. For each transposon identified in a *novel* insertion, the *in silico* annotation provided several TEs, each with an associated probability. Here, for each *novel* insertion, we used as TE identity that with highest estimated probability.

Confirmation of TE identity of novel insertions

To confirm *in silico* predictions of TE composition, we selected 129 predicted *novel* insertions in 12 DGRP genotypes including 13 for TE *blood*, 4 for *copia*, 7 for *Doc*, 7 for *gypsy5*, 7 for *H-element*, 5 for *hopper,* 3 for *F-element*, 25 for *I-element*, 5 for *Juan*, 19 for *mdg1*, 2 for *opus*, 17 for *pogo*, 10 for *roo* and 5 for *Transpac* elements. List of TE insertions and respective DGRP lines tested here can be found in **Supplementary Table 2.1**. We extracted gDNA from RAL-021, RAL-321, RAL-357, RAL-358, RAL-381, RAL-443, RAL-761, RAL-790, RAL-810, RAL-812, RAL-892 and RAL-908 flies, pool of ten males per sample. DNeasy Blood & Tissue Kit (Qiagen) was used for the extraction, following manufacturer's protocol. Sample concentration was measured in Nanodrop ND-1000 Spectrophotometer and samples were stored at -20°C until use.

Based on the genome sequence available for the DGRPs, we designed primers for the flanking region of each *novel* insertion and obtained amplicons by longPCR (Roche) using 4ng of gDNA in 15μl of reaction with 0.5μM primers, 2% DMSO, 1x buffer, 0.5mM dNTPs, 0.21μl of enzyme in 15μl of reaction. We used standard amplification program for longPCR with 10 min of elongation time; temperature of annealing and primer sequences can be found in **Supplementary Table 2.1**. Size of insertion was obtained by electrophoresis in 1% agarose gel and classified as "larger", "expected" or "smaller" taking into account the canonical TE sequences described for *Drosophila* transposons (version 9.42). We assessed TE identity by sequencing the amplicons with BigDye Terminator protocol (TermoFisher) or SUPREMErun™ (NZYTech) using the same forward primers used for amplification (**Supplementary Table 2.1**).

To detect presence of false negatives in Mackay *et al.* predictions of TE insertions we performed a PCR using as template gDNA from lines where insertions were predicted to be absent (**Supplementary Table 2.2**). As positive controls, we performed same PCRs using either gDNA from a line with predicted insertions (RAL-321; check for primers) or primers for gene Rpl32 (check for template). In negative control reactions for each primer pair gDNA was replaced by water. The gDNA extraction of RAL-109, RAL-161, RAL-237, RAL-350, RAL-

362, RAL-555, RAL-776 and RAL-908 was performed as described above. Each PCR reaction contained 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2, 0.4μM primers and 0.4ng of gDNA. The thermal cycling protocol was the same for all reactions: 10 min at 95°C; 35 cycles of 95°C for 30 sec, 60°C for 1 min and 72°C for 30 sec; 5 min at 72ºC. Primer sequences are described in **Supplementary Table 2.3**.

Genome Wide Association Study (GWAS)

We performed a GWAS analysis using the number of *novel* insertions of *blood, gypsy5, mdg1, roo, I-element, jockey, H-element*, and *pogo* as phenotypes. This information is available for 146 of the DGRP genotypes. The analyses were done on the DGRP2 website (http://dgrp2.gnets.ncsu.edu/)[13,31]. For the genes containing alleles found to associate to variation in number of *novel* insertions (p<7e-05), we ran gene ontology (GO) analyses using GO consortium algorithm, and Bonferroni correction for multiple testing[32].

RNA extraction and cDNA synthesis

To assess *blood, gypsy5,* and *pointed* expression we chose nine fly stocks with different number of *novel* insertions: RAL-021, RAL-237, RAL-321, RAL-357, RAL-358, RAL-375, RAL-391, RAL-790 and RAL-908. To eliminate *Wolbachia* from the lines RAL-021, RAL-237, RAL-321 and RAL-790, flies were treated with tetracycline (0.05mg/ml) for two generations, and gut flora content homogenized as previously described[33]. We only used for this study individuals without endosymbiont infection. Flies were kept in standard cornmeal food[34] at 25°C, 60% humidity and 12h day-night cycles.

We collected ovaries from seven days old females and pooled both ovaries from eight females per sample. Ovaries were obtained from two independent experiments, one with eight and another with four replicate samples per genotype. Dissections were done in fresh cold PBS 1x and ovaries were directly stored in 400µl of Trizol (Ambicon) at 4ºC until homogenized by pestels. Tissue collection and homogenization was done in the same day; samples were then

stored at -80ºC until RNA extraction. For RNA extraction, we used Direct-zol RNA Miniprep kit (Zymo Research) following manufacturer's instructions, and total RNA was eluted in 25µl of RNase-free water (Sigma). RNA purity and concentration was measured with Nanodrop ND-1000 Spectrophotometer; A600/A800 absorbance was around two for all samples and concentrations ranged from 200 to 800ng/µl. All RNA samples of the same genotype were processed on the same day.

For cDNA synthesis we removed contaminations of gDNA from 1µg of RNA using DNAse (Promega) treatment and following manufacturer's instructions. We then followed Reverse Transcription System (Promega) protocol using 0.02µM Oligo dT primers to produce cDNA. Reaction was incubated at 42°C for 60min and heated at 95°C for 5min. cDNA was diluted 1:10 in RNAse-free water and stored at -20°C until quantitative real-time PCR (qPCR).

qPCR

We measured levels of *gypsy5, pointed, Rpl32* and *Tbp* expression by qPCR using a CFX384 thermal cycler (BioRad). For each reaction, we used 5µl iQ™ SYBR® Green supermix (BioRad), 1µl of 4µM primers and 4µl of diluted cDNA (1:10). Primer sequences are described in **Supplementary Table 2.3**. We ran two technical replicates per sample, both in the same qPCR plate. The thermal cycling protocol was the same for all reactions: 2 min at 50°C; 10 min at 95°C; 40 cycles of 95°C for 30 sec, 60°C for 1 min and 72°C for 30 sec. In each plate, we ran standard curves (one standard curve for each gene or TE tested) with serial dilutions of known amplicon concentrations that were amplified by PCR using 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2, 0.4µM primers and 1µl of cDNA (1:10). PCR products were cleaned using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), and eluted in 15µl of buffer NE. We measured concentrations in Nanodrop ND-1000 Spectrophotometer. The minimum amount of nucleic acids that can be detected by qPCR for each primer pair was estimated by extending dilutions of our standards until the relationship between the known amplicon concentrations and the quantification cycles (Cqs)

reached a plateau. Samples that had Cq values outside the linear detection (**Supplementary Figure 2.1**) and with standard deviation above 0.5 for the technical replicates were eliminated from the analysis. We used default threshold settings to obtain the Cq for each reaction, and transformed Cq values into absolute target concentrations using the respective standard curves. This methodology allowed us to control for plate effects and for differences in primer efficiency. qPCR melting curves were analyzed to confirm specificity of amplified products and negative controls never showed detectable amplification. For the analysis of expression data, we used NormFinder[35] and geNorm[36] algorithms to check expression stability of two putative reference genes, *Rpl32* and *Tbp*, commonly used to normalize gene expression at different environmental conditions and genotypes[37,38]. Stability values obtained for *Rpl32* (NormFinder, 0.58; geNorm, 1.32) were overall better than those obtained for *Tbp* (NormFinder, 0.92; geNorm, 1.32). We therefore normalized TE expression to Rpl32.

Statistical analyses

All statistical analyses were done in R (version 3.0.3)[39].

For the analysis of *gypsy5* and *pointed* expression in different DGRP lines, we used linear models (lm) with lognormal distribution. Multiple comparisons were performed with least-squares means (lsmeans), and compact letter display (cld) was used to group genotypes that are not statistically significant ($p>0.05$). Data was obtained from two independent experiments (one with eight and another with four biological replicates) that were plotted and analyzed together.

Analysis of TE expression from available next-generation RNA sequencing data

To study the importance of Sin3A to regulating *blood* expression we analyzed available next-generation RNA sequencing libraries (GSE72172 and GSE68775) deposited on NCBI[40]. We extracted raw data for S2 (GSM1857008, GSM1857009 and GSM1857010), Sin3A-187HA (GSM1857011, GSM1857012

and GSM1857013), GFP RNAi (GSM1681056, GSM1681057 and GSM1681058) and Sin3A KD (GSM1681059, GSM1681060 and GSM1681061) samples. Those samples were obtained from four established cell culture lines: without any construct (S2), with GFP RNAi construct, overexpressing Sin3A-187 (Sin3A-187HA), and another down-regulating all isoforms (Sin3A KD). Reads were aligned to transposon canonical sequences (version 9.42) of *D. melanogaster.* We extracted the number of reads for each TE and analyzed the data using edgeR by fitting read counts to a negative binomial distribution. To evaluate the effect of overexpressing Sin3A-187 (Sin3A-187HA), we used S2 samples as negative controls; to evaluate the effect of knocking down Sin3A (Sin3A KD) isoforms we used GFP RNAi samples as negative controls. We plotted fold-change of TE expression with false discovery rate inferior to 0.1.

2.4 Results

Confirmation of TE insertions in DGRP lines

We studied a total of 129 *novel* insertions that included transposons with different sequence properties: retrotransposons with and without long terminal repeats (LTR and non-LTR elements), and DNA elements. For all 129 predicted insertions (100%), we confirmed presence of transposon sequences, and for 113 of them (87.6%) we confirmed the most likely transposon identity as per the *in silico* predictions[13] (correct TE identity, **Figure 2.1**). For eight of the 16 cases where we did not confirm the most likely predicted identity, our sequence matched that of elements with same sequence properties and seven matched sequence identity of the second most likely predicted TE (**Supplementary Table 2.4**). For the remaining nine cases, the inserted TE was not within the list of predicted identities for the specific positions.

For the characterization of TE insertions according to length (relative to the length for the TEs identified by our sequencing data) we observed that 58.9% showed proper size of insertion, 33.3% were smaller and 7.8% larger (**Figure 2.1**).

We also assessed the occurrence of false negatives in Mackay *et al.* predictions by testing presence of TE insertions in DGRP genotypes that were predicted to lack those elements (**Supplementary Table 2.2**). We detected presence of tested TEs in the genome of all lines assessed (**Supplementary Figure 2.2**).

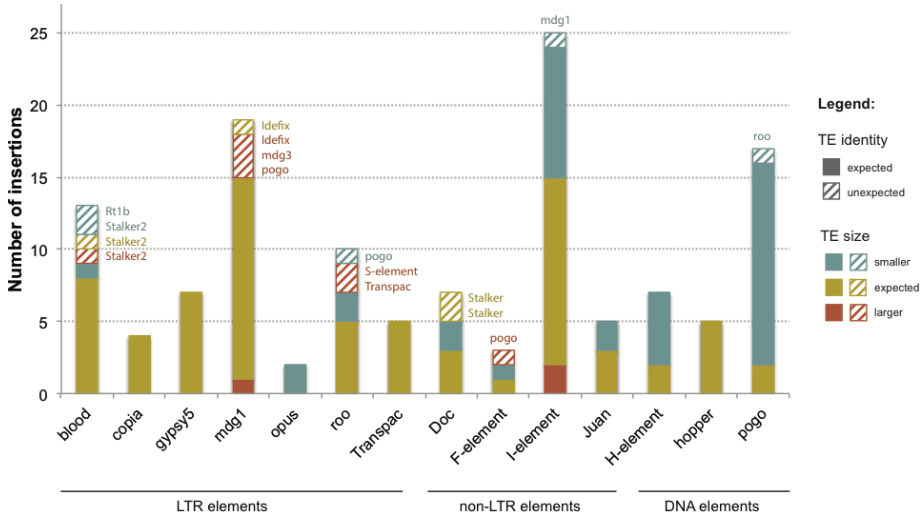


Figure 2.1. Characterization of novel TE insertions according to sequence identity and size. For each novel insertion we used as TE identity that with highest estimated probability of insertion[13]. We then characterized according to sequence identity and size insertions assigned for LTR (blood, copia, gypsy5, mdg1, opus, roo and Transpac), non-LTR (Doc, F-element, I-element and Juan) and DNA (H-element, hopper and pogo) elements. Novel TE insertions were grouped according to our ability to confirm TE identity estimated by Mackay et al. (expected and unexpected) and according to size (smaller, expected and larger). Bar charts show the number of blood, mdg1, opus, roo, Transpac, Doc, F-element, I-element, Juan, H-element, hopper and pogo insertions with expected and unexpected TE identity, and with smaller, expected and larger sizes. When TE identity was defined as unexpected we included the name of the transposon obtained instead. In those cases, TE size was characterized into smaller, expected or larger according to the canonical sequence for the TE we identified here.

Few variants at piRNA pathway genes are associated with variation in TE copy number

To establish the link between allelic variation segregating in natural populations and TE activity, we conducted a GWAS using as phenotype the number of *novel* insertions for eight frequent TEs (*blood, gypsy5, H-element, I-element, jockey, mdg1, pogo* and *roo*, **Supplementary Table 2.5**). The list included transposons with different sequence properties and tissue-specificity within ovaries (**Supplementary Figure 2.3**). The GWAS analyses revealed many alleles of small effect explaining differences in TE number, with 75.2% of those within or neighboring annotated protein-coding genes (**Figure 2.2** and **Supplementary Table 2.6**). Most of the loci we found to affect variation in TE number were TE-specific (**Supplementary Figure 2.4**), and only two were piRNA pathway genes: *vreteno* (*vret*) for *gypsy5* number (**Figure 2.2B**) and *Argonaute3* (*Ago3*) for *pogo* (**Figure 2.2H**).

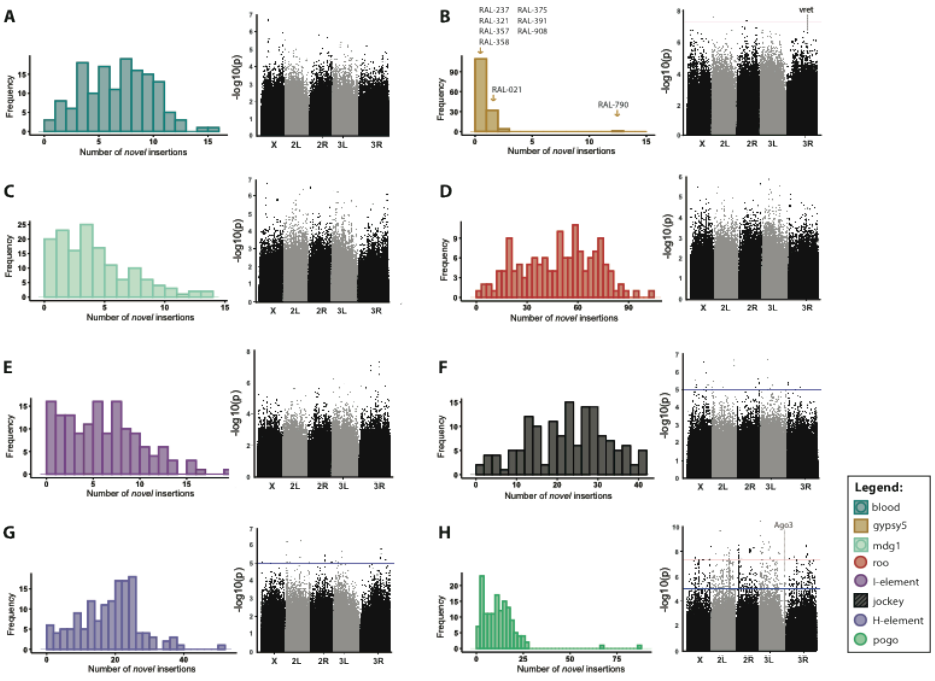


Figure 2.2. GWAS analyses using blood, gypsy5, mdg1, roo, I-element, jockey, H-element and pogo number of novel insertions as phenotype. Frequency distribution of and loci associated with blood **(A)**, gypsy5 **(B)**, mdg1 **(C)**, roo **(D)**, I-element **(E)**, jockey **(F)**, H-element **(G)** and pogo **(H)** copy number. Frequency distribution graphs represent the number of DGRP lines (y-axis) that contain a given number of novel insertions (x-axis); Manhattan plots illustrate the level of statistical significance measured by the negative log of the corresponding p-value for the association of each loci with the phenotype. Each dot represents a SNP; the five chromosomal arms (X, 2L, 2R, 3L and 3R) are represented by different colors. Loci are arranged according to their position in D. melanogaster chromosomes; positions are based on the assembly number five of D. melanogaster genome. The horizontal lines in Manhattan plots are significance thresholds (red, p-value=5e-08; blue, p-value=5e-05). Alleles affecting piRNA pathway genes (vret for gypsy5 and Ago3 for pogo) and DGRP lines used to analyze gypsy5 and pointed expression (RAL-021, RAL-237, RAL-321, RAL-357, RAL-358, RAL-375, RAL-391, RAL-790 and RAL-908) are highlighted.

Allelic variants associated with number of novel blood insertions

Our GWAS revealed 25 loci associated with *blood* copy number, 18 of which were within or neighboring annotated protein-coding genes (**Table 2.1** and **Supplementary Table 2.6**). The enrichment analysis of GO categories represented by those genes did not reveal any significantly under- or over-represented terms relative to “biological process” or “molecular function”.

In order to provide some sort of validation for our GWAS hits, we took advantage of a published RNAi screen for *blood* expression in *D. melanogaster* ovaries[41], and of available next-generation RNA sequencing data for *Drosophila* S2 cell lines overexpressing and down-regulating a transcription factor, Sin3A[40]. For the 14 genes (out of our list of 16 genes) that were specifically down-regulated in the ovaries[41], we were able to confirm that knocking down CG17839 and unc-13-4A decreased *blood* expression (**Table 2.1)**. On the other hand, down-regulation of Sin3A only reduced significantly the expression of *gypsy2*, while overexpression of Sin3A-187 affected the expression of numerous transposons, including the enhancement of *blood* and *roo* expression (**Figure 2.3**). Those results indicate that Sin3A is not essential but it is able to influence *blood* expression.

Table 2.1. Alleles associated with blood copy number

Allele position	Gene	Modifier	Molecular function	RNAi score[41]
X:6880579 X:6880226	*bou*	5'UTR (SNP) Upstream (SNP)	Unknown	0.919
2L:132010	*CG11455*	Intron (SNP)	NADH dehydrogenase activity	0.324
3L:14800130 3L:14801065	*CG17839*	Intron (SNP) Intron (SNP)	Unknown	2.725
X:6442508 X:6442604	*CG34417*	Intron (SNP) Intron (SNP)	Actin binding; structural constituent of cytoskeleton	-0.350
X:21328526	*CG42343*	Intron (SNP)	Unknown	NA
X:17427915	*CG43658*	Non synonymous (SNP)	Rho guanyl-nucleotide exchange factor activity	NA
2L:7614049	*CG6739*	Synonymous (SNP)	Unknown	0.755
2R:12475530	*CG8311*	Synonymous (SNP)	Unknown	0.071
3L:8272574	*Dscam4*	Intron (SNP)	Identical protein binding	0.804
2L:132010	*Gs1*	Upstream (SNP)	Glutamate synthase activity	0.026
3R:7418995	*Jupiter*	Intron (SNP)	Nucleic acid binding; microtubule binding; structural constituent of cytoskeleton; metal ion binding	0.064
2L:107257	*Nhe1*	Upstream (SNP)	Sodium:proton antiporter activity	-0.202
X:6880226 X:6880579	*ogre*	Intron (SNP) Intron (SNP)	Gap junction channel activity	-0.115
2L:107257	*Sam-S*	Intron (SNP)	Methionine adenosyltransferase activity; ATP binding	0.026
2R:8467572 2R:8468003	*Sin3A*	Intron (SNP) Non synonymous (SNP)	Transcription cofactor activity; protein heterodimerization activity; chromatin binding; transcription factor activity	0.726
3L:7279507 3L:7279514	*unc-13-4A*	Intron (SNP) Intron (SNP)	Unknown	2.047

Molecular functions were obtained from flybase and RNAi scores from Czech *et al.* For scores <-1.5 *blood* is considered to be up regulated and for scores >1.5 *blood* is considered to be repressed in the RNAi line. NAs represent genes that were not tested by RNAi; SNP stands for single nucleotide polymorphism. We obtained allelic position using the assembly number five of *D. melanogaster* genome.

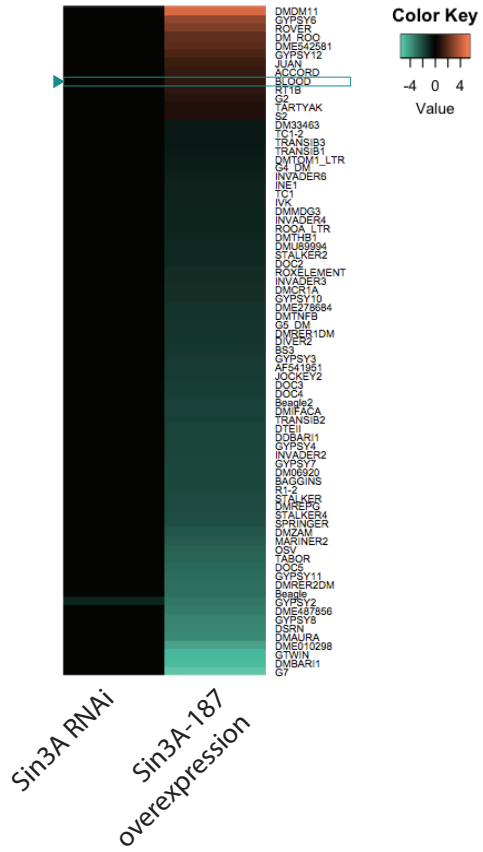


Figure 2.3. Sin3A is able to influence *blood* expression. In *Drosophila*, Sin3A is a single gene that produces several transcripts through alternative splicing. These transcripts are translated into proteins that can have different functions. The most well studied and abundant isoforms are Sin3A-187 and Sin3A-220. Both contain in their coding sequence loci that carry allelic variation contributing to variation in *blood* activity (**Supplementary Figure 2.5**)[40,42,43].

To study the importance of Sin3A in regulating *blood* expression we took advantage of available RNA-sequencing data obtained from two established cell culture lines: one overexpressing Sin3A-187 isoform and another down-regulating all isoforms (Sin3A RNAi)[40]. Heatmap shows fold change of expression with false discovery rate lower than 0.1. Relative expression of *blood* is highlighted by blue box.

Alleles associated with apparent burst of gyspsy5 activity in RAL-790

Contrary to what we have observed for *blood*, the variation of *gypsy5* copy number in DGRP lines was limited as most genotypes had zero to one *novel* insertions and only RAL-790 contained 12 (**Figure 2.2A and B**). This pattern suggested a burst of *gypsy5* activity in RAL-790. To learn more about the genes likely to be important for this phenomenon, we did an enrichment analysis for

gene ontology categories on the genes containing alleles associated with *gypsy5* copy number (**Supplementary Table 2.6**). This analysis revealed 2.9-fold enrichment for genes related to ovarian follicle cell development (GO:0030707, **Supplementary Table 2.7**); 42% coded for binding proteins, 19% for proteins with catalytic activity, 11% for transcription factors, 8% were receptors, 3% enzymes, and 3% molecules with structural activity (**Supplementary Figure 2.6**). The alleles at those loci that are found in RAL-790 are good candidates to explain the putative burst of *gyspy5* activity (**Supplementary Figure 2.7**). From the 22 genes related to follicle cell development, eight are important for border cell migration (*bbg, CG4552, jaguar, jing, kayak, lat, Mrtf* and *Ppcs*, **Table 2.2**). Gene *nclb* mediates follicle stem cell maintenance in the germarium and *Src64B* stabilizes female germline ring canals (**Table 2.2**). We decided to focus our analysis on the other 12 genes that are involved in the maturation and integrity of the epithelial layer around the oocyte, same cells where somatic transposons are usually expressed[29,30] (**Figure 2.4A** and **B**). Those were grouped according to their function in the Notch signaling pathway (**Figure 2.4a1** and **b1**), cell adhesion (**Figure 2.4a2**) and EGFR signaling pathway (**Figure 2.4a3**). Given that Pointed was previously described to be involved in the induction of *ZAM* expression[29], we decided to further analyze the presence of Pointed binding sites in *gypsy5* 5'LTR sequence, as well as the expression of *gypsy5* and *pointed* in different DGRP lines (**Figure 2.2B**). We confirmed that *ZAM* has one binding site for Pointed and observed that *gypsy5* has six (**Figure 2.5A**). As expected *gypsy5* expression is higher in RAL-790 when compared with that in other genotypes (**Figure 2.5B**). RAL-790 has also high *pointed* expression but so do other genotypes with few *novel* copies of *gypsy5* (**Figure 2.5C**).

Table 2.2. Alleles affecting genes related to follicle cell development

Allele position	Gene	Modifier	Function
3L:14423131	*bbg*	Intron (SNP)	JAK/STAT signaling and border cell migration[44]
3L:14507382		Intron (insertion)	
3L:14423141		Intron (SNP)	
2L:1131232	*CG4552*	Intron (SNP)	GTPase activator activity involved in border cell migration[45]
2L:1131223		Intron (SNP)	
2L:1131237		Intron (SNP)	
2R:9383533	*drk*	3'UTR (SNP)	Negative regulation of EGFR activity (flybase)
2R:17431581	*EGFR*	Intron (SNP)	Gurken-activated receptor activity; epidermal growth factor-activated receptor activity[46]
2R:17431587		Intron (SNP)	
2R:15030088	*enabled*	Intron (SNP)	Promotes local actin filament formation; key regulator of the epithelial actin cytoskeleton[47,48]
2L:18348210	*Fas3*	Intron (SNP)	Septate junction[49]
2L:2739179	*Hrs*	Upstream (SNP)	Ubiquitin binding; targets EGFR and Notch proteins for degradation[50]
3R:20079037	*jaguar*	3'UTR (SNP)	Required for correct spindle orientation[51]; border cell migration[52]
2R:2492662	*jing*	Intron (SNP)	Sequence-specific DNA binding; border cell migration (flybase)
3R:25608935	*kayak*	Intron (SNP)	Sequence-specific DNA binding (flybase); border cell migration[53]
3R:10523281	*kibra*	Downstream (insertion)	Signaling network that promotes follicle cell maturation[54]
2R:9089312	*Lat*	Downstream (SNP)	DNA binding; initiation of DNA replication; border cell migration (flybase)
2R:9934723	*mam*	Intron (SNP)	Co-activator of Notch signaling pathway[55]
2R:9888234		Intron (SNP)	
3L:2758889	*Mrtf*	Intron (SNP)	Transcriptional activator activity; border cell migration[56]
3L:2760262		Intron (SNP)	
2R:6763154	*nclb*	Upstream (SNP)	Chromatin binding; required for follicle stem cell maintenance[57]
2R:6763161		Upstream (SNP)	
X:6505272	*pigs*	Intron (SNP)	Negative regulation of Notch signaling pathway[58]
X:6505298		Intron (SNP)	
3R:19152988	*pointed*	Intron (SNP)	Transcriptional activator activity[29]
3R:14972497	*Ppcs*	Intron (SNP)	Ovarian follicle cell migration (flybase)
3R:14972498		Intron (SNP)	
3L:8557355	*rhea*	Synonymous (SNP)	Actin and integrin binding; important for oocyte-follicular cell contact[59]
3L:8551885		Intron (SNP)	
2R:16939241	*shotgun*	Synonymous (SNP)	Cell adhesion molecule[49,59]
2R:16945082		Upstream (SNP)	
X:12611438	*Smrter*	Intron (SNP)	DNA and chromatin binding[60]
X:12588203		Intron (SNP)	
3L:4626482	*Src64B*	Upstream (SNP)	Female germline ring canal formation[61]

We obtained allelic position using the assembly number five of *D. melanogaster* genome. SNP, single nucleotide polymorphism; UTR, untranslated region.

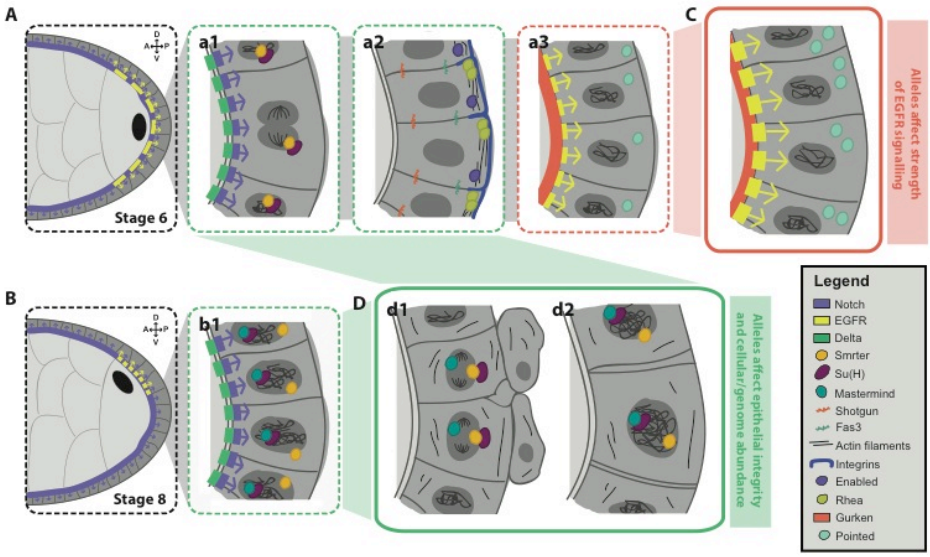


Figure 2.4. Model proposed for the effect of alleles associated with *gyspy5* burst in activity. Scheme illustrating genes important for follicular cell development in *D. melanogaster* ovaries **(A** and **B)**. At stage 6 and 8 of oocyte development, Notch is expressed in the follicular epithelia that protects the egg chamber. At stage 6, the interaction of Smrter with Su(H) inhibits Notch signaling and promotes replication of follicular cells **(a1)**[60]. Later in development, Smrter is dissociated from Su(H) that, consequently, recruits co-activators such as Mastermind **(b1)**[60]. This event results in the inhibition of mitotic cycles and follicular cells enter endoreplication to become polyploid[62]. The polarization of follicle cells is also essential to maintain epithelial functions throughout oogenesis; cell adhesion (Shotgun and Fas3) and actin-binding (Enabled and Rhea) proteins are important for the establishment and maintenance of cellular apical-basal polarity **(a2)**[47–49,59]. The interaction between oocyte and follicular cells is crucial for the establishment of anterioposterior and dorsoventral polarity in the oocyte, and is mainly mediated by waves of Gurken-EGFR signaling (**A** and **B)**[63]. At stage 6, an early signal of Gurken defines the anterioposterior axis **(A)**, whereas a later signal of Gurken at stage 8 results in dorsoventral polarity **(B)**. When the EGFR signals to the follicular cells it induces the expression of several genes including *pointed* **(a3)**[64]. We propose that, some alleles associated with *gypsy5* activity may have enhanced EGFR signaling pathway through effects in EGFR, Pointed, Hrs and Drk abundance and/or activity **(C)**. Also, alleles affecting Notch signaling (through effects in Smrter, Mastermind, Hrs, kibra and pigs) and cell polarity (through effects in Fas3, Shotgun, Rhea and Enabled) could have resulted in increased cell number **(d1)** or genomic content **(d2)**. Those effects could have a direct (through activation of *gypsy5* expression by the transcription factor Pointed, **C**) or an indirect (through increase in the number of cells expressing *gypsy5* or increase in the

number of *gypsy5* copies in polyploid nuclei, **D**) impact in *gypsy5* expression. In order to increase in copy-number, *gypsy5* needs to be expressed and to form viral-like particles that are able to migrate from follicle cells to the oocyte[29,30]. We further suggest that the disruption of epithelial integrity mediated by effects in cell cycle and cell-cell adhesion **(D)** could have benefited the movement of *gypsy5* viral-like particles to the oocyte. For more information about the alleles affecting genes related to follicle cell development see **Table 2.2**.

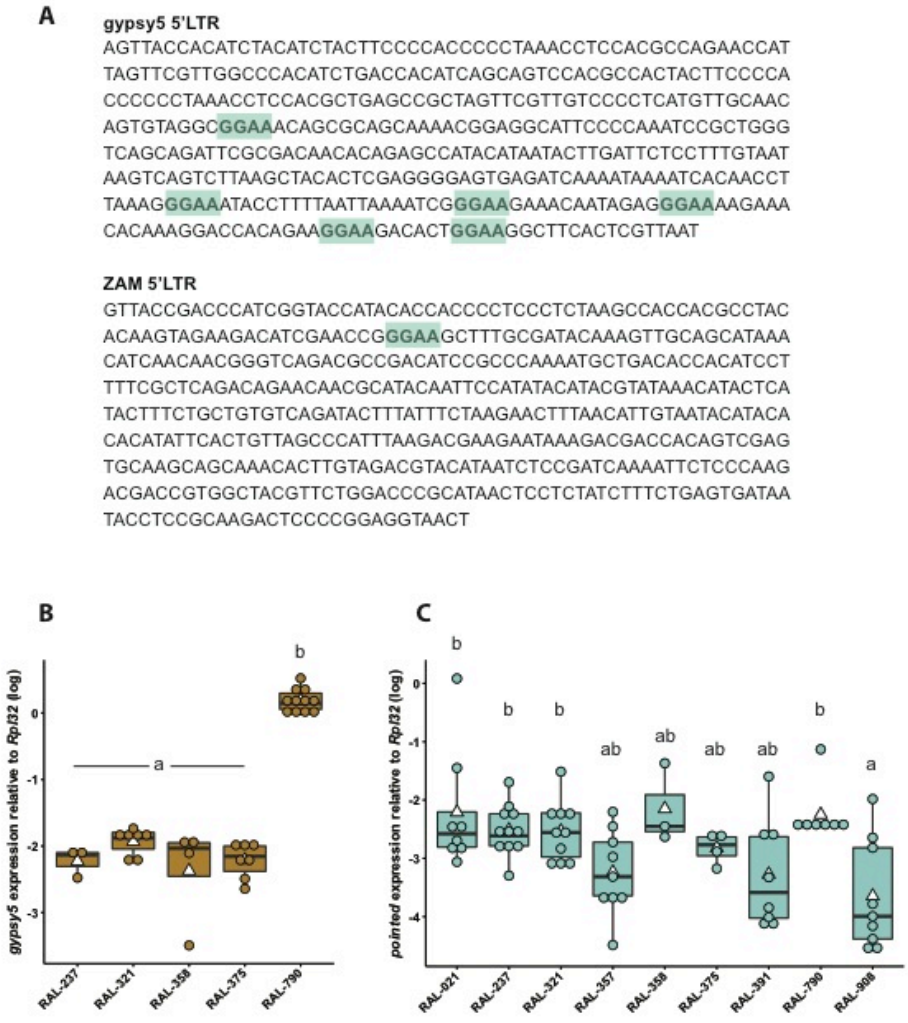


Figure 2.5. Alleles affecting *pointed* expression may explain burst in *gypsy5* activity. Pointed binding sites (GGAA) found in 5'LTR sequences of *gypsy5* and *ZAM* **(A)**. *Gypsy5* expression relative to *Rpl32* (logarithmic scale) obtained for RAL-237, RAL-

321, RAL-358, RAL-375 and RAL-790 **(B)**. We could not detect *gypsy5* expression in RAL-021, RAL-357, RAL-391, and RAL-908. Expression of *pointed* relative to *Rpl32* (logarithmic scale) in RAL-021, RAL-237, RAL-321, RAL-357, RAL-358, RAL-375, RAL-391, RAL-790 and RAL-908 **(C)**. We used linear model (lm) to evaluate the difference in *gypsy5* and *pointed* expression between DGRP lines. Multiple comparisons were performed using least-squares means (lsmeans). We used compact letter display (cld) to group genotypes that are not statistically significant ($p>0.05$), in the graphs represented by the same letter. Data was collected in two independent experiments; in the graph we show results for both experiments. Each circle in the boxplots represents a pool of eight pairs of ovaries; triangles correspond to the mean and black line to the median.

2.5 Discussion

DGRP genotypes: a library for natural TE insertions

TEs are active DNA sequences that constitute a large proportion of most genomes, making up for 4% of the *D. melanogaster* genome, for example[65]. The sequence diversity and repetitive nature of transposons makes their analysis challenging, especially when using next-generation sequencing data that provides fairly short sequencing reads[66,67]. For this reason, the empirical confirmation of *in silico* predictions for TE insertions is essential.

In this study, we confirmed presence of mobile elements at 129 positions predicted to have *novel* TE insertions[13]. The observation of 0% of false positives and our confirmation of 87.6% of the TE identities with highest probability[13], strongly suggests that the information available for TE insertions in DGRP genotypes is reliable, and this panel can therefore be used as a library for the natural variants of transposon insertions. The inability to confirm 12.4% of the more probable TE identities could, however, result from significant differences between our and previous[13] approaches to determine transposon identity. Although our estimation for identity was based on the score for the alignment with canonical TE sequences alone (giving greater weight to a better aligning TE), Mackay *et al.* predictions included the frequency of that element in the reference genome as well (giving greater weight to more frequent elements). In addition, our identities are mostly based in information for one side of TE

insertion tough Mackay *et al*. provide estimates for both sides. For the cases where identities with highest probability are only detected in one side of the insertion, and with second most likely TE being detected in both sides, it would be important to reconsider using *in silico* estimates from both sides (left and right) to assign TE identities for each position of insertion. As for the case of our four misidentified *blood* insertions, this new approach would provide better predictions of TE identities (estimated improvement of 3.1%).

For the characterization of TE insertions according to size, we observed that, although 58.9% had expected sequence length, 33.3% were smaller and 7.8% larger than the predicted sizes for the canonical sequences of the respective mobile elements. Larger insertions could result from the integration of more than one transposon in the same genomic position. In contrast, smaller insertions could derive from 1) sequence loss after TE mobilization and/or 2) the activity of incomplete and non-autonomous elements that share sequence properties with other autonomous transposons[22].

Regarding our results for the presence of false negatives in TE predictions by Mackay *et al.*, we suggest that the lack of genomic information for the chromosome 4 and possibly Y chromosome in the DGRP could explain our detection of specific TEs in genotypes predicted to lack them. Available data for transposon insertions in the fourth and Y chromosomes is of particular interest given the high prevalence of repetitive sequences in those chromosomes[68,69].

Importantly, and although Mackay *et al.* predictions had never been experimentally validated before, we only tested a very small proportion of insertions (129 of approximately 50 thousand *novel* insertions for all transposons).

Few SNPs affecting piRNA pathway genes are associated with TE copy number

A great deal of what is known about TE repression in the germline is grounded on studies that use *D. melanogaster* as a model organism[6,11,12]. Those describe the existence of piRNAs that constitute the hallmark for transposon inhibition in

the gonads and collectively map to clusters that are enriched for fragments of TE insertions. Notably, the piRNA biogenesis is mediated by specialized processes that differ between germline and somatic tissues of *Drosophila* ovaries[28].

In follicular cells, the production of those genomic guardians is independent of the expression of active transposons and constitutes the primary piRNA biogenesis[70]. The biogenesis step takes place in the cytoplasm, where the Piwi protein (only element of the PIWI-family expressed in follicular cells) is loaded with piRNAs[71]. Piwi-piRNA complexes then migrate to the nucleus to mediate transcriptional repression of mobile elements[72]. Only recently, primary biogenesis of piRNAs has been linked to TUDOR domain-containing proteins (such as vreteno) that modulate Piwi function through protein-protein interaction[73].

In germline cells, the architecture of piRNA biogenesis is complex, as it involves the primary biogenesis and a loop of amplification that is mostly driven by Aub and AGO3. Those two PIWI-family proteins are enriched around the nucleus where Aub-piRNA complexes target and direct TE transcripts for cleavage. As a consequence, sense piRNAs are produced and incorporated by AGO3, that has the potential to specifically trigger piRNA cluster processing and increase the abundance of antisense piRNAs loaded into Aub[6].

Interestingly, our GWAS for TE copy-number revealed very few alleles affecting piRNA pathway genes. In fact, we only found one single nucleotide polymorphism (SNP) in *vreteno* coding sequence, and another SNP in an intron of *AGO3*, that were associated with the activity of *gypsy5* and *pogo*, respectively. Those results are of particular interest given the close relationship established between the function of those proteins and the place of *gypsy5* and *pogo* expression. In other words, vreteno protein functions in the primary piRNA biogenesis and in the cell types where *gypsy5* is specifically expressed, whereas AGO3 amplifies the abundance of piRNAs in the germline where *pogo* is exclusively expressed. However, given that perturbations in the function of those two proteins is described to result in overall uncontrolled TE activity[73,74], we would expect those SNPs within vreteno and AGO3 sequences to be associated

with variation in the activity of several transposons. The TE-specific association of both loci for variation in *gypsy5* and *pogo* copy number was therefore unexpected and we further suggest that pleiotropic effects between those and other alleles could explain those results.

Alleles in CG17839, unc-13-4A and Sin3A may explain differences in blood activity

Data available from previous studies[40,41] allowed us to confirm that *CG17839, unc-13-4A* and *Sin3A* are able to regulate *blood* expression. Those results suggest that alleles affecting such genes could potentially influence *blood* expression and lead to changes in *blood* copy number. Although the molecular function of CG17839 and unc-13-4A is still unknown, Sin3A is recognized for its transcription factor activity. Sin3A lacks DNA-binding motifs but can act as scaffold for the assembly of numerous transcription factors[75]. We then propose that the non-synonymous mutation found to be associated with variation in *blood* copy number might affect Sin3A ability to interact with some of those factors, whereas the SNP in intronic sequence may affect its expression. Both loci could therefore impact *blood* expression through effects in the activity and/or expression of this protein. However, our analysis for the overexpression of Sin3A revealed effects in the expression of many transposons including *blood* and *roo*. How alleles in Sin3A could mediate specific effects in *blood* (as those were not found to be associated with variation in *roo* activity) is still largely unknown. We propose that the non-synonymous mutation may have affected the interaction between Sin3A and co-factors able to influence *blood* but not *roo* expression. This hypothesis needs, however, to be tested together with the functional validation of those loci.

Alleles associated with apparent burst of gypsy5 activity in RAL-790

The assessment of TE population dynamics in different species suggests that transposons move through waves of increased activity[76,77]. However, very few episodes of spontaneous bursts in transposition were empirically detected.

Those include a couple of examples from *D. melanogaster* isogenic lines maintained in laboratory conditions[78–80], and crosses between strains with incompatible TE composition[81,82]. Here, for the first time, we reported a burst in TE activity (*gypsy5*) in a genotype (RAL-790) that was collected from the wild and contained alleles segregating in natural populations. Some of the loci associated with variation in the number of *gypsy5* insertions were found to be within or close to genes related to follicle cell development. Given that *gypsy5* is specifically expressed in those cells, we propose a model suggesting that alleles affecting those genes could result in the activation of *gypsy5* expression and/or promote its mobilization to the oocyte (**Figure 2.4C** and **D**). In this model, the SNPs associated with *gypsy5* activity may have caused the intensification of EGFR signaling cascade that can result in more expression of the transcription factor *pointed* (**Figure 2.4C**). Also, effects in the Notch signaling pathway, together with problems in cell adhesion, could affect the abundance of follicular cells and the integrity of the epithelium, either by intensifying cellular division (**Figure 2.4d1**) or endoreplication (**Figure 2.4d2**). In summary, we suggest that higher expression of *pointed*, increase in follicular cell number or higher genome abundance could increase *gypsy5* expression. Furthermore, we propose that the disruption of the somatic epithelium could facilitate the movement of *gypsy5* viral-like particles from the follicular cells to the oocyte.

According to our model, RAL-790 should have higher expression of *gypsy5* and *pointed* when compared to that of other DGRP lines. Although that expectation was confirmed for TE expression, the levels of *pointed* expression were not enough to explain *gypsy5* burst in activity. We suggest that, alleles affecting the activity of other proteins necessary for the activation of Pointed (such as those that mediate post-translational modifications and activation through phosphorylation)[83] could influence the activity of this transcription factor without affecting its expression. However, further analyses are needed to clarify this hypothesis.

It is also important to note that, although *ZAM* also contained *cis*-regulatory elements for Pointed, we only observed higher copy-number of *gypsy5* in RAL-

790. We suggest that the number of regulatory elements for Pointed dictates the sensitivity to changes in that transcription factor: a transposon containing more binding sites would be more sensitive to increase in Pointed expression and/or activity.

In sum, our GWAS analysis for the number of *novel* TE insertions proved to be a powerful tool not only to find natural variants associated with, but also to obtain unanticipated candidate regulators of, transposon activity. However, our data provide less insight at the mechanistic level and future studies of the allelic effects for gene expression and/or protein function are critical for our comprehension of the genetic factors affecting TE activity in natural populations.

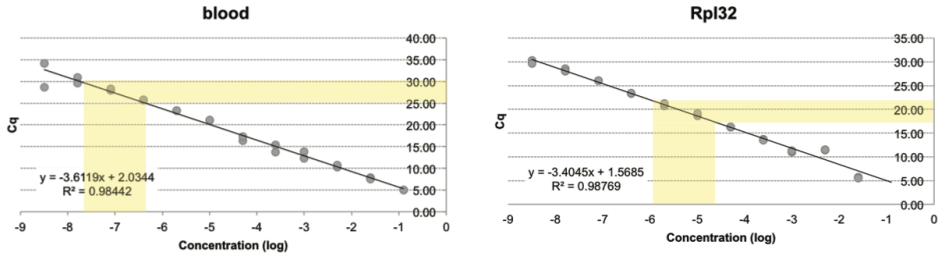
Acknowledgements

We would like to acknowledge Renato Alves for helping with bioinformatic tools and Daniel Sobral for aligning next generation sequencing reads against TE canonical sequences. Wen Huang and Michael Magwire for answering all our questions regarding the Drosophila Genetic Reference Panel, and to Oscar Ruiz and Elvira Lafuente for reading this chapter. We are also thankful to Casey Bergman for important feedback on this project.

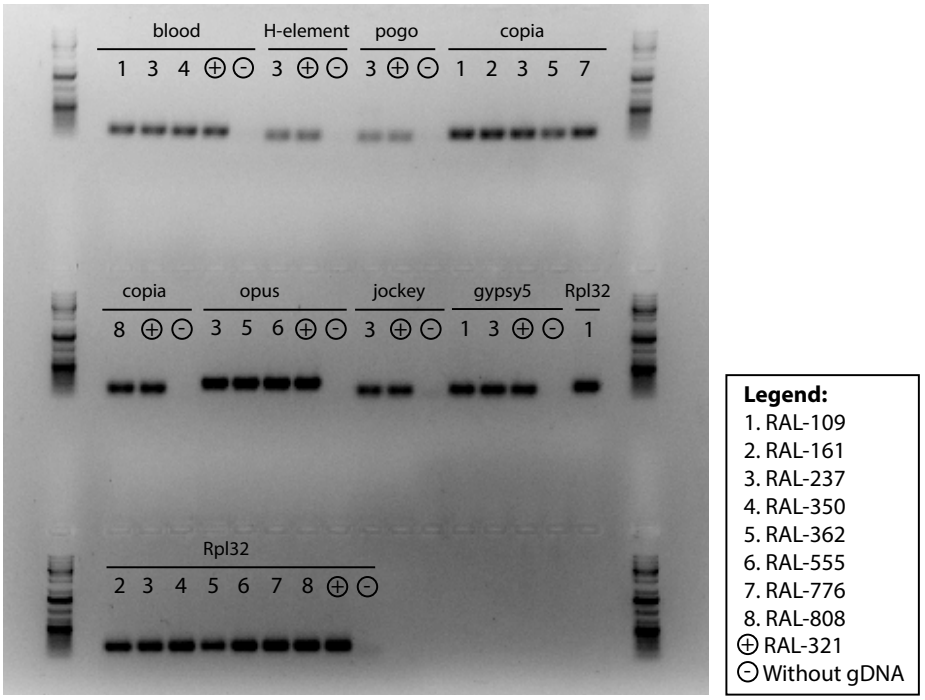
Author contributions

The author of this thesis performed some experiments to confirm TE identity, did statistical and bioinformatic analysis, and together with Patrícia Beldade designed all experiments. André Alves checked presence of false negatives, confirmed identity and size of some novel insertions, and repeated the experiment to confirm *gypsy5* and *pointed* expression. Ana Eugénio performed all other experiments.

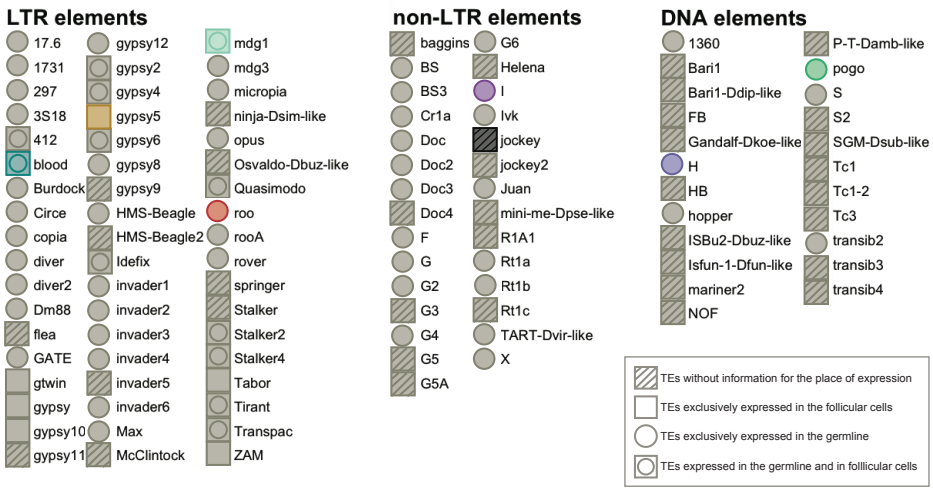
Supplementary Information



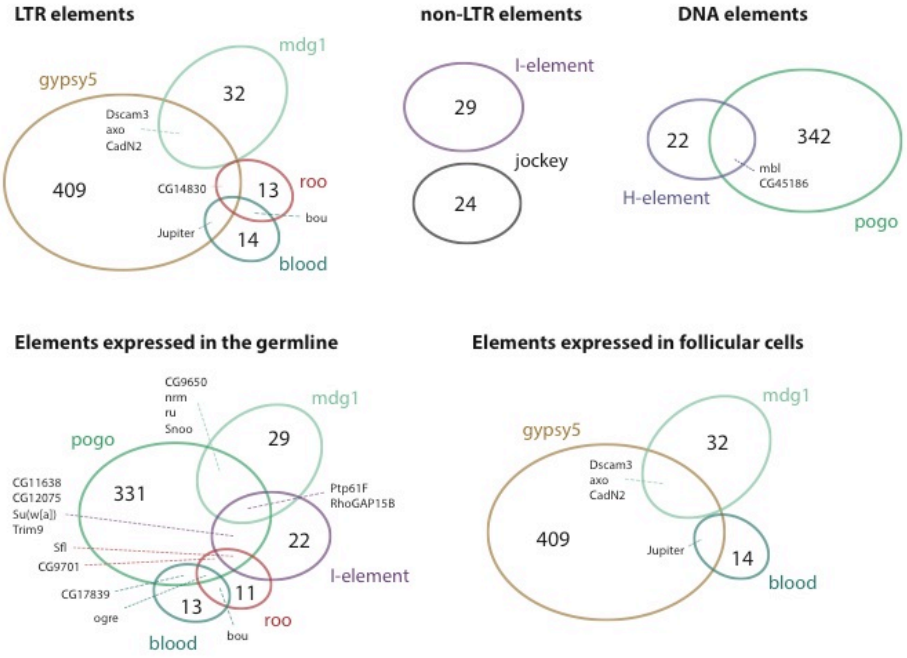
Supplementary Figure 2.1. Range of quantification cycles (Cqs) for *blood* and *Rpl32* that were used in the analysis. For each sample with known amplicon concentration we obtained the amount of cycles that were necessary to detect expression by qPCR using default threshold settings. We ran two technical replicates and checked that the relationship between the known amplicon concentrations (log) and the quantification cycles (Cq) was linear. The range of Cqs used to quantify TE or gene expression is within the linear amplification. Same methodology was used to detect expression of *gypsy5, Tbp* and *pointed*.



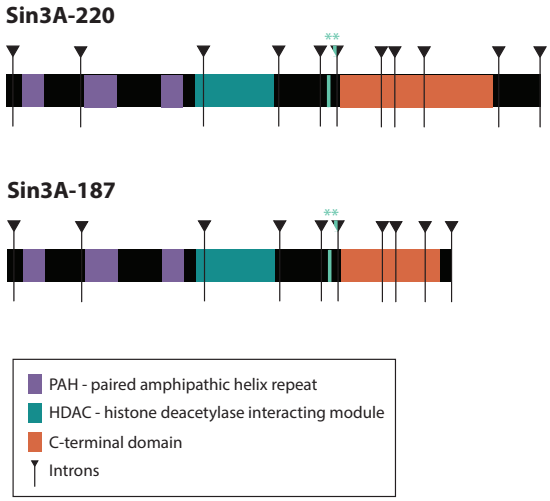
Supplementary Figure 2.2. Testing presence of false negatives in Mackay et al. predictions. We assessed the presence of blood, H-element, pogo, copia, opus and jockey insertions in DGRP genotypes that were predicted to lack those elements. To test primer efficiency we used the gDNA of RAL-321 (positive control). PCR reactions without gDNA were used as negative controls. For more information about TEs and DGRP lines tested see **Supplementary Table 2.2.**



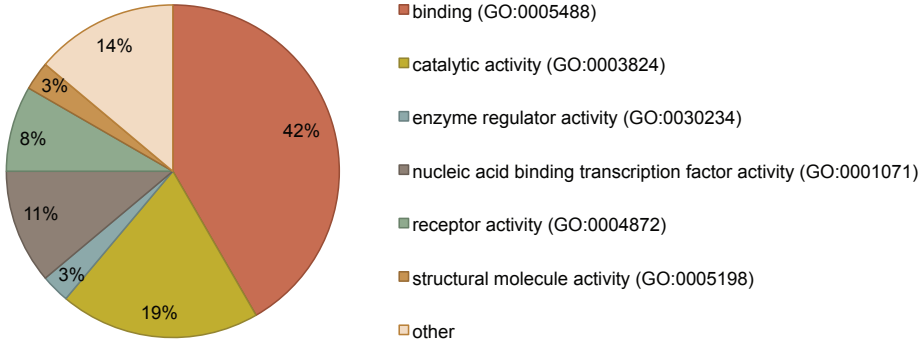
Supplementary Figure 2.3. Sequence properties and tissue-specificity of D. melanogaster transposons. All transposons found in DGRP genotypes are grouped according to their sequence properties: retrotransposons with or without LTR sequences (LTR and non-LTR elements) and DNA elements. We excluded all short interspersed nuclear elements (INE-1 and INE-1-Dbuz-like). Place of TE expression in fruit fly ovaries (expression in follicular cells, in the germline or in both cell types) was obtained from Malone et al[28]. TEs used for the GWAS analyses are highlighted with different color (blood, gyspy5, mdg1, roo, I-element, jockey, H-element and pogo).



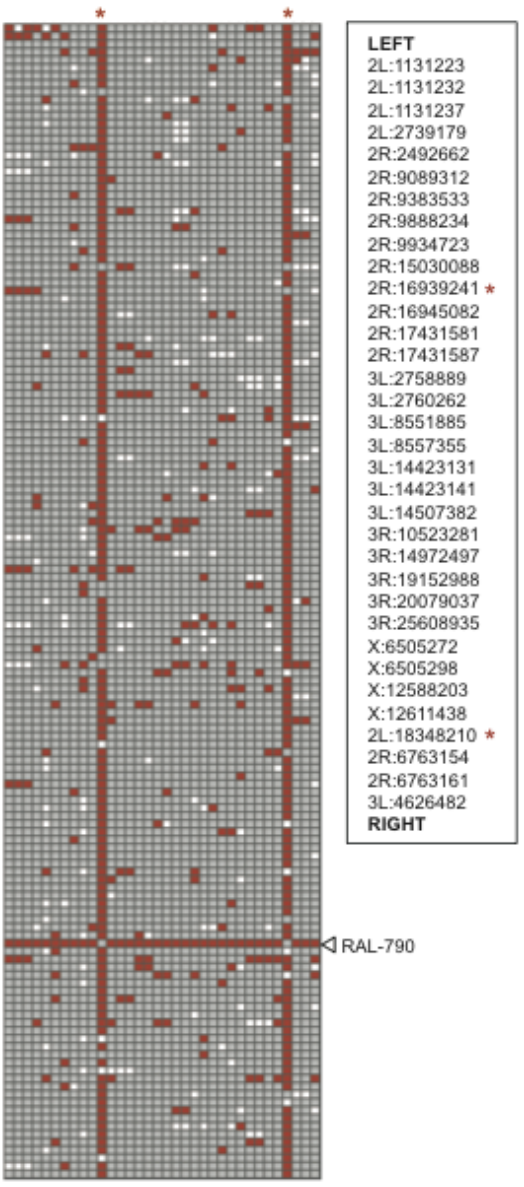
Supplementary Figure 2.4. Venn diagrams showing genes that contain alleles associated with TE activity. We grouped TEs according to their sequence properties (LTR, non-LTR and DNA elements) and place of expression in ovaries (germline and follicular cells). We only show the name of the genes that were found to have alleles associated with the activity of different transposons.



Supplementary Figure 2.5. Scheme illustrating loci in Sin3A isoforms that contain allelic variation associated with blood copy number. Scheme representing protein domains of Sin3A-220 and Sin3A-187 isoforms that were obtained using Pfam 30.0 protein database[84]. All predicted modules (paired amphipatic helix repeat, histone deacetylase interacting module and C-terminal domain) are important for Sin3A interaction with other proteins[75]. The two loci associated with blood activity that were found within Sin3A sequence are represented in light blue (*). One single nucleotide polymorphism is in coding and the other is intronic region.



Supplementary Figure 2.6. Molecular function of genes related with follicle cell development. Percentages inside chart represent the proportion of genes that share a molecular function.



Supplementary Figure 2.7. Distribution of alleles associated with gypsy5 activity that affect genes important for follicle cell development. Each column represents a locus; each line represents a DGRP genotype (RAL-790 is highlighted). Legend illustrates locus position ordered from left to right positions in the figure. Different colors in the figure represent presence of major (grey) or minor (red) allele in the corresponding DGRP genotype. An allele found in the reference genome is classified as major allele; an allele absent in the reference genome is classified as minor allele. Major alleles are more frequent in the DGRP population, alleles in the positions 2R:16939241 and 2L:18348210 (*) are exceptions of that. RAL-790 contains all alleles that are less frequent in the DGRP population. White squares illustrate positions without allelic information for the respective genotype.

Supplementary Table 2.1. Primers used to confirm identity and size of novel TE insertions.

	Position of insertion	DGRP line	Ta (ºC)	Forward primer sequence	Reverse primer sequence
Retrotransposons: LTR elements					
blood	2L:14056691..14056894	RAL-321	60	CGGGGATGTTGTCTTTCGG	TGCAATTCATCACCCCAAGT
	2R:4932328..4932331	RAL-321	58	CCTTGTCCATGAACACGGG	ACAACACTGCGGTTTCCA
	3L:1030460..1030467	RAL-321	58	TAAGTTTACCGCCAGCACG	GCGACACTTTTCCTCCTGT
	3L:22887401..22887413	RAL-321	62	AGTAAATAGTCGGGGCGGG	CGATGGCAATCCTCCATTTG
	3R:7387389..7387392	RAL-358	62	TCCAGCGGGCTCCAATTTA	CCGTCCCAGACTATATCACC
	2L:9310321..9310326	RAL-790	60	CAGCTCTCATGATGACGACG	GACCGAAATACACCGTTCCA
	2L:18683411..18683415	RAL-790	58	GACACCAGTACTTGACGCA	CAATGACCTGGCATCGCT
	2R:20107115..20107118	RAL-790	62	GACTCGAGTTCTTTACGCCA	AAGGACAAAGGATGCGGAT
	3L:5167028..5167031	RAL-790	60	TGTCGCATAAATCCACTCCG	CCCCATAATTCCCCGCAAT
	3R:7994453..7994456	RAL-790	60	GGTCAGAGCAAACCGGAAAA	CGTGTCGGCCCAGAATAT
	3R:19792545..19792549	RAL-790	60	ATCTGCTCTGGGGCCTCA	GGCCGTTGAGACTGAATTG
	3R:22107018..22107022	RAL-790	60	AGTCAGGGCATGTGGTACTA	CTAAAAGGACCTTCTGCCCC
	X:10922194..10922200	RAL-790	60	AGCATGCTTCAAAGGCAGG	AAGTGGAAGCAAGGAGTCC
copia	3R:4991260..4991264	RAL-810 RAL-812	60	TCCTCTCCCCCTCTCTGTCT	TTAAGCCCAACCACATAGCC
	3R:13161944..13161944	RAL-908	58	CACGTGTCCATAGCCCATTT	CTGCTTAACCATTGCGTCCT
	2L:4426305..4426310	RAL-908	60	CTCGAGAGTTCGGAAAGCAT	AGGACTCTGGACAGGTGGTG
gypsy5	2L:19806748..19807008	RAL-790	60	GCGTCTTAGCCATGATGAAGAA	TTGCCAAGTTTACGCTACG
	2R:15116726..15116809	RAL-790	60	GTTTTGTGTGATCGCGCTG	AGTGGGAGTTGAAGGGGTG
	3L:773447..773450	RAL-790	60	TGAACGTGAACTGAGAGTGG	CGGGCTCGCTGAATTGTC
	3L:3891219..3891222	RAL-790	58	GCTGAAAGAGGAGCCCAG	GGCAAGCTGGTGTATTGAAA
	3L:9736885..9736886	RAL-790	58	TTGGATGGGTAGGCAGTAGA	GCGTGGGAGTTATTGCGTA
	3R:2368919..2368922	RAL-790	58	TTTGCCGTTAGAAGCTGGAG	GCCACTGGACCCCACTTA
	3R:22031220..22031223	RAL-790	62	GCTAATACCGCCAGCCTTT	GATGGAGCTGGCCTTGAAA
mdg1	2L:1788665..1788852	RAL-357	60	TGACTTTCTTTGTGCGCCC	TCTCCTTAAACGTGGCTAGTG
	2L:15102151..15102178	RAL-357	60	GACCACCGTCGACTTTAAAA	GAGGAGCCGAGATCTGTC
	2R:19003049..19003053	RAL-357	60	TACGGCCTCCCCAAGTGAA	GGATATGACTGGTGCTTCTTATC
	2R:19347971..19347974	RAL-357	60	ACTTGGCCTAGGGAGCAATA	AGTTCAGTCCCAGAGGCC

Supplementary Table 2.1. Cont.

	Position of insertion	DGRP line	Ta (ºC)	Forward primer sequence	Reverse primer sequence
mdg1	3L:10095528..10095532	RAL-357	60	AGAGGCCTGTGAGATCCTTT	AAGTGTTGCCCCATTCGAA
	3L:20245853..20245856	RAL-357	60	TGCCCGTACATTTATGCAGC	CGTGTGTGTGGAGCGATC
	3R:5798180..5798183	RAL-357	60	AGGGCACACAGTTTTGGTAC	CTCGAATGGCAACCTGTACT
	3R:12117132..12117136	RAL-357	62	CCAGAACGTCGACCAACC	GTGACCTAACGCAATACACAA
	3R:23192134..23192145	RAL-357	62	CCTCAACAAGTCGGCGGG	CCCCATTCCAACGCACAA
	X:16472629..16472643	RAL-357	60	TAAAGGCGAGAGGAAAGAGC	TTTTGCCCCGTTCGAGTG
	2L:19305699..19305703	RAL-908	62	AACTCCACCGACTTGCTTTT	TGGCTCTGACATGACCTAGA
	2R:11793310..11793699	RAL-908	60	AAAGGAGCAGGAGAAGGAGT	GAGGCAAAACAACGAGCATC
	3L:676970..677090	RAL-908	60	ATATTCACGGTTTTGGCCAG	GAGGGGCCGAGTCCTAAAT
	3L:13514785..13514789	RAL-908	60	CGCGCTTTGTTAACGTGG	CATATTGGCTCGCGTTTTGT
	3R:5737883..5737886	RAL-908	60	GAACACGCCGAGGATAGC	CGAAAATCTACTCAGCCTGCT
	3R:8620803..8620806	RAL-908	60	CTGACTGAGAAATGGGGCG	TCACCAAAGTTGCAGTCGAA
	3R:16918669..16918673	RAL-908	62	ACCGGGTGCCAAGGAATG	CTTCCTCTTCGCTTCAGTGG
	3R:18045826..18045829	RAL-908	57	ACTCTACGGCAAATGAATGGG	GGCGTGACCCTTACTGAA
	X:5024909..5024914	RAL-908	60	ACAGAGAGCGACGAAAGC	CGATGCGATGTCATGTCC
opus	2L:7336324..7336327	RAL-908	60	GCATGACGATTACGTGGCTA	ACAACCAAACGCTTTTCACC
	2R:17950637..17950706	RAL-443	62	ATATGTCCTCGCCTGACCTG	GTTTCCACTGCACAGCCATA
roo	3R:17957511..17957520	RAL-358	62	CCCACATAGTTGCATGCCA	GTGTGTGGCAGTGGTAGT
	3R:7884988..7884992	RAL-358	60	CAGTGCCCCATTGTTACGA	CGAACTCATTGCAACTCCCT
	3R:17515331..17515334	RAL-358	60	TCGTTCTCTCAGCAACCAAC	AGAGAACCGAAACGCTACTG
	X:21270475..21270477	RAL-908	62	TTGGCGGGGAACAAATACAC	TGCCCCTTACCGCTCCAT
	2L:9204672..9204675	RAL-908	60	TCTTGTTTCCCCGCATCATG	TCTACCTCCGGAGCTCACA
	2L:9361897..9361901	RAL-908	60	CAACTGATGGGGTAAGCAATG	GCGAGCAAAGACAGCACTA
	2L:9705174..9705181	RAL-908	60	CGGTGCTTGAGGTGTCCC	CACTTCAAGCCTGGTTACCC
	2L:12191481..12191488	RAL-908	60	CACGACGTTAGAGCTGGAAC	CACGCATCAAGTCAGGGTT
	2L:12434248..12434266	RAL-908	60	GAATTTCGGGCTGGTTTTCA	CCATGATGTCCTGTGCTCC
	2L:12902268..12902274	RAL-908	60	GAATTTCGGGCTGGTTTTCA	CCATGATGTCCTGTGCTCC

Supplementary Table 2.1. Cont.

	Position of insertion	DGRP line	Ta (ºC)	Forward primer sequence	Reverse primer sequence
Transpac	2R:12108723..12108729	RAL-810	60	ATAAACCGCAGCAAAAGTGG	CCAATGGATTTTCGAGAGGA
	3L:12856836..12856842	RAL-908	60	CCCACTTCCTCTTCCACTCA	AGTCGACCAGGGACAATGAC
	X:345626..345630	RAL-357 RAL-892	58	TAACGATGGTGGCTGCTACA	AAGGAAAGCGATTCAAGACC
	X:15333499..15333504	RAL-443	60	CTGCAACTTTCCATGGCTTT	ACAGCTTTCCCCTTCTGGAT
Retrotransposons: non-LTR elements					
Doc	2L:11138677..11138734	RAL-357 RAL-892	60	AAAATCCATTCGGCAAACTG	TCGATCAGCGCCTAGTATCA
	3R:7873179..7873180	RAL-357 RAL-381	60	ATTGTCTGCGCAACTGTCTG	ATGAATTCGTCTGCCTGTCC
	2R:7936918..7936946	RAL-810	60	CGAAGACATCAGTCCTGCAA	CCGCTGACTGTGATTGCTAA
	3R:10301789..10301822	RAL-908	60	GCACGAGACTCACACAGGAA	TTATGGCCATTGTACGCTGA
	2L:1399467..1399478	RAL-908	60	TGCATCTGTGTGCGTATGTG	GCACTTTTTGCCTCTGTTCC
F-element	3R:6947766..6947992	RAL-810	60	TAGGCGCTGTTATTGAAACC	CAGTGAAAGTGGGTGCAAA
	3L:7996067..7996071	RAL-908	60	GGGATTTGCTCTTGCTCTTG	GCCATGGTCGAAACAAAACT
	3R:9994118..9994126	RAL-908	60	GCTTGTCAAAGGGTCCAAGA	TGTTATGTGCGCGAACTTGT
I-element	2R:10573276..10573286	RAL-357	60	TCCGTCGGCTCTTATTTGTC	CGTCTTACACTCGCAGCAAA
	2R:13861586..13861600	RAL-810	60	CCCAGATTCGCAATACCAAA	AACAAAAGCAACCACCAAGG
	3L:5068011..5068084	RAL-908	56	TCGAATTGATACAACCCCAAT	CTACTACGGCGGTGTTGGTT
	3R:11405981..11405987	RAL-357	60	GGCAGTGCAAACAAAAACAA	CTGAGGCCAAGGACTTATGC
	3R:14271683..14271693	RAL-810	56	ACCTCATAGGGGGTGCTTTT	TTGGAAGTGAAGGCTTTGAA
	2L:11388784..11388793	RAL-790	60	CTTGAAGACCGGGTACTTCT	GCTGGCATATCTTCTCCGAC
	2R:7961110..7961122	RAL-790	60	AACCCAGGGAGCTAAGTAGA	ATTGAATGTCCGGGGATCTT
	2R:12732718..12732730	RAL-790	60	GCCGAACAGCATATACCCT	TCAAGCGTGTTTCCTCGAT
	2R:14371197..14371209	RAL-790	60	TACCCGCCACTCAATTATCC	AGTATGGCGTCGAGTGTG
	3R:4277178..4277189	RAL-790	60	CTTTAAGCACCACGAGACGA	ACCCAAATGCAAAGCCGT
	3R:14671496..14671508	RAL-790	58	TAAATAAGTGCCTTCGCCCC	TGTATCTCGGCTGTCTCCA
	3R:21118597..21118611	RAL-790	62	CCGGAGCCTGGTAGTTCTT	TATCTTTCTCCCAGCCCGTC
	3R:21756397..21756400	RAL-790	60	CGCGCCTAGAACTATGCAA	ACACACTAGCAAGCACTGG
	2R:2648988..2649001	RAL-908	60	CCCCTCGTTTTCTACTGCTAC	GTGGGTGTGGAACTCTGTG
	2R:7224003..7224015	RAL-908	58	TAGCCTGCTTTTGTGTGGAG	CTTGCAAGTTGGTTTTGGGG

Supplementary Table 2.1. Cont.

	Position of insertion	DGRP line	Ta (ºC)	Forward primer sequence	Reverse primer sequence
I-element	3L:4501669..4501675	RAL-908	60	ACAGAAGTACAGTGAGCGT	GGGGAGGTTCAATTGGTCA
	3L:5068011..5068084	RAL-908	60	AGTCAATTCGCCTAGTACCAC	CAATTGGAGCTGCATCCTTT
	3L:10088380..10088381	RAL-908	60	AGGGCACTTTCCTCGAGA	ATATGCTTTGATACGGCGCT
	3L:22408430..22408442	RAL-908	60	GTCTTGACGCCTTGCCTAG	CACTGCATTTCAAACGCTCG
	3R:8071617..8071628	RAL-908	58	GAACACGCCGAGGATAGC	CGAAAATCTACTCAGCCTGCT
	3R:11192794..11192806	RAL-908	60	GCCTTCCACACGCATCTG	GATGCCCGCACTGAGAGA
	3R:22596610..22596622	RAL-908	60	ACCCAAGTTCTTGTCGCC	GCGGCGCATCAACTAATG
	3R:22850421..22850434	RAL-908	60	TCCGCTGGAGAAATTGCAT	AATCTAAAAGGGGCTGCCA
	X:302449..302464	RAL-908	60	TAAGTCACAACCCTACAGCA	TGACAGCAGTTGGGATCAA
	X:10027285..10027298	RAL-908	60	CGCTTACACTGTATTTGCCC	GCCACCGTCTCTACTTGC
Juan	2R:5572279..5572291	RAL-810	60	CTAACACGTTTCCGCCAAGT	TTCGAGGGTGTGGGTGTATT
	3L:17698435..17698491	RAL-357 RAL-761	60	TCAAGTCCCAGATGCACTCA	ATGTGGAACTTGGAGGATGC
	X:18924423..18924436	RAL-908	58	TCGAAGCCATTGCTATTTTTG	TGACACCTATTCCTCAGACTCG
	3L:16130441..16130456	RAL-443	60	CAATCGCCTAGATCGCTTGT	AGTAGCAGGTCGCCTTGAAA
DNA elements: TIR elements					
hobo	3R:10703373..10703379	RAL-357	60	CTCCCAAGGATTCTGTCCAA	AATGTTTCCCAAAGCTGACG
	3R:22133514..22133521	RAL-357	60	GGGTCTGAAAGCAGCTATGG	CATTGTTCTTGGCTGACGAA
	2R:12871101..12871113	RAL-810	60	TCAACGCTGAAAAGTATGCAA	GCAGATGATGTTGGCTTGAA
	3L:19555349..19555354	RAL-761 RAL-810 RAL-812	60	AGCTTTAGCCACAGCCACAT	GAGAGGCACGCAGGTAAGAC
	3R:22134885..22134892	RAL-908	60	CAAAGGCAGGGCTAACAAAA	CACAAGTGGGAGCATCAACA
hopper	2L:2390528..2390533	RAL-357	60	ACCCATCAGACTTCCACGAC	GGAATCGCCTACAGAAGCTG
	3R:13938897..13938901	RAL-357 RAL-812	60	TCGATTTGGCTGGAAACTCT	ATGCTGAACACGATGTGGAA
	3R:15010776..15010780	RAL-908	60	GGGTACAATCAAATCGAGCTTC	GCGAAAACTGCACTCAATCA
	X:4110151..4110442	RAL-810	60	CTTCGTTTCATTTGGCCATT	TGTGCCAAAAACACAGGCTA
pogo	2R:11767856..11767857	RAL-810 RAL-381	60	GGCTACGACATTTCCGTTGT	AACCTATTCCTTGCGGACCT
	3L:21388576..21388578	RAL-810 RAL-812	60	TTCAATACGGATTTGCCACA	GCAAAAATAAGGGCCATCCT
	3R:11292096..11292097	RAL-357	60	GTTGAGCAAACAGACCCACA	GGAGCCTCATAATCCGGTCT

Supplementary Table 2.1. Cont.

	Position of insertion	DGRP line	Ta (ºC)	Forward primer sequence	Reverse primer sequence
pogo	3R:2926760..2926761	RAL-357	60	AACTCGAATCTGGCTCGAAA	AGTGGCCTTATCGATTGGAA
	X:6015256..6015259	RAL-908 RAL-381	60	GATGTTTCGTGTGGCTGTTG	GCAGTCGCTGCAGTTTGATA
	3L:1571691..1571694	RAL-358	60	TGCACATGACTGGATTCACA	CACACGAACATTGCTCCGA
	2L:12470302..12470309	RAL-908	60	TTAGAAAGCAAGTACCGGCA	TTCTGCCATCGTGTTGCC
	2L:13770337..13770338	RAL-908	60	GTGGGGCCTCATAGATACGT	TATGCGCTGAGGTACACTTG
	2L:13780514..13780515	RAL-908	60	CCGGCCCATGTTAAGCTTTA	GAGGCAGCGGATCAATTCA
	2L:15151373..15151377	RAL-908	58	TGTTCGGGTGGTAAAAGCG	CAACGTTCCCAGGACACC
	2L:18860089..18860091	RAL-908	60	ACGCCTCGGATTTGACATC	CCGTTGGCATTTTGTGGATA
	2R:1749442..1749443	RAL-908	60	ATAGCACATTCAGCCACACG	TGCTGAATTCGGAAAGAGCT
	2R:5741587..5741594	RAL-908	60	GGTTTCGATTCGGTATTGGTTG	ATTGCTTCGTGTTAGGACCC
	2R:6426442..6426443	RAL-908	60	CTGAGTCGAGCTGGTAGGT	CGACATTTTCTGCGGCCG

Position of insertions is based on the assembly number five of *D. melanogaster* genome.
Ta stands for annealing temperature

Supplementary Table 2.2. DGRP lines with no insertions detected for *blood, copia, gypsy5, opus, jockey, H-element* and *pogo* elements.

DGRP line	*blood*	*copia*	*gypsy5*	*opus*	*jockey*	*H-element*	*pogo*
RAL-109	X	X					
RAL-161		X					
RAL-237	X	X	X	X	X	X	X
RAL-350	X						
RAL-362		X		X			
RAL-555				X			
RAL-776		X					
RAL-808		X					

Supplementary Table 2.3. Primers used to detect false negatives and for qPCR reactions.

Target	Forward primer sequence	Reverse primer sequence	Ref
blood	AACAATAGAAAGAAGCCACCGAAC	AGTCATGGACTATTGAGGGTGTTG	[73]
copia	TGCCAGAGAGCAAGTTCAGA	GCAAACCCAATTTGTCTCGT	
gypsy5	GCCCAGAGACAACGACAGAA	CTGTCTTTGCTGTCCCGGAT	
H-element	CATTAAGTCGGAAGGCCAAA	CTTGCTCTTCCGCTATCCAC	
jockey	GCGGATTAACAAGGGGCTCT	CCTGGGAGATAGATGCGCTG	
opus	CGAGGAGTGGGGAGAGATTG	TGCGAAAATCTGCCTGAACC	[85]
pogo	CCAGCGATAACGAAGAAAGC	GCTGCAAACCCATCCTTAAA	
pointed	GGGGCGTGCTGTTGTTGATG	TCGCTGGGACTGGGCTACTTC	[86]
Rpl32	ATGCTAAGCTGTCGCACAAATG	GTTCGATCCGTAACCGATGT	
Tbp	GGCAAAGAGTGAGGACGACT	GAGCCGACCATGTTTTGAAT	[38]

Ref, reference

Supplementary Table 2.4. Probabilities estimated by Mackay et al. for the identity of TE insertions.

Position of insertion	DGRP line	Most likely TE	Probabilities for LEFT reads	Probabilities for RIGHT reads
3L: 22887401..22887413	RAL-321	*blood*	*Rt1b*, 0.33 *Cr1a*, 0.28	*blood*, 0.93 *Rt1b*, 0.07
3R:7387389..7387392	RAL-358	*blood*	*blood*, 0.47 *Stalker2*, 0.42	*Stalker2*, 0.42 *1360*, 0.21
2L:14056691..14056894	RAL-321	*blood*	*blood*, 0.6 *Stalker2*, 0.39	*Stalker2*, 0.52 *INE-1*, 0.25
X:10922194..10922200	RAL-790	*blood*	*blood*, 0.6 *Stalker2*, 0.39	*Stalker2*, 0.35 *1360*, 0.23
2L:1788665..1788852	RAL-357	*mdg1*	*mdg1*, 0.58 *HMS-Beagle*, 0.19	Empty
3R:23192134..23192145	RAL-357	*mdg1*	Empty	*mdg1*, 0.86 *Transpac*, 0.13
2R:19347971..19347974	RAL-357	*mdg1*	*mdg1*, 0.45 *Cr1a*, 0.23	Empty
X:5024909..5024914	RAL-908	*mdg1*	*mdg1*, 0.83 *Transpac*, 0.12	*Idefix*, 0.79 *INE-1*, 0.21
X:21270475..21270477	RAL-908	*roo*	Empty	*roo*, 0.57 *pogo*, 0.43
2L:12191481..12191488	RAL-908	*roo*	*roo*, 0.81	*roo*, 0.55 *INE-1*, 0.29
3R:17515331..17515334	RAL-358	*roo*	*INE-1*, 0.81 *roo*, 0.17	*roo*, 0.99
3R:7873179..7873180	RAL-357	*Doc*	*Doc*, 0.96	*INE-1*, 0.49 *1360*, 0.3
3R:7873179..7873180	RAL-381	*Doc*	*Doc*, 0.97	*INE-1*, 0.69 *1360*, 0.16
3R:6947766..6947992	RAL-810	*F-element*	*INE-1*, 0.8	*F-element*, 0.9
3L:4501669..4501675	RAL-908	*I-element*	*INE-1*, 0.74 *I-element*, 0.12	*I-element*, 1
2R:6426442..6426443	RAL-908	*pogo*	*pogo*, 0.58 *roo*, 0.19	*pogo*, 0.69 *roo*, 0.12

We only include information for the insertions analyzed in this study and that contained a different TE identity than the one predicted by Mackay *et al.* Position of insertions is based on the assembly number five of *D. melanogaster* genome.

Supplementary Table 2.5. Number of *novel* insertions per TE per genotype. TEs that had more than ten *novel* insertions at least in one DGRP line (Maximum) are highlighted in violet. From those, we chose eight TEs for the GWAS analyses (highlighted in orange). This table is provided in digital format.

Supplementary Table 2.6. Alleles associated with *blood, gypsy5, mdg1, roo, I-element, jockey, H-element* and *pogo* activity. Table illustrates top alleles (single p-value<7e-05) associated with number of *novel* insertions for each TE. We obtained allelic position using the assembly number five of *D. melanogaster* genome. This table is provided in digital format.

Supplementary Table 2.7. Gene ontology (GO) analysis of genes containing alleles associated with *gypsy5* activity. GO analysis was performed to assess enrichment for certain biological processes in the group of genes containing loci with allelic variation affecting *blood* activity. Significant GO terms were obtained after Bonferroni correction for multiple testing. Enrichment for ovarian follicle cell development (GO:0030707) genes is highlighted in orange. This table is provided in digital format.

References

1. Barrón, M. G., Fiston-Lavier, A.-S., Petrov, D. A. & González, J. Population Genomics of Transposable Elements in Drosophila. *Annu. Rev. Genet.* **48,** 561–581 (2014).

2. Vicient, C. M. Transcriptional activity of transposable elements in maize. *BMC Genomics* **11,** 601 (2010).

3. González, J., Lenkov, K., Lipatov, M., Macpherson, J. M. & Petrov, D. a. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol.* **6,** 2109–2129 (2008).

4. Lisch, D. How important are transposons for plant evolution? *Nat Rev Genet* **14,** 49–61 (2012).

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

6. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128,** 1089–1103 (2007).

7. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (80-.).* **316,** 744–747 (2007).

8. Kawaoka, S. *et al.* The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* **15,** 1258–64 (2009).

9. Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E. & Blower, M. D. Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. *EMBO J.* **28,** 2945–2958 (2009).

10. Khurana, J. S. *et al.* Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell* **147,** 1551–1563 (2011).

11. Aravin, A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr. Biol.* **11,** 1017–1027 (2001).

12. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. **33,** 2052–2059 (2005).

13. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482,** 173–178 (2012).

14. McClintock, B. The origin and behaviour of mutable loci in maize. *Genetics* **36,** 344–355 (1950).

15. Aminetzach, Y. T., Macpherson, J. M. & Petrov, D. A. Pesticide

Resistance via Transposition-Mediated Adaptive Gene Truncation in Drosophila. *Science (80-.).* **309,** 764–767 (2005).

16. Daborn, P. J. *et al.* A Single P450 Allele Associated with Insecticide Resistance in Drosophila. *Science (80-.).* **2253,** 2253–2256 (2013).

17. Barroso-Batista, J. *et al.* The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.* **10,** (2014).

18. Guio, L., Barrón, M. G. & González, J. The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila. *Mol. Ecol.* **23,** 2020–2030 (2014).

19. Emera, D. & Wagner, G. P. Transposable element recruitments in the mammalian placenta: Impacts and mechanisms. *Brief. Funct. Genomics* **11,** 267–276 (2012).

20. Santos, M. E. *et al.* The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* **5,** 5149 (2014).

21. Hof, A. E. van't *et al.* The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534,** 102–105 (2016).

22. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

23. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12,** 615–627 (2011).

24. Curcio, M. J. & Garfinkel, D. J. New lines of host defense : inhibition of Ty1 retrotransposition by Fus3p. *Trends Genet.* **15,** 43–45 (1999).

25. Dai, J., Xie, W., Brady, T. L., Gao, J. & Voytas, D. F. Phosphorylation Regulates Integration of the Yeast Ty5 Retrotransposon into Heterochromatin. *Mol. Cell* **27,** 289–299 (2007).

26. Siebel, C. W. & Rio, D. C. Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. *Science (80-.).* **248,** 1200–1208 (1990).

27. Kibanov, M. V *et al.* A novel organelle, the piNG-body, in the nuage of Drosophila male germ cells is associated with piRNA-mediated gene silencing. *Mol. Biol. Cell* **22,** 3410–3419 (2011).

28. Malone, C. D. *et al.* Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137,** 522–535 (2009).

29. Meignin, C., Dastugue, B. & Vaury, C. Intercellular communication between germ line and somatic line is utilized to control the transcription of ZAM, an endogenous retrovirus from Drosophila melanogaster. *Nucleic Acids Res.* **32,** 3799–3806 (2004).

30. Touret, F., Guiguen, F. & Terzian, C. Wolbachia Influences the Maternal

Transmission of the gypsy Endogenous Retrovirus in Drosophila melanogaster. *MBio* **5,** 1–10 (2014).

31. Huang, W. *et al.* Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines Natural Variation in Genome Architecture Among 205 Drosophila melanogaster Genetic Reference Panel Lines. *Genome Res.* **24,** 1193–1208 (2014).

32. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

33. Chrostek, E. *et al.* Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis. *PLoS Genet.* **9,** 1–22 (2013).

34. Mendes, C. C. & Mirth, C. K. Stage-Specific Plasticity in Ovary Size Is Regulated by Insulin/Insulin-Like Growth Factor and Ecdysone Signalling in Drosophila. *Genetics* (2015). doi:10.1534/genetics.115.179960

35. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res.* **64,** 5245–5250 (2004).

36. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** 1–12 (2002).

37. Ponton, F., Chapuis, M.-P., Pernice, M., Sword, G. A. & Simpson, S. J. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in Drosophila melanogaster. *J. Insect Physiol.* **57,** 840–850 (2011).

38. Matta, B. P., Bitner-Mathé, B. C. & Alves-Ferreira, M. Getting real with real-time qPCR: A case study of reference gene selection for morphological variation in Drosophila melanogaster wings. *Dev. Genes Evol.* **221,** 49–57 (2011).

39. R-Core-Team. R - A language and environment for statistical computing. *https://www.r-project.org/* (2012).

40. Saha, N., Liu, M., Gajan, A. & Pile, L. A. Genome-wide studies reveal novel and distinct biological pathways regulated by SIN3 isoforms. *BMC Genomics* **17,** 111 (2016).

41. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50,** 749–61 (2013).

42. Sharma, V., Swaminathan, A., Bao, R. & Pile, L. A. Drosophila SIN3 is required at multiple stages of development. *Dev. Dyn.* **237,** 3040–3050 (2008).

43. Spain, M. M., Caruso, J. A., Swaminathan, A. & Pile, L. A. Drosophila SIN3 isoforms interact with distinct proteins and have unique biological functions. *J. Biol. Chem.* **285,** 27457–27467 (2010).

44. Aranjuez, G., Kudlaty, E., Longworth, M. S. & McDonald, J. A. On the Role of PDZ Domain-Encoding Genes in Drosophila Border Cell Migration. *Genes|Genomes|Genetics* **2,** 1379–1391 (2012).

45. Laflamme, C. *et al.* Evi5 promotes collective cell migration through its Rab-GAP activity. *J. Cell Biol.* **198,** 57–67 (2012).

46. Shvartsman, S. Y., Muratov, C. B. & Lauffenburger, D. a. Modeling and computational analysis of EGF receptor-mediated cell communication in Drosophila oogenesis. *Development* **129,** 2577–2589 (2002).

47. Gates, J. *et al.* Enabled and Capping protein play important roles in shaping cell behavior during Drosophila oogenesis. *Dev. Biol.* **333,** 90–107 (2009).

48. Baum, B. & Perrimon, N. Spatial control of the actin cytoskeleton in Drosophila epithelial cells. *Nat. Cell Biol.* **3,** 883–890 (2001).

49. Wu, V. M. & Beitel, G. J. A junctional problem of apical proportions: Epithelial tube-size control by septate junctions in the Drosophila tracheal system. *Curr. Opin. Cell Biol.* **16,** 493–499 (2004).

50. Jékely, G. & Rørth, P. Hrs mediates downregulation of multiple signalling receptors in Drosophila. *EMBO Rep.* **4,** 1163–8 (2003).

51. Petritsch, C., Tavosanis, G., Turck, C. W., Jan, L. Y. & Jan, Y. N. The Drosophila myosin VI jaguar is required for basal protein targeting and correct spindle orientation in mitotic neuroblasts. *Dev. Cell* **4,** 273–281 (2003).

52. Geisbrecht, E. R. & Montell, D. J. Myosin VI is required for E-cadherin-mediated border cell migration. *Nat. Cell Biol.* **4,** 616–620 (2002).

53. Dequier, E. *et al.* Top-DER- and Dpp-dependent requirements for the Drosophila fos/kayak gene in follicular epithelium morphogenesis. *Mech. Dev.* **106,** 47–60 (2001).

54. Genevet, A., Wehr, M. C., Brain, R., Thompson, B. J. & Tapon, N. Kibra Is a Regulator of the Salvador/Warts/Hippo Signaling Network. *Dev. Cell* **18,** 300–308 (2010).

55. Helms, W. *et al.* Engineered truncations in the Drosophila mastermind protein disrupt Notch pathway function. *Dev. Biol.* **215,** 358–74 (1999).

56. Somogyi, K. & Rørth, P. Evidence for tension-based regulation of Drosophila MAL and SRF during invasive cell migration. *Dev. Cell* **7,** 85–93 (2004).

57. Casper, A. L., Baxter, K. & Van Doren, M. No Child Left Behind Encodes a Novel Chromatin Factor Required for Germline Stem Cell Maintenance in Males But Not Females. *Development* **138,** 3357 –3366 (2011).

58. Yeh, T. H., Huang, S. Y., Lan, W. Y., Liaw, G. J. & Yu, J. Y. Modulation of cell morphogenesis by tousled-like kinase in the Drosophila follicle cell. *Dev. Dyn.* **244,** 852–865 (2015).

59. Fernández-Miñán, A., Cobreros, L., González-Reyes, A. & Martín-Bermudo, M. D. Integrins contribute to the establishment and maintenance of cell polarity in the follicular epithelium of the Drosophila ovary. *Int. J. Dev. Biol.* **52,** 925–932 (2008).

60. Heck, B. W. *et al.* The transcriptional corepressor SMRTER influences both Notch and ecdysone signaling during Drosophila development. *Biol. Open* **1,** 182–96 (2012).

61. Dodson, G. S., Guarnieri, D. J. & Simon, M. a. Src64 is required for ovarian ring canal morphogenesis during Drosophila oogenesis. *Development* **125,** 2883–2892 (1998).

62. Shcherbata, H. R., Althauser, C., Findley, S. D. & Ruohola-Baker, H. The mitotic-to-endocycle switch in Drosophila follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions. *Development* **131,** 3169–3181 (2004).

63. Van Eeden, F. & St Johnston, D. The polarisation of the anterior-posterior and dorsal-ventral axes during Drosophila oogenesis. *Curr. Opin. Genet. Dev.* **9,** 396–404 (1999).

64. Morimoto, a M. *et al.* Pointed, an ETS domain transcription factor, negatively regulates the EGF receptor pathway in Drosophila oogenesis. *Development* **122,** 3745–3754 (1996).

65. Kaminker, J. *et al.* The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol.* **3,** 1–20 (2002).

66. Linheiro, R. S. & Bergman, C. M. Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. *PLoS One* **7,** (2012).

67. Fiston-Lavier, A.-S., Barrón, M. G., Petrov, D. A. & González, J. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Res.* **43,** e22 (2014).

68. Miklos, G. L., Yamamoto, M. T., Davies, J. & Pirrotta, V. Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the beta-heterochromatin of Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2051–5 (1988).

69. Pimpinelli, S. *et al.* Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3804–3808 (1995).

70. Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res.* 1776–1785 (2009). doi:10.1101/gr.094896.109.In

71. Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461,** 1296–1299 (2009).

72. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell* **151,** 964–980 (2012).

73. Handler, D. *et al.* A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30,** 3977–3993 (2011).

74. Li, C. *et al.* Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137,** 509–521 (2009).

75. Grzenda, A., Lomberk, G., Zhang, J. S. & Urrutia, R. Sin3: Master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789,** 443–450 (2009).

76. Vieira, C. *et al.* A comparative analysis of the amounts and dynamics of transposable elements in natural populations of Drosophila melanogaster and Drosophila simulans. *J. Environ. Radioact.* **113,** 83–6 (2012).

77. Kofler, R., Nolte, V. & Schlötterer, C. Tempo and Mode of Transposable Element Activity in Drosophila. *PLoS Genet.* **11,** e1005406 (2015).

78. Gerasimova, T. I., Matjunina, L. V, Mizrokhi, L. J. & Georgiev, G. P. Successive transposition explosions in Drosophila melanogaster and reverse transpositions of mobile dispersed genetic elements. *EMBO J.* **4,** 3773–9 (1985).

79. Biemont, C., Aouar, A. & Arnault, C. Genome reshuffling of the copia element in an inbred line of Drosophila melanogaster. *Nature* **330,** 401–404 (1987).

80. Biemont, C., Arnault, C., Heizmann, A. & Ronsseray, S. Massive Changes in Genomic Locations of P Elements in an Inbred Line of Drosophila melanogaster. **77,** 485–488 (1990).

81. Kidwell, M., Kidwell, J. & Sved, J. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traitsincluding mutation, sterility and male recombination. **85,** 1973–1975 (1976).

82. Brennecke, J. *et al.* An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (80-.).* **1387,** 1387–1392 (2008).

83. Tootle, T. L. & Rebay, I. Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. *BioEssays* **27,** 285–298 (2005).

84. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44,** 279–285 (2015).

85. Specchia, V. *et al.* Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463,** 662–665 (2010).

86. Xia, Y., Midoun, S. Z., Xu, Z. & Hong, L. Heixuedian (heix), a potential melanotic tumor suppressor gene, exhibits specific spatial and temporal expression pattern during drosophila hematopoiesis. *Dev. Biol.* **398,** 218–230 (2015).

Chapter III

Wolbachia, temperature and *Drosophila melanogaster* genetic background influence transposable element activity

3.1 Abstract

Transposable elements (TEs) are important contributors to phenotypic diversification[1–3] and represent an important source of adaptive genetic variation[4,5]. However, a large proportion of insertions is eliminated due to deleterious effects[6], and hosts benefit from mechanisms that specifically target TEs in the germline, and that are able to control the production of heritable genetic variation[7–11]. Despite recent advances in the understanding of those processes, the environmental and genetic factors that trigger transposition in *Drosophila melanogaster* remain poorly described[12]. In addition, studies about the importance of biotic[13–15] and abiotic[16–18] perturbations in TE activity provided conflicting results as each assess effects in different transposons and in different genetic backgrounds. In this work, we overcame those problems by testing GxExE interactions in the activity of different transposons. In particular, we studied the effect of a biotic (*Wolbachia*) and an abiotic (temperature) factor in the expression of retrotransposons (*blood, mdg1* and *roo*) in the ovaries of four different fully sequenced, wild-derived genotypes (genotypic factors)[19]. How changes in temperature during adult life influenced TE responses to the presence of *Wolbachia* (ExE interaction) was also included in the analysis. Interestingly, we observed genotype- and TE-dependent effects of *Wolbachia* and temperature in *blood, mdg1* and *roo* expression. More specifically, the presence of endosymbiont and thermal perturbations were able to reduce, induce or confer no effect in the expression of transposons. We discuss the influence of host transcription factors as mediators of those responses.

3.2 Introduction

Transposable elements (TEs) are considered important drivers of genome evolution. They constitute a large proportion of host DNA and represent a major source of genetic variation[6,20]. Although recent studies recognized that TEs are important contributors to phenotypic diversification[1–3], and that several insertions have adaptive value[4,5], a large proportion is still eliminated due to putative deleterious effects[6].

When over-active in the germline, TEs can damage genomic integrity and lead to sterility[21]. As a consequence, hosts have evolved strategies to specifically recognize mobile elements and control transposition in germinal cells[7–11]. Although recent advances have uncovered many aspects of such mechanisms, the environmental and genetic factors that trigger jumping remain largely unknown[12]. Also, and though transposon insertions influence host adaptation to environmental perturbations[22], there is no consensus about the importance of specific stresses, biotic[13–15] and abiotic[16–18], in *D. melanogaster* transposition rates. Inconsistencies obtained for TE response to heat shock[16–18] and scarce studies on transposon activity in flies infected with natural pathogens[14] highlight the demand for further attention to this subject.

D. melanogaster genome contains several insertions of different mobile elements that can be grouped into two main classes[6,23]. Those classify TEs according to sequence properties, and differentiate retrotransposons from DNA elements[23]. Retrotransposons move via an RNA intermediate, and share structural commonalities with retroviruses; DNA elements move via a non-replicative *cut-and-paste* mechanism, and are usually flanked by terminal inverted repeats[24]. They both express proteins that are crucial to complete each cycle of transposition, and to mediate integration in host genome[25]. DNA elements code for a transposase protein, while retrotransposons and retroviruses usually express a viral-particle coat, reverse transcriptase, ribonuclease H and integrase[25]. Traditionally, retrotransposons are divided into two main groups based on the presence or absence of long terminal repeats (LTRs). In *D. melanogaster*, most copies belong to the LTR group[6] that can contain *cis*-regulatory elements within 5' LTR or TE open reading frames[22]. Those regulatory elements are necessary to trigger expression of transposons, and are typically recognized by host transcription factors[26–28]. Interestingly, experiments done in plant species proved importance of such elements in mediating TE response to specific environmental stresses[22,26–30]. This idea was never empirically tested in metazoans.

Here, we propose to study the effect of the interaction between two environmental factors (one biotic and another abiotic) in the activity of transposons of *D. melanogaster* ovaries. For the biotic factor, we tested the importance of *Wolbachia* – an obligatory intracellular bacteria that infects several arthropod and nematode species[31–33]. The study of the effect of this maternally transmitted endosymbiont in the activity of transposons is of particular interest given 1) *Wolbachia*'s ability to increase host survival to RNA viral infection[34,35] and 2) the structural commonalities between retroviruses and retrotransposons. In accordance with these two observations, we hypothesize that the presence of *Wolbachia* in *D. melanogaster* ovaries could reduce retrotransposon expression, similar to that observed for viral load[34]. Also, and given that some *Wolbachia*-mediated host phenotypes (such as male killing in *D. bifasciata*[36]) are sensitive to temperature, we decided to further test to which extent thermal perturbations during adult life influenced *Wolbachia*-induced TE responses. The expression of three retrotransposons (*blood, mdg1* and *roo*) was assessed in ovaries collected from four fully sequenced, wild-derived genotypes. We observed genotype- and TE-dependent effects of *Wolbachia* and temperature in *blood, mdg1* and *roo* expression. More specifically, the presence of endosymbiont and thermal perturbations were able to reduce, induce or confer no effect in the expression of transposons. We further suggest that host transcription factors may mediate those responses.

3.3 Materials and Methods

Drosophila melanogaster *stocks and maintenance*

Oregon-R stock was obtained from Rui Martinho laboratory, and was naturally cleared of *Wolbachia* infection. Flies from the Drosophila Genetic Reference Panel (DGRP)[19] (RAL-021, RAL-237, RAL-321 and RAL-790) that were naturally infected with the same strain of *Wolbachia (wMel-like* genotype[37]), were purchased from Bloomington stock center. To obtain genetically identical genotypes infected or not with *Wolbachia*, those flies were treated with tetracycline (0.05mg/ml) for two generations, and gut flora content homogenized

as previously described[38]. Flies were kept in standard cornmeal food[39] at 25°C, 60% humidity and 12h day-dark cycles, unless otherwise mentioned.

Temperature treatments

To assess TE response to temperature change at different time points, we collected 0-8h old Oregon-R flies and kept five females with two males for three days at 25ºC. Males were then removed and females placed at 25ºC (control condition) or 29ºC (experimental condition). Ovaries were dissected at day one, two, three and four after temperature change, for RNA extraction. We pooled eight pairs of ovaries per sample, four replicates per condition.

To test the effect of *D. melanogaster* genetic background in TE response to temperature and *Wolbachia*, we collected 0-8h old RAL-021, RAL-237, RAL-321 and RAL-790 flies infected or not with endosymbiont. For each genotype and *Wolbachia* status, five females were kept with two males for three days, at 25ºC. Males were then removed and females placed at different temperatures, 25ºC for control, 21ºC and 29ºC for experimental conditions. Ovaries were dissected at day four after thermal change for RNA extraction. We pooled eight pairs of ovaries per sample, six to eight replicates per condition.

RNA extraction and cDNA synthesis

Ovaries were dissected in fresh cold PBS 1x, and directly stored at 4ºC in 400µl of Trizol (Ambicon) until homogenized by pestels. Tissue collection and homogenization was done in the same day; samples were stored at -80ºC until RNA extraction. We polled eight pairs of ovaries per sample. For RNA extraction we used Direct-zol RNA Miniprep kit (Zymo Research) and total RNA was eluted in 25µl of RNase-free water (Sigma). RNA purity and concentration of each sample was measured with Nanodrop ND-1000 Spectrophotometer; A600/A800 absorbance was around two for all samples, and concentrations ranged from 200 to 800ng/µl. All RNA samples of the same genotype were processed on the same day.

For cDNA synthesis we removed contaminations of gDNA from 1μg of RNA using DNAse (Promega) treatment. For cDNA synthesis, we followed Reverse Transcription System (Promega) protocol, using 0.02μM Oligo dT primers. Reaction was incubated at 42°C for 60 min and heated at 95°C for 5 min. cDNA was diluted in RNAse-free water (1:10) and stored at -20°C until quantitative real-time PCR (qPCR).

qPCR

We measured levels of mRNA expression by qPCR in CFX384 thermal cycler (BioRad). For each reaction we used 5μl iQ™ SYBR® Green supermix (BioRad), 1μl of 4μM primers and 4μl of diluted cDNA (1:10). Primer sequences are described in **Supplementary Table 3.1**. We run two technical replicates per sample, per plate. Thermal cycling protocol was the same for all reactions: 2 min at 50°C; 10 min at 95°C; 40 cycles of 95°C for 30 sec, 60°C for 1 min and 72°C for 30 sec. We used default threshold settings to obtain the quantification cycle (Cq) for each reaction. qPCR melting curves were analyzed to confirm specificity of amplified products and negative controls never showed detectable amplification.

To test the dynamic of TE expression in Oregon-R ovaries upon exposure to 25ºC and 29ºC, we used Pfaffl method[40] that corrects expression for the efficiency of each primer pair. We used *Rpl32* as reference gene, given that the values obtained for expression stability using NormFinder[41] and geNorm[42] algorithms were better for that gene (*Rpl32:* NormFinder=0.04, geNorm=0.60; *Tbp:* NormFinder=0.05, geNorm=0.61). Fold-change of expression was determined for each TE, using as normalizer the average relative expression obtained for samples collected in day one, and kept at 25ºC.

To assess levels of TE and gene expression for DGRP samples, we ran in each plate standard curves (one standard curve for each gene or TE tested) with serial dilutions of known amplicon concentrations. To obtain amplicons for standard curves, we amplified target regions by PCR using 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2, 0.4μM primers (**Supplementary**

Table 3.1) and 1μl of cDNA (1:10). PCR products were cleaned using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), eluted in 15μl of buffer NE, and concentrations measured in Nanodrop ND-1000 Spectrophotometer. The minimum amount of nucleic acids that can be detected by qPCR for each primer pair was estimated by extending dilutions of our standards until the relationship between the known amplicon concentrations and the quantification cycles (Cqs) reached a plateau. Samples that had Cq values outside the linear detection and with standard deviation above 0.5 for the technical replicates were eliminated from the analysis. To obtain levels of expression we transformed Cq values into absolute target concentrations using respective standard curves. This methodology allowed us to control for plate effects, and for differences in primer efficiency. For the analysis of expression data, we used NormFinder[41] and geNorm[42] algorithms to check expression stability of two reference genes commonly used to normalize gene expression at different environmental conditions and genotypes, *Rpl32* and *Tbp*[43,44]. Stability values obtained for *Rpl32* (RAL-021: NormFinder=0.26; geNorm=0.36; RAL-237: NormFinder=0.24; geNorm=0.22; RAL-321: NormFinder=0.07; geNorm=0.19; RAL-790: NormFinder=0.03; geNorm=0.13) and *Tbp* (RAL-021: NormFinder=0.33; geNorm=0.33; RAL-237: NormFinder=0.15; geNorm=0.20; RAL-321: NormFinder=0.09; geNorm=0.21; RAL-790: NormFinder=0.08; geNorm=0.13) were similar in both algorithms. We then decided to use *Tbp* as control gene for RAL-021, and *Rpl32* for RAL-237, RAL-321 and RAL-790 as those seem to be more stable for the respective DGRP lines (**Supplementary Figure 3.1**).

gDNA extraction

gDNA from RAL-021, RAL-321 and RAL-790 male adults was extracted using DNeasy Blood & Tissue Kit (Qiagen). Tissues were homogenized in 180μl of buffer ATL, using pestels. We used 200μl of buffer AE to elute gDNA and treated all samples with 0.2mg/ml of RNAse (Invitrogen) for 20 min at 55ºC. gDNA concentration was measured in Nanodrop ND-1000 Spectrophotometer, and samples stored at -20°C until nested PCR.

Nested PCR

We designed primers for the flanking region of *blood* or *roo* insertions predicted within *CG31751, CG10073, CG31974, CG8920, CG4374, CG5440, CG11319, CG2022, CG5681, CG17883, CG33966, CG5210, CG9194, pot, Sema-1a, Act42A, Dg, Taf12, Fas2, Topors, tws* and *dnc* genes (**Supplementary Figure 3.2A**). Primer sequences are described in **Supplementary Table 3.2**. We obtained amplicon by longPCR (Roche): 4ng of RAL-021, RAL-321 or RAL-790 gDNA, 0.5μM primers, 2% DMSO, 1x buffer, 0.5mM dNTPs, and 0.21μl of enzyme per 15μl of reaction. We used standard amplification program for longPCR with 10 min of elongation time, and 60ºC as annealing temperature. Size of insertion was obtained by electrophoresis, and 1μl of the product of longPCR (1:100) used as template for a second PCR reaction: 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2 and 0.4μM of primers for *blood* or *roo* (**Supplementary Table 3.1**). This methodology allowed us to confirm whether *blood* or *roo* insertions that were predicted to have landed inside genes and that were putatively expressed together with protein-coding mRNAs contained sequences detected by qPCR.

Statistical analyses

All statistical analyses were done in R (version 3.0.3).

To assess the effect of temperature and time of exposure (categorical fixed variables) in TE activity, we used linear models (lm) with lognormal distribution (Oregon-R dataset). Interaction between temperature and time was removed if not contributing to explain differences in expression. We tested the significance ($p<0.05$) for the effect of temperature at each time point by using least-squares means (lsmeans). When time of exposure influenced TE expression, regardless of the effect of temperature, we obtained an estimate and significance for that effect, using same models described above and including time of exposure as continuous fixed variable.

To test for the effect of *Wolbachia* and temperature (categorical fixed variables) on relative TE and gene expression (dependent variables) in RAL-021, RAL-237,

RAL-321 and RAL-790 ovaries, we used lm with lognormal distribution. When the interaction between those two factors (*Wolbachia* and temperature) was statistically significant we used lsmeans to test for the effect of *Wolbachia* at each temperature. Each transposable element, gene and genotype was analyzed separately.

Characterize place of TE insertions

We extracted available information for *blood, mdg1* and *roo* insertions of RAL-021, RAL-237, RAL-321 and RAL-790 that were deposited online (http://www.hgsc.bcm.tmc.edu/projects/dgrp/[19]) and used Variant Effect Predictor tool[45] (Ensembl) to further characterize insertions according to their genomic position (e.g. coding vs. non-coding). For those that landed inside or in the neighbor regions of gene sequences, we run a gene ontology (GO) analysis to assess enrichment for insertions affecting genes with particular molecular functions. We used GO consortium algorithm, and Bonferroni correction for multiple testing[46].

Prediction of transcription factor binding sites

We searched for transcription factor binding sites in *blood, mdg1* and *roo* canonical sequences using Consite[47] and TFBIND[48] algorithms. We only considered *cis*-regulatory elements that were predicted by both programs, as previously described[49]. As Consite and TFBIND use databases for mammalian binding sites, we only show transcription factors that are known to recognize conserved *cis*-regulatory elements.

3.4 Results

Wolbachia *status and temperature perturbations influence TE activity in a genotype-dependent way*

In order to assess the dynamic of TE response to temperature, we exposed adult females to 29ºC (and compared them with control females that were maintained at 25ºC) for different periods of time. We observed that TE

expression was affected by temperature, and that those effects were dependent on the time of exposure and on the TE analyzed regardless of sequence properties (**Supplementary Figure 3.3**). Given that more TEs responded to temperature after four days of exposure, we then decided to use this time point to analyze the effect of *Wolbachia* on TE activity of four genotypes upon exposure to different temperatures (21ºC, 25ºC and 29ºC).

Interestingly, the dynamic of TE response to environmental factors (biotic and abiotic) was specific for each mobile element and genotype analyzed (**Supplementary Figure 3.4** and **Figure 3.1**). Although the expression of *blood, mdg1* and *roo* in one genotype (RAL-790) was robust to environmental perturbations, it responded to one or both factors for the other genetic backgrounds analyzed (RAL-021, RAL-237 and RAL-321). Also, and when there was an effect, mobile elements of the same genotype responded in the same way to each environment (*Wolbachia* or temperature) but not between them. In particular, we observed three types of responses that were genotype-, TE- and environment-specific, and included decrease, increase and no effect in TE expression (**Figure 3.1**). Only for one of the genotypes analyzed (RAL-321) we have observed *Wolbachia*-mediated responses in TE expression that were dependent on temperature (**Supplementary Figure 3.4**).

To understand to which extent environmental factors also affect oocyte progression, we assessed the expression of *CG15279* that is mostly expressed in later stages of oocyte development[50] (**Figure 3.2A**). We observed genotype-dependent effects of *Wolbachia* status and temperature perturbations in the expression of *CG15279* (**Figure 3.2A** and **Supplementary Figure 3.5**). In particular, for most genotypes (RAL-237, RAL-321 and RAL-790) we detected effects of both *Wolbachia* and temperature, and for only one genotype (RAL-021) we have observed an effect for the interaction between these two factors.

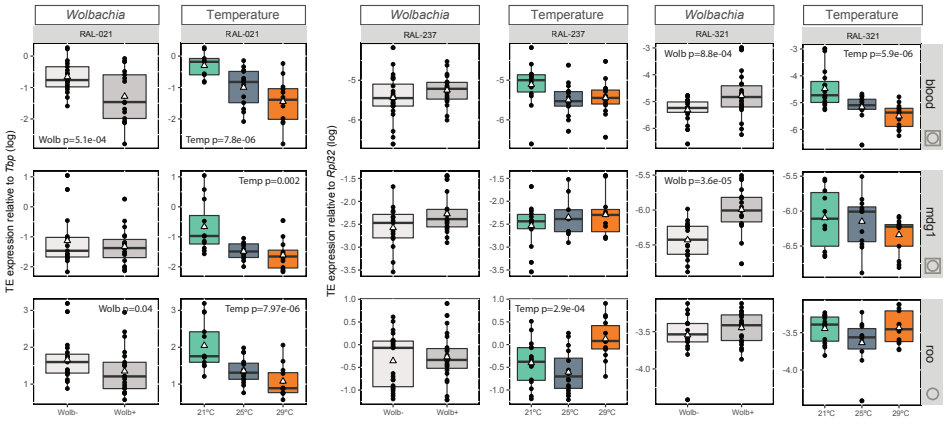


Figure 3.1. Effect of *Wolbachia* and temperature on TE expression of RAL-021, RAL-237 and RAL-321. Relative expression of *blood, mdg1* and *roo* retrotransposons, in RAL-021, RAL-237 and RAL-321 ovaries. We tested the effect of *Wolbachia* and temperature (categorical fixed variables) in relative TE expression using linear models with lognormal distribution. Data for each TE and genetic background was analyzed separately; statistically significant factors are shown in boxplots with respective p-values. Data for RAL-790 is not included in boxplots as there was no effect for the presence of *Wolbachia*, temperature perturbation or interaction between both (**Supplementary Figure 3.4**). Each circle in the boxplots represents a pool of eight ovary pairs; triangles correspond to the mean and black line to the median. *Blood* and *mdg1* are expressed in the germline and somatic tissue of ovaries whereas *roo* is only expressed in the germline (see **Supplementary Figure 2.3** for more details).

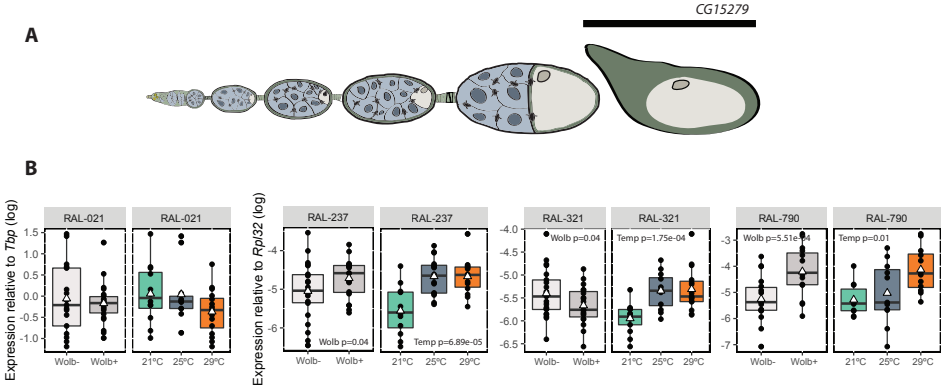


Figure 3.2. Effect of *Wolbachia* and temperature on ovary development. To test the effect of *Wolbachia* and temperature in oocyte progression we assessed the expression of *CG15279*, a gene restricted to later stages of oocyte development (black box)[50] **(A)**, in

the ovaries of RAL-021, RAL-237, RAL-321 and RAL-790 flies **(B)**. The effect of *Wolbachia* status and temperature perturbation (categorical fixed variables) was tested using linear models with dependent variable (expression relative to *Rpl32*) in logarithmic scale. Data for each genetic background was analyzed separately; statistically significant factors are shown in boxplots with respective p-values. Each circle in the boxplots represents a pool of eight pairs of ovaries; triangles correspond to the mean, and black line to the median.

Places of blood, mdg1 *and* roo *insertions*

To check whether the quantification of *blood, mdg1* and *roo* expression by qPCR could be influenced by the detection of gene-derived mRNAs including TE insertions we characterized TE integration sites in RAL-021, RAL-237, RAL-321 and RAL-790 genotypes. Despite the general enrichment of insertions in intergenic regions, upstream or downstream of a gene, and within intronic sequences (**Figure 3.3** and **Supplementary Table 3.3**), we were still able to detect several insertions in places that will be present in processed mRNAs (insertions in 5' or 3' UTR and in exons). Those included several *roo* insertions in three DGRP lines (RAL-021, RAL-321 and RAL-790) and one *blood* insertion in RAL-790 (**Figure 3.3**). For those TE insertions present in mature mRNAs we found that 70.6% were within genes upregulated in response to environmental perturbations (modENCODE treatment expression data: *CG8920, pot, Act42A, Dg, Taf12, CG17883, Fas2, Topors, tws, dnc, CG5210* and *CG33966*), and 47.1% were inside genes expressed in ovaries (modENCODE tissue expression data: *CG8920, Act42A, Dg, Taf12, CG17883, Topors, tws* and *CG5210*). Although most of those insertions were of incomplete transposons, all of them contained sequences that were amplified by primers used to quantify expression of transposons (**Supplementary Figure 3.2B**).

We further assessed whether *blood, mdg1* and *roo* preferentially insert in regions containing genes that share a particular function and found enrichment for insertions near or inside neuropeptide (fold enrichment=8.43, p-value=5.92e-04) and G-protein coupled receptors (fold enrichment=7.9, p-value=1.05e-03) activity (**Supplementary Table 3.4**).

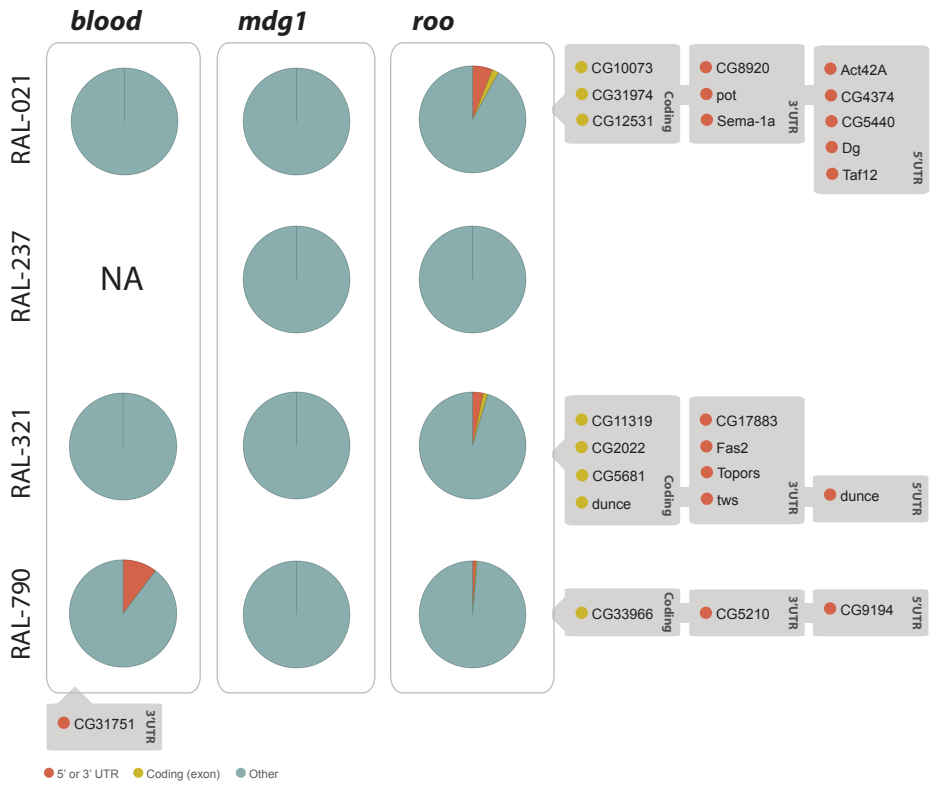


Figure 3.3. Place of *blood, mdg1* and *roo* insertions in RAL-021, RAL-237, RAL-321 and RAL-790. Pie charts represent proportion of *blood, mdg1* and *roo* insertions, in RAL-021, RAL-237, RAL-321 and RAL-790 genotypes, that are inserted in regions excluded from mature mRNAs (intergenic regions, upstream or downstream of a gene, introns, and non-coding transcripts). Those are grouped in pie charts as "Other" category. Pie charts also include proportion of TE sequences present in processed mRNAs (insertions in 5' or 3' UTRs and in exons) of several genes. We included in the analysis all insertions predicted for *blood* (number of insertions: RAL-021=8, RAL-321=5, RAL-790=9), *mdg1* (number of insertions: RAL-021=12, RAL-237=1, RAL-321=10, RAL-790=7), and *roo* (number of insertions: RAL-021=77, RAL-237=4, RAL-321=90, RAL-790=66) elements. No *blood* insertions were detected in RAL-237 (NA).

Predictions of transcription factor binding sites in blood, mdg1 *and* roo *sequences*

As regulatory elements within TE sequences are able to influence TE expression in response to environmental factors[26–28] we decided to search for the *cis*-

regulatory elements within *blood, mdg1* and *roo* canonical sequences that are recognized by different transcription factors. Interestingly, the number and nature of such regulatory elements greatly differed between retrotransposons. We detected binding sites for CREB, Dorsal, E2F1, Fork head, Pebbled, Cap-n-collar, PAX-6, Slow border cells and Seven up transcription factors (**Table 3.1**). All of those, excepting *PAX-6*, are induced in response to specific environmental stimuli (**Supplementary Figure 6A**), while only *Dorsal, E2F1, Fork head, Pebbled,* and *Cap-n-collar* are expressed in ovaries (**Supplementary Figure 6B**). We further propose that the transcription factor Cap-n-collar could be influencing *bood* and *roo* expression in response to *Wolbachia* status, whereas E2F1 and/or Fork head would affect *blood, mdg1* and *roo* response to temperature in RAL-021 (**Figure 3.4**).

Table 3.1. Number of transcription factor binding sites found in *blood, mdg1* and *roo* sequences.

Transcription factor	*blood*	*mdg1*	*roo*	Function
CREB	5	1	2	Dorsal/ventral pattern formation
Dorsal	-	9	18	Response to stress
E2F1	8	7	8	Cell proliferation, ovary follicle cell development, apoptosis, response to external biotic stimulus
Fork head	66	50	74	Defense response to virus
Pebbled	-	1	2	Tissue homeostasis
Cap-n-collar	2	-	6	Oxidative stress
PAX-6	1	-	-	Eye development
Slow border cells	3	7	-	Regulation of JAK-STAT cascade
Seven up	10	2	5	Nervous system development, response to stimulus

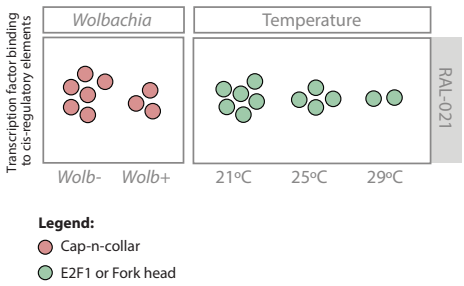


Figure 3.4. Model proposed to explain TE expression in RAL-021. In our model, the binding of the transcription factor Cap-n-collar to the respective *cis*-regulatory elements would be mainly affected by *Wolbachia* status regardless of the temperature (less binding in *Wolbachia* positive ovaries). The binding of E2F1 or Fork head would be mainly affected by temperature perturbations (less binding with increase in temperature).

3.5 Discussion

Effects of Wolbachia *and temperature in TE expression are dependent on* Drosophila *genotype*

To study the ability of retrotransposons to respond to biotic and abiotic environmental factors we obtained four *D. melanogaster* genotypes, each with different *Wolbachia* status, and exposed them to thermal perturbations. We generally observed TE-, genotype- and environmental-specific effects in *blood, mdg1* and *roo* expression. However, and given that environmental perturbations can affect many aspects of oocyte progression[51,52], those responses could be associated with developmental problems in the ovaries. Although for RAL-237 and RAL-321 genotypes we could not exclude the hypothesis that environmentally mediated TE responses derive from problems in oocyte development (detected by effects in the expression of *CG15279*), we did so for RAL-021 that will be the focus of our discussion.

To explain *blood, mdg1* and *roo* expression in RAL-021 we checked the presence of *cis*-regulatory elements within their canonical sequences that may be critical for retrotransposon response to environmental stressors. The regulatory elements previously found in TE sequences are often recognized by host transcription factors that are able to regulate transposon expression in an environmental specific manner[27–30,53,54]. Here, we proposed that *blood* and *roo* response to the presence of *Wolbachia* in RAL-021 ovaries might be mostly mediated by Cap-n-collar, and that *blood, mdg1* and *roo* expression upon thermal perturbations would be influenced by E2F1 or Fork head activity. However, this model still needs to be validated and further analysis for the activity of those transcription factors in the female gonads upon environmental challenges would help to clarify their function in modulating TE activity.

A previous genome-wide study reported that genomic differences between DGRP lines not only include variation in single nucleotide polymorphisms, but also in TE insertions[19]. Given that complex chromatin-remodeling modifications occur in response to environmental challenges[55], we cannot ignore the impact that those might have had in affecting the expression of TEs that are inserted in

different genomic places, specially when in regions that become more accessible upon stress. The enrichment of *blood, mdg1* and *roo* insertions near or inside genes related with G-protein coupled receptor activity, that transduce environmental signals and whose expression can be enhanced by external stimuli[56–58], is suggestive that chromatin reorganization in response to stress might have contributed to influence the expression of those elements. Accordingly, we propose three not mutually exclusive scenarios where TE activity can be affected by changes in 1) transcription factor expression, 2) transcription factor binding affinity to respective *cis*-regulatory elements, and/or 3) chromatin organization. Finally, we suggest that the genetic background can affect the relative contribution of each of those processes to mediate TE-specific responses to environmental challenges.

Environmental-mediated expression of* roo *in RAL-021 can be affected by gene-derived mRNAs

In agreement with a previous study reporting over-representation of TE insertions in non-coding regions[59], we observed an enrichment of *blood, mdg1* and *roo* in intergenic and intronic sequences. However, many transposons were still inserted in genomic locations that, when transcribed, are included in processed mRNAs of genes[60]. As a consequence, detection of transposon expression by qPCR could incorporate both, TE-specific and gene-derived mRNAs. To exclude any ambiguity due to insertions in such regions, we have taken advantage of the available information for TE insertions in the DGRP lines, and characterized *blood, mdg1* and *roo* insertions in RAL-021, RAL-237, RAL-321, and RAL-790 genotypes. We observed that only *blood* and *roo* were inserted in places likely to be co-expressed with host genes. Given that those insertions contained sequences amplified by primers used in our quantification for TE expression, we cannot exclude the hypothesis that our data for *roo* activity in RAL-021 could be affected by changes in the expression of some genes.

Effect of* Wolbachia *in TE expression and viral load may be related

Wolbachia is an endosymbiont able to spread in natural populations of *D. melanogaster*[61], and to confer protection against viruses[34,35]. We hypothesized that the presence of this endosymbiont in *D. melanogaster* ovaries could also reduce retrotransposon expression, similar to that observed for viral load[34]. Although the mechanism underlying the tripartite interaction between host, *Wolbachia* and virus is still largely unknown, recent studies have uncovered a possible association between *Wolbachia*-induced oxidative stress and its ability to promote host survival upon viral infection[62,63]. It is the balance between the production and destruction of reactive oxygen species (ROS) that reflects the degree of oxidative stress[64]. As one possibility, we hypothesize that *Wolbachia* may be able to reduce host ability to establish antioxidant defences through negative effects in transcription factors that promote cellular responses to oxidative stress resulting in increased levels of ROS.

Interestingly, *blood* and *roo* expression in RAL-021 ovaries followed that of our prediction and could be associated with presence of *cis*-regulatory elements for a transcription factor that mediates cellular responses to oxidative stress (Cap-n-collar)[65]. This observation suggests that TE and viral responses to *Wolbachia* status may share important players possibly related to the induction of oxidative stress.

Environmental perturbation as a driver for the production of genetic variation

Insertions of mobile elements influence host adaptation and response to environmental perturbations[22]. Accordingly, it has previously been suggested that, when exposed to stressful conditions, organisms could benefit from the production of novel adaptive genetic variation through the activation of transposons[66]. Here, we proposed to explore the effect of biotic and abiotic factors in *D. melanogaster* TE activity, using expression in the germline as a proxy for transposition rates. Also, and given the replicative properties of retrotransposons, we suggest that increased TE activity in the germline can

culminate in higher mutation burden that is transmitted to the next generation. Interestingly, we observed that the ability to produce heritable genetic variation is not only dependent on the environment, but also on the genetic background. Although some genotypes were robust to environmental changes and had no potential to trigger environmentally mediated TE activity (RAL-790) or reduced TE expression in response to perturbation (RAL-021 response to *Wolbachia* status), we observed that one DGRP line (RAL-021) was able to promote *blood, mdg1* and *roo* activity in response to thermal challenges (increase in TE expression with decrease in temperature). This study provides evidence that some environmental perturbations, but not all, may promote production of heritable genetic variation that can increase populations' ability to adapt to new environments (evolvability).

In summary, the current study shows that the regulation of TE activity can be very complex, as genotypic and environmental factors greatly influence the activity of retrotransposons. Additionally, specific time-dependent responses to external stimuli were also observed. We consequently propose that differences in the *D. melanogaster* genetic background, time of exposure to stress, age of adults and TE analyzed, could be at the basis for the previous inconsistencies found in TE responses to heat shock[16–18].

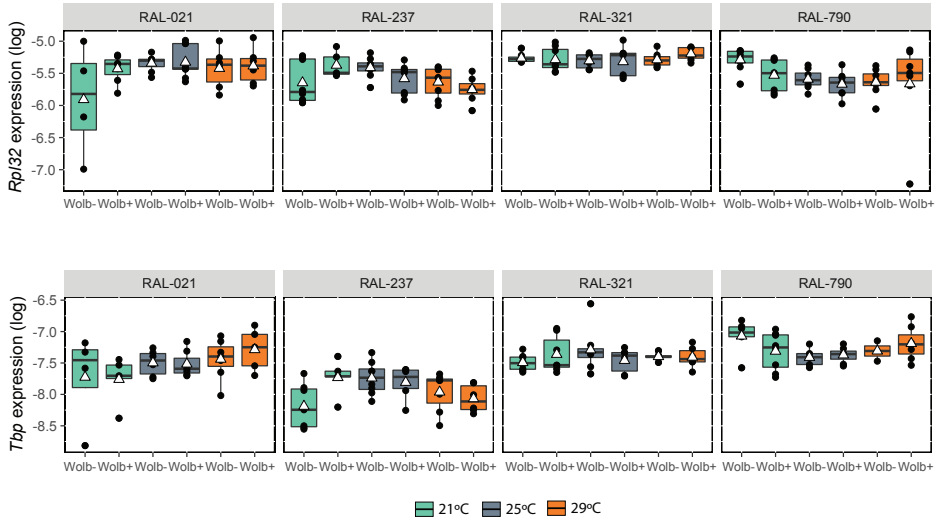
Acknowledgements

DGRP lines were obtained from Bloomington fly stock center, and Oregon-R flies from Rui Martinho's lab. The authors thank to Alisson Gontijo for providing primers against CG15279, and members of Alisson Gontijo, Chisten Mirth and Élio Sucena laboratories for suggestions and important discussions. We would also like to acknowledge Oscar Ruiz for reading this chapter.

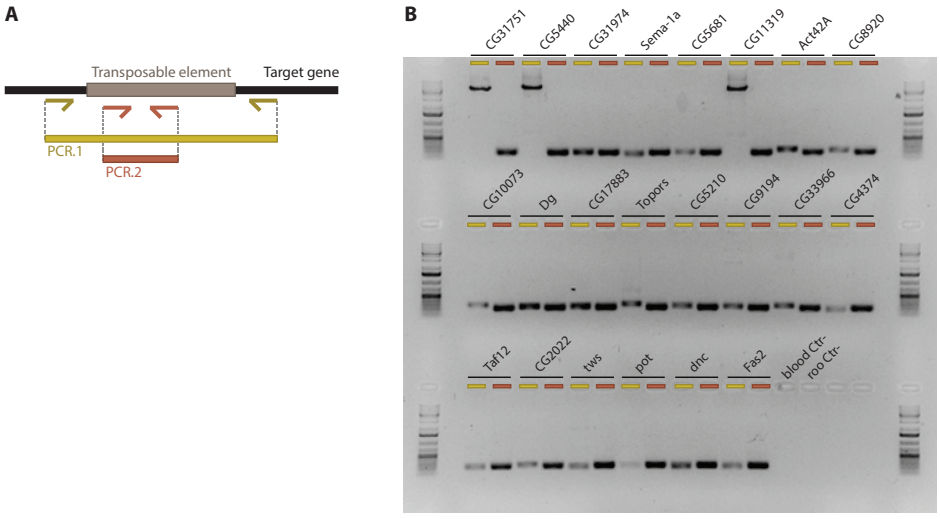
Author contributions

The author of this dissertation obtained data for Oregon-R, did statistical and sequence analysis, and wrote this chapter. Ana Eugénio obtained data for DGRP lines; Marta Marialva and Patrícia Beldade designed all experiments.

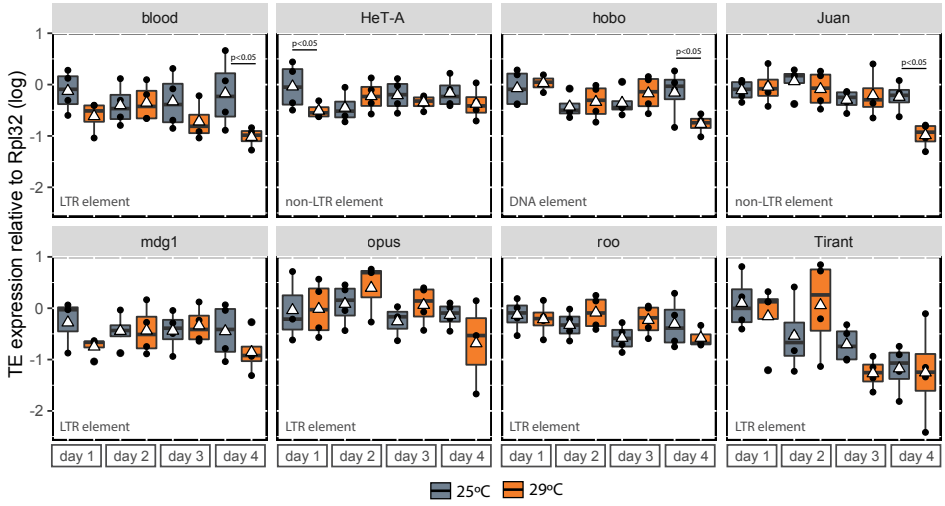
Supplementary Information



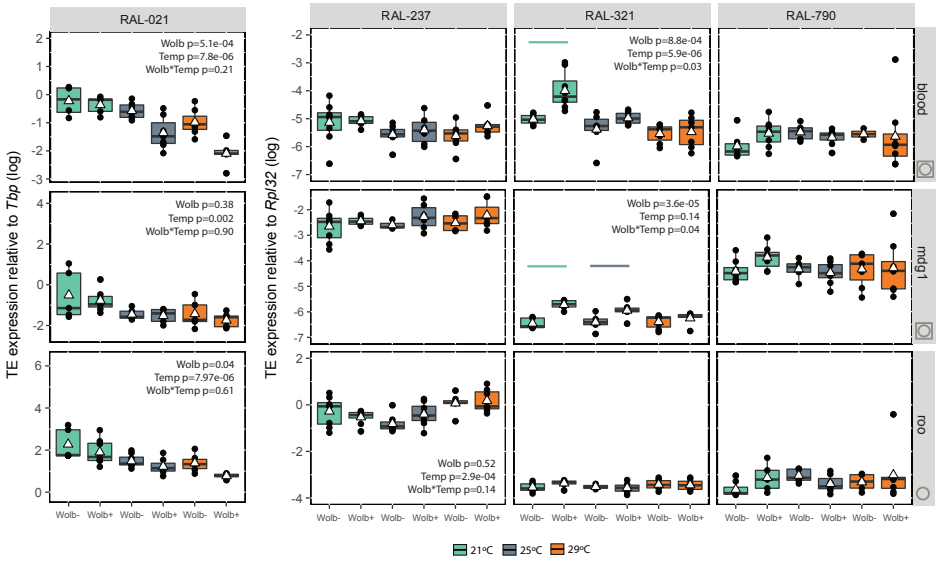
Supplementary Figure 3.1. Expression of control genes (*Rpl32* and *Tbp*). *Rpl32* and *Tbp* expression (logarithmic scale, log) in four DGRP lines (RAL-021, RAL-237, RAL-321 and RAL-790) infected (Wolb+) or not (Wolb-) with *Wolbachia* and exposed to different temperatures (21ºC, 25ºC and 29ºC). Each dot in box plots represents a biological replicate (pool of eight pairs of ovaries); triangles represent the mean.



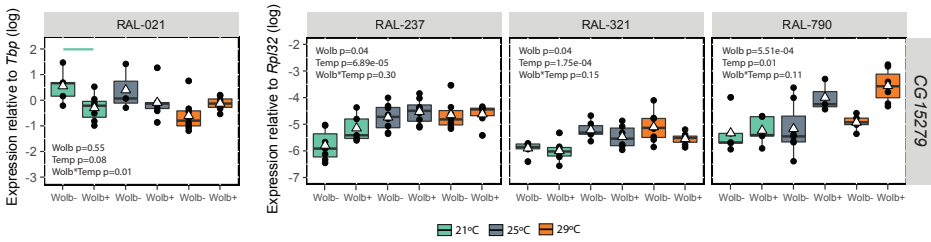
Supplementary Figure 3.2. Confirmation of TE insertions included in mature mRNAs. Scheme of nested PCR to amplify TE insertions that are within gene sequences **(A)**. Primers used for the first PCR reaction (PCR.1) were designed to recognize gene sequences around the insertion (**Supplementary Table 3.2**); product of first PCR was used as template for the second PCR reaction (PCR.2) using primers designed to recognize nucleotide sequences inside the TE (**Supplementary Table 3.1**). Agarose gel (1%) with amplicons obtained from the first (yellow) and second (red) PCR reactions **(B)**. *Blood* and *roo* negative controls (Ctr-) were performed using a PCR reaction without gDNA but with primers amplifying respective TE sequences (**Supplementary Table 3.1**). We used GeneRuler™ 1kb DNA ladder (Thermo Fisher Scientific).



Supplementary Figure 3.3. Effect of temperature on TE expression at different time points. Relative expression (Pfaffl method) of *blood, Het-A, hobo, Juan, mdg1, opus, roo* and *Tirant* elements, at day 1, 2, 3 and 4 after temperature change. *Blood, mdg1, opus, roo* and *Tirant* are retrotransposons with LTRs (LTR elements), *HeT-A* and *Juan* are retrotransposons without LTR sequences (non-LTR elements), *hobo* is a DNA element. We assessed the effect of temperature and time (categorical fixed variables) using linear models with dependent variable (relative TE expression) in logarithmic scale. The effect of temperature at each time point was tested using least-squares means ($p<0.05$). Statistical significant comparisons obtained from this analysis are highlighted in the boxplots. We also assessed the effect of time (continuous fixed variable) in TE expression (logarithmic scale) by using linear models (TEs with expression that is affected by the age of females: *Juan*, estimate=-0.185, p-value=0.002; *Tirant*, estimate=-0.430, p-value=0.0003). Each circle in the boxplots represents a pool of eight pairs of ovaries; triangles correspond to the mean, and black line to the median.



Supplementary Figure 3.4. Effect of *Wolbachia* and temperature on TE expression of RAL-021, RAL-237, RAL-321 and RAL-790. Relative expression of *blood, mdg1* and *roo* retrotransposons, in RAL-021, RAL-237, RAL-321 and RAL-790 ovaries. We tested the effect of *Wolbachia,* temperature and the interaction between both (categorical fixed variables) in relative TE expression using linear models with lognormal distribution. Data for each TE and genetic background was analyzed separately; statistically significant factors are shown in boxplots with respective p-values. When the interaction between *Wolbachia* and temperature was statistically significant (for the case of *blood* and *mdg1* expression in RAL-321) we used least-squares means (lsmeans) to test for the effect of *Wolbachia* at each temperature. In that case, colored lines represent statistically significant effects of *Wolbachia*. Each circle in the boxplots represents a pool of eight ovary pairs; triangles correspond to the mean and black line to the median. *Blood* and *mdg1* are expressed in the germline and somatic tissue of ovaries whereas *roo* is only expressed in the germline (see **Supplementary Figure 2.3** for more details).



Supplementary Figure 3.5. Effect of *Wolbachia* and temperature on the expression of *CG15279*. Relative expression of *CG15279* in RAL-021, RAL-237, RAL-321 and RAL-790 ovaries. We tested the effect of *Wolbachia,* temperature and the interaction between both factors (categorical fixed variables) in relative expression using linear models with lognormal distribution. Data for each genetic background was analyzed separately; statistically significant factors are shown in boxplots with respective p-values. Each circle in the boxplots represents a pool of eight ovary pairs; triangles correspond to the mean and black line to the median.

Supplementary Figure 3.6. Transcription factor expression. Expression of *CREB, dorsal, E2F1, fork head, Pebbled, cap-n-collar, PAX-6, slow border cells* and *seven up* transcription factors in response to several environmental perturbations (cold, heat, cadmium, copper, zinc, ethanol, caffeine, paraquat, rotetone and Sindbis virus) **(A)**. Expression of transcription factors in ovaries collected from mated females **(B)**. Expression in response to environmental stimuli was obtained from modENCODE treatment expression data[67]; expression in ovaries was acquired from modENCODE tissue expression data[67]. No expression, 0-3 reads; low expression, 4-10 reads; moderate expression, 11-25 reads; moderately high expression, 26-50 reads; high expression, 51-100 reads.

Supplementary Table 3.1. Primers used for qPCR

Target	Forward primer sequence	Reverse primer sequence	Ref
blood	AACAATAGAAAGAAGCCACCGAAC	AGTCATGGACTATTGAGGGTGTTG	[68]
Het-A	CGCGCGGAACCCATCTTCAGA	CGCCGCAGTCGTTTGGTGAGT	[69]
hobo	CATTAAGTCGGAAGGCCAAA	CTTGCTCTTCCGCTATCCAC	
Juan	GGGGCAAAATTCTCAATGAA	GCGGAATATATGTGGGTTGC	
mdg1	GTCAGAAGGAGGCCATTCAGGAATTT	GTTGCTGGCGGTTTCTGTTATTGTCAA	[70]
opus	CGAGGAGTGGGGAGAGATTG	TGCGAAAATCTGCCTGAACC	[7]
roo	CGTCTGCAATGTACTGGCTCT	CGGCACTCCACTAACTTCTCC	[7]
Tirant	AAATGTTTGCCCCATCTCAG	GGGTCAGTGTGGCGTTATTT	
CC15279	TCGTCTATCCAAATTGGTCTTACTC	AATGGCCACGATCATCCA	
Rpl32	ATGCTAAGCTGTCGCACAAATG	GTTCGATCCGTAACCGATGT	
Tbp	GGCAAAGAGTGAGGACGACT	GAGCCGACCATGTTTTGAAT	[44]

Ref, reference

Supplementary Table 3.2. Primers used to confirm TE insertions

Target	Forward primer sequence	Reverse primer sequence
Act42A	ATATAACGCCGCTCGTTCTC	ATCTCGCCTGAGCTCTACCA
CG10073	TTACGTTTGGACGTGGTTGA	CCATTCACACCACGTTTGTC
CG11319	TGACAACGATGAGGAACCAA	ATCGTCCCTGGTGTAGATCG
CG17883	CCCTCCACACCAGTTTTCTT	CAGTCACATGGTTCGATTGG
CG2022	TGACACGTTTTTCAGGGATG	GCGGGTAGAGATTATGCCTTA
CG31751	CCGTTAGGCTTTTCCGTTTT	TTTTGTTTGCGATCATCTCG
CG31974	AACAAGGCGCTAATTCGATG	AAGAACACGGGTGCAAAAAC
CG33966	ATGTCGATGTGGACAAAGCA	ATGAGACCAGCAGCTCGATT
CG4374	CGCCTTGATGTCGATCTTCT	TGCTTCACCAGTTCCAATGA
CG5210	AGAGGACTCGTCGTGGCTAA	GTGCTGCATGTTGGCTTAAA
CG5440	ACGGCACGATAAAGAAAAGG	ATGATCGTCGAGGTCCACTC
CG5681	TGAACGGATGTCCAAGTCAA	CCCCTGCACCATGATTAGTT
CG8920	TATTGGGCCTTTCGAATCTG	GGATTGGTTCTGCCATGTCT
CG9194	CGACCTCCAAACATTTTCGT	AGCAGATGGCAGCTACACAG
Dg	GTTTCTTCGGCACATCACCT	CAACTGCGGAATGAGACTGA
dnc	AAATGCAACACTTGCCTTCC	CACTTCCGATTTCGTTGGTT
Fas2	AGACGGTCGGTGAAACTTTG	GTTCATTCTTCCGCTTCTGC

Supplementary Table 3.2. Cont.

Target	Forward primer sequence	Reverse primer sequence
pot	CAATTTGCCATCGGGTTAAT	GACACACAATCTCGCCAAAA
Sema-1a	AGAAGCGCATTCATATTTGC	AACTTGGGTGTGTGGTGTCA
Taf12	AGCTCAAACTGAGCGGACAT	TCCGTTTTCTGGTCCGTAAG
Topors	TATGCAAATTAGCCGCACAA	CAGAAGCATCTTCCCCGATA
tws	GAAGAGCAACGACATCAGCA	ATGGAGTAGGGACGTTGTGG

Supplementary Table 3.3. Characterization of *blood, mdg1* and *roo* insertions in RAL-021, RAL-237, RAL-321 and RAL-790. This table is provided in digital format.

Supplementary Table 3.4. Gene ontology analysis for the genes containing *blood, mdg1* and *roo* insertions. This table is provided in digital format.

References

1. Lisch, D. How important are transposons for plant evolution? *Nat Rev Genet* **14,** 49–61 (2012).

2. Emera, D. & Wagner, G. P. Transposable element recruitments in the mammalian placenta: Impacts and mechanisms. *Brief. Funct. Genomics* **11,** 267–276 (2012).

3. Santos, M. E. *et al.* The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat. Commun.* **5,** 5149 (2014).

4. González, J., Karasov, T. L., Messer, P. W. & Petrov, D. a. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. *PLoS Genet.* **6,** 33–35 (2010).

5. González, J., Lenkov, K., Lipatov, M., Macpherson, J. M. & Petrov, D. a. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol.* **6,** 2109–2129 (2008).

6. Kaminker, J. *et al.* The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol.* **3,** 1–20 (2002).

7. Specchia, V. *et al.* Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463,** 662–665 (2010).

8. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. a. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12,** 246–258 (2011).

9. Juliano, C., Wang, J. & Lin, H. Uniting Germline and Stem Cells: the Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu. Rev. Genet.* 1–26 (2011). doi:10.1146/annurev-genet-110410-132541.Uniting

10. Luteijn, M. J. & Ketting, R. F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.* **14,** 523–34 (2013).

11. Izumi, N. *et al.* Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. *RNA J.* **19,** 896–901 (2013).

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

13. Touret, F., Guiguen, F. & Terzian, C. Wolbachia Influences the Maternal Transmission of the gypsy Endogenous Retrovirus in Drosophila melanogaster. *MBio* **5,** 1–10 (2014).

14. Jouan-Dufournel, I., Cosset, F., Contamine, D., Verdier, G. & Biémont, C. Transposable Elements Behavior Following Viral Genomie Stress in Drosophila melanogaster Inbred Line. *J. Mol. Evol.* **43,** 19–27 (1996).

15. Nabirochkin, S. D. *et al.* Oncoviral DNAs induce transposition of endogenous mobile elements in the genome of Drosophila melanogaster. *Mutat. Res.* **403,** 127–36 (1998).

16. Junakovic, N., Franco, C. Di, Barsanti, P. & Palumbo, G. Transposition of Copia-like Nomadic Elements Can Be Induced by Heat Shock. *J. Mol. Evol.* **24,** 89–93 (1986).

17. Ratner, V. A., Zabanov, S. A., Kolesnikova, O. V & Vasilyeva, L. A. Induction of the mobile genetic element Dm-412 transpositions in the Drosophila genome by heat shock treatment. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5650–4 (1992).

18. Arnault, C., Loevenbruck, C. & Biémont, C. Transposable element mobilization is not induced by heat shocks in Drosophila melanogaster. *Naturwissenschaften* **84,** 410–4 (1997).

19. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482,** 173–178 (2012).

20. Biémont, C. & Vieira, C. Genetics: junk DNA as an evolutionary force. *Nature* **443,** 521–524 (2006).

21. Khurana, J. S. *et al.* Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell* **147,** 1551–1563 (2011).

22. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **22,** 1503–1517 (2013).

23. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

24. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **12,** 615–627 (2011).

25. Kazazian Jr., H. H. Mobile Elements: Drivers of Genome Evolution. *Science (80-.).* **303,** 1626–1632 (2004).

26. Beguiristain, T., Grandbastien, M.-A., Puigdomènech, P. & Casacuberta, J. M. Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants. *Plant Physiol.* **127,** 212–221 (2001).

27. Woodrow, P., Pontecorvo, G., Ciarmiello, L. F., Fuggi, A. & Carillo, P. Ttd1a promoter is involved in DNA-protein binding by salt and light stresses. *Mol. Biol. Rep.* **38,** 3787–3794 (2011).

28. Salazar, M., González, E., Casaretto, J. A., Casacuberta, J. M. & Ruiz-Lara, S. The promoter of the TLC1.1 retrotransposon from Solanum chilense is activated by multiple stress-related signaling molecules. *Plant Cell Rep.* **26,** 1861–1868 (2007).

29. Takeda, S., Sugimoto, K., Otsuki, H. & Hirochika, H. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* **18,** 383–393 (1999).

30. Kimura, Y. *et al.* OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* **42,** 1345–1354 (2001).

31. Werren, J. H., Windsor, D. & Guo, L. Distribution of Wolbachia among neotropical arthropods. *Proc. R. Soc. London* **262,** 197–204 (1995).

32. Jeyaprakash, A. & Hoy, M. A. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **9,** 393–405 (2000).

33. Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A. & Werren, J. H. How many species are infected with Wolbachia? - A statistical analysis of current data. *FEMS Microbiol. Lett.* **281,** 215–220 (2008).

34. Teixeira, L., Ferreira, A. & Ashburner, M. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6,** 20753–2763 (2008).

35. Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Wolbachia and virus protection in insects. *Science (80-.).* **322,** (2008).

36. Hurst, G. D. D., Johnson, A. P., Schulenburg, J. H. G. v. d. & Fuyama, Y. Male-Killing Wolbachia in Drosophila: A Temperature-Sensitive Trait With a Threshold Bacterial Density. *Genet.* **156 ,** 699–709 (2000).

37. Richardson, M. F. *et al.* Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. *PLoS Genet.* **8,** e1003129 (2012).

38. Chrostek, E. *et al.* Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis. *PLoS Genet.* **9,** 1–22 (2013).

39. Mendes, C. C. & Mirth, C. K. Stage-Specific Plasticity in Ovary Size Is Regulated by Insulin/Insulin-Like Growth Factor and Ecdysone Signalling in Drosophila. *Genetics* (2015). doi:10.1534/genetics.115.179960

40. Pfaffl, M. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29,** 2002–2007 (2001).

41. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res.* **64,** 5245–5250 (2004).

42. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative

RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** 1–12 (2002).

43. Ponton, F., Chapuis, M.-P., Pernice, M., Sword, G. A. & Simpson, S. J. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in Drosophila melanogaster. *J. Insect Physiol.* **57,** 840–850 (2011).

44. Matta, B. P., Bitner-Mathé, B. C. & Alves-Ferreira, M. Getting real with real-time qPCR: A case study of reference gene selection for morphological variation in Drosophila melanogaster wings. *Dev. Genes Evol.* **221,** 49–57 (2011).

45. Mclaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* 1–14 (2016). doi:10.1186/s13059-016-0974-4

46. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

47. Sandelin, A., Wasserman, W. W. & Lenhard, B. ConSite : web-based prediction of regulatory elements using cross-species comparison. **32,** 249–252 (2004).

48. Tsunoda, T. & Takagi, T. Estimating transcription factor bindability on DNA. *Bioinformatics* **15,** 622–630 (1999).

49. Guio, L., Barrón, M. G. & González, J. The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila. *Mol. Ecol.* **23,** 2020–2030 (2014).

50. Tootle, T. L., Williams, D., Hubb, A., Frederick, R. & Spradling, A. Drosophila eggshell production: identification of new genes and coordination by Pxt. *PLoS One* **6,** e19943 (2011).

51. Delpuech, J. *et al.* Phenotypic plasticity and reaction norms in temperate and tropical populations of Drosophila melanogaster: ovarian size and developmental temperature. *Evolution (N. Y).* **49,** 670–675 (1995).

52. Drummond-Barbosa, D. & Spradling, A. C. Stem cells and their progeny respond to nutritional changes during Drosophila oogenesis. *Dev. Biol.* **231,** 265–78 (2001).

53. Servant, G., Pennetier, C. & Lesage, P. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. *Mol. Cell. Biol.* **28,** 5543–5554 (2008).

54. Servant, G. *et al.* Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. *Nucleic Acids Res.* **40,** 5271–5282 (2012).

55. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Genet.* (2011). doi:10.1038/nrg3055

56. Marinissen, M. J. & Gutkind, J. S. G-protein-coupled receptors and signaling networks: emerging paradigms. **6147,** 368–376 (2001).

57. Harmar, A. J. Protein family review family. *Genome Biol.* **2,** 1–7 (2001).

58. Lattin, J. E. *et al.* Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res.* **4,** 1–13 (2008).

59. Cridland, J. M., Macdonald, S. J., Long, A. D. & Thornton, K. R. Abundance and Distribution of Transposable Elements in Two Drosophila QTL Mapping Resources. *Mol. Biol. Evol.* **30,** 2311–2327 (2013).

60. Deloger, M. *et al.* Identification of expressed transposable element insertions in the sequenced genome of Drosophila melanogaster. *Gene* **439,** 55–62 (2009).

61. Corby-Harris, V. *et al.* Geographical Distribution and Diversity of Bacteria Associated with Natural Populations of Drosophila melanogaster. *Appl. Environ. Microbiol.* **73,** 3470–3479 (2007).

62. Pan, X. *et al.* Wolbachia induces reactive oxygen species (ROS) - dependent activation of the Toll pathway to control dengue virus in the mosquito Aedes aegypti. *PNAS* **109,** 23–31 (2012).

63. Wong, Z. S., Brownlie, J. C. & Johnson, N. Oxidative Stress Correlates with Wolbachia-Mediated Antiviral Protection in Wolbachia-Drosophila Associations. *Appl. Environ. Microbiol.* **81,** 3001–3005 (2015).

64. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408,** 239–247 (2000).

65. Sykiotis, G. P. & Bohmann, D. Keap1 / Nrf2 Signaling Regulates Oxidative Stress Tolerance and Lifespan in Drosophila. *Cell* **14,** 76–85 (2007).

66. Fablet, M. & Vieira, C. Evolvability, epigenetics and transposable elements. *Biomol. Concepts* **2,** 333–341 (2011).

67. Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512,** 393–399 (2014).

68. Handler, D. *et al.* A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30,** 3977–3993 (2011).

69. Vagin, V. V, Sigova, A., Li, C., Gvozdev, V. & Zamore, P. D. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science (80-.).* **313,** 320–324 (2006).

70. Navarro, C., Bullock, S. & Lehmann, R. Altered dynein-dependent transport in piRNA pathway mutants. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 9691–9696 (2009).

Chapter IV

TE response to stressful conditions is dependent on the type of stress and *Wolbachia* status: a study of the effect of temperature, cadmium and oxidative stress

4.1 Abstract

Environmental perturbations reducing fitness – stressful conditions[1] – are universal to all natural habitats. Upon exposure to threats, populations can benefit from responses that increase genetic variation[2,3], including those driven by transposition[4,5]. Despite growing evidence that suggest transposable elements (TEs) as important mediators of adaptive genetic variation[6–8], the environmental factors that influence mobilization remain poorly understood[9]. Here, we aimed at studying the impact of different environmental cues, including biotic and abiotic factors, on transposition in the germline. We exposed *Drosophila melanogaster* adult females to stressful conditions that included several temperatures, levels of oxidative stress and doses of heavy metal. We then assessed expression of five TEs (*blood, Juan, pogo, roo* and *Tirant*) in ovaries. Given the ability of *Wolbachia* to affect many aspects of host biology, and to spread in natural populations by maternal transmission[10–12], we further assessed the impact of this endosymbiont in affecting TE response to the aforementioned abiotic factors. We observed that TE transcription levels depended on the environment, *Wolbachia* status, and TE analysed. Although TE responses to oxidative stress were not robust and had significant differences between independent experiments, we observed robust responses of *blood* and *Tirant* to changes in temperature, and of *Juan* and *Tirant* to cadmium. We propose that the main repressor of TE activity in the germline (the piRNA pathway) is not affected by environmental perturbations and controls the activity of most mobile elements, whereas few transposons would specifically respond to perturbations (namely by effects in the activity of transcription factors). This environment-specific TE activity in ovaries that resulted in new transposon insertions transmitted to the next generation may produce genetic variation that is adaptive to the inducing environment.

4.2 Introduction

Wild populations are often exposed to environmental perturbations that can shape evolution by acting as agents of natural selection, involved in the sorting

of genetic variation that results in trans-generational changes in allelic frequencies, and as factors influencing the production of novel genetic variation, through effects on recombination[13,14] or mutation rates[2,3]. Despite growing evidence that transposable elements (TEs) constitute an important source of adaptive genetic variation[6–8], little is known about the environmental factors that influence TE mobilization[9].

Most of the information regarding the regulation of mobile elements in metazoan species describes a conserved pathway that controls TE activity in the germline[15]. Those organisms specifically repress TEs by producing small RNAs homologous to TE sequences and that interact with PIWI family proteins (piRNAs)[15]. Interestingly, recent studies revealed commonalities between the mechanisms that control TE activity and the pathways triggered in sub-optimal conditions[16]. Those included the induction of heat shock responses and the piRNA pathway that fails to contain TE activity in the absence of Hsp90 (a heat shock protein involved in protein folding)[17–19]. As many environmental changes result in misfolded proteins[20], it is conceivable that, in stressful conditions, the recruitment of chaperones such as Hsp90 for mis-folding correction, would result in relaxed piRNA surveillance and in increased transposon activity. If this happens, we should see deregulation of different TEs, including retrotransposons and DNA elements[21]. In this work, we aimed at studying the impact of several environmental cues, including biotic and abiotic factors, on the activity of retrotransposons and DNA elements in the germline of *Drosophila melanogaster*. We also assessed whether our challenges could affect the piRNA pathway.

D. melanogaster is able to colonize diverse habitats[22] and have adapted to several biotic and abiotic stresses known to trigger heat shock responses and to produce excessive amount of reactive oxygen species (ROS)[23–25]. These environmental conditions include infection[10,26,27], temperature perturbation [28–30] and exposure to heavy metals[31]. We therefore decided to evaluate TE expression after exposure to different types of environmental perturbation that included temperature changes as well as exposure to an oxidant agent and an

heavy metal. Also, given that *Wolbachia* – a bacterial symbiont common in insects – is able to influence many aspects of host biology, and to spread in natural populations by maternal transmission[10–12], we assessed the impact of the presence of this endosymbiont on TE response to the abiotic factors. Those effects were tested in fruit fly ovaries, because transposition occurring in the germline can add to heritable genetic variation in a population.

We observed that transposon activity is environment-specific, and suggest that temperature, and ingestion of cadmium, may influence ability to produce heritable genetic variation by changes in TE expression. In addition, and contrary to our expectations, piRNA pathway seems to be robust to environmental change, and keep controlling the bulk of mobile elements, while only few specific TEs respond to perturbations. We discuss that transcription factors related with particular pathways triggered by environmental change might directly affect TE activity, and that *Wolbachia* may play an important role in affecting transposon sensitivity to activation.

4.3 Materials and Methods

Drosophila melanogaster stocks and maintenance

All experiments were done in a genetic background resultant from a cross between Harwich females (with and without *Wolbachia*) and w[1118] males (free of *Wolbachia*). w[1118] and Harwich stocks were provided by Luís Teixeira. Harwich flies were naturally infected with wMel *Wolbachia* strain. To obtain a Harwich line free of infection, we treated flies with tetracycline (0.05mg/ml added to their food), for two generations. Gut flora content was homogenized between tetracycline treated and non-treated flies, as previously described[32]. We crossed $SpnE^{616}$/TM3 and $SpnE^{100.37}$/TM3 flies (provided by Vítor Barbosa) to obtain a positive control ($SpnE^{616}$/ $SpnE^{100.37}$) and negative controls ($SpnE^{616}$/TM3 or $SpnE^{100.37}$/TM3) for Bicaudal-D aggregate formation in ovaries. We kept flies in standard cornmeal food[33] at 25°C, 60% humidity and 12h-12h day-night cycles, unless otherwise mentioned.

Temperature, cadmium and oxidative stress treatment

For temperature, cadmium and oxidative stress treatments we collected eight-hour-old flies (with or without *Wolbachia*) and kept five females with two males per vial, at 25ºC. After three days, males were removed, females transferred to new tubes, and placed in all different experimental conditions. Females were then kept at each experimental condition for four days (**Supplementary Figure 4.1**). For temperature treatments, females were placed either at 21ºC, 25ºC or 29ºC. For exposure to heavy metals, flies were fed with one of three different doses of cadmium (0mM, 0.05mM, or 0.1mM), and for exposure to oxidative stress-inducing conditions, females were fed with different doses of paraquat (0mM, 1.25mM, or 2.5mM), a commonly used drug to trigger the production of ROS[34]. Cadmium and paraquat solutions (Sigma) were directly mixed in the fly food. To confirm that females fed, we included 2% of blue vegetable coloring in the food and checked for the presence of blue color in the fly abdomen and guts. We assessed fecundity (number of eggs laid per female) at day four after environmental perturbation to test whether changing temperature, exposing flies to heavy metal doses (cadmium) or to an oxidant agent (paraquat) are stressful for *D. melanogaster* females (with or without *Wolbachia*) (**Supplementary Figure 4.1A**). Also, after four days of exposure to each experimental condition, ovaries were dissected for RNA extraction for assessment of TE expression (see below and **Supplementary Figure 4.1B**). To confirm effect of temperature on TE copy number in the next generation, we collected embryos laid on day four after temperature change, for genomic DNA (gDNA) extraction (see below and **Supplementary Figure 4.1C**).

RNA extraction and cDNA synthesis

Ovaries were dissected in fresh cold PBS 1x, and directly stored at 4ºC in 400μl Trizol (Ambicon) until homogenization using pestels that was done in the same day. Before homogenization we polled 8-10 pairs of ovaries per sample, and used 6-8 replicates per condition, per experiment. Homogenized samples were stored at -80ºC until RNA extraction. For RNA extraction we used Direct-zol RNA

Miniprep kit (Zymo Research), and total RNA was eluted in 25μl of RNase-free water (Sigma). RNA purity and concentration for each sample, was measured with Nanodrop ND-1000 Spectrophotometer; A260/A280 absorbance was around two for all samples, and concentrations ranged from 200 to 1000ng/μl. All RNA samples of each experiment were processed on the same day. We cleaned 1μg of RNA from gDNA contamination, using DNAse (Promega) treatment, following manufacturer's instruction. For cDNA synthesis, we followed Reverse Transcription System (Promega) instructions, using 0.02μM of Oligo dT primers. Reaction was incubated at 42°C for 60min, and heated at 95°C for 5min. cDNA was then diluted in RNAse-free water (1:10) and stored at -20°C until quantitative real-time PCR (qPCR).

gDNA extraction

gDNA from embryos was extracted using DNeasy Blood & Tissue Kit (Qiagen). For each sample, we pooled all eggs laid by a group of 20 females (**Supplementary Figure 4.1**), six replicates per condition. Tissues were homogenized in 180μl of buffer ATL, using pestels. We used 200μl of buffer AE to elute gDNA and treated all samples with 0.2mg/ml of RNAse (Invitrogen) for 20 min at 55ºC. gDNA concentration was measured in Nanodrop ND-1000 Spectrophotometer, and samples stored at -20°C, until qPCR.

qPCR

We measured levels of TE expression and TE copy number by qPCR: CFX384 (BioRad) or ABI QuantStudio-384 (Applied Biosystems) thermal cyclers. For each reaction we used 5μl of iQ™ SYBR® Green supermix (BioRad), 0.4μM primers, 4μl of diluted cDNA (1:10) or gDNA (4ng/μl). We assessed the expression of five TEs (*blood, Juan, pogo, roo* and *Tirant*) and three control genes (*Rpl32, Tbp* and *vasa*). Primer sequences are described in **Supplementary Table 4.1**. We ran two technical replicates per sample, per plate. The thermal cycling protocol was the same for all reactions: 2 min at 50°C; 10 min at 95°C; 40 cycles of 95°C for 30s, 60°C for 1 min and 72°C for 30s.

In each plate we ran reactions on serial dilutions of known target amplicon concentrations to build standard curves. To obtain standards, we amplified target regions of each primer pair (**Supplementary Table 4.1**) by PCR using 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2, 0.4μM primers and 1μl of cDNA (1:10). PCR products were cleaned using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel), and eluted in 15μl of buffer NE. We measured concentrations in Nanodrop ND-1000 Spectrophotometer, and did 13 serial dilutions (1:5) from 0.125ng/μl until 6.4e-10ng/μl. For each primer pair, we estimated the minimum amount of nucleic acids detectable by qPCR by extending dilutions of our standards until the relationship between the known amplicon concentrations and the quantification cycles (Cqs) reached a plateau.

We used default threshold settings of the qPCR machines to obtain the Cq for each reaction. Cq values for all targets and experimental samples were within the linear amplification. We transformed Cq values into absolute target amplicon concentrations using the respective standard curves. This methodology allowed us to control for plate effects and primer efficiency. qPCR melting curves were analyzed to confirm specificity of amplified products, and negative controls never showed detectable amplification.

For the analysis of expression data, we used NormFinder[35] and geNorm[36] algorithms to check expression stability of two reference genes commonly used to calculate relative gene expression at different environmental conditions, *Rpl32* and *Tbp*[37,38]. We also tested *vasa* expression stability, which is a germline-specific gene[39]. For both algorithms used, stability values obtained for *vasa* (NormFinder, 0.39; geNorm, 0.84) were better than those for *Rpl32* (NormFinder, 0.57; geNorm, 0.95) and *Tbp* (NormFinder, 0.69; geNorm, 1.09). However, *vasa* expression in the second experiment for Paraquat treatments is more variable than that observed for *Rpl32* and *Tbp* (**Supplementary Figure 4.2**). We therefore decided to use *Tbp* (NormFinder, 0.46; geNorm, 0.40) as reference gene for that dataset instead of *vasa* (NormFinder, 0.81; geNorm, 0.62) or *Rpl32* (NormFinder, 0.63; geNorm, 0.47). The median of relative expression in control samples (25ºC, 0mM of cadmium, or 0mM of paraquat),

free of *Wolbachia*, was used to calibrate each treatment (temperature, cadmium and paraquat) per experiment and transposon (Pffafl method)[40].

To assess TE copy number in the genomic DNA of embryos, we used *Rpl32* (a known single copy gene) as reference.

Antibody staining of ovaries

Ovaries were dissected in fresh cold PBS 1x, and fixed in 200µl of 3.2% formaldehyde (Sigma), 0.5% NP40 (Sigma) diluted in PBS 1x, and 600µl of heptane (Sigma) for 20 min with gentle (400 rpm) shaking. To avoid unspecific antibody (Ab) binding and permeabilize tissues, ovaries were kept for 1 hour at 400 rpm, in 1% BSA (Sigma), 0.01% TritonX (Sigma) diluted in PBS 1x. We then added a cocktail of anti-Bicaudal-D 4C2 (1:10) and anti-Bicaudal-D 1B11 (1:10) (Developmental Studies Hybridoma Bank) solutions and incubated overnight at 4ºC. Alexa488 anti-mouse was used as secondary Ab, and diluted (1:500) in PBS 1x, containing 0.2% Tween (Sigma) (PBT). Ovaries were kept in that solution for 2 hours, in the dark at 400 rpm. Nuclei were then stained with Toto-3 (1:500) for 7 min in the dark. We performed washes with PBT between all steps mentioned above; last washes of the protocol were done with PBS. All protocol was performed at room temperature, unless stated otherwise. Ovaries were mounted in glycerol, and images acquired in Leica SP5 inverted confocal.

Statistical analyses

All statistical analyses were done in R (version 3.2.1).

To test the effect of abiotic factors (temperature, cadmium and paraquat) and *Wolbachia* status (categorical fixed variables) on fecundity, we used general linear models with negative binomial distribution (glm.nb). The interaction between treatment and presence of *Wolbachia* was included in the analysis; each abiotic factor was analyzed separately. Given that *Wolbachia* status was never a factor contributing to explain the data, we then assessed differences between treatments using non-parametric Wilcoxon rank test. For temperature treatments (21ºC, 25ºC and 29ºC) we performed all pairwise comparisons

whereas for cadmium and paraquat treatments we only compared the fecundity of treated (fed with 0.05mM or 0.1mM of cadmium, and with 1.25mM or 2.5mM of paraquat) with non-treated females (fed with 0mM of cadmium and paraquat).

We assessed the effect of temperature, cadmium and paraquat treatments (categorical fixed variables) on the relative expression of TEs (*blood, Juan, pogo, roo,* and *Tirant*) using linear models (lm) with lognormal distribution. As these data were obtained in two independent experiments (each with at least six replicate measurements in pools of ovaries), we also tested for robustness of transposon response, by including experiment as a categorical fixed factor. Flies with and without *Wolbachia* were analyzed separately. Model output for each TE, gave significance for the effect of factors and interaction between them (treatment and experiment). When the effect of interaction between treatment and experiment was significant ($p<0.05$), we considered that the effect of the treatment conditions on TE activity was not robust. In that case, and to evaluate the differences between experiments, we performed pairwise comparisons using non-parametric Wilcoxon rank test. For the treatments that mediated robust TE responses, we also performed pairwise comparisons using non-parametric Wilcoxon rank test but analyzed both experiments together.

To test the effect of thermal challenges (21ºC, 25ºC and 29ºC) on the copy number of *blood, Juan, roo* and *Tirant* we performed pairwise comparisons using non-parametric Wilcoxon rank test.

4.4 Results

Wolbachia affects TE response to temperature, but not to cadmium or paraquat treatments

To test whether our environmental perturbations were able to reduce *D. melanogaster* fitness, we assessed female fecundity at day four after perturbation. We observed that females experiencing changes to lower temperature (21ºC), or fed with either cadmium (0.05mM and 0.1mM) or paraquat (1.25mM and 2.5mM) treatments laid fewer eggs than control females

that were maintained at 25ºC without doses of cadmium or paraquat (0mM) in their diet (**Figure 4.1**). Presence of *Wolbachia* did not influence fecundity.

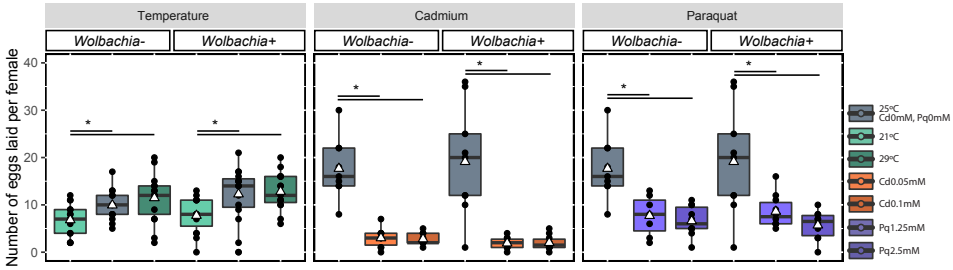


Figure 4.1. Effect of temperature, cadmium and paraquat treatments in fecundity. Fecundity was measured as the number of eggs laid per female, at day four after environmental perturbation. Adult flies with and without *Wolbachia* were exposed to several temperatures (21ºC, 25ºC and 29ºC), or fed with different doses of cadmium (0mM, 0.05mM and 0.1mM) and paraquat (0mM, 1.25mM and 2.5mM) (see **Supplementary Figure 4.1A** for more details about experimental setup). Given that *Wolbachia* status was never a factor contributing to explain the data (glm.nb model that included the effect of biotic and abiotic factors, as well as the interaction between both), we tested the effect of abiotic perturbations, for each *Wolbachia* status, by performing pairwise comparisons with non-parametric Wilcoxon rank test. For the temperature treatments we performed all paired comparisons; for cadmium and paraquat conditions we compared fecundity of treated females (fed with 0.05mM or 0.1mM of cadmium, and with 1.25mM or 2.5mM of paraquat) with that of non-treated (fed with 0mM of cadmium and paraquat). Statistically significant comparisons ($0.01<p<0.05$) are represented in the boxplots (*). Each circle in the boxplots represents a single female; triangles correspond to the mean and black line to the median.

To address the question of how different transposons may be affected by stressful conditions, we further assessed the expression of *blood, Juan, pogo, roo* and *Tirant* upon exposure to temperature perturbations and doses of cadmium and paraquat. In order to see the ability of the endosymbiont to modulate those responses we tested effects of abiotic factors separately for each *Wolbachia* status. We observed that, when there was an effect, most TE responses to perturbation were not consistent between independent

experiments, and were therefore considered not robust (**Supplementary Figure 4.3** and **Table 4.1**). While focusing on the robust responses we detected a *Wolbachia*-dependent effect of temperature on the expression of *blood* and *Tirant* and *Wolbachia*-independent *Juan* and *Tirant* responses to doses of cadmium (**Figure 4.2**). For the effect of oxidative stress we observed that none of the mobile elements analyzed here were influenced by any dose of paraquat.

In order to confirm that females were feeding normally upon exposure to cadmium and paraquat treatments, we included 2% of blue vegetable coloring in the food and checked for the presence of blue color in the fly abdomen and guts. We observed that when exposed to cadmium, but not paraquat, females stop eating and are therefore under starvation (**Supplementary Figure 4.4**).

Table 4.1. Statistics (p-values) for the effect of abiotic factors in flies infected or not with *Wolbachia*

		Wolbachia-			*Wolbachia+*		
	TEs	Treatment	Experiment	Treat*Exp	Treatment	Experiment	Treat*Exp
Temperature	*blood*	0.02 (*)	0.47	0.86	<0.001 (***)	<0.001 (***)	<0.001 (***)
	Juan	0.13	0.04 (*)	0.07	<0.001 (***)	0.17	0.06
	pogo	<0.001 (***)	<0.001 (***)	<0.001 (***)	0.59	0.17	0.25
	roo	0.12	0.19	0.47	0.02 (*)	<0.001 (***)	0.004 (**)
	Tirant	0.003 (**)	0.009 (**)	0.07	<0.001 (***)	<0.001 (***)	0.39
Cadmium	*blood*	0.87	0.42	0.02 (*)	0.37	0.86	0.21
	Juan	0.04 (*)	<0.001 (***)	0.06	0.03 (*)	0.003 (**)	0.62
	pogo	<0.001 (***)	<0.001 (***)	0.02 (*)	0.002 (**)	0.04 (*)	0.03
	roo	0.97	0.29	0.18	0.66	0.38	0.58
	Tirant	<0.001 (***)	0.33	0.05	<0.001 (***)	0.82	0.15
Paraquat	*blood*	0.66	0.61	0.48	0.47	0.23	0.04 (*)
	Juan	0.17	0.59	0.37	0.002 (**)	0.06	<0.001 (***)
	pogo	0.89	0.86	0.19	0.001 (**)	<0.001 (***)	0.69
	roo	0.60	0.73	0.75	0.15	0.06	0.002 (**)
	Tirant	0.85	0.01 (*)	0.63	0.002 (**)	0.74	0.007 (**)

Effect of each treatment (Treat: temperature, cadmium and paraquat) and experiment (Exp), on relative expression of *blood, Juan, pogo, roo* and *Tirant*, was tested using linear models (lm) with lognormal distribution. Model output gave significance (p-value, p) of each factor (treatment and experiment) and interaction between them (Treat*Exp). Interaction between treatment and experiment gives robustness of TE response: if interaction is statistically significant (p<0.05) TE response to environmental perturbation is considered not robust. Data collected for females infected (*Wolbachia*+) and not infected (*Wolbachia*-) with endosymbiont was analyzed separately. P-value<0.001 (***); 0.001<p-value<0.01 (**); 0.05<p-value<0.01 (*).

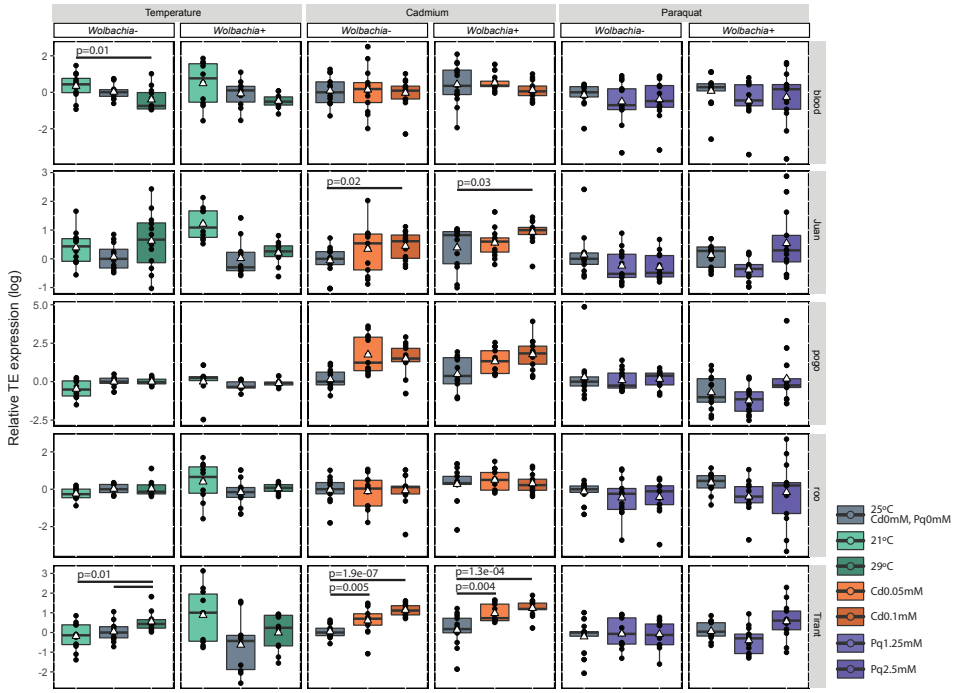


Figure 4.2. Effect of *Wolbachia* in TE response to temperature, cadmium and paraquat. Relative expression of *blood, Juan, pogo, roo* and *Tirant* in ovaries collected from females infected (*Wolbachia*+) or not (*Wolbachia*-) with endosymbiont and that experienced different environmental treatments. Those included several temperatures (21ºC, 25ºC and 29ºC), doses of cadmium (Cd0mM, Cd0.05mM and Cd0.1mM) or paraquat (Pq0mM, Pq1.25mM or Pq2.5mM). Data was collected in two independent experiments that are plotted together. For more information about experimental setup see **Supplementary Figure 4.1B**). TE expression is shown in logarithmic scale (log) and is relative to *vasa* expression excepting that of paraquat treatments obtained in the second experiment and normalized against *Tbp*. Relative expression of transposons was then calibrated for each treatment and experiment with the median expression of control flies (those free of *Wolbachia*, and exposed to 25ºC, Cd0mM and Pq0mM) (Pfaffl method[40]). We performed pairwise comparisons using non-parametric Wilcoxon rank test for the environmental treatments that were robust and previously shown to influence TE expression (see **Supplementary Figure 4.3** and **Table 4.1**). Data for each abiotic factor and *Wolbachia* status was analyzed separately. We only show statistics (p-values, p) in the boxplots when comparisons were statistically significant ($p<0.05$). All multiple comparisons for *Juan* and *Tirant* expression between temperature treatments, as well as for *pogo* expression in paraquat treatments (both in females infected with *Wolbachia*) were not statistically significant. Each circle in the boxplots represents a pool of 8-10 pairs of ovaries; triangles correspond to the mean and black line to the median.

Higher TE expression in ovaries has the potential to increase TE copy number in the next generation

To understand to which extent changes in the expression of transposons in ovaries mediate the production of novel genetic variation transmitted to the next generation, we assessed total copy number of retrotransposons (*blood, Juan, roo* and *Tirant*) in the gDNA of the F1. The gDNA was obtained from embryos that were laid by females with different *Wolbachia* status and exposed to several temperatures (21ºC, 25ºC and 29ºC). For the relationship established between copy number and expression we observed three main types of responses: increase in the number of copies 1) with robust increase, 2) with increase that is not robust and 3) without changes in expression (**Figure 4.3**). We detected (in average) more 61 copies of *blood* in embryos free of infection after mothers were exposed to 21ºC, more 7 copies of *Juan* and 34 copies of *roo* in embryos infected with endosymbiont after mothers experienced change to 29ºC and 21ºC, respectively.

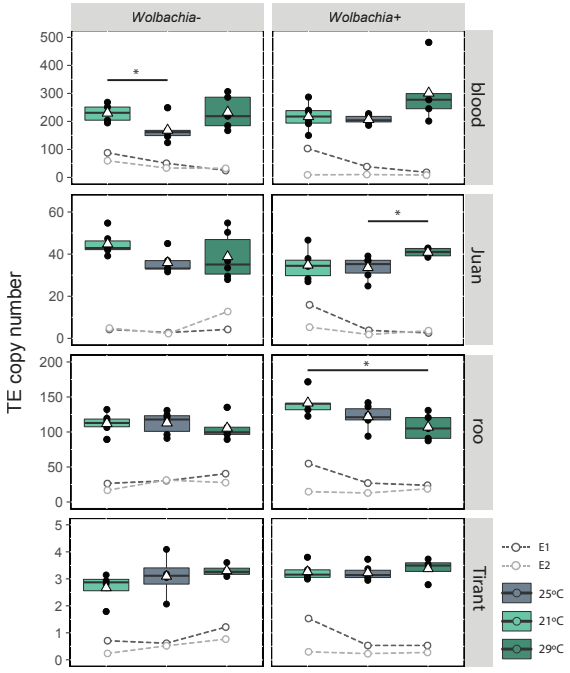


Figure 4.3. Effect of *Wolbachia* and temperature in the copy number of retrotransposons. Copy number of *blood, Juan, roo* and *Tirant* was assessed in the gDNA of eggs that were laid by females with (*Wolbachia+*) and without (*Wolbachia-*) endosymbiont infection, and exposed to thermal perturbations (21ºC, 25ºC and 29ºC). TE copy number represented in the graphs is relative to the abundance of *D. melanogaster* genome using *Rpl32* (a known single copy gene) as reference. To test the effect of temperature challenges in the total number of insertions we performed pairwise

comparisons using non-parametric Wilcoxon rank test. Statistically significant comparisons ($0.01<p<0.05$) are shown in boxplots (*). Each colored circle in the boxplots represents a pool of eggs laid by 20 females; triangles correspond to the mean, and black line to the median. We also include the average expression of *blood, Juan, roo* and *Tirant* per temperature and *Wolbachia* status (represented by open circles and dashed lines). For more information about expression data go to **Supplementary Figure 4.3**. E1, experiment 1; E2, experiment 2.

piRNA pathway is robust to environmental perturbation

When the biogenesis of piRNAs is disrupted, transposons become active, and the distribution of dynein motor complex proteins is affected. In those conditions, proteins such as Dynein Heavy Chain, Dynamitin, Egalitarian and Bicaudal-D form of large cytoplasmic aggregates in egg chambers within adult ovaries[41]. Here, in order to test to which extent our environmental perturbations were affecting TE expression through problems in the piRNA pathway we looked at the formation of Bicaudal-D aggregates in the ovaries of females exposed to several temperatures, levels of oxidative stress and doses of cadmium. We confirmed that Bicaudal-D forms aggregates in piRNA pathway mutant egg chambers (SpnE mutant: $SpnE^{616}/SpnE^{100.37}$) but not in ovaries collected from wild-type flies ($SpnE^{616}$/TM3 and $SpnE^{100.37}$/TM3) (**Figure 4.4**). Also, Bicaudal-D antibody staining in egg chambers collected from females with different *Wolbachia* status and exposed to several abiotic factors (temperature, cadmium and paraquat) was similar to that obtained in *wild-type* controls (**Figure 4.4**).

Wolbachia+
Wolbachia-
29°C
25°C
21°C
Cd 0.1mM
Cd 0.05mM
Cd/Pa 0mM
Wild-type
Spnf mutant
Bicaudal-D
DNA
Merge
Toto-3
Pa 2.5mM
Pa 1.25mM

Figure 4.4. Absence of Bicaudal-D aggregates in ovaries exposed to temperature, cadmium and paraquat treatments. Stage 8 egg chambers stained with anti-Bicaudal-D (Bic-D, green) and Toto-3 (DNA, blue). Bic-D aggregates formed in piRNA pathway mutant ($SpnE^{616}/SpnE^{100.37}$) that are not observed in $SpnE^{616}$/TM3 or $SpnE^{100.37}$/TM3 wild-type controls are highlighted by white arrows. No Bic-D aggregates were observed in ovaries collected from females, infected (*Wolbachia*+) or not (*Wolbachia*-) with endosymbiont, and exposed to several temperatures (21ºC, 25ºC and 29ºC), or fed with different doses of cadmium (Cd 0mM, Cd 0.05mM or Cd 0.1mM) and paraquat (Pq 0mM, Pq 1.25mM or Pq 2.5mM). Same results were obtained in all egg chambers analyzed per ovary (number of ovaries: $SpnE^{616}/SpnE^{100.37}$ mutant, n=18; $SpnE^{616}$/TM3 and $SpnE^{100.37}$/TM3, n=14; *Wolbachia*- 21ºC, n=7; 25ºC, n=6; 29ºC, n=8; Cd0mM, n=14; Cd0.05mM, n=6; Cd0.1mM, n=4; Pq0mM, n=14; Pq1.25mM, n=4; Pq2.5mM, n=5; *Wolbachia*+ 21ºC, n=6; 25ºC, n=7; 29ºC, n=8; Cd0mM, n=10; Cd0.05mM, n=8; Cd0.1mM, n=5; Pq0mM, n=17; Pq1.25mM, n=10 and Pq2.5mM, n=10).

4.5 Discussion

TE response to sub-optimal conditions is dependent on the environment

This study aimed at empirically testing the idea that organisms benefit from the production of novel genetic variation when exposed to stressful environments (environments that reduce fitness). For this purpose, we tested the effect of several sub-optimal conditions in TE activity, and used ovarian expression as a proxy for transposon ability to jump and to produce heritable genetic variation. We found that not all stressful conditions were able to influence TE expression and that although several transposons increase activity in sub-optimal conditions many elements were irresponsive or decreased expression. Those responses were dependent on factors such as transposon analyzed, type of treatment (temperature, cadmium and paraquat) and presence of *Wolbachia*. We suggest that *D. melanogaster* success in invading new ecosystems may have relied on their ability to increase genetic variation in response to some, but not all, environmental conditions. In addition, given high prevalence of *Wolbachia* in natural populations[10], we propose that this endosymbiont may have played an important role in manipulating those responses without significant effects in fecundity.

Effect on the robustness of TE expression suggests transient activation in response to specific and environmentally regulated transcription factors

The study of the dynamic interplay between the transient activation of a particular transcription factor and consequent effects in gene expression led to the discovery that, although protein activation is linear, the transcriptional outcome is bimodal[42]. In other words, cells exposed to osmotic stressful conditions that linearly influence transcription factor activation were able to shift gene expression from a repressed to an activated state with high and stochastic variation in the transition between those two states. Similar results were observed for our analysis of TE expression, as striking differences in activity were found for the response to some environmental perturbations assessed in independent experiments. In those cases, the effect of exposure to sub-optimal conditions in TE activity was described as not robust. We further suggest that, transposon response to environmental stimuli is bimodal and driven by specific transcription factors (**Figure 4.5**). In addition, we propose that *Wolbachia* plays an important role in shifting the transcriptional sensitivity of transposons to external cues.

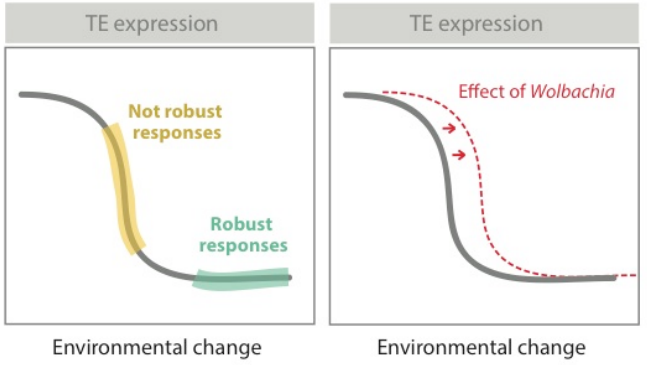


Figure 4.5. Model for TE expression in response to environmental perturbations. We propose that the expression of transposons in response to stressful conditions is sigmoidal. This means that for some environments TE expression is very variable (not robust responses) but for other conditions is robust. In addition, we suggest that the presence of *Wolbachia* is able to shift the transcriptional sensitivity of transposons to external cues.

piRNA pathway is not affected by environmental perturbations

PiRNAs are main defenders of host genomes against transposon activity[15]. Problems in this homology-dependent silencing pathway, can result in uncontrolled TE jumping that drive severe DNA damage and sterility[39]. Although the presence of Hsp90 is crucial to maintain the piRNA biogenesis operational[17–19] and links this pathway with responses to stress, we observed that the piRNA surveillance is robust to environmental changes. We suggest that, in response to threats, individuals benefit from the ability to produce new genetic variation, without compromising reproduction due to an overall uncontrolled TE activity. We hypothesize that fruit flies stimulate moderate transposition in response to perturbation (namely by effects in the activity of transcription factors), while the majority of mobile elements is still controlled by a functional piRNA pathway.

Increased TE expression in response to stress can mediate the production of heritable genetic variation

D. melanogaster is able to colonize diverse habitats, and to survive in dynamic environments that frequently fluctuate[22]. Upon exposure to threats, populations can benefit from responses that increase genetic variation[2,3], including those produced by transposition[4,5]. The assessment of retrotransposon copy number in the gDNA of F1 embryos (laid by females exposed to environmental perturbations) confirmed that increase of TE expression in fruit fly ovaries has the potential to mediate the production of heritable genetic variation. However, that was only true for the more frequent elements in *D. melanogaster* genome, and the relative contribution with new insertions was also dependent on the TE abundance. In other words, we observed that a more abundant TE was able to produce more novel insertions, when compared with a less abundant element. Notably, induced expression of *blood* (with 169 average number of copies at 25ºC) resulted in 61 more copies, whereas *roo* (with 113 average number of copies at 25ºC) produced about to 34 extra insertions, and *Juan* (with 36 average number of copies at 25ºC) 7 new copies that were transmitted to the next generation. We can therefore argue that the consequences for genome

architecture, and the ability to increase genetic variation, both driven by transposon mobilization, are positively correlated with TE abundance.

TE-specific responses to sub-optimal conditions can increase the adaptive potential

TEs have specific target preferences[21] and transposition into genomic regions that are near or within coding sequences can influence many aspects of gene function[43–45]. Previous studies showed that mobile elements preferentially insert into promoters of stress response genes, and that are able to synergistically trigger gene expression in response to threats[46] or increase tolerance to stimuli[47]. In the context of environmentally induced variation, we suggest that a precise response of certain elements can increase the adaptive potential of the novel genetic variation produced (evolvability). Specifically, we propose that *blood, Juan,* and *roo* are able to create variants adaptive to changes in temperatures, while *Juan* and *Tirant* would produce novel insertions more prone to confer resistance to cadmium. Despite evidence that heat shock genes are typically evolvable due to *P-element* insertions[47], this idea was never empirically tested for *blood, Juan, roo* and *Tirant*. The assessment of natural populations for insertions of those elements may help to clarify to what extent they contribute for the evolvability of genes that confer tolerance to different temperatures and presence of cadmium in the diet.

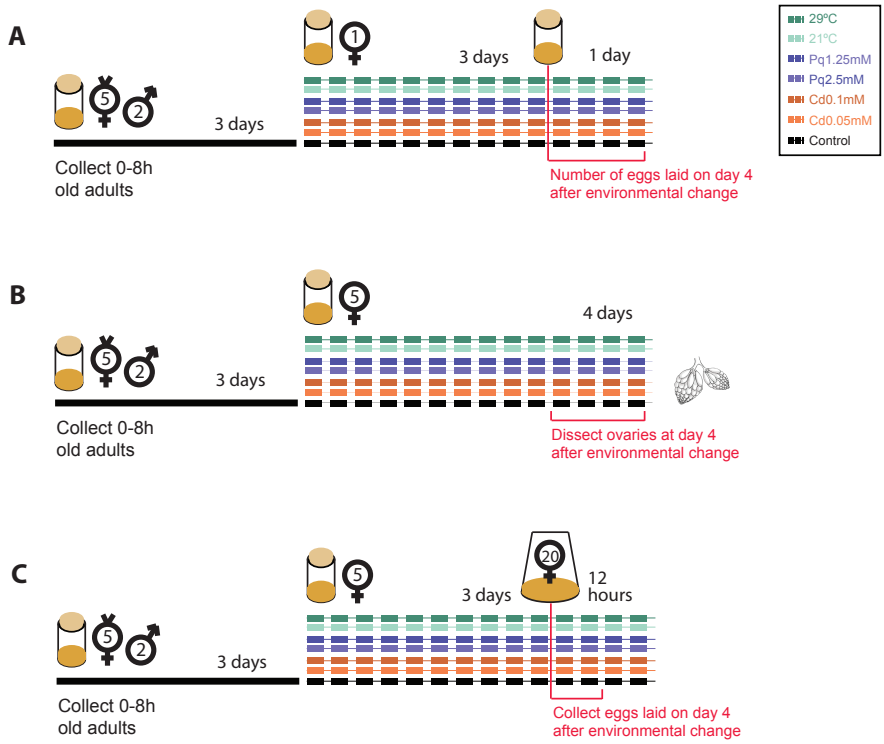
Author contributions

The author of this thesis performed experiments to evaluate fecundity and TE expression in response to stress, did statistical analysis and wrote this manuscript. Marta Marialva, together with Patrícia Beldade, also designed all experiments of this chapter. Joana Carvalho performed experiments to assess TE copy number and Ana Eugénio did antibody stainings.

Acknowledgements

We thank to David Duneau for helping with statistical analysis, and Triin Laos for valuable assistance with antibody staining. We are also thankful to Alisson Gontijo and Luis Teixeira for relevant feedback on this project. $SpnE^{616}$/TM3 and $SpnE^{100.37}$/TM3 stocks were provided by Vitor Barbosa; Harwich and w[1118] lines were obtained from Luis Teixeira laboratory.

Supplementary Information



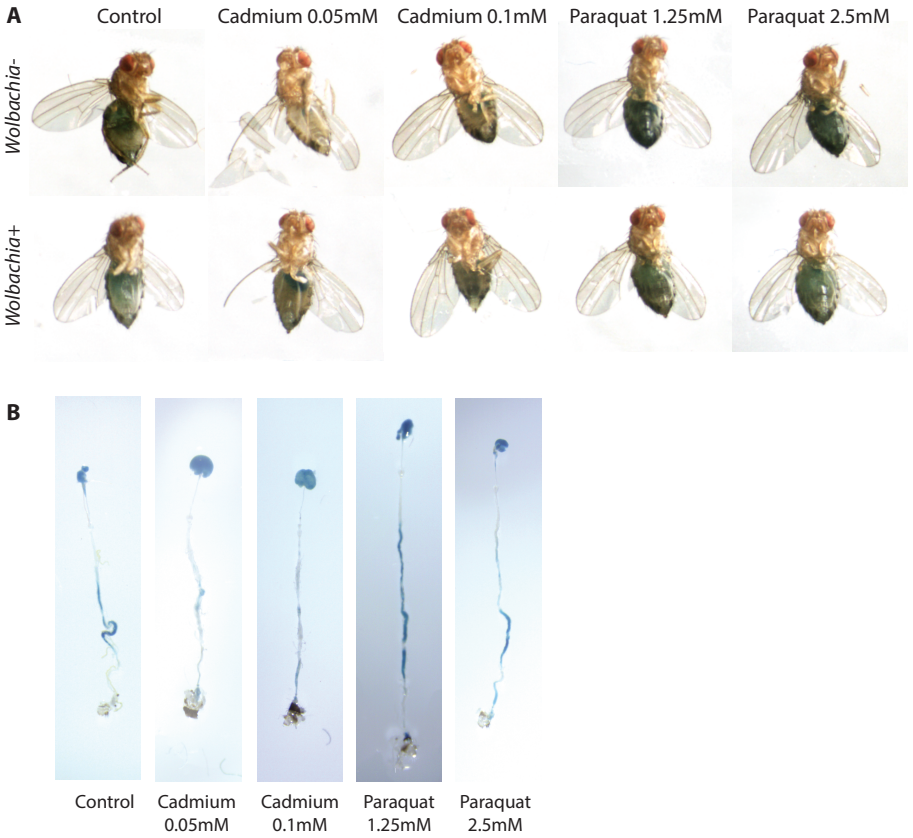
Supplementary Figure 4.1. Experimental setup for temperature, cadmium and paraquat treatments. Experimental setup for the assessment of fecundity **(A)**, TE expression in ovaries **(B)** and TE copy number in the next generation **(C)**. For all experiments we collected eight-hour-old adults and kept five females (with or without *Wolbachia*) with two males in the same vial, for three days. We then removed males and placed females in all experimental conditions for temperature (21ºC, 25ºC or 29ºC), cadmium (0mM, 0.05mM or 0.1mM), and paraquat (0mM, 1.25mM or 2.5mM) treatments. For the assessment of fecundity we kept one female per vial in each environmental condition for three days **(A)**. Females were then placed in new vials (at day three) and the number of eggs laid per female was determined at day four after environmental challenge. For the assessment of TE expression in ovaries, we kept five females per vial and dissected ovaries at day four after environmental perturbation **(B)**. For the assessment of TE copy number in the next generation we also kept five females per vial for four days **(C)**. At day four, we pooled 20 females per cage, let them lay eggs for 12 hours and then collected embryos for DNA extraction.

Experiment 1
Experiment 2
Temperature
Cadmium
Paraquat
Wolbachia-
Wolbachia+
Rpl32 expression (log)
Tbp expression (log)
vasa expression (log)
21°C
25°C
29°C
Cd0mM
Cd0.05mM
Cd0.1mM
P0mM
P1.25mM
P2.5mM

Supplementary Figure 4.2. Expression of control genes. Expression (in logarithmic scale, log) of *Rpl32, Tbp* and *vasa* in ovaries collected from females, infected (*Wolbachia+*) or not (*Wolbachia-*) with endosymbiont, and exposed to different abiotic treatments (temperature, cadmium and paraquat). Data was obtained from two independent experiments (Experiment 1 and 2). For more information about experimental setup see **Supplementary Figure 4.1B**). Each circle in the boxplots represents a pool of 8-10 pairs of ovaries; triangles correspond to the mean and black line to the median.

Temperature
Cadmium
Paraquat
Wolbachia-
Wolbachia+
E1
E2
blood
Juan
pogo
roo
Tirant
Relative TE expression (log)
25°C Cd0mM, Pq0mM
21°C
29°C
Cd0.05mM
Cd0.1mM
Pq1.25mM
Pq2.5mM

Supplementary Figure 4.3. *Wolbachia*-dependent effect of temperature, cadmium and paraquat in TE expression. Relative expression of *blood, Juan, pogo, roo* and *Tirant* in ovaries collected from females infected (*Wolbachia*+) or not (*Wolbachia*-) with endosymbiont and that experienced different environmental treatments. Those included several temperatures (21ºC, 25ºC and 29ºC), doses of cadmium (Cd0mM, Cd0.05mM and Cd0.1mM) or paraquat (Pq0mM, Pq1.25mM or Pq2.5mM). Data was collected in two independent experiments (Experiment 1 and 2). For more information about experimental setup see **Supplementary Figure 4.1B**). TE expression is shown in logarithmic scale (log) and is relative to *vasa* expression excepting that of paraquat treatments obtained in the second experiment and normalized against *Tbp*. Relative expression of transposons was then calibrated for each treatment and experiment with the median expression of control flies (those free of *Wolbachia*, and exposed to 25ºC, Cd0mM and Pq0mM) (Pfaffl method[40]). We tested the effect of treatment, experiment and the interaction between those two factors (categorical fixed variables) by using linear models (lm) with logarithmic distribution of the data. Interaction between treatment and experiment gives robustness of TE response: if interaction is statistically significant ($p<0.05$) TE response to environmental perturbation is considered not robust. Data for each abiotic factor and *Wolbachia* status was analyzed separately. We only show statistics (p-values, p) in the boxplots when at least one of the factors was statistically significant ($p<0.05$) for the respective environmental cue. If the interaction between experiment and treatment was statistically significant we performed pairwise comparisons using non-parametric Wilcoxon rank test. For more details about statistical analysis go to **Table 4.1**. Each circle in the boxplots represents a pool of 8-10 pairs of ovaries; triangles correspond to the mean and black line to the median. $0.001<$p-value<0.01 (**); $0.05<$p-value<0.01 (*).



Supplementary Figure 4.4. Feeding assay after ingestion of cadmium and paraquat. Female abdomens **(A)** and guts **(B)** with blue staining to confirm that adults are still feeding four days after beginning of cadmium (0.05mM and 0.1mM) and paraquat (1.25mM and 2.5mM) treatments (n=5 for all conditions).

Supplementary Table 4.1. Primers used for PCR and qPCR

Target	Forward primer sequence	Reverse primer sequence	Ref
blood	AACAATAGAAAGAAGCCACCGAAC	AGTCATGGACTATTGAGGGTGTTG	[48]
Juan	GGGGCAAAATTCTCAATGAA	GCGGAATATATGTGGGTTGC	
pogo	CCAGCGATAACGAAGAAAGC	GCTGCAAACCCATCCTTAAA	
roo	CGTCTGCAATGTACTGGCTCT	CGGCACTCCACTAACTTCTCC	[17]
Tirant	AAATGTTTGCCCCATCTCAG	GGGTCAGTGTGGCGTTATTT	
Rpl32	ATGCTAAGCTGTCGCACAAATG	GTTCGATCCGTAACCGATGT	
Tbp	GGCAAAGAGTGAGGACGACT	GAGCCGACCATGTTTTGAAT	[38]
vasa	AGCTTCCGGGATTCATTTTT	TCAGCGCTCGTAAAA TGTTG	[39]

Ref, reference

References

1. Armbruster P, Reed DH. Inbreeding depression in benign and stressful environments. *Heredity (Edinb)*. 2005;95(3):235-242. doi:10.1038/sj.hdy.6800721.

2. Pal C, Maciá MD, Oliver A, Schachar I, Buckling A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*. 2007;450(7172):1079-1081. doi:10.1038/nature06350.

3. Bjedov I, Tenaillon O, Gérard B, et al. Stress-induced mutagenesis in bacteria. *Sicence*. 2003;42(5):1404-1409. doi:10.1080/10409230701648494.

4. Junakovic N, Franco C Di, Barsanti P, Palumbo G. Transposition of Copia-like Nomadic Elements Can Be Induced by Heat Shock. *J Mol Evol*. 1986;24:89-93. http://link.springer.com/article/10.1007/BF02099955.

5. Ratner VA, Zabanov SA, Kolesnikova O V, Vasilyeva LA. Induction of the mobile genetic element Dm-412 transpositions in the Drosophila genome by heat shock treatment. *Proc Natl Acad Sci U S A*. 1992;89(12):5650-5654. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49350&tool=pmcentrez&rendertype=abstract.

6. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov D a. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol*. 2008;6(10):2109-2129. doi:10.1371/journal.pbio.0060251.

7. González J, Macpherson JM, Petrov D a. A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila melanogaster. *Mol Biol Evol*. 2009;26(9):1949-1961. doi:10.1093/molbev/msp107.

8. González J, Karasov TL, Messer PW, Petrov D a. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. *PLoS Genet*. 2010;6(4):33-35. doi:10.1371/journal.pgen.1000905.

9. García Guerreiro MP. What makes transposable elements move in the Drosophila genome? *Heredity (Edinb)*. 2012;108(5):461-468. doi:10.1038/hdy.2011.89.

10. Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE, Promislow DEL. Geographical Distribution and Diversity of Bacteria Associated with Natural Populations of Drosophila melanogaster. *Appl Environ Microbiol*. 2007;73(11):3470-3479. doi:10.1128/AEM.02120-06.

11. Werren JH, Baldo L, Clark ME. Wolbachia: master manipulators of

invertebrate biology. *Nat Rev Microbiol*. 2008;6(10):741-751. doi:10.1038/nrmicro1969.

12. Teixeira L, Ferreira A, Ashburner M. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol*. 2008;6(12):20753-22763. doi:10.1371/journal.pbio.1000002.

13. Jackson S, Nielsen DM, Singh ND. Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila. *BMC Evol Biol*. 2015;15(1):175. doi:10.1186/s12862-015-0452-8.

14. Abdullah MF, Borts RH. Meiotic recombination frequencies are affected by nutritional states in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 2001;98(25):14524-14529. doi:10.1073/pnas.201529598.

15. Juliano C, Wang J, Lin H. Uniting Germline and Stem Cells: the Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu Rev Genet*. 2011;(45):1-26. doi:10.1146/annurev-genet-110410-132541.Uniting.

16. Fablet M, Vieira C. Evolvability, epigenetics and transposable elements. *Biomol Concepts*. 2011;2(5):333-341. doi:10.1515/bmc.2011.035.

17. Specchia V, Piacentini L, Tritto P, et al. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature*. 2010;463(7281):662-665. doi:10.1038/nature08739.

18. Gangaraju VK, Yin H, Weiner MM, Wang J, Huang XA, Lin H. Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat Genet*. 2011;43(2):153-158. doi:10.1038/ng.743.

19. Izumi N, Kawaoka S, Yasuhara S, et al. Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. *RNA J*. 2013;19:896-901. doi:10.1261/rna.037200.112.

20. Rutkowski DT, Kaufman RJ. A trip to the ER: coping with stress. *Trends Cell Biol*. 2004;14(1):20-28. doi:10.1016/j.tcb.2003.11.001.

21. Wicker T, Sabot F, Hua-Van A, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8(12):973-982. doi:10.1038/nrg2165-c4.

22. Jean DR, Capy P. Genetic variation of Drosophila melanogaster natural populations. *Elsevier Publ*. 1988;4(88):106-111.

23. Betteridge DJ. What Is Oxidative Stress? *Metabolism*. 2000;49:3-8.

24. Buchon N, Silverman N, Cherry S. Immunity in Drosophila melanogaster - from microbial recognition to whole-organism physiology. *Nat Rev Immunol*. 2014;14(12):796-810. doi:10.1038/nri3763.

25. Shamovsky I, Nudler E. New insights into the mechanism of heat shock response activation. *Cell Mol Life Sci*. 2008;65(6):855-861. doi:10.1007/s00018-008-7458-y.

26. Dupas S, Carton Y, Poiriè M. Genetic dimension of the coevolution of virulence–resistance in Drosophila – parasitoid wasp relationships. *Heredity (Edinb)*. 2003;90(1):84-89. doi:10.1038/sj.hdy.6800182.

27. Corby-Harris V, Promislow DEL. Host ecology shapes geographical variation for resistance to bacterial infection in Drosophila melanogaster. *J Anim Ecol*. 2008;77(4):768-776. doi:10.1111/j.1365-2656.2008.01399.x.

28. Krebs R a, Feder ME. Natural variation in the expression of the heat-shock protein Hsp70 in a population of Drosophila melanogaster and its correlation with tolerance of ecologically relevant thermal stress. *Evolution (N Y)*. 1997;51(1):173-179. doi:10.2307/2410970.

29. Tantawy AO, Mallah GS. Heat resistance and geographical variation in Drosophila melanogaster and D. simulans. *Evolution (N Y)*. 1961;15:1-14.

30. Hoffmann AA, Anderson A, Hallas R. Opposing clines for high and low temperature resistance in Drosophila melanogaster. *Ecol Lett*. 2002;5:614-618.

31. Maroni G, Wise J, Young JE, Otto E. Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster. *Genetics*. 1987;117(4):739-744. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203245&tool=pmcentrez&rendertype=abstract.

32. Chrostek E, Marialva MSP, Esteves SS, et al. Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis. *PLoS Genet*. 2013;9(12):1-22. doi:10.1371/journal.pgen.1003896.

33. Mendes CC, Mirth CK. Stage-Specific Plasticity in Ovary Size Is Regulated by Insulin/Insulin-Like Growth Factor and Ecdysone Signalling in Drosophila. *Genetics*. 2015. doi:10.1534/genetics.115.179960.

34. Rzezniczak TZ, Douglas LA, Watterson JH, Merritt TJS. Paraquat administration in Drosophila for use in metabolic studies of oxidative stress. *Anal Biochem*. 2011;419(2):345-347. doi:10.1016/j.ab.2011.08.023.

35. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res*. 2004;64(15):5245-5250. doi:10.1158/0008-5472.CAN-04-0496.

36. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3(7):1-12. doi:10.1186/gb-2002-3-7-research0034.

37. Ponton F, Chapuis M-P, Pernice M, Sword GA, Simpson SJ. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in Drosophila melanogaster. *J Insect Physiol*. 2011;57(6):840-850. doi:10.1016/j.jinsphys.2011.03.014.

38. Matta BP, Bitner-Mathé BC, Alves-Ferreira M. Getting real with real-time qPCR: A case study of reference gene selection for morphological variation in Drosophila melanogaster wings. *Dev Genes Evol*. 2011;221(1):49-57. doi:10.1007/s00427-011-0356-6.

39. Khurana JS, Wang J, Xu J, et al. Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell*. 2011;147(7):1551-1563. doi:10.1016/j.cell.2011.11.042.

40. Pfaffl M. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29:2002-2007.

41. Navarro C, Bullock S, Lehmann R. Altered dynein-dependent transport in piRNA pathway mutants. *Proc Natl Acad Sci U S A*. 2009;106(24):9691-9696. doi:10.1073/pnas.0903837106.

42. Pelet S, Rudolf F, Nadal-ribelles M, Nadal E De, Posas F, Peter M. Transient Activation of the HOG MAPK Pathway Regulates Bimodal Gene Expression. *Science (80-)*. 2011;332:732-735.

43. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9(5):397-405. doi:10.1038/nrg2337.

44. Rebollo R, Romanish MT, Mager DL. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. *Annu Rev Genet*. 2011;46(1):120913153128008. doi:10.1146/annurev-genet-110711-155621.

45. Cowley M, Oakey RJ. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. *PLoS Genet*. 2013;9(1). doi:10.1371/journal.pgen.1003234.

46. Feng G, Leem YE, Levin HL. Transposon integration enhances expression of stress response genes. *Nucleic Acids Res*. 2013;41(2):775-789. doi:10.1093/nar/gks1185.

47. Chen B, Walser JC, Rodgers TH, et al. Abundant, diverse, and consequential P elements segregate in promoters of small heat-shock genes in Drosophila populations. *J Evol Biol*. 2007;20(5):2056-2066. doi:10.1111/j.1420-9101.2007.01348.x.

48. Handler D, Olivieri D, Novatchkova M, et al. A systematic analysis of

Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* 2011;30(19):3977-3993. doi:10.1038/emboj.2011.308.

Chapter V

Wolbachia reduces severity of hybrid dysgenesis in a genotype-dependent way

5.1 Abstract

Hybrid dysgenesis is a syndrome characterized by problems in the fertility of hybrids[1] that is triggered by parental incompatibilities in transposable element (TE) composition[2]. In *Drosophila melanogaster* this phenomenon has been described for three transposable elements, namely *hobo, I-element and P-element*, and is usually associated with an increase in mutation rate[1–3]. The invasion of *P-element* transposon in natural populations has been associated with a concurrent global replacement of *Wolbachia* strains[4]. *Wolbachia* are maternally transmitted α-proteobacteria that increase their own fitness by influencing the host's ability to reproduce[5]. However, the effect of this endosymbiont on hybrid dysgenesis, which is classified as a syndrome of sterility, has never been empirically tested. Here, we characterized a panel of 86 natural genotypes, and observed four main phenotypes according to the percentage of gonadal dysgenesis they induce when crossed with laboratory stocks. Those phenotypes grouped genotypes according to their susceptibility to and activity of *P-elements*. After controlling for the host genetic background, we further assessed the effect of the endosymbiont in the severity of hybrid dysgenesis induced by four natural genotypes. We generally observed that *Wolbachia* is able to reduce, in a genotype-dependent way, the severity of dysgenic phenotypes in the number of eggs laid, and proportion of eggs that are able to develop until becoming adults. We propose that the endosymbiont *Wolbachia* may have facilitated *P-element* invasion in nature by increasing the fitness of dysgenic hybrids.

5.2 Introduction

Transposable elements (TEs) are diverse and active DNA sequences that constitute a large proportion of host genomes[6,7]. The profile of transposons can however greatly differ between species[7–9] and populations[10,11]. When inter- or intraspecific crosses between mating pairs with incompatible TE composition occur, hybrids have reproductive problems that result from uncontrolled TE activity in their germline[2,12–14]. This phenomenon is commonly referred to as

hybrid dysgenesis and may have significantly contributed for species diversification[15–18].

In *Drosophila melanogaster* natural populations, a recent invasion of *P* and *I elements* have been described to induce gonadal dysgenesis but no associated events of speciation were observed so far[1,2,19]. Flies containing *P-element* insertions are referred to as P strain; flies without *P-element* insertions in the genome are classified as M strain. When M strain females mate with P strain males, the progeny is sterile (dysgenic) due to *P-element* mobilization in the germline. On the contrary, a reciprocal cross between P strain females and M strain males ends up with genetically identical and fertile descendants (non-dysgenic)[1]. In contrast to M strain, P strain flies express and maternally transmit PIWI interacting small RNAs (piRNAs) that are necessary to epigenetically restrain *P-element* activity[2]. The maternal transmission of piRNAs is crucial for the repression of transposons in the germline of hybrids. Interestingly, as dysgenic females age, *P-elements* can be silenced through *de novo* production of piRNAs and though they are initially unable to restrain *P-element* activity[20].

In the wild, the invasion of *P-element* transposon has been associated with a global replacement of *Wolbachia* strains[4]. *Wolbachia* are intracellular α-proteobacteria that infect numerous arthropod and nematode species[21,22]. This maternally transmitted endosymbiont is well known as a reproductive parasite that affects several aspects of host fitness[5]. Although hybrid dysgenesis is a syndrome characterized by problems in the fertility of hybrids[1], the effect of *Wolbachia* in this phenomena has never been empirically tested.

In the current study, we addressed this question by first characterizing a panel of 86 natural genotypes according to *Wolbachia* status and to the percentage of gonadal dysgenesis they induce when crossed with laboratory stocks. After controlling for the host genetic background, we focused on the effect of endosymbiont in the severity of hybrid dysgenesis induced by four natural genotypes of *D. melanogaster*. Overall, we observed that *Wolbachia* is able to reduce, in a genotype-dependent way, the severity of dysgenic phenotypes in the number of eggs laid per female, and in the proportion of eggs that develop

until becoming adults. Our analysis suggests that the endosymbiont may have promoted *P-element* invasion in natural populations by increasing fitness of dysgenic hybrids.

5.3 Materials and Methods

Drosophila melanogaster stocks and maintenance

We established a panel of 86 *D. melanogaster* lines that were collected in different months of 2014, at Yesiloz in Turkey (**Table 5.1** and **Supplementary Table 5.1**). Each stock derived from mated females followed by three generations of full-sibling inbreeding. Harwich and w[1118] stocks (with and without *Wolbachia*) were provided by Luis Teixeira. We kept flies in standard cornmeal food[23] at 21°C, 55% humidity and 12h day-dark cycles, unless otherwise mentioned.

Table 5.1. Temporal sampling of isofemales

Time of collection	Number of lines	Average temperature (ºC)	Presence of *Wolbachia* (%)
June	15	19.6	73.3
July	7	22.6	71.4
August	22	22.5	81.8
September	20	18.7	80
October	22	13.4	77.3

All samples were collected in Yesiloz, Turkey (2014) at 40.23º latitude, 32.26º longitude and 704m of altitude. Average temperature was calculated based on our daily measurements that were obtained from data loggers iButton hygrochron DS 1923 (Maxim Integrated).

Gonadal dysgenesis tests

To group each genotype according to the ability to induce hybrid dysgenesis we performed two crosses, A (w[1118] females x tested males) and A* (tested females x Harwich males) (**Figure 5.1A**), and scored the percentage of undeveloped ovaries (percentage of gonadal dysgenesis) for each tested genotype and cross (**Supplementary Table 5.1**). We scored more than 35 F1 females per genotype and cross-type (A and A*). Those females derived from three independent crosses. Simultaneously, and as a control, w[1118] females

were mated with Harwich males which resulted in 100% dysgenic ovaries. Each genotype was classified as M, Q, P and P' according to the percentage of gonadal dysgenesis obtained for each cross-type (**Figure 5.1A**). Crosses and gonadal dysgenesis tests were performed at 29ºC using w[1118] and Harwich stocks free of *Wolbachia* infection. For more information about experimental procedures see Onder B. and Kasap O. (2014)[24].

gDNA extraction

gDNA was extracted from the 86 isofemale lines that were derived from natural populations (**Supplementary Table 5.1**), and from three w[1118] stocks, 1) cleared of *Wolbachia* infection, 2) infected with *wMel* strain, and 3) infected with *wMelCS* strain. We used DNeasy Blood & Tissue Kit (Qiagen) for the extraction and pooled ten adults per sample. Tissues were homogenized using pestels in 180µl of buffer ATL. We used 200µl of buffer AE to elute gDNA and treated all samples with 0.2mg/ml of RNAse (Invitrogen) for 20 min at 55ºC. Sample concentration was measured in Nanodrop ND-1000 Spectrophotometer and stored at -20°C.

PCR to characterize lines according to* Wolbachia *infection

We performed multiplex PCR reactions to characterize all 86 natural genotypes (**Supplementary Table 5.1**) according to presence of *Wolbachia*; w[1118] stocks were used as positive controls. We were able to distinguish *wMel* from *wMelCS* strains using primer sequences described in **Supplementary Table 5.2**. Each reaction contained 0.25U of GoTaq (Promega), 1x reaction buffer, 1.5mM MgCl2, 0.4µM of primer mix and 1µl of 4µM gDNA. Primer mix solution contained all primers described in **Supplementary Table 5.2**. Thermal cycling protocol was the same for all reactions: 10 min at 95°C; 40 cycles of 95°C for 30s, 60°C for 1 min and 72°C for 30s; 5 min at 72ºC.

Assessment of fecundity and percentage of eclosion

We selected four *D. melanogaster* lines (2.203, 2.214, 4.104 and 6.51, **Supplementary Table 5.1**) that were naturally infected with *wMel* strain of *Wolbachia.* To obtain stocks free of infection we treated those lines with tetracycline (0.05mg/ml) for two generations. Gut flora content was homogenized between tetracycline treated and non-treated flies as previously described[25]. We induced hybrid dysgenesis by crossing tested females (with or without *Wolbachia*) with Harwich males; F1 non-dysgenic flies were obtained by crossing Harwich females (with or without *Wolbachia*) with tested males. We then assessed fecundity (number of eggs laid per female), and percentage of eclosion (proportion of eggs that develop until becoming adults) for the F1 females collected from both crosses. We kept one female with two males per tube and egg lay was counted daily for ten days. After 15 days we counted number of adults eclosed per vial. Flies were maintained at 25ºC during all experiment.

Statistical analyses

All statistical analyses were done in R (version 3.0.3)[26].

To test correlation between the average temperature at the time of collection and percentage of gonadal dysgenesis for cross A and A* we performed a Pearson correlation analysis. P-values provide significance of correlation; r provides the signal and degree of correlation.

To assess the effect of *Wolbachia* in the ability to induce hybrid dysgenesis (comparison between non-dysgenic and dysgenic crosses for each *Wolbachia* status) and in each type of cross (comparison between presence and absence of endosymbiont for each type of cross) we performed pairwise comparisons using the non-parametric Wilcoxon rank test. Those tests were performed for the total number of eggs laid and total percentage of eclosion assessed from day one to day five, from day six to day ten, and from day one to day ten. Each genotype was analyzed separately.

To assess the effect of *Wolbachia* and type of cross (categorical fixed variable with four levels: non-dysgenic *Wolbachia*-, non-dysgenic *Wolbachia*+, dysgenic *Wolbachia*- and dysgenic *Wolbachia*+) as a function of time (continuous fixed variable) we used general linear models (glm) including the interaction between both factors in the model. We assigned poisson distribution to test effects in fecundity and binomial distribution to test effects in the proportion of eggs that develop until becoming adults. Each genotype was analyzed separately; we obtained significance values by performing same pairwise comparisons described above.

5.4 Results

Characterization of natural genotypes according to their ability to induce hybrid dysgenesis

Our characterization of natural genotypes collected from Yesiloz showed that 63.95% were unable to induce hybrid dysgenesis in both cross-types, 23.26% behaved as M strains, 6.98% as P strains and 5.81% showed undeveloped ovaries in both crosses (cross A and A*, **Figure 5.1A and B**). Environmental factors such as temperature can affect the severity of dysgenic traits[27,28]. However, this effect was only observed for few *D. melanogaster* lines and upon exposure to constant and unnatural laboratory environments. In agreement with a previous description reporting no correlation between the percentage of gonadal dysgenesis and several (natural) environmental variables such as temperature, humidity, rainfall and altitude[24], we also found no association for temperature in our population (Pearson correlation: cross A, r=-0.08, p-value=0.48; cross A*, r=-0.02, p-value=0.85) (**Figure 5.1C**).

For the characterization of *Wolbachia* status we observed that the majority of the natural lines analyzed here were infected with the endosymbiont (77%, **Figure 5.1D**). Most genotypes were infected with *wMel* strain, and only one out of the 67 infected genotypes contained the variant *wMelCS* (**Supplementary Table 5.1**). We found no difference in the percentage of gonadal dysgenesis between infected and non-infected flies for both crosses (cross A and A*, **Figure 5.1D**).

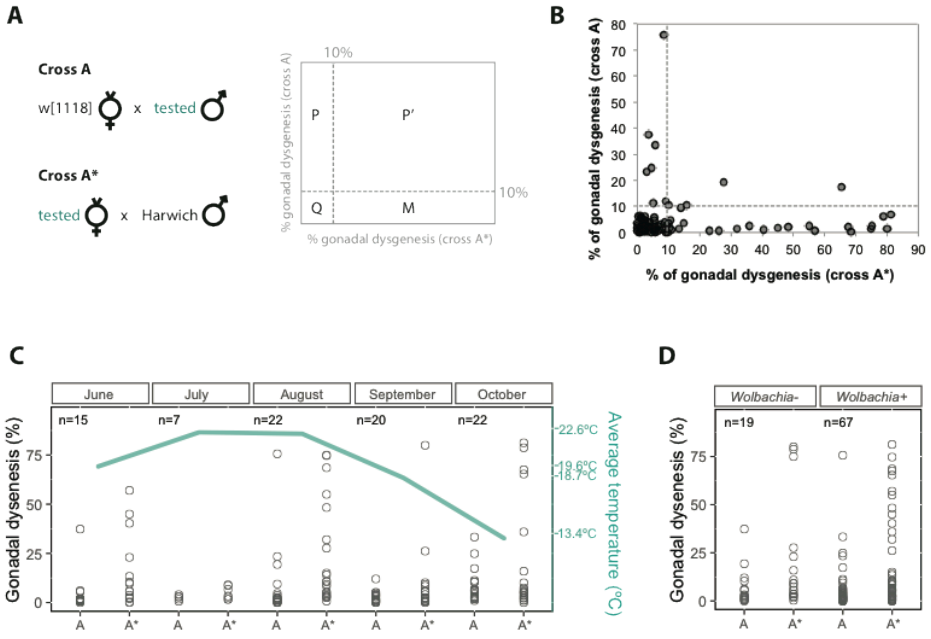


Figure 5.1. Characterization of *D. melanogaster* natural genotypes according to their ability to induce hybrid dysgenesis. Crosses performed (cross A and A*) to classify genotypes according to the percentage of gonadal dysgenesis **(A)**. w[1118] genotype does not have *P-element* insertions and was used as M strain; Harwich genotype contains *P-element* insertions and was used as P strain. Tested lines were classified into four cytotypes (Q, M, P and P') depending on the percentage of gonadal dysgenesis observed for both crosses. Genotypes that had less than 10% of undeveloped ovaries in both crosses (A and A*) were grouped into the Q cytotype whereas genotypes with more than 10% dysgenic ovaries for both crosses were classified as P'. M strains are those with high percentage (>10%) of undeveloped ovaries in cross A*; P strains are those with high percentage (>10%) of undeveloped ovaries in cross A. Classification of 86 genotypes according to the percentage of gonadal dysgenesis in cross A and A* **(B)**. Percentage of gonadal dysgenesis obtained for crosses A and A* established using genotypes that were collected in different months (June, July, August, September and October) **(C)**. Average temperature (ºC) at the time of collection is highlighted in blue. Percentage of gonadal dysgenesis obtained for crosses A and A* established using genotypes infected or not with *Wolbachia* **(D)**. Each dot represents a genotype **(B, C** and **D)**; total number of genotypes collected per month **(C),** or with a given *Wolbachia* status **(D)** are represented by n.

Wolbachia *reduces severity of hybrid dysgenesis in a genotype-dependent way*

To study the effect of endosymbiont in hybrid dysgenesis we chose four M genotypes naturally infected with *wMel* (**Supplementary Table 5.1**) and obtained correspondent *Wolbachia* free stocks. This allowed us to assess the effect of endosymbiont for each host genetic background. Overall, *Wolbachia* was able to reduce severity of hybrid dysgenesis though the relative effects in fecundity and percentage of eclosion were dependent on the genotype (**Figure 5.2**). We observed that those effects of *Wolbachia* were only observed in the context of hybrid dysgenesis, as no significant effects for the presence of endosymbiont (in fecundity and percentage of eclosion) were detected for non-dysgenic crosses regardless of the genetic background. For the females 2.203, we only observed phenotypes of dysgenic syndrome (reduction in the fecundity of dysgenic in comparison with non-dysgenic females) for flies free of *Wolbachia* infection (**Figure 5.2A**). This effect was dependent on the age of females, as it seems that the fecundity of dysgenic flies without endosymbiont was only reduced in the early life of those females (**Figure 5.2A** and **Supplementary Figure 5.1**). For the genotype 2.214, we observed symptoms of dysgenic syndrome for flies with and without *Wolbachia* though deleterious effects in fecundity were more severe for dysgenic females free of endosymbiont infection (comparison between dysgenic females with and without *Wolbachia*) (**Figure 5.2B**). Interestingly, the fecundity of dysgenic flies infected with endosymbiont was mostly affected in early time points (until five days old) but the fecundity of those without infection was reduced for all ten days assessed (**Supplementary Figure 5.1**). For the genotype 4.104, we observed that the dysgenic females with *Wolbachia* lay more eggs than the respective non-dysgenic flies (**Figure 5.2C**) and that this effect is only detected in their first five days of life (**Supplementary Figure 5.1**). Interestingly, for all genotypes described above (2.203, 2.214 and 4.104) we did not observe effects in the proportion of eggs that develop until becoming adults (**Figure 5.2A** to **C**). On the contrary, for the 6.51 genotype we observed effects for the percentage of eclosion but not for

fecundity (**Figure 5.2D** and **Supplementary Figure 5.1**). In this case, the phenotype of dysgenic syndrome was only observed for females free of *Wolbachia*, and the percentage of eclosion of eggs laid by dysgenic females without endosymbiont was reduced during their early life (until six days old) but not later (from seven to ten days old) (**Figure 5.2D**).

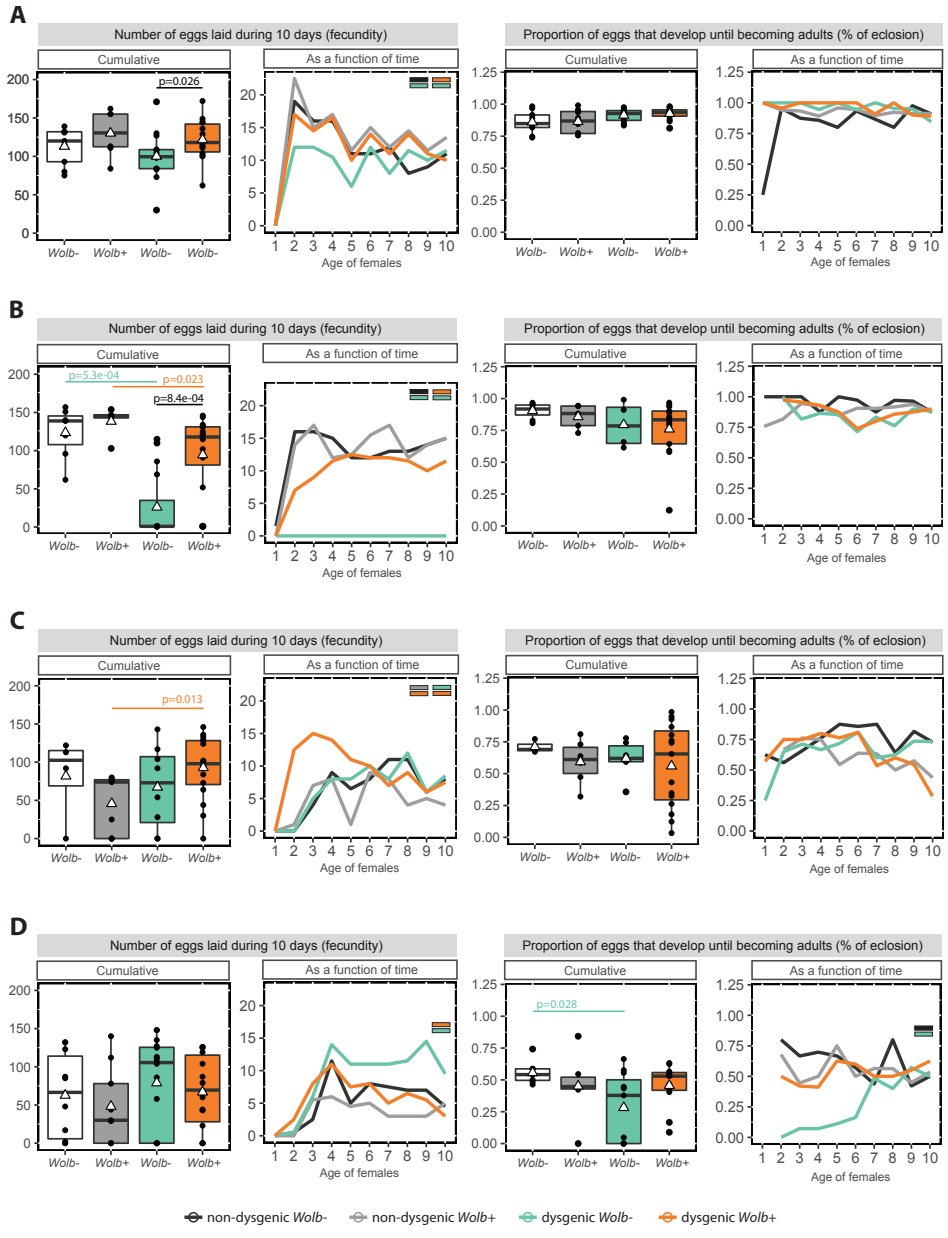


Figure 5.2. Effect of *Wolbachia* in hybrid dysgenesis. We assessed the effect of *Wolbachia* in the context of hybrid dysgenesis using four M genotypes: 2.203 **(A)**, 2.214 **(B)**, 4.104 **(C)** and 6.51 **(D)**. For each genotype we obtained non-dysgenic and dysgenic F1 females infected (*Wolb+*) and not infected (*Wolb-*) with endosymbiont. We then assessed the number of eggs laid for ten days (fecundity) and the proportion of those that develop until becoming adults (percentage of eclosion). The general effect of *Wolbachia* was tested using the total number of eggs laid during ten days and the total proportion of those that develop until becoming adults (cumulative). Pairwise comparisons were performed using Wilcoxon rank test; statistically significant comparisons are shown in the boxplots with respective p-values. Each dot in boxplots (cumulative) represents data for a single female; triangles represent the mean and line the median. We also assessed the effect of endosymbiont and type of cross (categorical fixed variable with four levels: *wild-type Wolbachia-*, *wild-type Wolbachia+*, dysgenesis *Wolbachia-* and dysgenesis *Wolbachia+*) as a function of time (age of females that is used as continuous fixed variable) in fecundity and percentage of eclosion using general linear models (glm). The interaction between those two factors was included in the analysis and pairwise comparisons were performed to test the effect of *Wolbachia* per type of cross and the effect of *Wolbachia* in the ability to induce hybrid dysgenesis (compare non-dysgenic with dysgenic F1 females for each *Wolbachia* status) over time. We represent by colored boxes the comparisons that are statistically significant ($p<0.05$) and that interact with the age of females; each color represents type of cross and *Wolbachia* status (black: non-dysgenic *Wolbachia* negative; gray: non-dysgenic *Wolbachia* positive; blue: dysgenic *Wolbachia* negative; orange: dysgenic *Wolbachia* positive). Lines in graphs (as a function of time) connect the median for the number of eggs laid and the proportion of eggs that develop until becoming adults, at each time point.

5.5 Discussion

Natural genotypes show high variability in the severity of gonadal dysgenesis

A previous analysis of natural genotypes for their ability to induce hybrid dysgenesis when crossed with laboratory stocks revealed high variation in *P-element* activity and susceptibility[24]. We obtained similar results for our characterization of *D. melanogaster* flies collected in Yesiloz, Turkey. According to the percentage of gonadal dysgenesis observed for hybrids, genotypes were classified into Q, M, P and P'. As we previously described, flies containing *P-element* insertions are referred to as P strain, and flies without *P-element*

insertions in the genome are classified as M strain. Although 23% of our lines behaved phenotypically as M strains, we cannot exclude the possibility that those genotypes contain inactive and/or incomplete *P-element* insertions. This phenomena was previously observed for other fruit fly populations[24] and a more comprehensive analysis of *P-element* insertions in our panel would help to clarify the frequency of this. Interestingly, we also observed flies that were resistant or susceptible to *P-element* activity, regardless of the mating scheme (Q and P' cytotype, respectively). We propose that Q genotypes should have inactive *P-element* insertions and express piRNAs with sequence homology to this transposon, though P' genotypes should have TE incompatibility with Harwich and w[1118] lines for other transposons than *P-elements*. As a consequence, P' males would induce gonadal dysgenesis when mated with w[1118] males through *P-element* activation, and P' females would induce gonadal dysgenesis when crossed with Harwich males through the activation of another TE.

Wolbachia *may confer genotype-dependent protection against* P-element *jumping*

A global characterization of *Wolbachia* strains infecting *D. melanogaster* stocks revealed that *wMel* has been replacing *wMelCS* strains in the wild[4]. In agreement with those results, we also observed that 98.5% of the isofemale lines containing *Wolbachia* were infected with *wMel* strain. Given the strong association between the replacement of *Wolbachia* strains and the colonization of *D. melanogaster* genomes by *P-elements*[4], we suggested that *wMel* might have facilitated *P-element* invasion by mitigating dysgenesis phenotypes. In particular, and taking into account that fecundity and hatch rates are restored as dysgenic females age[20], we further hypothesized that the recovery of dysgenic flies would be faster when infected with *wMel*. We observed that, when there was a significant effect for the genotypes 2.203 and 6.51, *wMel* benefited dysgenic females and that those behaved as non-dysgenic hybrids immediately after eclosion. We also observed that the rate of recovery with age for 2.214

dysgenic hybrids is faster when in the presence of *Wolbachia*. We therefore suggest that the presence of endosymbiont is able to manipulate *D. melanogaster* development in the context of hybrid dysgenesis for 2.203 and 6.51 genotypes, and to influence the rate of recovery with age for the genotype 2.214. In both scenarios the presence of endosymbiont seems to improve the fitness of dysgenic females though the mechanism that mediates this host-TE-*Wolbachia* interaction is still unknown.

As previously described, *Wolbachia* genes are differently expressed during *D. melanogaster* development[29]. It would be interesting to do a similar analysis, for both endosymbiont and host genes, upon induction of dysgenesis syndrome. This would help to clarify which genes and processes are involved in *Wolbachia*-mediated recovery from dysgenesis. Also, the sequencing of piRNA clusters of dysgenic and non-dysgenic hybrids (infected or not with *Wolbachia*) would allow us to assess whether *wMel* is able to protect host genomes from *P-element* activity by promoting the production of piRNAs homologous to *P-element* sequences through novel insertions in clusters during *D. melanogaster* development.

Interestingly, we observed that the effect of *wMel* in dysgenic traits, such as fecundity and percentage of eclosion, was dependent on host genotype. Although our analysis of two polymorphic markers allowed us to discriminate among *Wolbachia* strains, it is still possible that each fruit fly stock (2.203, 2.214, 4.104 and 6.51) was infected by *wMel* strains that differ in their genomic sequence by containing different single nucleotide polymorphisms, for example. As a consequence, we cannot exclude the hypothesis that allelic variation in *Wolbachia* or the interaction between those and *Drosophila* genotypes is at the basis for the different phenotypes observed. The sequencing of *Wolbachia* genotypes would be important to disentangle those relationships.

Recent studies using natural populations document TEs as important sources of genetic variation[30–32]. The ability to produce novel alleles, including those that result from transposon activity, can dictate the capacity of an organism to adapt

to new environments[33]. We can therefore suggest that the colonization of *D. melanogaster* genomes by *P-elements* could have played an important role for the dispersion of fruit fly populations and the colonization of new habitats. Also, the close relationship established between hosts and their maternally transmitted endosymbionts could explain how *Wolbachia* could have (indirectly) benefited from *P-element* invasion. In other words, novel genetic variation produced by increased activity of *P-elements* could directly benefit host fitness in new environmental conditions; higher fitness of infected hosts results in higher transmission of *Wolbachia* to the next generation. On the other hand, as *Wolbachia* reduces the severity of dysgenesis traits it benefits *P-element* spread in natural populations. The nature of those interactions established between host, endosymbiont and TE seem to have contributed for the fast spread of *P-elements* and for the replacement of *Wolbachia* strains in *D. melanogaster* natural populations.

Acknowledgements

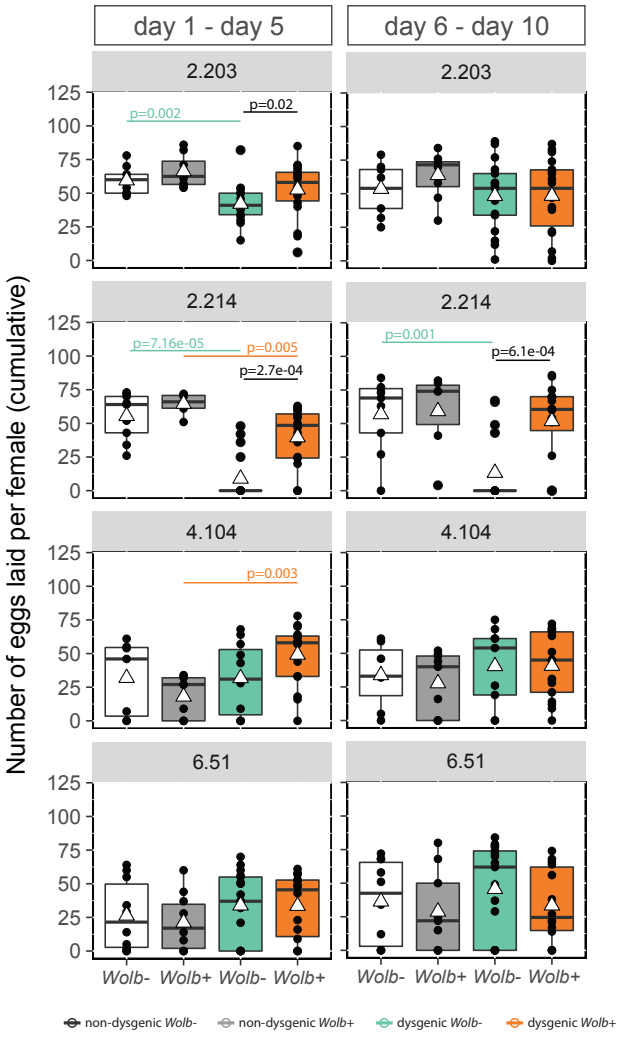
We obtained w[1118] and Harwich lines from Luis Teixeira. We also thank to Alisson Gontijo, Élio Sucena, and Luis Teixeira for all the productive discussions and input in the project, and to Liliana Vieira (fly facility) for providing great amounts of *D. melanogaster* food. We would also like to acknowledge Oscar Ruiz for reading this chapter and for providing significant input on the study.

Author contributions

This project was performed in collaboration with Banu Önder (BÖ) laboratory that established the panel of 86 *D. melanogaster* lines and characterized them according to percentage of gonadal dysgenesis.

The author of this dissertation characterized the panel of 86 genotypes according to the presence of *Wolbachia*, did statistical analysis and wrote this manuscript. Ana Eugénio performed experiments for the effect of *Wolbachia* in hybrid dysgenesis over time. Marta Marialva, Banu Önder and Patrícia Beldade designed experiments.

Supplementary Information



Supplementary Figure 5.1. Effect of *Wolbachia* in hybrid dysgenesis is affected by the age of dysgenic females. We assessed the effect of *Wolbachia* in the context of hybrid dysgenesis using four M genotypes: 2.203, 2.214, 4.104 and 6.51. For each genotype we obtained non-dysgenic and dysgenic F1 females infected (*Wolb+*) and not infected (*Wolb-*) with endosymbiont. We then assessed the total number of eggs laid per female during the first (day one to day five) and the following (day six to day ten) five days. Pairwise comparisons were performed using Wilcoxon rank test; statistically significant comparisons are shown in the boxplots with respective p-values. Each dot represents data for a single female; triangles represent the mean and line the median.

Supplementary Table 5.1. Information of *D. melanogaster* stocks derived from natural populations of Yesiloz, Turkey. GD stands for the percentage of gonadal dysgenesis scored for each genotype when performing cross A and cross A* (for more information about crosses see **Figure 5.1A**). Average temperature was calculated using on our daily measurements that were obtained from data loggers iButton hygrochron DS 1923 (Maxim Integrated). Genotypes selected to test the effect of *Wolbachia* in hybrid dysgenesis are highlighted in blue. This table is provided in digital format.

Supplementary Table 5.2. Primers used to characterize *Wolbachia* infection.

Target	Forward primer sequence	Reverse primer sequence	Ref
wMel (IS5 WD516/7)	CCATCAAGGTCTCTTTCA	TGCAAGGAAAACTAAACCAG	[4]
wMelCS (IS5 WD1310)	AGGAGAACTGGTCTACGC	TGTTGCTGAGCTTTGCT	[4]
Rpl32	ATGCTAAGCTGTCGCACAAATG	GTTCGATCCGTAACCGATGT	

Ref, reference

References

1. Kidwell, M., Kidwell, J. & Sved, J. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traitsincluding mutation, sterility and male recombination. **85,** 1973–1975 (1976).

2. Brennecke, J. *et al.* An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (80-.).* **1387,** 1387–1392 (2008).

3. Yannopoulos, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P. & Louis, C. hobo is responsible for the induction of hybrid dysgenesis by strains of Drosophila melanogaster bearing the male recombination factor 23.5MRF. *Cell* **49,** 487–495 (1987).

4. Riegler, M., Sidhu, M., Miller, W. J. & O'Neill, S. L. Evidence for a global Wolbachia replacement in Drosophila melanogaster. *Curr. Biol.* **15,** 1428–1433 (2005).

5. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6,** 741–751 (2008).

6. Kaminker, J. *et al.* The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol.* **3,** 1–20 (2002).

7. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

8. Lerat, E., Burlet, N., Biémont, C. & Vieira, C. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene* **473,** 100–109 (2011).

9. Granzotto, A., Lopes, F. R., Vieira, C. & Cararеto, C. M. A. Vertical inheritance and bursts of transposition have shaped the evolution of the BS non-LTR retrotransposon in Drosophila. *Mol. Genet. genomics* **286,** 57–66 (2011).

10. Vieira, C. & Biémont, C. Transposable element dynamics in two sibling species: Drosophila melanogaster and Drosophila simulans. *Genetica* **120,** 115–123 (2004).

11. Vieira, C. *et al.* A comparative analysis of the amounts and dynamics of transposable elements in natural populations of Drosophila melanogaster and Drosophila simulans. *J. Environ. Radioact.* **113,** 83–6 (2012).

12. Labrador, M., Farré, M., Utzet, F. & Fontdevila, A. Interspecific Hybridization Increases Transposition Rates of Osvaldo. *Mol. Biol. Evol* **16,** 931–937 (1999).

13. Josefsson, C., Dilkes, B. & Comai, L. Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. *Curr. Biol.* **16,** 1322–1328 (2006).

14. Akkouche, A. *et al.* Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. *EMBO Rep.* **14,** 458–64 (2013).

15. Pascale, E., Valle, E. & Furano, A. V. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. *Proc. Natl. Acad. Sci.* **87,** 9481–9485 (1990).

16. Dobigny, G. *et al.* LINE-1 amplification accompanies explosive genome repatterning in rodents. *Chromosom. Res.* **12,** 787–793 (2005).

17. de Boer, J. G., Yazawa, R., Davidson, W. S. & Koop, B. F. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* **8,** 422 (2007).

18. Ray, D. A. *et al.* Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. *Genome Res.* **18,** 717–728 (2008).

19. Rebollo, R., Horard, B., Hubert, B. & Vieira, C. Jumping genes and epigenetics: Towards new species. *Gene* **454,** 1–7 (2010).

20. Khurana, J. S. *et al.* Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell* **147,** 1551–1563 (2011).

21. Taylor, M. J. & Hoerauf, A. Wolbachia bacteria of filarial nematodes. *Parasitol. Today* **15,** 437–442 (1999).

22. Chandler, J. A., Lang, J., Bhatnagar, S., Eisen, J. A. & Kopp, A. Bacterial communities of diverse Drosophila species: Ecological context of a host-microbe model system. *PLoS Genet.* **7,** (2011).

23. Mendes, C. C. & Mirth, C. K. Stage-Specific Plasticity in Ovary Size Is Regulated by Insulin/Insulin-Like Growth Factor and Ecdysone Signalling in Drosophila. *Genetics* (2015). doi:10.1534/genetics.115.179960

24. Onder, B. S. & Kasap, O. E. P element activity and molecular structure in Drosophila melanogaster populations from Firtina Valley, Turkey. *J. Insect Sci.* **14,** 16 (2014).

25. Chrostek, E. *et al.* Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis. *PLoS Genet.* **9,** 1–22 (2013).

26. R-Core-Team. R - A language and environment for statistical computing. *https://www.r-project.org/* (2012).

27. Schaefer, R. E., Kidwell, Z. M. G. & Fausto-sterling, A. Hybrid dysgenesis in Drosophila melanogaster: morphological and cytological studies of ovarian dysgenesis. *Public Health* 1141–1152 (1979).

28. Dramard, X., Heidmann, T. & Jensen, S. Natural epigenetic protection against the I-factor, a Drosophila LINE retrotransposon, by remnants of ancestral invasions. *PLoS One* **2,** e304 (2007).

29. Gutzwiller, F. *et al.* Dynamics of Wolbachia pipientis Gene Expression Across the Drosophila melanogaster Life Cycle. *Genes|Genomes|Genetics* **5,** 2843–2856 (2015).

30. González, J., Lenkov, K., Lipatov, M., Macpherson, J. M. & Petrov, D. a. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol.* **6,** 2109–2129 (2008).

31. González, J., Karasov, T. L., Messer, P. W. & Petrov, D. a. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. *PLoS Genet.* **6,** 33–35 (2010).

32. Cridland, J. M., Macdonald, S. J., Long, A. D. & Thornton, K. R. Abundance and Distribution of Transposable Elements in Two Drosophila QTL Mapping Resources. *Mol. Biol. Evol.* **30,** 2311–2327 (2013).

33. Fablet, M. & Vieira, C. Evolvability, epigenetics and transposable elements. *Biomol. Concepts* **2,** 333–341 (2011).

Chapter VI

Discussion and Perspectives

6.1 Brief overview and conclusions

Phenotypic variation is a universal property of biological systems and heritable phenotypic variation is the fuel of evolutionary changes[1]. The environment contributes to shaping phenotypic variation by affecting: mutation that produces novel genetic variants, development that translates genotype into phenotype[2], and selection that eliminates unsuitable phenotypes. This project focused on the generation of genetic variation and how environmental and genetic factors may affect mutation driven by transposable element (TE) activity in the gonads.

Transposable elements (TEs) are important sources of genetic variation and studies concerning transposon population dynamics advanced that multiple waves of increased TE activity have occurred during evolution[3,4]. However, there are several caveats with respect to our understanding of the genetic and environmental factors that mediate such busts in transposition.

This thesis aimed at providing new insights about the genetic basis for TE activity (**Chapter 2**) and the importance of biotic and abiotic factors in affecting transposon expression in *D. melanogaster* ovaries (**Chapter 3 and 4**). In particular, we investigated the importance of *Wolbachia* on transposon responses to temperature (**Chapter 3 and 4**), as well as to exposure to heavy metals and oxidative stress (**Chapter 4**). We also addressed to what extent those effects were dependent on genotype by analysing the impact of genetic-by-environment-by-environment (GxExE) interactions on TE dynamics (**Chapter 3**). Most of our studies were done on lab-established genotypes that were derived from the wild and were characterized for TE composition (**Chapter 2** and **3**). We additionally assessed the importance of *Wolbachia* in affecting transposon dynamics in the context of TE invasion of natural populations (**Chapter 5**). This work unravelled the complexity of TE regulation in *Drosophila* ovaries and underscored the importance of both environmental and genetic factors in mediating the production of new genetic variation through transposon activity.

6.1.1 Genotypic and environmental effects in TE expression

Despite the high prevalence of transposons in genomes[5] and their impact on fitness[6], little is known about the factors controlling TE mobilization. Several studies have shown that environmental factors including temperature can affect transposition[7–9]. However, published data is controversial and there is no systematic and broad analysis of this effect including multiple TEs, genetic backgrounds, and environmental factors. This project aimed at filling that gap to provide new insight into the environmentally induced TE activity.

Effect of* Wolbachia *in TE activity

Wolbachia is a maternally transmitted endosymbiont that favours its own transmission by manipulating host reproduction[10]. It also influences other aspects of host's biology as described by its ability to confer protection against viral infections[11,12]. Given the capability of *Wolbachia* to reduce loads of RNA viruses and the structural commonalities between retroviruses and retrotransposons, we hypothesized that the presence of *Wolbachia* in *D. melanogaster* ovaries could reduce retrotransposon expression (similar to that observed for viral load[11]).

While characterizing the effect of *Wolbachia* in the activity of retrotransposons on different genotypes, we observed two types of responses that were dependent on genotype and element analysed (**Chapter 3** and **4**). In particular, TE expression varied from *Wolbachia*-mediated decrease or no effect in the expression of retrotransposons (**Figure 6.1**).

Similarly to what has been described for the tripartite interaction between *Wolbachia*, host and virus[13,14], we propose that the *Wolbachia* and host genotypes, as well as the interaction between both, may explain our results for the relationship established between *Wolbachia*, host and TEs. Although all *D. melanogaster* genetic backgrounds used in this study were naturally infected with the same strain of endosymbiont (*wMel*), we propose that single nucleotide polymorphisms (SNPs) in other loci than the ones used to define *Wolbachia* strains may differ between our lines. Those SNPs together with host genotypes

could influence *D. melanogaster* ability to induce oxidative stress when infected with endosymbiont[15]. We propose that the increase in levels of reactive oxygen species (ROS) that occurs when organisms are under oxidative stress and that has been described to mediate *Wolbachia* protection against viruses[15], could also be associated with TE activity. This idea gained special attention after our discovery of *cis*-regulatory elements in *Wolbachia*-responsive TEs (*blood* and *roo* in **Chapter 3**) that are recognized by a transcription factor involved in the establishment of antioxidant responses, Cap-n-collar[16].

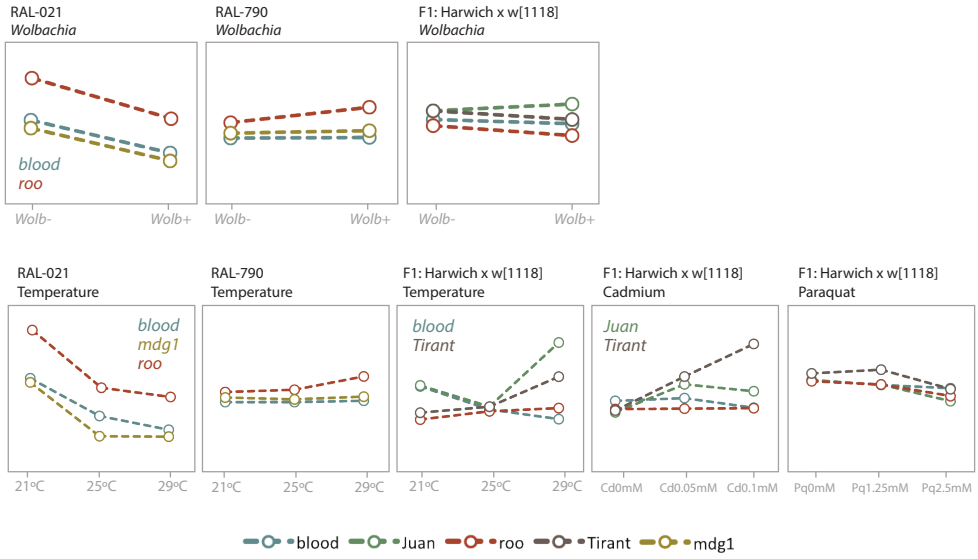


Figure 6.1 Overview of TE expression. Representation of average TE expression in response to biotic (*Wolbachia* status) and abiotic (temperature, cadmium and paraquat) factors. Expression of *blood, mdg1* and *roo* in RAL-021 and RAL-790 (**Chapter 3**); expression of *blood, Juan, roo* and *Tirant* in the F1 females obtained from a cross between Harwich females with w[1118] males (**Chapter 4**). Data from F1 females to see effect of temperature perturbations, levels of oxidative stress and paraquat doses was obtained from flies free of *Wolbachia* infection. On the other hand, to see effect of *Wolbachia* we represent data collected at 25ºC. TEs that have expression influenced by a given environmental factor are highlighted inside boxes.

Effect of stressful environments in TE activity

Stressful conditions are environmental perturbations that reduce individual fitness[17]. Here, we aimed at studying the impact of such challenges (including thermal perturbations, levels of oxidative stress and doses of cadmium) in mediating TE responses in the germline (**Chapter 4**).

In addition to the previously described genotype- and TE-dependent responses to the presence of *Wolbachia*, we also detected environment-specific effects in transposon expression (**Chapter 3** and **4**) (**Figure 6.1**). Although TEs were irresponsive to an increase in oxidative stress (**Chapter 4**), they were able to respond to the presence of cadmium in fly diet (**Chapter 4**) and to thermal perturbations (**Chapter 3** and **4**). The observation that the expression of TEs (including *blood* and *roo*) was not affected by an increase in ROS seems, however, to contradict our previous hypothesis that *Wolbachia*-induced increase in oxidative stress mediates a decrease in TE activity. Nonetheless, we combine both observations and further propose that transposons do not respond directly to an increase of ROS, but are affected by a putative inhibition of antioxidant responses mediated by the presence of endosymbiont (**Figure 6.2**). More importantly, we suggest that *Wolbachia* is able to reduce Cap-n-collar activity that results in the maintenance of higher levels of ROS and lower TE expression.

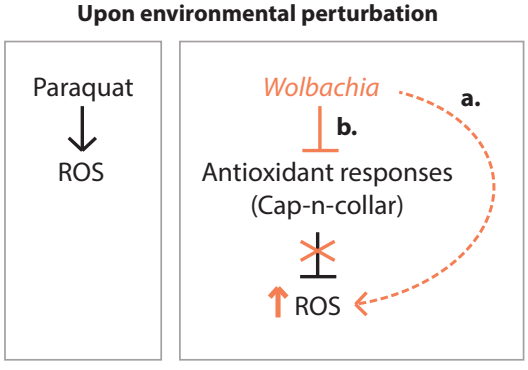


Figure 6.2 Production of reactive oxygen species (ROS) upon environmental perturbation. In normal physiological conditions, ROS are kept at low levels. When fed with paraquat – a commonly used herbicide that promotes redox reactions[18] – or upon infection with *Wolbachia*[15] the levels of ROS increase. For *Wolbachia*-mediated increase in ROS we propose that the presence of endosymbiont may be able to directly increase oxidative stress (a) and to keep levels of ROS at high concentrations through inhibition of the antioxidant responses (b).

Regarding the responses to temperature perturbations, we also proposed that changes in the activity of transcription factor(s) could influence TE expression. In more details, we suggested E2F1 and/or Fork head as putative players in the thermal sensitivity of *blood, mdg1* and *roo* (**Chapter 3**). Interestingly, the expression of *E2F1* seems to be inhibited by Sin3A[19], another transcription factor found to contain allelic variation associated with *blood* copy number (**Chapter 2**). This interaction between Sin3A activity and *E2F1* expression may explain our results for the levels of *blood* in cell lines over-expressing or down-regulating Sin3A (**Figure 6.3**). In other words, we propose that Sin3A is able to directly influence *blood* activity and mediate more expression of *blood* in lines over-expressing *Sin3A* (**Figure 6.3A**). In the absence of this transcription factor, *E2F1* is expressed and allows *blood* activity in cell lines knocking down *Sin3A* (**Figure 6.3B**).

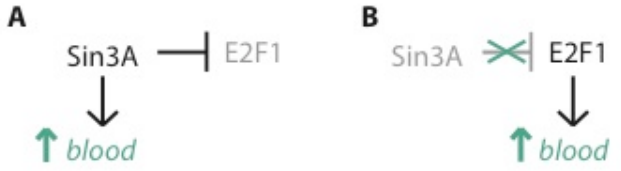


Figure 6.3 Model for the activation of *blood* expression by Sin3A and E2F1. In this model we include Sin3A ability to inhibit expression of the transcription factor E2F1[19], and further propose that Sin3A is able to (directly) induce expression of *blood*. When the activity of Sin3A is reduced, the expression of E2F1 is activated and *blood* expression may be induced by E2F1.

All responses to environmental and genetic factors that we documented and summarized above were element-dependent and varied even for TEs with common properties. Conversely, the same transposon responded in a different manner to different types of environmental perturbation even when these are described to activate the same (protection) mechanisms[20]. Given the unanticipated varieties and the TE-specific responses to environmental and genetic components we finally conclude that transposon activity is precisely regulated (mostly by transcription factors) in a context dependent way.

Lack of evidence that the piRNA pathway is involved in genotype- and environment-specific TE activity

The genetic mechanism specialized in controlling TE activity in the germline directs the production of Piwi-interacting small RNAs (piRNAs) that share sequence homology with mobile elements[24]. The functional dependence between this pathway and the mechanisms influenced by sub-optimal conditions (such as heat shock responses) led us to hypothesize that in stressful environments piRNA surveillance would be relaxed due to the recruitment of stress-responsive proteins (like Hsp90) to mediate other functions[25–27]. This would result in the deregulation of different TEs (including retrotransposons and DNA elements) in response to stress[5].

In addition to the aforementioned TE-specific effects of environmental perturbations (**Chapter 3** and **4**), we also did not find evidence of problems in the piRNA pathway upon exposure to sub-optimal conditions (**Figure 4.4**). Furthermore, the assessment of allelic variation segregating in natural populations found almost no association between alleles affecting piRNA pathway genes and number of TE insertions (**Chapter 2**).

In order to understand which other biological processes and pathways may be triggered by temperature, cadmium and paraquat treatments, we took advantage of available RNA-sequencing data[28], and performed a more comprehensive analysis for TE and gene expression (**Figure 6.4**). It is though important to note that, this dataset was obtained from a different genetic background, and the fold change of expression was assessed by comparing dosages of the treatment, rather than using control samples. Those features could explain the differences obtained for TE response, when comparing both datasets (RNA-sequencing in **Figure 6.4** and qPCRs in **Chapter 4**). In agreement with our results, this analysis also did not find effects in piRNA pathway genes upon exposure to stress. Instead, we observed enrichment for cellular responses that were unique to each environmental perturbation and that could explain the TE-specific effects also observed in this dataset (**Figure 6.4**). For a comprehensive explanation of

why the piRNA pathway seems to be robust to environmental and genetic factors, one should take into account that any perturbation of this pathway could lead to strong deleterious effects due to genomic instability resultant from uncontrolled TE activity[29–31].

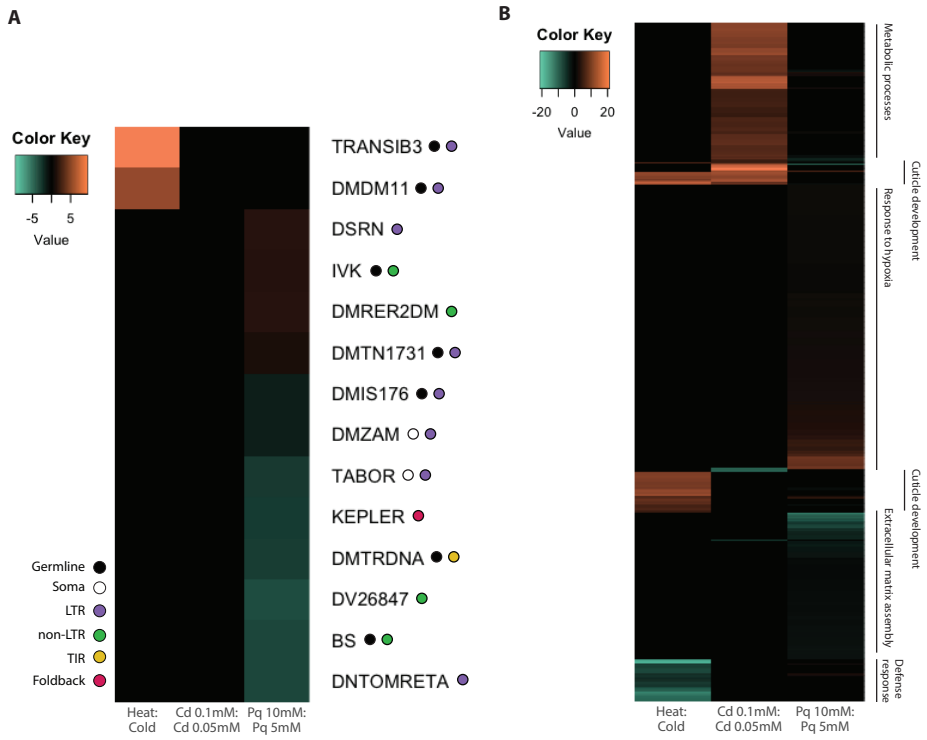


Figure 6.4. TE and gene expression in response to temperature, cadmium and paraquat. Analysis of available next-generation RNA sequencing data[28] for TE **(A)** and gene **(B)** expression in response to temperature (SRR103710, SRR103711, SRR103724, SRR103723), cadmium (SRR124260, SRR103725, SRR103744, SRR103745) and paraquat (SRR103718, SRR124259, SRR103722, SRR103721) treatments. Those libraries were deposited on modENCODE website. Reads were aligned to transposon canonical sequences (version 9.42, **A**) and gene transcripts (release 6, **B**) of *D. melanogaster*. We extracted the number of reads aligned for each TE and gene, and analyzed the data using edgeR by fitting read counts to a negative binomial distribution. We limited our analysis to data collected from the same genotype, and thus assessed differences in expression using cold, lower dosages of cadmium (0.05mM) and paraquat (5mM), as controls for temperature, heavy metals and oxidative stress treatments, respectively. Difference in expression for temperature, cadmium and paraquat treatments

is relative to cold, 0.05mM of cadmium (Cd) and 5mM of paraquat (Pq), respectively. Heatmap shows fold change of expression with false discovery rate inferior to 10%. Fourteen transposons with different sequence properties and places of expression, respond to environmental perturbations **(A)**. Colored circles represent places of TE expression: germline in black and somatic tissue of ovaries in white; and sequence properties: retrotransponsons with long terminal repeats (LTRs) in violet, retrotransposons without LTRs (non-LTRs) in green, DNA elements with terminal inverted repeats (TIR) in yellow and foldback element in pink. For some elements we have no information about the place of expression[32]. For deregulated genes, we performed a gene ontology analysis using GO consortium algorithm[33], and Bonferroni correction for multiple testing **(B)**. We found enrichment for several biological processes that included cuticle development (fold enrichment=7.3, p-value=3.5e-03) and defense response (fold enrichment=8.73, p-value=4.54e-02) for temperature, chitin (fold enrichment=13.56, p-value=1.15e-18), aminoglycan (fold enrichment=11.79, p-value=1.36e-18), glucosamine (fold enrichment=12.71, p-value=5.97-18), amino sugar (fold enrichment=12.61, p-value=7.27e-18), and gluthatione (fold enrichment=10.66, p-value=7.87e-08) metabolic process for cadmium, and response to hypoxia (fold enrichment=8.09, p-value=4.27e-05) and extracellular matrix assembly (fold enrichment=24.02, p-value=7.20e-06) for paraquat treatment.

Interestingly, a gene ontology analysis for the genes up-regulated upon exposure to cadmium revealed enrichment for those related with nutrient reservoir activity (**Figure 6.4**). Given that in our experimental setup we also found evidence that flies stop eating when exposed to doses of cadmium (**Supplementary Figure 4.4**), we cannot exclude the hypothesis that the observed TE responses to the presence of this heavy metal are due to effects of starvation or interaction between those two factors.

6.1.2 TE ability to produce genetic variation in response to environmental factors

It is well established that environmental stressors can act as mutagens and higher mutation rates might help cope with environmental challenge[34]. Increased genetic variation in the progeny increases chances of having variants capable of thriving under new conditions. This can explain the advantage of mutator alleles in experimental populations evolving under unstable conditions[35], and the recent

finding on increasing rates of recombination in gamete production upon exposure to parasites[36]. In this study, we proposed to investigate the effect of environmental perturbation on another source of genetic variation, transposons[37]. We hypothesized that perturbation-driven TE mobilization in the germline could help populations coping with environmental perturbation by increasing genetic variability in the progeny. TE expression in the gonads was used here as a proxy for the activity of mobile elements (**Chapter 3** and **4**).

We observed that increased TE expression in ovaries could add new insertions to the profile of standing genetic variation that is transmitted to the next generation (**Chapter 4**). However, those results were not general for all transposons analysed and possible additional effects at other stages of TE "life cycle" can extend the display of outcomes observed for expression. In other words, although TE expression in the germline can be used as a proxy for activity, those results need to be cautiously interpreted. Further research on the variants obtained in following generations, including the number and places of novel insertions, is essential to understand how environmental and genetic factors shape TE population dynamics.

6.1.3 Ability of *Wolbachia* to promote *P-element* invasion in natural populations

Traditionally, during the process of TE invasion of natural populations, mobile elements behave as parasites. They can ultimately reduce host fitness through an increase in the number of deleterious mutations and in the genomic instability of gonads[21]. This phenomenon is often referred to as hybrid dysgenesis and is triggered by parental incompatibilities in TE composition that result in uncontrolled transposon activity in the gonads[21,38]. In those circumstances, hosts benefit from protective systems against TE mobility and that are able to restore ability to reproduce.

Interestingly, a recent invasion of *D. melanogaster* populations by a DNA element (*P*-element) that is associated with hybrid syndrome correlates with the

replacement of *Wolbachia* strains[39–41]. In this context, we hypothesized that *Wolbachia* might have played an important role in the initial step of *P-element* spread in the wild (**Chapter 5**). This idea seems to contradict the aforementioned hypothesis suggesting that *Wolbachia* is capable of reducing the activity to retrotransposons similarly to its function in reducing loads of retroviruses[11,12] (**Chapter 3**). However, it is important to note that the presence of endosymbiont in *D. melanogaster* does not confer protection against DNA viruses[11] and that *P-element* is a DNA transposon without sequence similarities with retrotransposons.

In our analysis, we propose that the presence of *Wolbachia* could have promoted TE colonization of *D. melanogaster* genome by increasing host fitness trough genotype-specific effects on fecundity and/or percentage of eclosion (**Figure 5.2**). Although the processes that mediate host-*Wolbachia* interaction in the context of hybrid dysgenesis are still unknown, we propose that endosymbiont may have restored hybrid fertility similarly to its ability to suppress Sex-lethal mutant phenotype[42,43]. This last phenomenon was recently proven to be mediated by a *Wolbachia* coded gene (the toxic manipulator of oogenesis, TomO) that restores the production of germ cells. The importance of TomO in affecting oogenesis of hybrids still needs to be validated.

6.2 Discussion of shortcomings and ideas for future work

A major finding of this dissertation relates to the importance of genetic and environmental factors, and the interaction between the two, in mediating TE dynamics in wild-derived genotypes, where TE expression is usually low (**Chapter 2**, **3** and **4**), and in hybrids that have uncontrolled TE activity (**Chapter 5**). The extension of the analyses performed in this thesis to other genotypes, stresses, and times of exposure could help shed light into the complex regulation of TE mobilization. Also, the usage of other techniques and/or experimental setups would help clarify the importance of stressful environments to produce novel genetic variation in nature.

6.2.1 Limitations of the qPCR methodology

Quantitative real-time PCR (qPCR) is a powerful tool to assess expression in a relatively fast manner. This technique depends on the usage of reference genes (such as *Rpl32* and *Tbp*) that are considered to provide a good measure (and control) to the total RNA in the sample[44,45]. However, finding control genes suitable for all treatments and genotypes used in this thesis proved to be difficult (**Supplementary Figure 3.1** and **Supplementary Figure 4.2**). We suggest that further studies should include other technics that are less dependent on the usage of control genes. Those could include next-generation RNA sequencing analysis that not only overcome the limitations mentioned above but also allow for the analysis of TE and gene de-regulation in response to perturbation (**Figure 6.4**).

6.2.2 Validate importance of host transcription factors to mediate TE responses to stressful conditions

We found that mobile element activity was dependent on the genotype, environment and TE analysed (**Chapter 2**, **3** and **4**). Some of those responses could be associated to *cis*-regulatory elements found within transposon sequences that need validation (**Chapter 3**). To understand the contribution of transcription factor activity to regulating TE activity it would be important to measure transposon expression and mobilization in genotypes with over- and under-activation of transcription factors such as Cap-n-collar, E2F1 and Fork head, and with TE insertions containing mutated *cis*-regulatory elements.

6.2.3 Failure at properly testing the importance of stressful environments to produce adaptive genetic variation

Despite being more active when hosts are exposed to some environmental perturbations (**Chapter 3** and **4**), TEs are also able to insert into specific places. This is found across different species and mobile elements, including *Drosophila* P-element bias towards HSP promoters[46,47] and tobacco retrotransposon Tnt1 insertions in nitrate reductase gene that drive the appearance of chlorite-

resistant phenotypes[48]. It is possible that, transposons triggered by biotic or abiotic factors mediate *de novo* insertions that influence host ability to respond to those perturbations, and, thus, their evolvability[47]. In addition to the study of fitness traits, the assessment of the type of genetic variation that is produced in individuals born from parents that were exposed to environmental perturbation versus control conditions could help clarify the contribution of TE-environment interaction to adaptation to novel conditions.

The lack of connection between genotype, phenotype and fitness could benefit from more integrative analysis including genome wide association studies (GWAS) and experimental evolution experiments. In the context of this essay, and to clarify what are the alleles contributing to TE responses to stressful conditions, we propose a GWAS analysis exposing the complete panel of DGRP lines[49] to environmental perturbations and using how much TEs respond as phenotype. This approach would also help to understand the relative contribution of mutations in sensing, signal transduction or effector processes involved in transposon activity. It is, however, important to note that GWAS analyses provide information about genotype-to-phenotype map and should be complemented with other types of analyses to study phenotype-to-fitness map (such as experimental evolution experiments). Although this was never empirically tested, increased TE activity in response to stress is believed to mediate individual evolvability by driving the production of adaptive variation[37,47]. It would therefore be interesting to address this question by establishing two types of populations, one composed by genotypes that contain TEs responsive to environmental perturbations, and another composed by genotypes whose mobile elements are robust to perturbation. Those two populations could then be used to perform experimental evolution experiments under environments differing in thermal regime and/or exposure to heavy metals. This type of experiment would help clarify whether populations more prone to activate TEs in stressful environments (plastic TE activity) adapt faster to environmental perturbations themselves. Moreover, identifying positively selected TE insertions

could add further to our understanding of the contribution of novel insertions to adaptation.

In sum, this study provides original experimental evidence on the complexity of transposon responses to environmental and genetic factors. The insights gained, together with future research, will shed light onto our understanding of the dynamics of TE activity in natural populations and of the intricate interactions established between host genomes, mobile elements and respective environments.

Acknowledgements

I would like to thank to Daniel Sobral (bioinformatics unit) for assistance with RNA sequencing data and to Bahtiyar Yılmaz for reading the last version of this chapter and for providing important feedback to improve it.

References

1. Stern, D. L. Evolutionary Developmental Biology and the Problem of Variation. *Evolution (N. Y).* **54,** 1079 (2000).
2. Beldade, P., Mateus, A. R. A. & Keller, R. A. Evolution and molecular mechanisms of adaptive developmental plasticity. *Mol. Ecol.* **20,** 1347–1363 (2011).
3. Lu, C. *et al.* Miniature Inverted – Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in Oryza sativa. *Mol Biol Evol* **29,** 1005–1017 (2012).
4. Kofler, R., Nolte, V. & Schlotterer, C. Massive bursts of transposable element activity in Drosophila. *bioRxiv* (2014). doi:10.1101/010231
5. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).
6. Barroso-Batista, J. *et al.* The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.* **10,** (2014).
7. Ratner, V. A., Zabanov, S. A., Kolesnikova, O. V & Vasilyeva, L. A. Induction of the mobile genetic element Dm-412 transpositions in the Drosophila genome by heat shock treatment. *Proc. Natl. Acad. Sci. U. S.*

A. **89,** 5650–4 (1992).

8. Bouvet, G. F., Jacobi, V., Plourde, K. V. & Bernier, L. Stress-induced mobility of OPHIO1 and OPHIO2, DNA transposons of the Dutch elm disease fungi. *Fungal Genet. Biol.* **45,** 565–578 (2008).

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

10. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6,** 741–751 (2008).

11. Teixeira, L., Ferreira, A. & Ashburner, M. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6,** 20753–2763 (2008).

12. Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Wolbachia and virus protection in insects. *Science (80-.).* **322,** (2008).

13. Reynolds, K. T., Thomson, L. J. & Hoffmann, A. A. The effects of host age, host nuclear background and temperature on phenotypic effects of the virulent Wolbachia strain popcorn in Drosophila melanogaster. *Genetics* **164,** 1027–1034 (2003).

14. Chrostek, E. *et al.* Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis. *PLoS Genet.* **9,** 1–22 (2013).

15. Wong, Z. S., Brownlie, J. C. & Johnson, N. Oxidative Stress Correlates with Wolbachia-Mediated Antiviral Protection in Wolbachia-Drosophila Associations. *Appl. Environ. Microbiol.* **81,** 3001–3005 (2015).

16. Sykiotis, G. P. & Bohmann, D. Keap1 / Nrf2 Signaling Regulates Oxidative Stress Tolerance and Lifespan in Drosophila. *Cell* **14,** 76–85 (2007).

17. Armbruster, P. & Reed, D. H. Inbreeding depression in benign and stressful environments. *Heredity (Edinb).* **95,** 235–242 (2005).

18. Rzezniczak, T. Z., Douglas, L. A., Watterson, J. H. & Merritt, T. J. S. Paraquat administration in Drosophila for use in metabolic studies of oxidative stress. *Anal. Biochem.* **419,** 345–347 (2011).

19. Zhang, Y., Akinmade, D. & Hamburger, A. W. The ErbB3 binding protein Ebp1 interacts with Sin3A to repress E2F1 and AR-mediated transcription. *Nucleic Acids Res.* **33,** 6024–6033 (2005).

20. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Genet.* (2011). doi:10.1038/nrg3055

21. Khurana, J. S. *et al.* Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell* **147,** 1551–1563 (2011).

22. Li, W. *et al.* Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat. Neurosci.* **16,** 529–531 (2013).

23. Seroude, L., Brummel, T., Kapahi, P. & Benzer, S. Spatio-temporal analysis of gene expression during aging in Drosophila melanogaster. *Aging Cell* **1,** 47–56 (2002).

24. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128,** 1089–1103 (2007).

25. Specchia, V. *et al.* Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463,** 662–665 (2010).

26. Gangaraju, V. K. *et al.* Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat. Genet.* **43,** 153–158 (2011).

27. Izumi, N. *et al.* Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. *RNA J.* **19,** 896–901 (2013).

28. Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512,** 393–399 (2014).

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

30. Handler, D. *et al.* A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30,** 3977–3993 (2011).

31. Haase, A. D. A Small RNA-Based Immune System Defends Germ Cells against Mobile Genetic Elements. *Stem Cells Int.* 1–9 (2015).

32. Malone, C. D. *et al.* Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137,** 522–535 (2009).

33. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

34. Ram, Y. & Hadany, L. The evolution of stress-induced hypermutation in asexual populations. *Evolution (N. Y).* 1–14 (2012). doi:10.1111/j.1558-5646.2012.01576.x

35. Giraud, A. *et al.* Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science (80-.).* **291,** 2606–2608 (2001).

36. Singh, N. D. *et al.* Fruit fly diversify their offspring in response to parasite infection. *Science (80-.).* **349,** 747–750 (2015).

37. Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. Adaptation to Global Change : A Transposable Element – Epigenetics Perspective.

Trends Ecol. Evol. **xx,** 1–13 (2016).

38. Brennecke, J. *et al.* An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (80-.).* **1387,** 1387–1392 (2008).

39. Kidwell, M., Kidwell, J. & Sved, J. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traitsincluding mutation, sterility and male recombination. **85,** 1973–1975 (1976).

40. Engels, W. R. Invasions of P elements. *Genetics* **45,** 11–15 (1997).

41. Riegler, M., Sidhu, M., Miller, W. J. & O'Neill, S. L. Evidence for a global Wolbachia replacement in Drosophila melanogaster. *Curr. Biol.* **15,** 1428–1433 (2005).

42. Ote, M., Ueyama, M. & Yamamoto, D. Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in Drosophila Sex-lethal Article Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in Drosophila Sex-lethal Mutants. *Curr. Biol.* **26,** 1–10 (2016).

43. Sullivan, W. Endosymbiosis: The Remarkable Healing Powers of Wolbachia. *Curr. Biol.* **26,** 797–799 (2016).

44. Ponton, F., Chapuis, M.-P., Pernice, M., Sword, G. A. & Simpson, S. J. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in Drosophila melanogaster. *J. Insect Physiol.* **57,** 840–850 (2011).

45. Matta, B. P., Bitner-Mathé, B. C. & Alves-Ferreira, M. Getting real with real-time qPCR: A case study of reference gene selection for morphological variation in Drosophila melanogaster wings. *Dev. Genes Evol.* **221,** 49–57 (2011).

46. Walser, J.-C., Chen, B. & Feder, M. E. Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. *PLoS Genet.* **2,** e165 (2006).

47. Fablet, M. & Vieira, C. Evolvability, epigenetics and transposable elements. *Biomol. Concepts* **2,** 333–341 (2011).

48. Feschotte, C., Jiang, N. & Wessler, S. R. Plant Transposable Elements: Where Genetics Meets Genomics. *Nat. Rev. Genet.* **3,** 329–341 (2002).

49. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482,** 173–178 (2012).

ROAD TO NOWHERE

“We’re on a road to nowhere
Come on inside
Taking that ride to nowhere
We’ll take that ride
I’m feeling okay this morning
And you know
We’re on the road to paradise
Here we go, here we go”

Talking Heads

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt