Unravelling the role of transposable elements in the eukaryotic stress response

Vivien Horváth

TESI DOCTORAL UPF / 2020

Thesis supervisor

Dra. Josefa González Pérez

Institute of Evolutionary Biology (CSIC-UPF)

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

Acknowledgments

First of all, I would like to thank the help of my supervisor and mentor Josefa González. Thank you for giving me the opportunity first, to come to your lab as a Master student, and then for staying and doing the PhD. Thank you for your patience, and for teaching and guiding me during these last five years. I truly learned a lot and you showed me how is a good scientist.

También quiero dar las gracias a los miembros del Gonzalez Lab. En primer lugar, a Miriam quien me enseñó todo desde el principio. Muchas gracias por tu paciencia y dedicación, que siempre me respondías mis preguntas y resolvías mis dudas. ¡Siempre te tendré como ejemplo a seguir!

También muchas gracias a ti Judit, aunque has llegado más tarde al lab, pero has sido como mi "madre del laboratorio" (¡aunque la edad no la tienes!) desde el primer momento que llegaste. Gracias por guiarme y enseñarme como trabajar y pensar como científica. Gracias por tus consejos el nivel personal y científico. Y también muchas gracias por convencerme a llevar casco en la bici.

Gracias a Gaby también, quien con su paciencia infinita me respondió todas mis preguntas y explico todas las cosas que no sabia sobre bioinformática, y también gracias por tus consejos. Además, gracias por los partidos de volley, aunque éramos "casi que no" pero eran partidos muy divertidos (casi siempre).

Sara, a ti te tengo que dar las gracias también por tu paciencia, ayuda y consejos. Gracias por enseñarme todo lo que se de estadística, y no cansarte de explicarme la misma cosa mil veces.

Santi, muchas gracias por tu apoyo, ayuda, y por escuchar y responder todas mis dudas y preguntas, especialmente durante los últimos cuatro meses de la quaranthesis.

Gracias a Carlos por todo, pero sobretodo por los videos y fotos de gatos, aunque ya sabes cual es mi gato favorito.

I also want to thank you Llew, for answering all my questions and for helping me whenever you could.

También quiero dar las gracias a los miembros anteriores del Gonzalez Lab. A Lain, Maite y especialmente a José, por su contribución en el primer parte de esta tesis y por su ayuda en los primeros años. Y también a Laura, por su ayuda y consejos.

I also would like to thank to Gianfranco Anfora for his help and for giving me the opportunity of staying three months in the beautyfoul San Michelle (Edmund Mach Foundation). Also, thanks to Mirco Rodheriego, Eugenio Aprea and Omar Rota Stabelli for their help during my stay in the Edmund Mach Foundation. Moreover, thanks to the members of the lab for being so welcoming and nice, special thanks to Valeria, Imane and for my hiking partners Karen, Valerio, Aceisha and Becky.

Gracias a las chicas de cuquí house, Maria, Anna, Silvia, Lidia y Miriam. Gracias por compartir estos últimos años, por todos los viajes, sushi n1s, bravas y escape rooms (la cúpula no). Tambien gracias por vuestro apoyo, consejos científicos y no científicos durante este tiempo, especialmente a ti María, porque nos entendemos.

Tambien quiero mencionar la gente del IBE, quienes mejoraron mis días: Elena Parra, Núria, Alfonso, Michelle, Aleksandra, Cristina, Vicente, Anna, Pablo, Blanca, Isabel. Especialmente a Pepi Cruz, quien me ayudo muchas veces y siempre resolvió mis dudas.

Fuera del ambiente científico quiero dar las gracias a las Patris, Marisol y Maydo, mis compañeras de entreno. Me lo pase muy bien con vosotras, y gracias por vuestro apoyo especialmente en los últimos meses.

Szeretnem meg köszönni a családomnak a támogatást, mert a segítségük nélkül nem tudtam volna eljutni oda, ahol most vagyok. Különösképpen az anyukámnak, aki a kezdetektől támogatott, valamint a nagyszüleimnek, akikre mindig számíthattam. Neked is koszi Szimi, mert mindig bármivel fordulhattam hozzad. Petike, rólad nem is beszelve. Lassan régebben ismerlek mint magamat, köszönöm, hogy mindig ott vagy meg akkor is ha az utóbbi években csak virtuálisan vagy a legjobb baratom.

Y ahora quiero dar las gracias a ti Alex, que has estado a mi lado durante todo este tiempo. Gracias por ser como eres, por apoyarme y ayudarme en cada momento, y por entender lo importante que era para mi el doctorado. ¡Gracias por aguantarme sobretodo en los últimos meses, eres un héroe! Tengo muchas ganas de empezar nuestra nueva aventura en Suecia, gracias por apoyarme y compartir todo esto conmigo.

Abstract

Environmental stress is a major aspect of life and a limiting factor for the survival and distribution of species. Understanding the organismal stress response and the developed coping mechanisms is indispensable in order to mitigate the negative effects of stress. A lot of studies focus on the role of single nucleotide polymorphisms, while transposable elements (TEs) which are also very powerful mutagens are largely overlooked. In this thesis, we studied the contribution of TEs to the eukaryotic stress response, first by using a molecular mechanism specific approach. By studying six different stress responses in humans and *Drosophila melanogaster*, we showed that TEs can regulate the expression of stress response genes by adding transcription factor binding sites. In the second part, we chose a stress specific approach, and described the transcriptomic and physiological basis of *D. melanogaster* desiccation tolerance. Moreover, we found that TEs might be involved in desiccation stress response but they do not seem to be the main mutations fuelling this stress response. Overall, we showed that transposable elements are relevant players in eukaryotic stress response, however their effect could be stress specific.

Resumen

El estrés ambiental es un aspecto importante de la vida y un factor limitante en la supervivencia y distribución de las especies. Entender la respuesta a estrés a nivel de organismo y los mecanismos desarrollados para enfrentarlo son indispensable para mitigar sus efectos negativos. Muchos estudios se han centrado en el rol de los polimorfismos de nucleótido único, mientras que otros potentes mutágenos, como son los elementos transponibles (TEs), han sido ignorados. En la primera parte de esta tesis, hemos estudiado la contribución de los TEs en la respuesta eucariótica a estrés utilizando un enfoque específico al mecanismo molecular. Al estudiar seis tipos de respuesta a estrés diferentes en humanos y en *Drosophila melanogaster* hemos demostrado que los TEs pueden regular la expresión de genes de respuesta a estrés gracias a la adición de sitios de unión de factores de transcripción. En la segunda parte, hemos escogido un enfoque específico de estrés y hemos descrito las bases transcriptómicas y fisiológicas de la tolerancia a la desecación en *D. melanogaster*. Además, hemos observado que los los TEs podrían estar involucrados en la respuesta a estrés por desecación, aunque no parecen ser la principal mutación causante de la respuesta a este estrés. En conjunto, se muestra que los TEs tienen un rol importante en la respuesta eucariótica a estrés, sin embargo, su contribución parecería ser específica del estrés.

.

Preface

Organisms have to face increasingly stressful environmental conditions, which can even put their existence at risk. In order to overcome these harsh circumstances, they have developed stress response mechanisms. Understanding these mechanisms is essential in order to mitigate the impact of stress and to protect species. In case of perturbations, the stress response mostly manifests in alterations in transcriptional regulation and results in changes in the expression of specific genes. The studies aiming at discovering the genomic variants responsible for these alterations in gene expression are mostly focusing on single nucleotide polymorphisms. Even though transposable elements (TEs) have also been shown to be powerful mutagens and to regulate gene expression, their role in regulating the eukaryotic stress response genes is still not extensively studied. The aim of this work was to characterize the role of TEs as gene regulators in seven different stress responses in humans and *Drosophila melanogaster*, and to describe the genomic and physiological bases of *D. melanogaster* desiccation stress response.

In Section 3.3. we analysed the role of TEs in immune, hypoxia, oxidative, xenobiotic, heat shock, and heavy metal stress regulatory networks in humans and *Drosophila melanogaster*. Precisely, we identified transcription factor binding sites (TFBSs) and transcription factor binding motifs (TFBMs) embedded within TEs. In the human genome we found that transposable elements are enriched for *NFE2L2* TFBSs, and that transcription factors such as *CREB1* and *NF-κB* have slightly more TFBMs within TEs than expected. Meanwhile, in the Drosophila genome TEs were enriched for *caudal*, *dorsal*, *HSF* and *tango* binding sites, however, the % of TFBMs appears to be small. Considering the *D. melanogaster* population frequencies of transposable elements with predicted binding motifs and/or binding sites, we showed that those containing three or more binding motifs/sites are more likely to be functional. For a representative subset of these TEs, we performed in vivo transgenic reporter assays in *D. melanogaster* in different stress conditions, and found that TEs can regulate the expression of immune stress response genes.

In Section 3.4. we aimed to describe the transcriptomic and physiological bases of desiccation stress response, and the role of TEs in it. First, we subjected to desiccation stress conditions 74 European *D. melanogaster* strains from five different climatic regions, and found that the strains from cold semi-arid climates are more tolerant compared to strains from hot summer mediterranian climate zones. Moreover, the variation in the survival of the strains correlates with

the interaction of altitude and evaporation. We also found, that the tolerant strains have lower water content and lose less water during desiccation stress. The lower level of water loss in the tolerant strains can be explained by a higher decrease in the respiration rate in desiccation stress conditions, and a more favorable cuticular hydrocarbon composition. When identifying the desiccation stress responsive genes, we found that genes involved in several metabolic processes are down-regulated while the genes involved in response to stimulus and environmental sensing are up-regulated. Moreover, we identified a list of candidate transposable element insertions possibly affecting the expression of the nearby genes, mostly in a strain specific manner.

Overall, this work highlights the complexity of the eukaryotic stress response, and suggests that TEs are relevant players in the regulation of stress response genes, however their contribution is stress specific.

Table of contents

Acknowledgements I

Abstract V

Preface IIV

1. Introduction 1

1.1. Defining stress 3

1.2. The impinge of biotic and abiotic stress factors on organisms 3

1.3. Transcriptional regulation during eukaryotic stress response 6

1.3.1. *Cis-regulatory* evolution as an adaptive driving force 7

1.4. Transposable elements, a historical overview 8

1.4.1. Transposable element classification 9

1.4.2. Transposable element load in different organisms 11

1.4.3. Transposable elements as double-edged swords in genome variation 12

1.4.3.1. Transposable element activation due to environmental stress 13

1.4.4. Transposable elements as source of coding ad non-coding RNA 14

1.4.5. Transposable elements as "batteries" of gene regulation 16

1.4.6. Transposable elements expand the *cis-regulatory* module repertoire in *D. melanogaster* 18

2. Objectives 21

3. Results 25

3.1. The structure of the research 27

3.2. Thesis advisor's report about authorship and impact factor of the publications of this doctoral thesis 29

3.3. Diverse families of transposable elements affect the transcriptional regulation of stress- response genes in *Drosophila melanogaster* 31

3.4. The transcriptomic and physiological bases of desiccation tolerance in natural European *D. melanoagaster* populations 49

4. Discussion **81**

4.1. TEs can harbor TFBSs in different contexts and organisms 83

4.1.1. TEs adding stress related transcription factor binding sites 84

4.1.2. Are TEs more prone to be involved in immune gene regulation compared to other stresses? 85

4.1.3. TEs adding development related TFBSs 87

4.1.4. TEs adding metabolic pathway and disease related TFBSs 88

4.2. The importance of functional experiments when evaluating the role of TEs as gene regulatory elements 88

4.3. Novel gene regulators: tRNA derived small RNAs 90

4.4. The role of plastic responses in insect survival due to climate change 92

5. Conclusions **95**

6. References **99**

7. Annexes **119**

7.1. Supplementary material: Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster* 121

7.2. Supplementary material: The transcriptomic and physiological bases of desiccation tolerance in natural European *D. melanoagaster* populations 142

SECTION 1

INTRODUCTION

1. INTRODUCTION

1.1. Defining stress

Stress is a major aspect of life and understanding its effect on organisms is crucial (Straalen and Roelofs 2012). Since stress is a word very commonly used both in every day conversations and in science, establishing a succinct definition to capture its scientific basis is challenging (Kültz 2020). The word stress has been used now for many centuries, however the first time it gained a scientific meaning was in the 1660s. Robert Hooke, who was working on the field of mechanics and material sciences, used the word *stress* to describe the law of elasticity (Hooke's law) (Kültz 2020). Some scientist debate that biological stress should be defined similarly to the physical concept of stress. However, as described in van Straalen & Roelofs 2012 "Most biologists consider stress as an internal state, brought about by a hostile environment or negative social interaction" (Straalen and Roelofs 2012). From a biological point of view, stress is a very important factor in nature, by being a key driver of adaptation and phenotypic evolution, and as an ultimate architect of species diversity and distribution (Hoffmann and Parsons 1989, Loeschcke et al. 2004, Maggert 2019). In the context of adaptation and phenotypic evolution to environmental challenges, stress has been described in several ways, for instance Koehn & Bayne (1989) described stress "as an environmental change that results in reduction of net energy balance" (Koehn and Bayne 1989). Another exact definition from Hoffmann and Parsons (1991) states that environmental stress is when an "environmental factor causes a change in a biological system, which is potentially injurious" (Hoffmann and Parsons 1991). Bijlsma and Loeschcke (1997) completed this definition by saying that this potentially injurious process has some fitness consequences (Bijlsma and Loeschcke 1997). In summary, in this thesis stress is considered as a multi-faced force that affects the homeostasis of species and thus often causes fitness consequences, hence, it is a central force for the evolution of life.

1.2. The impinge of biotic and abiotic stress factors on organisms

When talking about stress in a biological context, we can distinguish between biotic and abiotic stresses. Biotic stress can be brought about by factors such as competition, predation or parasitism, which can trigger for instance the immune stress response (Relyea 2005) (Figure 1.1). This trait is very relevant, since organisms have to adapt to new pathogens every time when colonizing new environments. Local adaptation to habitat-specific pathogens is common in immune response both in terrestrial and aquatic species (Tinsley

and Majerus 2006, Bryan-Walker et al. 2007, Scharsack et al. 2007, Lazzaro et al. 2008, Fumagalli et al. 2011, Juneja et al. 2016). For instance, certain genetic variants such as the sickle cell, α+- and β+-thalassemia at the hemoglobin loci, as well as variants at ABO, GYPA, GYPB, GYPE, and G6PD loci, has been shown to confer resistance to malaria in African human populations (Kwiatkowski 2005, Fumagalli et al. 2011).

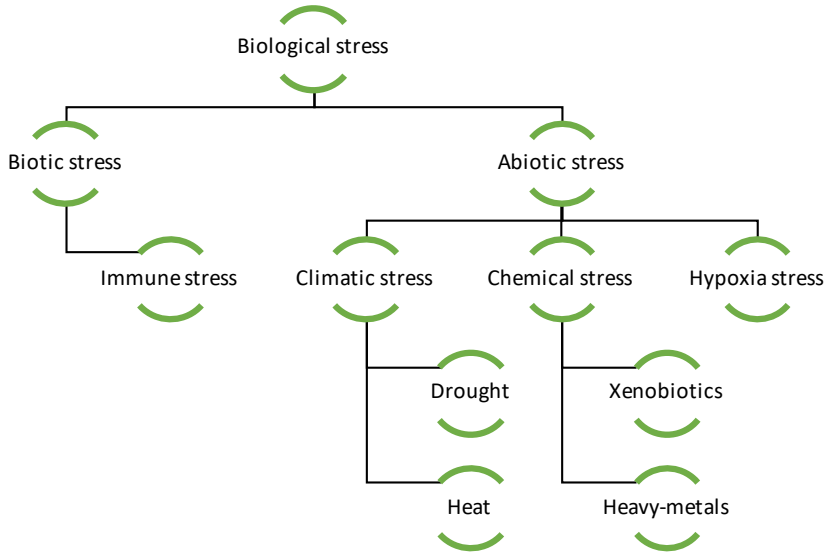


Figure1.1. Representation of the biological stress factors under the scope of the present thesis

Abiotic stress is mostly caused by environmental factors such as climatic factors or chemical components (Lindgren and Laurila 2005, Sørensen et al. 2005) (Figure 1.1). They are major players in shaping the distribution and abundance of species and have been proved to be important drivers of adaptation (Bubliy and Loeschcke 2005). Drought and heat are one of the most important environmental constraints, and are predicted to become more severe with global climate change (Acuña-Galindo et al. 2015) (Figure 1.1). Dry environmental conditions are an extremely relevant problem for insects, since they have highly permeable membranes and a high surface area to volume ratio, thus maintaining water balance is crucial for them (Edney 1977, Chown et al. 2011). Even though harsh, dry environmental conditions are challenging for insects, there are several examples of species that had invaded deserts and dry environments, such as the cactophilic Drosophila species (Gibbs 2002). It has been described during the years, that thanks to three main physiological mechanisms insects could adapt to these environmental conditions (Chown et al. 2011). They can i) increase their bulk water storage, which favours the water balance in desiccation conditions; ii) reduce the rate of water loss, which can happen through respiration, excretion or cuticular transpiration; iii)

tolerate a great amount of water loss (Chown 2002, Gibbs et al. 2003). The process of adaptation to dry environments is a complex albeit very important trait, and little is known about the molecular changes underlying it (Telonis-Scott et al. 2012, Rajpurohit et al. 2018). Moreover, it has also been reported that drought events hastened the extinction of some populations of the checkerspot butterflies (Mclaughlin et al. 2002). Drought and heat are not only candidate traits likely to be important in shaping species resilience and distribution under climate change, but are also one of the most important environmental factors in crop production, with direct consequences on human populations (Chown et al. 2011, Kellermann et al. 2012, Acuña-Galindo et al. 2015).

Another important group of abiotic stress factors are the chemicals, such as pesticides and heavy metals. The effect of these compounds, together with the increasing expansion and impact of human populations on the biosphere are getting more and more severe (Palumbi 2001) (Figure 1.1). As a consequence of these abiotic stress factors, the changes in organisms around us is accelerating, thus affecting a huge amount of species, including disease organisms and agricultural pests (Palumbi 2001). There are examples of adaptation to pesticides in several organisms, such as adaptation of *Drosophila melanogaster* strains to DDT or the adaptation of the malaria vector *Anopheles coluzzi* to permethrin (Palumbi 2001, Daborn et al. 2002, Kim et al. 2018, Main et al. 2018). Moreover, the adaptation of weeds to herbicides has also been reported, being one of the most problematic areas in crop production (Heap 1997, Oerke 2005, Baucom 2019). Since a lot of species that gain resistance can be parasites or pests, their presence can cause serious health and economic problems, and moreover, ecologically relevant species can disappear.

Another group of chemicals, the heavy metals, are among the pollutants of greatest importance and concern in the world nowadays, and are also strong drivers of selection (Gadd 2010, Dixit et al. 2015). There are some metals which are nutrient metals, such as copper and zinc, and thus are necessary for essential growth and metabolic functions, however the excessive amount of metal may disrupt the proper biological pathways (Hoostal et al. 2008). Heavy metals are especially dangerous, because they do not degrade. It's accumulation in plants and animals can get into the food chains and can cause for instance oxidative stress, by accumulating reactive oxygen species (ROS) (Cavalcanti Luna et al. 2015). There are various documented cases for the evolution of resistance to heavy metals in several organisms (Meharg et al. 1993, Mckenzie et al. 1994). One example of adaptation of human populations to heavy metal occurred in San Antonio de los Cobres, Argentina, where people

have adapted to high level of the acutely toxic arsenic (Schlebusch et al. 2015). In another study by checking the extracellular enzyme activity (EEA) profiles of microbe communities in polluted and non-polluted areas of Lake Erie, the authors found local adaptation to copper, arsenic and cadmium (Hoostal et al. 2008).

There are some abiotic stress factors which reduce the level of oxygen, and thus cause hypoxia stress (Bickler and Buck 2007, Zhao et al. 2020). One example of adaptation to these environments is in human populations on the Tibetan Plateau, Andean Altiplano, and the Semien plateau of Ethiopia at high altitudes (>2500m). Populations living in these places are constantly exposed to hypoxia, which cause insufficient supply of oxygen to vital organs. However, these populations managed to adapt to these harsh conditions and are there since thousands of years (Beall 2006). When comparing these populations at high altitudes with populations at low altitudes, genomic signatures of adaptation have been found (Lachance and Tishkoff 2013).

With the ongoing climate change and expanding human populations, and due to the increasing amount of the previously described biotic and abiotic stresses, habitats of several species are becoming increasingly stressful (Wray 2003). When there is a stressful environment, organisms in order to survive can go through behavioural (shift in the habitat, migration and dispersal to zones with optimal conditions), physiological (phenotypic plasticity) or adaptive evolutionary changes (Travis et al. 2013, Seebacher et al. 2015). Adaptive evolution is suggested to play the most important role in the survival of species during challenging environmental conditions, mostly through changes in the expression of genes, which then manifests through heritable phenotypic variations (Tuğrul et al. 2015).

1.3. Transcriptional regulation during eukaryotic stress response

Gene regulation in eukaryotes is extremely complex and fundamental for every biological processes both in the absence of a stress and under stress conditions (Lelli et al. 2012). In case of perturbations, a cascade of internal alterations is activated, which mostly manifests in alterations in transcriptional regulation and results in changes in the expression of specific genes (Tuğrul et al. 2015, Buffry et al. 2016). Stress is noted by a stress-specific regulatory sensing system and the stress signal is transduced via a stress sensing network (Kültz 2005). Then, a group of proteins known as transcription factors (TF), will be translocated to the nucleus and will bind to specific DNA sequences, which will promote the gene expression of target genes (van Straalen and Roelofs 2012). The region where these specific DNA

sequences can be found are called *cis-regulatory* regions, such as core promoters that are in near proximity of the genes, and are the ones orchestrating the changes in the gene expression in *cis* (Spitz and Furlong 2012). There are also various *cis-regulatory* modules that are localized at greater distance from the transcription start site, such as insulators, silencers, tethering elements and enhancers (Spitz and Furlong 2012). Promoters and enhancers are one of the main factors in driving gene expression, and in eukaryotes usually consist of hundred and several thousand base pairs (bps) in length (Yao et al. 2015). They harbour transcription factor binding sites (TFBSs) - which typically consist of 6-12 bps - to which the TFs can bind, thus serve as operational platforms to recruit TFs. These previously mentioned internal processes are providing great basis on which adaptive evolution can act (Spitz and Furlong 2012, Tuğrul et al. 2015).

1.3.1. *Cis-regulatory* evolution as an adaptive driving force

There are several ways how adaptive evolutionary forces can affect the expression of a gene. For instance, adaptive mutations can be located in the coding region of a target gene (structural evolution) or in the regulatory regions (regulatory evolution). The mutations in the genes encoding the transcription factors and in their regulatory regions are considered as *trans-regulatory* changes. Meanwhile mutations in the regulatory regions of the target genes can be responsible for *cis-regulatory* evolution (Long et al. 2016) (Figure 1.2). Most of the evolutionary changes that may represent the predominant genetic basis for phenotypic evolution, are thought to happen in the *cis-regulatory* regions such as promoter and enhancer sequences (Wray 2003). It is proposed, that they evolve much faster than coding sequences, thus we expect a higher genetic variation on which evolutionary forces can act (Gerhart and Kirschner 1997, Stern 2000, Carroll et al. 2001, Wilkins 2002). An essential factor that underlies this variability in the promoter and enhancer regions are the TFBSs. It is thought that evolutionary changes in gene expression mostly concerns substitutions (such as SNPs), deletions, inversions and insertions in the TFBSs (van Straalen and Roelofs 2012). There are examples of the evolution of enhancers underlying phenotypic evolution. For instance, a polymorphism in the Hox/Pax-responsive enhancer has been proved to play a role in the evolution of the vertebrate spine (Guerreiro et al. 2013). Another example in *Drosophila biarmipes* shows that the evolution of a spot on the wing involved modifications of an ancestral *cis-regulatory* element of the *yellow* pigmentation gene (Gompel et al. 2005).

As mentioned before, changes in the regulatory networks are known to be important for organismal evolution, however the mechanisms and genetic variations underlying the emergence of new regulatory elements are still not fully understood (Ivancevic and Chuong 2020). Apart from deletions, inversions and single nucleotide polymorphisms, transposable elements (TEs) have been proved to be particularly powerful players in the generation of genetic variation (Capy et al. 2000, Bennetzen and Wang 2014, Makarevitch et al. 2015, Schrader and Schmitz 2019). Thus, unlike in the early years, TEs are no longer recognized as the negligible fraction of genomes, but they are considered as potential contributors to evolutionary adaptation (Feschotte 2008). However, the genome wide role of transposable elements in the *cis-regulatory* module repertoire of the most relevant environmental stresses is poorly studied in most of the species, including both humans and *Drosophila melanogaster*.

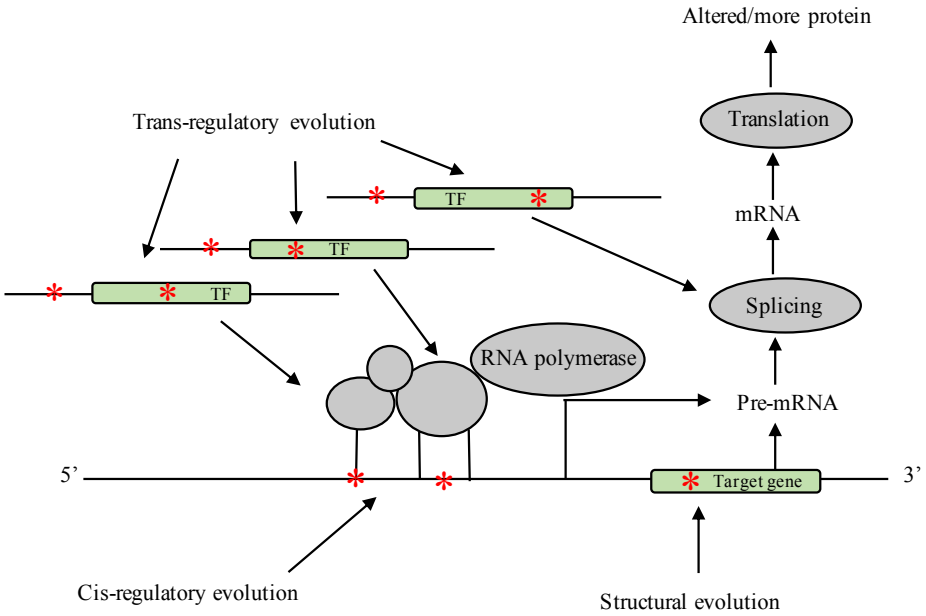


Figure1.2. Schematic view of adaptive evolutionary changes affecting a target gene. The red stars represent the mutations. TF (Transcription factor coding gene). Figure modified from van Straalen and Roelofs 2012.

1.4. Transposable elements, a historical overview

Transposable elements are mobile genetic units, which have the ability to change their position within a genome (Schrader and Schmitz 2019). They can proliferate their copy number and contribute to genome expansion by changing their location in the genome (Percharde et al. 2020). They were discovered by Barbara McClintock in the 1940s, who also called them "controlling elements". This expression originated from McClintock's discovery,

that these mobile elements seemed to have a role in the changing color pattern of maize (*Zea Mayes*) kernels (Mcclintock 1950, Mcclintock 1953). She also suspected, that the activation of these "controlling elements" can be triggered by some kind of genomic stress (Mcclintock 1956). For the discovery of these elements she was awarded with the Nobel Prize in Physiology or Medicine (1983), being the first women to win an unshared Nobel prize in this category. The idea of a dynamic genome, controlled by these mobile elements was first disregarded in the scientific community, since at that time the accepted scientific consensus was that genes are precisely aligned and fixed along the chromosomes (Schrader and Schmitz 2019). However, there were some pioneers who further developed the "controlling element" hypothesis, such as Britten and Davidson in the late 1960s, who developed the so called "Gene-Battery" model (Britten and Davidson 1969, Chuong et al. 2016). This idea conjectured that the amplification of transposable elements in the genome could spread regulatory elements to drive the evolution of gene-regulatory networks (Britten and Davidson 1969). Despite the efforts to highlight the role of TEs in gene regulation, they were still considered as parasitic "selfish elements" which have no beneficial effects on the host organism (Ohno 1972, Doolittle and Sapienza 1980, Orgel et al. 1980). This assumption was further boosted by the discovery of deleterious mutations caused by TEs (Kazazian et al. 1988). Since the late 90s up to date, there are an increasing number of evidences for TE-mediated beneficial genetic modifications, which made the reputation of TEs to change (Volff 2006, Feschotte 2008, Oliver and Greene 2009, Fedoroff 2012). Even though TEs are often referred as double-edged swords, nowadays whole genome studies shed light on their diversity and pervasive nature, and they cannot be marginalized anymore (Hurst and Werren 2001).

1.4.1. Transposable element classification

Nowadays, in the third-generation sequencing era, more and more high-quality genomes are available, a huge part of which are composed of transposable elements (Wicker et al. 2007). Although there is a constantly emerging amount of data, the identification and annotation of TEs is still a challenging task, and the classification of these elements is in a constant flux (Bourque et al. 2018). It is necessary to have a common classification system for different species, thus making possible the comparative and evolutionary studies of TEs between taxa. There are several approaches to classify TEs, one of which is based on their replication capability, which classifies TEs into autonomous and non-autonomous elements.

Autonomous elements encode all the domains necessary for moving around in the genome, such as open reading frames (ORFs) and regulatory sequences. However, non-autonomous TEs lack this capacity to generate their own regulatory sequences, so in order to move, they depend on the autonomous insertions.

In 1989, Finnegan introduced another classification system, which divided the TEs in two classes based on their transposition intermediate (Finnegan 1989). Class I elements or also called retrotransposons, mobilize through a "copy and paste" mechanism, with an RNA intermediate, which is reverse-transcribed into a cDNA copy, and is integrated elsewhere in the genome (Boeke et al. 1985) (Figure 1.3). Class II contains several unrelated groups of TEs, which have different transposition mechanisms, the only common thing being the absence of an RNA intermediate in the transposition process (Figure 1.3). Thus, inside the Class II group we can find TEs that use a tyrosine recombinase (*Crypton*), a transposase (TIRs), replication initiation-like protein (*Helitron*), and TEs that use a protein-primed B-type DNA polymerase (*Maverick*) (Finnegan 1989).

The next unified classification systems the so-called "Wicker" and "Repbase" systems were based on the classification suggested by Finnegan (Finnegan 1989). The basis of the Wicker classification are enzymatic and mechanistic criteria. He further divided the two main classes in orders (based on their insertion mechanism, structure and encoded proteins) and superfamilies (based on their replication strategy and on the presence and size of target site duplications), and different families (based on sequence conservation) (Wicker et al. 2007, Kapitonov and Jurka 2008)(Figure 1.3). The Repbase classification system apart from suggesting a unified nomenclature, divides TEs in two types: Type 1 (DNA transposons) and Type 2 (retrotransposons). These TE types are composed of five major classes and the classification is based on enzymology, structural similarities and sequence relationships. Each class of TEs is composed of small number of superfamilies and clades, and each superfamily consist of several families (Kapitonov and Jurka 2008).

Note, that there are new suggestions for improvement of the classification systems which have been proposed. For instance, Piegu et al 2015 proposed to use a different system, which not only considers the sequence homology, structural features and target site duplications but also the evolutionary origin of TEs (Piégu et al. 2015). The same authors also suggested that the scientific community should consider a classification which includes both prokaryotic and eukaryotic TE classes (Piégu et al. 2015). Arkhipova also proposed changes in the classification system, which is based on the replicative, integrative, and structural

components of TEs, and integrates different aspects of all the existing classification systems (Arkhipova 2017).

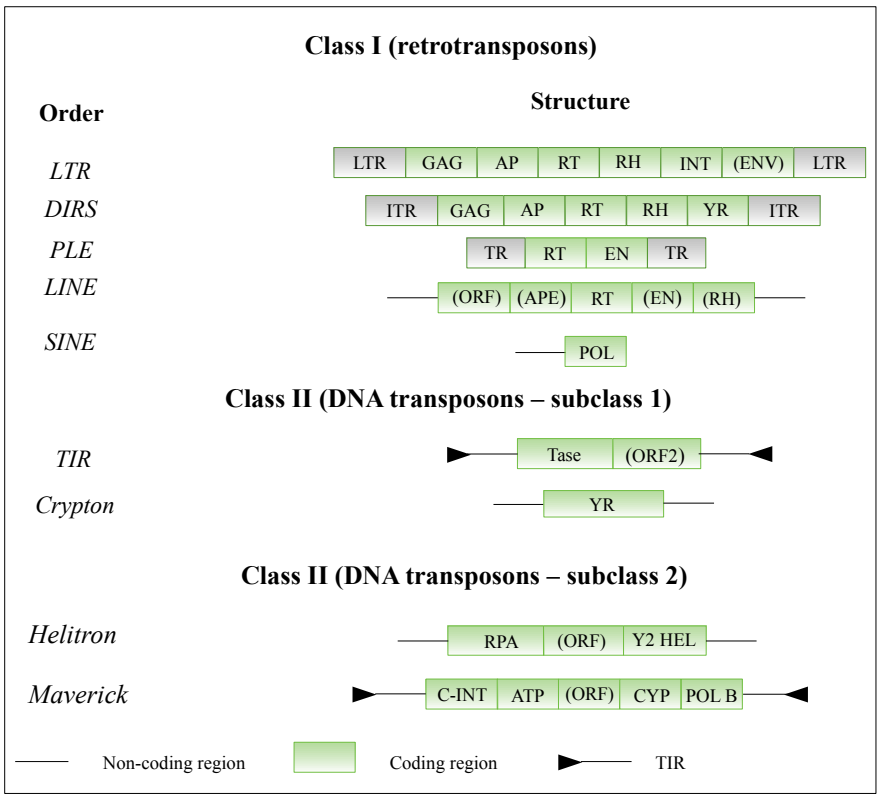


Figure 1.3. Classification of eukaryotic transposable elements based on Wicker et al 2007. The genetic structure of every order of TEs is represented. Components shown in brackets are not common for all the TEs from the same order. AP: Aspartic proteinase, APE: Apurinic endonuclease, ATP: Packaging ATPase, C-INT: C-integrase, CYP: Cysteine protease, EN: Endonuclease, ENV: Envelope protein, GAG: Capsid protein, HEL: Helicase, INT: integrase, ORF: Open reading frame, POL: Polymerase III promoter, POL B: DNA polymerase B, RH: RNase H, RPA: Replication protein A, RT: Reverse transcriptase, Tase: Transposase, YR: Tyrosin recombinase, Y2: YR with YY motif

1.4.2. Transposable element load in different organisms

TEs have been found virtually in all eukaryotic and in almost all prokaryotic species investigated, with the exception of *Plasmodium falciparum* (Gardner et al. 2002, Wicker et al. 2007, Hua-Van et al. 2011). They represent a considerable albeit variable fraction of the genomes, ranging from ~1% to almost 90% (Touchon and Rocha 2007, Ambrozová et al. 2011) (Figure 1.4). The genomic abundance and composition of TEs can vary even among species belonging to the same phylogenetic group (Sessegolo et al. 2016). Also, the

abundance of TE classes is not the same among species. For instance, in the human and mouse genome the most abundant and active TE family are LINE and SINE, respectively. In *C. elegans* the most common TE group is the DNA transposons, while in *Drosophila melanogaster* the LTR elements compose the biggest fraction of the genome (Lander et al. 2001, Hua-Van et al. 2011, Sessegolo et al. 2016).

There are a lot of studies investigating the abundance of different elements in different genomes, however in most cases the annotated TE content is bias toward insertions annotated in the reference genomes (Guio and González 2019). Thanks to the new long-read sequencing and annotation tools, there are a lot of TEs that can be annotated, and were unrecognised until now (Flutre et al. 2011, Bao et al. 2015). As examples, we can mention the TE content in the human genome, which was predicted to be around 40-45%, and later on de Koning et al (2011) found that at least 66-69% of the human genome consists of TEs (De Koning et al. 2011). Something similar happened in *D.melanogaster* and *D.buzzatii,* where thanks to third-generation sequencing techniques, ~37% more TE insertions were annotated compared to the previously used short-read sequencing techniques (Rius et al. 2016, Chakraborty et al. 2018).

As the examples show, the appearance of the long-read sequencing techniques represents a novel opportunity to uncover until now hidden genetic variability and transposable element load in all types of genomes.

1.4.3. Transposable elements as double-edged swords in genome variation

The contribution of transposable elements to the genetic and genomic variability of organisms is extraordinary and versatile (Schrader and Schmitz 2019). Either by inserting in regulatory or coding regions or by nonhomologous ectopic recombination, they can generate a great diversity of mutations (Bourque et al. 2008). However, most of these changes will have negative or neutral effects on the host fitness. One of the early examples identified where TEs can be harmful was described in humans. Kazazian and co-workers identified that a *LINE-1* element inserted into the 14th exon of the factor VIII gene causes haemophilia A (Kazazian et al. 1988). Moreover, transposable elements have been associated with several age-related neurodegenerative disorders, such as Alzheimer's disease, and also with different type of cancers (Miki et al. 1992, De Cecco et al. 2019).

Even though the relatively rare beneficial and adaptive changes have to be searched *like needles in a haystack*, there is a growing number of evidences pointing to the fact that transposons can have beneficial effects on the organism (Volff 2006, Oliver and Greene 2009, Fedoroff 2012, Casacuberta and González 2013). In the last decades, the role of TEs in genome evolution has been revisited and today it is well recognized that TEs have been involved in several cases of adaptive evolution (Chuong et al. 2017).

1.4.3.1. Transposable element activation due to environmental stress

Stress conditions have been repeatedly associated with the activation of TEs, which has been thought to be an evidence for the adaptive role of TEs in stress response (Mcclintock 1956, Capy et al. 2000, Fablet and Vieira 2011, Chénais et al. 2012, Casacuberta and González 2013, Negi et al. 2016, Rey et al. 2016). Both biotic and abiotic stresses have been shown to promote TE activation, increasing both TE transcriptional and mobilization rates (Horváth et al. 2017). The process of TE activation is thought to lead to an increase in the mutation rate and generating variability, upon which natural selection can act (Cowley and Oakey 2013).

The activity of TEs have been shown in several organisms and in different stress conditions. For instance, in the yeast *Schizosaccharomyces pombe* it has been found that TE mobility was greatly increased due to stresses such as heavy metals, caffeine, and the plasticizer phthalate (Esnault et al. 2019). Another example in the tomato plant showed that the accumulation of transcripts and transposition intermediates of the *Rider* retrotransposon family in the form of extrachromosomal DNA, is triggered by drought stress (Benoit et al. 2019). In some examples, the underlying mechanisms of TE activation has been described, such as in Van Meter et al (2014). In this study they found that the longevity regulating protein Sirtuin 6 (SIRT6), which in normal conditions silence LINE-1 elements, under stress conditions gets relocated to DNA damage sites, and thus LINE1 elements can be transcribed (Van Meter et al. 2014). In Drosophila it has been shown that the expression of TEs increases by heat shock at a post transcriptional level, due to the effect of the inducible chaperone Hsp70 on Piwi-interacting RNA (piRNA) biogenesis (Cappucci et al. 2019). Interestingly, besides TE activation, stress has also been related with TE repression (Menees and Sandmeyer 1996, Trivedi et al. 2014, Horváth et al. 2017).

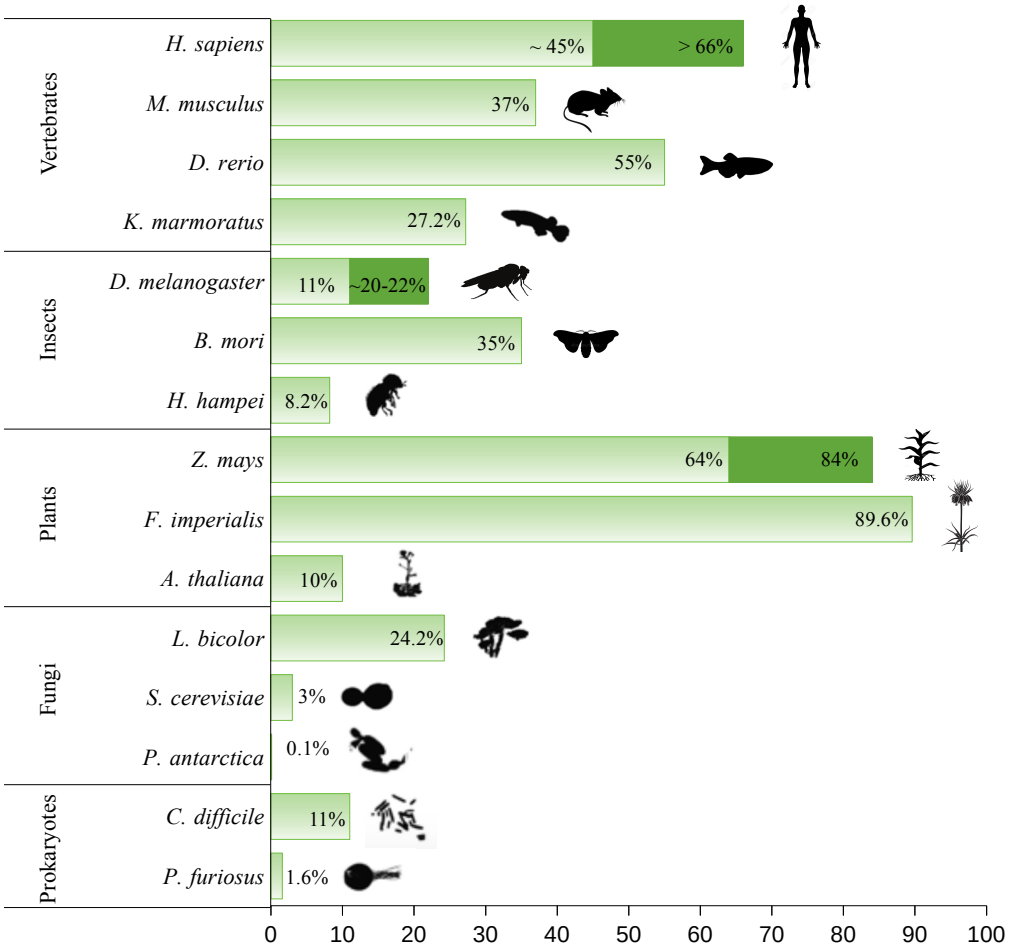


Figure 1.4. Transposable element abundance (%) in the genome of different organisms. Figure adapted from Guio and Gonzalez (2019)

1.4.4. Transposable elements as a source of coding and non-coding RNA

One of the ways for TEs to have adaptive potential is when functional proteins generated from the TE coding sequences benefit the host. This process has been discovered a long time ago (Brosius 1991, Britten 1996), and has been called using several names, such as domestication (Miller et al. 1997), co-option (Sarkar et al. 2003) or exaptation (Brandt et al. 2005). Exaptation is mostly used in specific cases, where the evolved trait has a different usage of nucleotides or a different function, compared to its original form (Schrader and Schmitz 2019).

In some occasions TE derived proteins have been repurposed as part of the defence system against retroviruses or TEs themselves (Jangam et al. 2017). In those cases, the potential deleterious effect of invasive genetic elements like retroviruses and TEs was prevented with

the recruitment of TE proteins. For instance in one of the examples in *Arabidopsis thaliana*, they found that *MAIL1* and *MAIN* genes define an alternative silencing pathway by encoding a Ty3/gypsy retrotransposon-related plant mobile domain (Ikeda et al. 2017). Another example of retroelement gene domestication is the case of *L1TD1* gene, whose protein-coding sequence is almost entirely derived from a LINE-1 retroelement. This domain was co-opted from the open reading frame 1 (ORF1) of *LINE-1* elements by the ancestor of eutherian mammals and it is thought that at first it was acquired for genome defense against *LINE-1* elements (Mclaughlin et al. 2014). However, *L1TD1* has also been shown to play an important role in the maintenance of pluripotency (Mclaughlin et al. 2014). In other cases, TE-derived proteins have been domesticated to catalyse important cellular mechanisms (Miller et al. 1997, Jangam et al. 2017). A spectacular well-known example for TE domestication is the adaptive immune system, where *Rag1* and *Rag2* antigen receptor proteins derive from TEs. These receptors initiate the assemble of the gene segments that generate immunoglobulin and T cell receptors in vertebrates, known as the V(D)J recombination, which is a conserved process of jawed vertebrates (Agrawal et al. 1998). A recent study provided the definitive evidence for the transposon exaptation of *Rag* antigen receptors (Huang (Huang et al. 2016). In this work, that was done in the cephalochordate lancelet, the authors found that *Rag1* and *Rag2* descend from ProtoRAG, an ancestral *Transib* transposon that was transmitted vertically through chordate and vertebrate evolution (Huang et al. 2016). Another extraordinary example of TE domestication is in Drosophila, where the control of telomeres is orchestrated by two domesticated retrotransposons, *HeT-A* and *TART* (Pardue and Debaryshe 2003). Moreover, in a recent study they found that *TART-A* has captured a portion of the *nxf2* gene, which is involved in suppressing *TART-A* activity via the piRNA pathway and that *TART-A* produces abundant piRNAs, some of which are antisense to the *nxf2* transcript (Ellison et al. 2020).

TEs may also be substantial contributors to the non-protein coding RNA repertoire of the genomes, for instance it has been shown that they are major components of long non-coding RNAs in human and mouse genomes (Kapusta et al. 2013). Some of these TE derived long non-coding RNAs have been shown to be important players in different cellular functions, such as the *HERVH* human endogenous retrovirus, which has been showed to be a nuclear long non-coding RNA required to maintain stem cell pluripotency (Lu et al. 2014). Moreover, TE derived microRNAs and other small RNAs derived from TEs can also adopt regulatory roles serving host cell functions (Piriyapongsa and Jordan 2007, Mccue and Slotkin 2012).

For example, it has been described in human brain tissues, that LINE-2 elements are mayor depository of functional microRNAs (Petri et al. 2019).

1.4.5. Transposable elements as "batteries" of gene regulation

TEs can be beneficial for their host also by their regulatory potential, which echo the visionary predictions of Barbara McClintock and David and Britten (Mcclintock 1956, Britten and Davidson 1969). The contribution of TEs to alterations in gene expression is relevant for the short-term adaptation to environmental changes (Moschetti et al. 2020).
There is now mounting evidence that newly inserted TEs can be a rich source of material for changes in gene expression (Feschotte 2008, Chuong et al. 2017). They can modulate gene expression in myriad ways in *cis* and *trans*, and also post transcriptionally (Figure 1.5). Among others, TEs can spread insulator sequences (Lunyak et al. 2007, Schmidt et al. 2012, Wang et al. 2015), repressive elements (Lippman et al. 2004, Rebollo et al. 2011) or immense amounts of promoters and enhancers (Bejerano et al. 2006, Jacques et al. 2013, Chuong et al. 2016, Thompson et al. 2016, Trizzino et al. 2017), and can also add TFBSs (Wang et al. 2007, Kunarso et al. 2010, Sundaram et al. 2014, Ito et al. 2017, Bourque et al. 2018, Sun et al. 2018, Ullastres et al. 2019) (Figure 1.5). They can add TFBSs in the proximity of promoter-less coding sequences or they can juxtaposition to existing promoters (Moschetti et al. 2020). TFBS can be generated by point mutations and develop to functional binding sites inside TEs by time, or TEs can have pre-existing binding sites (Feschotte 2008). By acting as regulatory modules, transposons can regulate the expression of one gene or can also re-wire and fine-tune whole regulatory networks (Chuong et al. 2017). One of the classroom examples of adaptation to environmental changes fuelled by the effect of a transposable element on one gene, is the change in the colour polymorphism of the peppered moth (*Biston betularia*) (Van't Hof et al. 2016). The dark coloured phenotype in this species evolved through the intronic insertion of a large, tandemly repeated, transposable element into the first intron of the gene *cortex*. The insertion results in the increase in the transcript abundance of the affected cortex gene (Van't Hof et al. 2016). Another great example, is in the African oil palm (*Elaeis guineensis)*, where a LINE retrotransposon (*Karma*) inserts into the intron of the gene *DEFICIENS,* important for flowering. The methylation level of the inserted TE ultimately controls whether or not the plant bear oil-rich fruit (Ong-Abdullah et al. 2015).

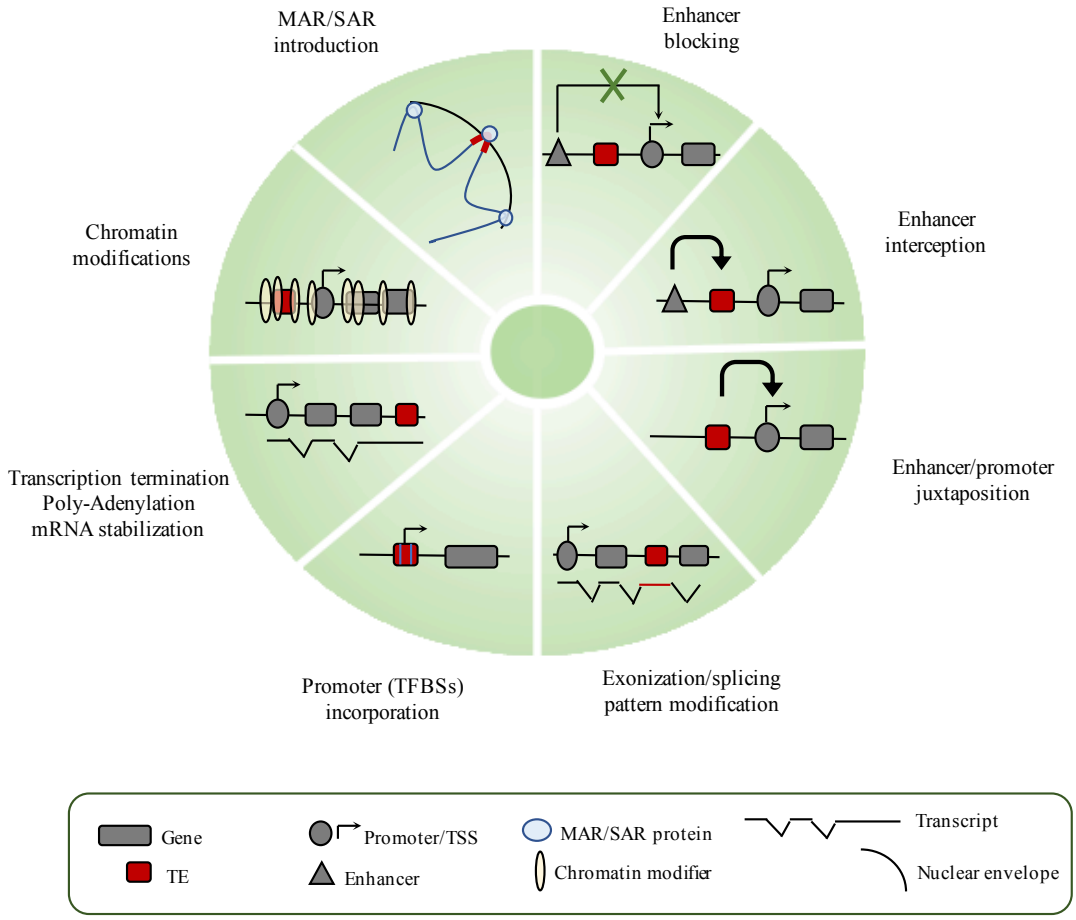


Figure 1.5. Representation of the key regulatory effects of TEs. Adapted from Bourque et al 2018.

There are several examples of TEs regulating whole networks of genes, where they have provided the building blocks for the assembly and remodeling of *cis-regulatory* networks during evolution (Reviewed in (Bourque et al. 2018). These modifications include very diverse processes, such as pregnancy, innate immunity and neocortex development in mammals, or the response to abiotic stress in maize (Kunarso et al. 2010, Lynch et al. 2011, Shen et al. 2011, Wang et al. 2014, Makarevitch et al. 2015, Notwell et al. 2015, Chuong et al. 2016). In one outstanding example, they showed that ERVs shape the transcriptional network related to the interferon (IFN) response (Chuong et al. 2016). They studied the evolution of gene regulatory networks induced by the pro-inflammatory cytokine, interferon gamma (IFNG), and found that lineage-specific ERVs have dispersed numerous IFN-inducible enhancers. Moreover, by deleting a group of ERV elements using CRISPR-Cas9 genome editing, the involvement of ERVs in essential immune functions, including activation of the AIM2 inflammasome was found (Chuong et al. 2016).

In another work, the authors analysed ChIP-Seq data from 26 transcription factors (TFs) in human and mouse and they found that 2-40% of the binding sites of these TFs reside in TEs (Sundaram et al. 2014). They also found that 66% of TE-derived binding events were cell type specific, and most of them were species-specific, however there were some which were conserved between human and mouse (Sundaram et al. 2014).

Another example can be found in maize, where Makarevitch and colleagues did gene expression analysis in several stress conditions (cold, heat, high salt, UV stress) and checked if there were any transposons associated with the changes in the expression. They found, that between four and nine different TE families are associated with up-regulated gene expression. These results suggest that TEs may provide local enhancer activities that stimulate stress-responsive gene expression (Makarevitch et al. 2015). Even though there are several examples of TEs causing adaptive evolution, the role of TEs in this process remains a matter of intense interrogation.

There are no studies checking the genome wide contribution of TEs to specific stress-related enhancer repertoire, neither in humans nor in Drosophila. The contribution of TEs to adaptation and gene expression has been studied in a great variety of species, however undoubtedly Drosophila is among the most powerful model organisms to study TE-related adaptation and their effects on stability and evolution on genes and genomes (Moschetti et al. 2020).

1.4.6. Transposable elements expand the *cis-regulatory* module repertoire in *D. melanogaster*

Drosophila melanogaster is a good model to study the contribution to regulatory sequences by eukaryotic TEs, because it has 30% of its TE repertoire potentially active, compared to humans and other mammals, where there are very few active copies (Moschetti et al. 2020)2020). Moreover, it is a great model to study adaptive changes, since it is originated from southern Africa and has spread all over the world relatively recently (19,000 years ago), suggesting that it had to adapt fast to different environmental conditions (David and Capy 1988, Li and Stephan 2006, Arguello et al. 2019, Sprengelmeyer et al. 2020).

The studies of TEs in the Drosophila genome started in the 70s, with studies hypothesizing that repetitive sequences have been inserted at new sites *in vitro* and *in vivo* in the *D. melanogaster* genome (Strobel and Dignam 1978, Potter et al. 1979). However, later on more studies followed, and the Drosophila P-M hybrid dysgenesis and the instability of an eye-

color phenotype has also been associated with transposable elements (Engels 1979, Rasmuson et al. 1980, Bingham et al. 1982, Rubin et al. 1982). Moreover, TE related studies also led to the discovery of widely used insertional mutagenesis tools such as the P-elements from *D. melanogaster* genome (Rubin and Spradling 1982, Spradling and Rubin 1982). Nowadays there are more and more studies in Drosophila species including *D. melanogaster*, showing the importance of the regulatory functions of TEs in these organisms (Moschetti et al. 2020). For instance, there are several examples of TE-derived *cis-regulatory* elements. TEs have been shown to provide insulators (Spana et al. 1988, Roseman et al. 1995, Cai and Levine 1997, Conte et al. 2002), silencers (Zatsepina et al. 2001, Lerman and Feder 2005), alternative TSS (Merenciano et al. 2016) altered splicing events (Ding et al. 2016) or Poly-A signals (Marsano et al. 2005, Mateo et al. 2014). An interesting example is the adaptive insertion of a *POGON1* DNA transposon within the 3′ UTR of the gene *CG11699* in *D. melanogaster*. The insertion disrupts one of the two polyadenylation signals of the gene resulting in a shorter 3′ UTR, elevated mRNA levels and increased resistance to xenobiotic stress (Mateo et al. 2014).

Even though there are several examples of TEs acting as *cis-regulatory* factors, the most common case is when TEs act as enhancers or promoters (Tanda and Corces 1991, Brönner et al. 1995, Wilson et al. 1998, Deprá et al. 2009, Batut et al. 2013, Clemmons and Wasserman 2013). A very well-known example is the DDT resistance in *D. melanogaster*, where the insertion of an *Accord* transposable element in the 5' end of a cytochrome P450 gene (Cyp6g1) causes over-expression of this gene and thus, production of a large amount of biotransformation enzyme that can degrade DDT (Daborn et al. 2002). Another example of the contribution of TEs to the TFBS repertoire and thus the regulation of the nearby genes has been shown in a recent work of Ullastres and co-workers (2019). They performed enhancer assays and identified several TEs that add functional TFBSs and change the expression of the reporter gene in immunity stress conditions (Ullastres et al. 2019). In another recent study authors found that TEs could contribute to the regulation of gene expression under malathion insecticide exposure, by remodelling the *cis-regulatory* network in *D. melanogaster* (Salces-Ortiz et al. 2020).

Despite the existing studies of TEs adding TFBSs and regulating stress response genes, the contribution of TEs to specific and biologically relevant stress response networks e.g. heat shock, heavy metal or desiccation is still not clear.

SECTION 2

OBJECTIVES

2. OBJECTIVES

The objectives of the present thesis are:

1. **To test if transposable elements (TEs) contribute to the genome wide distribution of stress related transcription factor binding sites (TFBSs) in humans and *Drosophila melanogaster*, and if these binding sites are functional.**

   We will predict transcription factor binding sites (TFBS) and transcription factor binding motifs (TFBM) using available data in humans and *D. melanogaster*. Moreover, we will choose candidate insertions which can potentially act as enhancers in *D. melanogaster* and functionally validate them.

2. **To understand the transcriptomic and physiological basis of desiccation resistance in European natural *Drosophila melanogaster* populations and the role of TEs in this process.**

   We will subject natural *D. melanogaster* strains to dry environmental conditions and look for differences in their survival rates. Once we discover the tolerance level of different strains, we will look for the responsible physiological and molecular mechanisms. We will investigate the TEs which could possibly play a role in adaptation to desiccations tress response.

SECTION 3

RESULTS

3. RESULTS

3.1. The structure of the research

The results of the present thesis are structured in two sections. Section 3.3. corresponds to a published article, while Section 3.4. is a manuscript.

Section 3.3. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*

In this section, we identified transcription factor binding sites and binding motifs embedded within TEs in humans and *D. melanogaster*. We found that in the human genome transposable elements are enriched for *NFE2L2* TFBSs, and that transcription factors such as *CREB1* and *NF-κB* have slightly more TFBMs within TEs than expected. Meanwhile, in the Drosophila genome TEs were enriched for *caudal*, *dorsal*, *HSF* and *tango* binding sites, however, the percentage of TFBMs appears to be small. Moreover, by doing enhancer reporter assays in *D. melanogaster*, we showed that TEs can affect the expression of the reporter gene in immune stress conditions.

Section 3.4. The transcriptomic and physiological basis of desiccation tolerance in natural *Drosophila melanogaster* populations

In this section, we investigated the transcriptomic and physiological basis of *D. melanogaster* desiccation stress response and identified possible candidate transposable element insertions which might affect the gene expression in response to stress. We subjected to desiccation stress 74 natural *D. melanogaster* strains from five different climate zones, and found differences in the tolerance of the strains among them. We also described, that the tolerant and sensitive strains differ in physiological traits relevant for desiccation stress. Moreover, we showed that genes related to stress response and environmental sensing are up-regulated, while genes related to diverse metabolic processes are down-regulated after desiccation stress conditions. Finally, we generated a list of mostly strains specific candidate TE insertions possibly involved in the regulation of desiccation stress response genes.

The two sections (3.3. and 3.4.) and further directions are discussed in Section 4 (Discussion) and conclusions of this thesis are presented in Section 5 (Conclusions). References quoted in Section 1 (Introduction) and Section 4 (Discussion) are shown in Section 6 (References), while Sections 3.3. and 3.4. contain their own references.

3.2. Thesis advisor's report about authorship and impact factor of the publications of this doctoral thesis

Publication (Section 3.3)

Villanueva-Cañas JL*, **HorváthV***, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Res*. 2019; 47 (13): 6842-6857. doi:10.1093/nar/gkz490

Impact factor (2019): 11.501

5 Year impact factor: 11.797

Authorship: Vivien Horváth is a co-first author with equal contribution with Villanueva-Cañas JL

Contribution: In this work Vivien Horváth contributed to the design of the experimental part of the research, performed experiments, analysed and interpret the data and drafted the manuscript.

(Section 3.4) (In preparation)

HorváthV, Guirao-Rico S, Salces-Ortiz J, Rech G.E, Aprea E, Rodeghiero M, Gianfranco A, González J. The transcriptomic and physiological basis of desiccation tolerance in natural European *Drosophila melanogaster* populations.

Impact factor (2019): NA

5 Year impact factor: NA

Authorship: Vivien Horváth is a first author

Contribution: In this work Vivien Horváth contributed to the design of the project, performed the experiments, analysed and interpret the data and drafted the manuscript.

3.3. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*

Villanueva-Cañas JL, **Horvath V**, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Res*. 2019;47(13):6842-6857. https://doi.org/10.1093/nar/gkz490

6842–6857 *Nucleic Acids Research, 2019, Vol. 47, No. 13* *Published online 8 June 2019*
doi: 10.1093/nar/gkz490

Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*

José Luis Villanueva-Cañas†, Vivien Horvath†, Laura Aguilera and Josefa González*

Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain

Received December 19, 2018; Revised May 20, 2019; Editorial Decision May 21, 2019; Accepted May 22, 2019

ABSTRACT

Although transposable elements are an important source of regulatory variation, their genome-wide contribution to the transcriptional regulation of stress-response genes has not been studied yet. Stress is a major aspect of natural selection in the wild, leading to changes in the transcriptional regulation of a variety of genes that are often triggered by one or a few transcription factors. In this work, we take advantage of the wealth of information available for *Drosophila melanogaster* and humans to analyze the role of transposable elements in six stress regulatory networks: immune, hypoxia, oxidative, xenobiotic, heat shock, and heavy metal. We found that transposable elements were enriched for caudal, dorsal, HSF, and tango binding sites in *D. melanogaster* and for NFE2L2 binding sites in humans. Taking into account the *D. melanogaster* population frequencies of transposable elements with predicted binding motifs and/or binding sites, we showed that those containing three or more binding motifs/sites are more likely to be functional. For a representative subset of these TEs, we performed *in vivo* transgenic reporter assays in different stress conditions. Overall, our results showed that TEs are relevant contributors to the transcriptional regulation of stress-response genes.

INTRODUCTION

Transposable elements (TEs) represent a large portion of eukaryotic genomes. They are repetitive sequences that have the ability to move around in the genome making new copies of themselves in the process. Some TEs contain regulatory sequences such as promoters, transcription start sites (TSSs) and transcription factor binding sites (TFBSs) that can affect the expression of nearby genes (1,2). Multiple examples of individual TE copies affecting gene expression have been described in a wide-range of organisms (3). More recently, genome-wide approaches have been used to explore the overall contribution of TEs to gene regulation (4,5). In particular, several studies have found that TEs contain binding sites for a variety of transcription factors (TFs) involved in very relevant cellular processes such as cell pluripotency, placenta development or immune response (6–9). These studies also found that it is one or a few TE families the ones that contribute more to the TFBS repertoire.

Most of the genome-wide approaches aimed at identifying TFBSs in TEs are based on the analysis of chromatin immunoprecipitation sequencing (ChIP-seq) that provide experimental evidence for the binding of a particular TF to a discrete genomic region. However, ChIP-seq only provides information for the binding sites occurring in the particular conditions in which the experiment is performed. Because it is impossible to assay all tissue types and developmental stages under all conditions, combining binding site predictions using ChIP-seq with transcription factor binding motif (TFBM) predictions using bioinformatic tools should help identify a more complete dataset of binding sites (10). The binding profiles for an increasing number of TFs are available in dedicated databases such as JASPAR, including the newer ones based on hidden Markov models named transcription factor flexible models (TFFM) (11,12). Several genomic features, such as chromatin accessibility or epigenetic marks, are often used to evaluate the regulatory potential of the genomic sequences containing TFBS (13,14). In any case, functional validation of the identified TFBSs is needed to conclude that the predicted binding sites are functional.

Most stress-related TFs are conserved across organisms (15). Stress is a major aspect of natural selection in the wild that leads to changes in the transcriptional regulation of a variety of genes. Both in humans and in the fruitfly *Drosophila melanogaster*, adaptation to high altitude, toxic environments, high temperature environments, and pathogen exposure has already been described (16,17). These adaptations are related with hypoxia, xenobiotic, heavy-metal, oxidative, heat, and immune stress. However,

*To whom correspondence should be addressed. Tel: +34 93 2309637; Email: josefa.gonzalez@ibe.upf-csic.es
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

the contribution of TEs to the binding sites of stress-related TFs remains largely unexplored.

In this work, we used bioinformatic tools to predict the presence of binding motifs in TEs, and available ChIP-seq data to identify binding sites for several TFs involved in six stress responses (Table 1) (18–25). Enhancer and/or promoter features such as open chromatin regions, active histone marks and co-binding of stress-related proteins, and other genomic features such as location regarding nearby genes and function of nearby genes were also investigated. Besides genomic information, population-level information was also used to identify the subset of TEs more likely to contain functional TFBSs. Finally, *in vivo* enhancer assays were performed for a diverse set of TEs. Our results showed that TEs are likely to contribute a significant fraction of stress-related transcription factor binding sites in humans and in *D. melanogaster*.

MATERIALS AND METHODS

Trancription factor binding motifs (TFBMs) predictions based on PWMs

To determine the relative contribution of TEs to the six stress regulatory networks analyzed, we first quantified the presence of motifs for several stress-related transcription factors (TFs) in *D. melanogaster* and in humans, at a genome-wide level (Table 2). We then checked how many of those TFBMs were located in the TEs annotated in the reference genomes of the two species. To test whether there was enrichment of TFBMs in TEs, we used the binomial test and Bonferroni correction. We compared the number of motifs/sites predicted in TEs with the number expected if motifs/sites were distributed randomly in the genome, taking into account that 5.45% of the *D. melanogaster* genome and 45.5% of the human genome are TEs. Besides the *P*-value, we also considered the fold enrichment, as the number of motifs/sites predicted is very high.

We made motif predictions using TFBSTools (26) against version 6.04 of *D. melanogaster* genome, including all 5416 annotated TEs, and against version hg38 of the human genome. We downloaded the repetitive elements track from UCSC for hg38. After filtering out low complexity regions, simple repeats and other non-TE sequences such as snRNA and tRNA, we ended up with a dataset of 4 510 651 annotated TEs. These TEs belong to 1084 different families, and 36 superfamilies.

Each TF motif has a different length and different information content, thus we will obtain more predictions just by chance for shorter motifs, or motifs with several positions with low information content such as DEAF1 or caudal (Supplementary Figure S1). Thus, a single threshold score for all TFs will not suffice. We calculated an adjusted score threshold for each TF, which takes into account the background nucleotide frequencies of the genome and the relation between false positives and false negatives (Table 1). This threshold has a relation between the false-positive rate and the false-negative rate of 1000 ($f_{nr}/f_{pr} = 1000$). The threshold calculation was done using the 'motifs' library included in BioPython. Motif plots (PWM, TFFM) were done using the *ggseqlogo* R package (27).

In *D. melanogaster*, we also extracted the coordinates of genes in the areas surrounding TEs from the Flybase annotation (28). We obtained the gene structure (promoters, UTR's, exons, and introns) along with the parental relations between genes and their transcripts parsing the Flybase annotation with an in-house script (28).

Construction of TFFMs for *D. melanogaster* and predictions based on TFFMs

We built a TFFM for each of the four datasets with available ChIP-seq data in *D. melanogaster* (Supplementary Table S1) and downloaded the TFFMs built for stress-related TFs in humans (11). TFFMs were constructed using a seed motif that is trained with ChIP-seq peaks enriched for motifs of the desired TF (see ChIP-seq data processing section below). For each peak detected by *MACS2*, we extracted 500-bp, 250-bp region at each side of the summit, using an in-house python script and we used *RepeatMasker* to mask the repetitive regions. With those sequences, we run *meme-chip* (29) to obtain the enriched motifs. *Meme-chip* uses the central 100 bp of input sequences to look for motifs and the rest of the sequence as background. We also enabled the *-centrimo-local* option and used the *JASPAR_CORE_2016* as a target database for *CentriMo* and *TOMTOM* (30). In addition, we limited the number of sequences to pass to *MEME* suite to 2000 (*-nmeme*). Taking the best motif found by *MEME* suite, and the sequences from the ChIP-seq peaks we generated the TFFMs.

We used the different TFFMs to run predictions in each of the TEs annotated in *D. melanogaster* and human genomes and kept predicted TFBMs with a score better than 0.90. If two predictions were overlapping, we kept the one with the best score. We also predicted TFBMs using TFFMs in a set of background sequences matching the GC content of the TEs. We generated the background sequences using BEDtools (bedtools random -l 1000 -n 2000000) and the BiasAway script to adjust for GC content (31). The ratio TE / background was obtained dividing the number of TFBMs every 10 kb in both datasets. If we found the same number of TFBMs in background sequences and in TEs, the ratio is 1. A higher ratio means we found more TFBMs in TEs and a ratio lower than one means that we found more TFBMs in background sequences.

ChIP-seq data processing

For *D. melanogaster*, we processed the raw data instead of just using the TFBSs regions reported by the authors to ensure fair comparisons across datasets, and to overcome one of the main limitations of ChIP-seq traditional pipelines: the use of uniquely mapping reads that make it very difficult to detect binding regions in TEs. Multi-mapping read allocation allows the detection of binding regions in repetitive regions and improves detection of peaks in mappable regions (32). This approach is based on allocating multi-reads or reads that map to multiple location as fractional counts weighting every alignment.

We found high quality ChIP-seq datasets for four *D. melanogaster* TFs. We classify a ChIP-seq as high quality if (i) it has good quality reads, (ii) it has no major red flags

Table 1. Transcription factors analyzed in this study. Description of the stress-related transcription factors, including the identifier (ID) of the position weight matrix (PWM) or transcription factor flexible model (TFFM) used. The stresses analyzed were: HSE: Heat Shock element; ARE: Antioxidant response element; HRE: Hypoxia response element; IRE: Immunity response element; MRE: Metal response element; and XRE: Xenobiotic response element. In D. melanogaster, only TFFM IDs are provided for models built based on a D. melanogaster PWMs. *For HSF, HIF1, MTF-1, and XBP1 a vertebrate PWM was used.

	Drosophila melanogaster				Human			
Stress	Transcription Factors	PWM ID	Score threshold (PWM)	TFFM ID	Transcription factors	PWM ID	Score threshold (PWM)	TFFM ID
HSE/ARE/HRE/ IRE/MRE/XRE	HSF (18)	MA0486.2*	10.04	NA	HSF1 (18)	MA0486.2	10.09	TFFM0048.1
ARE/IRE	DL (19)	MA0022.1	8.48	TFFM0158	NFKB1 (20)	MA0105.4	10.35	
HRE	HIF1 (HIF1A, tango-HIF1B) (21)	MA0259.1*	9.43	NA	EGR1 (22)	MA0162.2	9.79	TFFM0020.1
					SP1 (20)	MA0079.3	10.48	TFFM0097.1
MRE	MTF-1 (23)	PB0044.1*	9.27	NA	–	–	–	–
IRE	CAD (19)	MA0216.2	10.12	TFFM0159	–	–	–	–
	DEAF1 (24)	MA0185.1	8.33	NA	–	–	–	–
	NUB (78)	MA0197.2	8.99	NA	–	–	–	–
	XBP1 (25)	MA0844.1*	9.69	NA	–	–	–	–
ARE/XRE	CNC (20)	MA0530.1	9.97	NA	–	–	–	–
ARE	–	–	–	–	NFE2L2 (20)	MA0150.2	9.56	TFFM0071.1
ARE/HRE	–	–	–	–	NRF1 (20)	MA0506.1	10.04	TFFM0082.1
	–	–	–	–	CREB1 (20)	MA0018.2	7.96	TFFM0012.1
HRE/XRE	–	–	–	–	AP1 (FOS) (20)	MA0476.1	10.46	TFFM0032.1

Description of the stress-related transcription factors, including the identifier (ID) of the position weight matrix (PWM) or transcription factor flexible model (TFFM) used. The stresses analyzed were: HSE: Heat Shock element; ARE: Antioxidant response element; HRE: Hypoxia response element; IRE: Immunity response element; MRE: Metal response element; and XRE: Xenobiotic response element. In *D. melanogaster*, only TFFM IDs are provided for models built based on a *D. melanogaster* PWMs. *For HSF, HIF1, MTF-1 and XBP1 a vertebrate PWM was used.

Table 2. Prediction of binding motifs (TFBMs) and binding sites (TFBSs) in *D. melanogaster* and humans

	TFBMs					TFBSs			
	PWMs			TFFMs		Chip-seq			
Transcription factors	Number (TEs/Genome)	%	*P*-value	TEs	Ratio TE / back-ground	Number (TEs/genome)	%	*P*-value	Merged TFBMs/ TFBSs
(A) *D. melanogaster*									
CNC	1832/ 34 558	5.3	1	–	–	–	–		1573
DEAF1	10 735/ 219 557	4.89	$5.72e^{-31}$	–	–	–	–		9042
MTF-1*	2223/ 29 964	7.42	$2.62e^{-45}$	–	–	–	–		1839
NUB	8666/ 181 721	4.77	$5.70e^{-38}$	–	–	–	–		7335
XBP1*	528/ 10 402	5.08	0.86	–	–	–	–		458
caudal	7068/ 123 046	5.74	$5.87e^{-05}$	1519	0.64	5907 / 35 630	16.58	$>1e^{-323}$	8567
dorsal	5427/ 116 125	4.67	$7.46e^{-32}$	4579	1.16	985 / 2883	34.17	$>1e^{-323}$	7555
HSF*	480/ 7354	6.52	$6.78e^{-4}$	734	1.86	1643 / 4493	36.57	$>1e^{-323}$	2191
tango (HIF1B)*	2754/ 62 228	4.43	$3.32e^{-30}$	1119	1.97	4349 / 15 238	28.54	$>1e^{-323}$	4382
Total	39 713/ 784 955	5.06	$2.2e^{-16}$	7995	–	12 884 / 58 244	22.33	$2.2e^{-16}$	42 942
(B) Humans									
CREB1	1 462 850/ 2 434 226	60.10	$3.95e^{-322}$	308 156	0.89	2317/ 15 908	14.56	$3.95e^{-323}$	1 627 554
EGR1	434 593/ 1 169 693	37.15	$2.77e^{-322}$	196 187	1.15	9972/ 36 982	26.96	$3.95e^{-323}$	509 377
FOS	324 072/ 747 204	43.37	$1.36e^{-309}$	630 618	0.89	45 748/ 92 352	49.54	$4.4e^{-130}$	370 407
HSF1	83 286/ 211 771	39.33	$1.18e^{-322}$	338 290	0.69	343/ 1432	23.95	$3.82e^{-63}$	325 915
NFE2L2	298 168/ 571 695	52.16	$1.98e^{-322}$	377 740	0.95	639/ 744	85.89	$1.42e^{-115}$	505 947
NFKB1	30 447/ 49 199	61.89	$3.95e^{-323}$	180 383	1.47	12 638/ 28 678	44.07	$4.62e^{-6}$	161 213
NRF1	26 327/ 127 953	20.58	$7.9e^{-323}$	28 857	0.88	259/ 4511	5.74	$3.95e^{-323}$	37 708
SP1	903 287/ 1 929 185	46.82	$1.53e^{-279}$	138 185	1.94	4463/ 15 104	29.55	$3.95e^{-323}$	847 478
Total	3 563 030/ 7 240 926	45.54	$2.2e^{-16}$	2 198 416	–	76 379 / 195 711	39.02	$2.2e^{-16}$	4 385 599

*TFs for which a vertebrate PWM was used.

Number of PWMs and ChIP-seq peaks (TFBSs) predicted in TEs/number predicted in the genome. For TFFMs, the number of predictions in TEs, and the ratio of predictions in TE versus background sequences is given. The merged TFBMs/TFBSs column shows the number of unique motifs/sites after considering the overlapping of coordinates between PWM, TFFM and ChIP-seq peaks predictions.

(FastQC inspection), (iii) it includes an 'input' in the experimental setup and (iv) the cross-correlation profile (SPP package) yielded a clear fragment length to continue with the analysis. ChIP-seq experiments for each TF and its corresponding control were downloaded from NCBI (Supplementary Table S2). We mapped the reads to version 6.04 of *D. melanogaster* genome using *Bowtie* (*-v* = 2, *m* = 99) (33). We used *CSEM* to assign multi-mapping reads (32). For each sample, we run a cross-correlation analysis using *SPP R* package for ChIP-seq experiment quality assessment and choose an appropriate fragment length for running *MACS2* peak calling software (34). The peak calling with *MACS2* was done using the *BAM* files processed with *CSEM* for the ChIP-seq experiment (*-t*) and the input as control. We enabled the *–no-model* and *-extsize* 200 parameters. For one experiment (caudal), we used the calculated fragment length (123) instead of 200, because it yielded a higher number of peaks with identifiable motifs. For each TF, we merged peaks retrieved from replicas or different developmental stages into one single file using BEDtools.

For humans, we downloaded eight TFs ChIP-seq datasets from the ENCODE project along with the TFFMs that were constructed based on them (Table 1). All the *narrowPeak* files coordinates were converted from hg19 to hg38. We calculated the overlap with our set of human TEs with an in-house python script and BEDTools.

TE family enrichment

To calculate the enrichment score we use, the following formula as in Sundaram *et al.* (2014) (5):

lor = log2((Number of TFBS in all TE copies / Total length of TE family (Kb)) / (Number of TFBS in the genome / genome size (Kb)))

In *D. melanogaster*, we removed nested TEs to avoid counting twice the same TFBS, ending up with a dataset of 3768 TEs. We focused on the 55 families with high copy number: at least 20 genomic copies. We also required a total length for the TE family of 1 Kb. For the ChIP-seq family enrichment, we also required a minimum of five ChIP-seq peaks in a family. We only consider peaks to belong to a TE if the peak overlaps at least 75% with the TE. In humans, we followed the same strategy used in *D. melanogaster,* but we required a family to have at least 50 copies. In total, we analyzed 1084 families. In both species, we used a threshold of 1.5 in lor score, which equals 2.83 more TFBSs in TEs than in the rest of the genome.

Overlap of TFBMs and TFBSs

We used BEDTools (35) and an in-house python script to merge the TFBMs/TFBSs coordinates from the three different sources, PWM, TFFM and ChIP-seq peaks, into single regions.

Open chromatin and CBP binding experimental data

We collected up to 12 ATAC-seq and FAIRE-seq experiments and one ChIP-seq CBP experiment (36,37). We converted the coordinates to the v6 *D. melanogaster* genome and checked which TEs overlap with known open-chromatin regions or contain a CBP peak. Overlapping with open chromatin regions and permutation tests were done using *regioneR* (38).

Epigenetic marks experimental data

The histone modification regions come from ChIP-seq data with very high coverage (39). The peaks were called by Jung *et al.* (2014) (39) using 100 million uniquely mapping reads for H3K4me3, H3K36me3 and the input. We converted the ChIP-seq peak coordinates to v6.04 of the *D. melanogaster* genome. These experiments were done in the *Oregon* strain. However, in this work we focused on the TEs annotated in the reference strain (y^1;cn^1, bw^1, sp^1). To obtain a list of TEs present in the Oregon strain, we run the presence module from *T-lex2* (40) with DNA-seq data from modENCODE for the three experiments done with the Oregon strain (SRP045325). We consider a TE to overlap an epigenetic mark if it shares nucleotides with the TE and also with the nucleotides located left or right of that TE (±1000 bp). Currently, we can only estimate with confidence the presence of 3894 out of the 5416 TEs annotated in the reference genome. We found that 2798 of these TEs were present in the *Oregon* strain; for those TEs, we analyzed the presence of epigenetic marks.

Evidence of selection

We used the list of TEs with evidence of selection reported in Rech *et al.* (2019) (41). In addition, we also considered TEs with evidence of positive selection based on *iHS, H12*, *nSL* and/or F_{ST} and located in low recombination regions that were not included in Rech *et al.* (2019) and were identified using exactly the same procedure (Supplementary Table S3) (41).

TFBS ratio

The TFBS ratio was calculated dividing the expected TFBS in a TE given its length, using the *glm* from Supplementary Figure S2, by the number of TFBS found in a TE. For example, a TFBS ratio of 1.2 means that we find 20% more TFBS than expected in a given TE.

In vivo enhancer assays

Fly husbandry. Flies were kept at 25°C, with 12-h light and dark cycles, and 60% humidity. DGRP (Drosophila Genetic Reference Panel) strains were used for generating the transgenic constructs (42).

Construct design. For three TEs, *FBti0019012*, *FBti0061428* and *FBti0019309*, we amplified only the TFBS containing part of the TE. For *FBti0019197* and *FBti0019985*, we amplified all the TE sequence. For *FBti0019978*, *FBti0019082*, *FBti0061578*, *FBti0019386* and *FBti0019453*, we amplified the intergenic region containing the TE and the intergenic region without the TE. In both cases, the intergenic region was the 500 bp region on both sides of the insertion. Finally, for *FBti0018880* we cloned three regions: only the TFBS containing part of the TE, the intergenic region with the TE and the intergenic region

without the TE. We checked the polymorphism in several DGRP lines with and without these insertions using the online database POPDROWSER, and chose the two most similar strains for the amplification (43). For the fixed TEs, we amplified the two sides of the insertion separately and joined them with a PCR step.

We also generated transgenic flies to be used as positive controls for immunity (44), heat-shock (45) and for oxidative stress (46) (Supplementary Table S4). As negative controls, we generated transgenic flies with the empty vectors. The primers used to amplify all the regions under study are reported in Supplementary Table S4.

Genomic DNA was extracted with the Puregene Cell and Tissue Kit (QIAGEN) and expand high fidelity Taq DNA polymerase was used for DNA amplification (Sigma).

Embryo microinjections. We purified the vector with the GeneEluteTM Plasmid Miniprep kit (Sigma) and prepared the injection mix at 6 μg vector concentration diluted with injection buffer (5 mM KCl, 0.1 mM sodium phosphate, pH 6.8) We microinjected the constructs with the *Eppendord Femtojet 4i* microinjector into a *D. melanogaster* strain with a stable docking site (Bloomington Stock number: 24749). Flies were crossed until homozygous flies for the insertion were obtained. The insertion of the construct was verified by PCR and sequencing. We generated three independent stocks that were used as biological replicates for the qPCR experiments.

Stress experiments

All experiments were performed with three biological replicates of thirty 5 to 8 day-old females.

Oxidative stress. Flies were placed on 1.5% agar and 5% sucrose with (stress) and without (non-stress) 10 mM Paraquat (Fisher Scientific) and kept at 25°C for 12 h. After that, guts were dissected, flash frozen in liquid nitrogen and stored at −80°C until RNA extraction.

Xenobiotic stress response. Scintillation vials (Labbox) were coated with a solution containing 200 μl of acetone and 50 μg dichlorodiphenyltrichloroethane (DDT) mixture. Each vial was rolled until the acetone evaporated. The vials were sealed with cotton balls soaked with 1 ml of 5% sucrose solution as a source of food and water (47). Flies were then kept at 21°C for 1,5 h because the efficiency of the DDT is higher at lower temperature (48). RNA was extracted from the whole fly.

Immune stress. Flies were infected with *Pseudomonas entomophila*, a gram-negative bacteria that infects *D. melanogaster* in the wild (49). Prior to the infection, flies were starved for 2 h. Then, they were placed in vials containing food and a piece of filter paper soaked with 1.25% of sucrose and bacterial pellet. The bacterial preparation was adjusted to a final OD600 = 50–100 (50). Flies were placed to the optimal infection conditions of the bacteria (29°C and 65% humidity) for 10–12 h. The non-infected flies were exposed to LB medium and 1.25% sucrose on the filter paper. After 10–12 h depending on the strain, guts were dissected, flash frozen in liquid nitrogen and stored at −80°C until RNA extraction.

Heat-shock stress. Flies were placed in empty vials in a water bath at 36°C followed by a 1 h recovery time at room temperature (25°C) (51). After the treatment, flies were flash frozen with liquid nitrogen. The non-treated flies were kept at room temperature for the same period of time and were flash frozen with liquid nitrogen. Samples were stored at −80°C until the RNA extraction. RNA was extracted from the whole fly.

RNA extraction and cDNA synthesis. RNA was extracted using GenElute™ Mammalian Total RNA Miniprep Kit (Sigma Aldrich). We treated the RNA with DNAse I (Thermo Fisher Scientific) after the extraction. cDNA was synthesized from 500 to 1000 ng of RNA using the NZY First-Strand cDNA Synthesis Kit (NZYTECH).

qRT-PCR. Expression was measured using SYBR Green master mix (BioRad) on an iQ5 Thermal cycler. Results were analyzed using the ddCT method and the two-tailed Student's *t*-test (52).

RESULTS

We focused on six evolutionary conserved stress responses that are relevant for *D. melanogaster* and human adaptation: heat-shock, oxidative, hypoxia, immune, xenobiotics and heavy-metal stress (Table 1). Through literature searches, we identified the transcription factors (TFs) involved in these six stress responses. In *Drosophila*, we analyzed nine TFs available in JASPAR (12). When the *D. melanogaste*r motif was not available, we used the vertebrate motif, as stress-related TFs are thought to be highly conserved (15). For example, there is functional data showing that human MTF-1 can restore to a large extent metal tolerance to flies lacking their own MTF-1 gene (53). Indeed, we found that genes that have previously been reported as heavy-metal responsive in *D. melanogaster* contained binding motifs for MTF-1 predicted with the human motif (Supplementary Table S2). In humans, we analyzed the eight TFs that were available in the ENCODE project, and PWMs were downloaded from HomerMotifDB (54) (Table 1).

Besides predicting transcription factor binding motifs (TFBMs), when available we used ChIP-seq data to identify transcription factor binding sites (TFBSs). All the TFBS predictions generated in this work are available at http://dx.doi.org/10.20350/digitalCSIC/8590

TEs contain stress-related TFBMs in *D. melanogaster* and in humans

We used two different approaches to identify TFBMs: Position Weight Matrices (PWMs) and Transcription Factor Flexible Models (TFFMs) (Table 1, see 'Materials and Methods' section). While PWMs consider that the nucleotides within the TFBMs are independent, TFFMs take into account nucleotide interdependencies and allow for gaps, which improve the identification of some TFBMs (11).

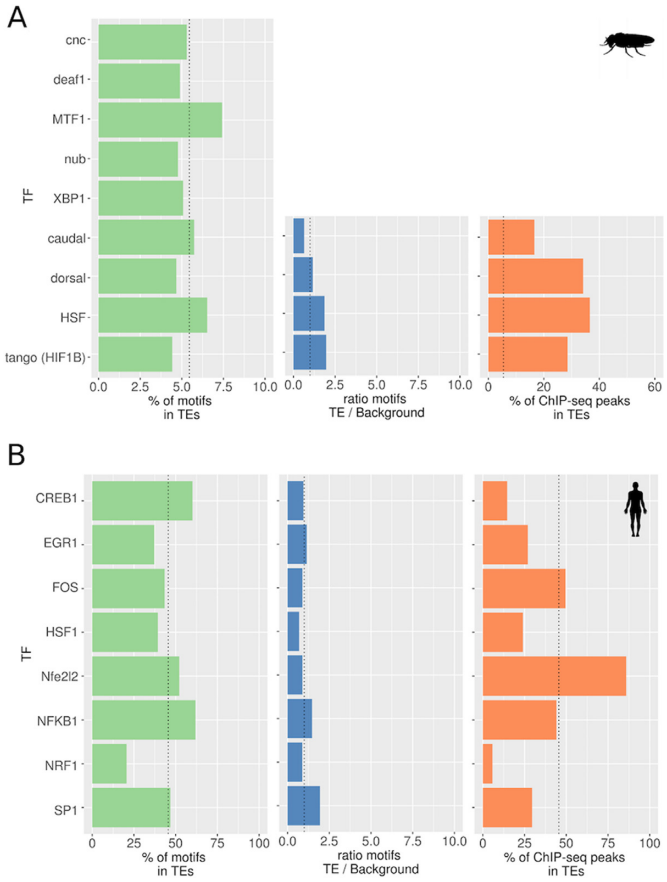


Figure 1. Percentage of transcription factor binding motifs (TFBMs) and ChIP-seq peaks (TFBSs) located in TEs in (**A**) *Drosophila melanogaster* and in (**B**) Humans. In green, motif predictions using position weight matrix (PWMs). The vertical dotted line depicts the expected percentage of motifs in TEs in *D. melanogaster* (5.45%) and human (45.54%). In blue, ratio of number of motifs predicted in TEs and number of motifs predicted in background sequences with the same properties than TEs. The expected ratio is 1 (vertical dotted line). In orange, percentage of ChIP-seq peaks located in TEs. The expected percentages of TFBSs falling in TEs are represented as vertical dotted lines as in the PWM predictions.

PWMs predictions. For all TFs, we predicted motifs using PWMs with the software TFBSTools (26), and we adjusted the score threshold for each TF (Table 1 and Supplementary Figure S1A; see 'Materials and Methods' section). Overall, the percentage of TFBMs in TEs appears to be small in *D. melanogaster* (4.43–7.42%, Table 2A). The 5416 TEs annotated in the reference genome represent 5.45% of the euchromatic fraction of the *D. melanogaster* genome, and on average 5.06% of TFBMs are located in TEs suggesting that overall TEs contained a similar number of TFBMs than expected if motifs were randomly distributed in the genome (Figure 1A and Table 2A). Only MTF-1 motifs were slightly enriched in TEs (1.4-fold enrichment, P-value = $2.62e^{-45}$, Table 2A). We tested whether removing the *INE-1* family from the analyses affected these results. While the majority of *D. melanogaster* TE families are active or have been recently active, and contain from a few to 100 copies, the *INE-1* family contains ∼2000 copies and has been inactive for the past ∼3- 4.6 million years (55–57). We found that overall, non-*INE-1* TEs were not enriched for TFBMs either (Supplementary Figure S3).

In humans, we also focused on the TEs annotated in the reference genome: 4 510 651 TEs. We found that the percentage of predicted TFBMs inside TEs was quite variable, from 21% to 62% (Figure 1B and Table 2B). Some TFs such as CREB1 or NFKB1 have slightly more TFBMs within TEs than expected considering that TEs constitute 45.5% of the human genome (1.3-fold enrichment, P-value = $3.95e^{-322}$ and 1.4-fold enrichment, $3.95e^{-323}$, respectively, Figure 1B and Table 2B).

TFFMs predictions. For *D. melanogaster*, we constructed TFFMs for the four TFs for which ChIP-seq data are available (Supplementary Table S1 and see 'Materials and Methods' section). The number of predicted binding motifs in *D. melanogaster* TEs for all TFs was smaller compared to the PWM predictions except for HSF (Table 2A). This can be partially explained because this motif has one gap at positions 9–10 (Supplementary Figure S1), and as mentioned before, TFFMs are able to handle small gaps. In addition, PWM predictions for HSF were made using the human PWM, while for the TFFM we used *D. melanogaster* ChIP-seq data. We also predicted motifs in a set of background sequences and estimated the ratio of predictions in TEs versus the background sequences (see 'Materials and Methods' section). The ratio was between 0.64 and 1.97 depending on the TF analyzed (see 'Materials and Methods' section; Table 2A and Figure 1A).

In humans, TFFMs were available for all eight TFs analyzed (Table 1 and Supplementary Figure S1B). The number of TFBMs predicted using TFFMs was quite variable (Table 2B). Similar to the results obtained with *D. melanogaster*, the ratio also varied depending on the TF analyzed (see 'Materials and Methods' section; Table 2B and Figure 1).

Overall, our results suggest that TEs contain a variable number of bindings sites for the TFs studied in *D. melanogaster* and in humans. Only for some TFs, we did find a slight enrichment of binding sites in TEs (Figure 1 and Table 2).

TEs are enriched for some stress-related TFBSs in *D. melanogaster* and humans

Not all the predicted TFBMs will be actively bound by their corresponding TFs (58,59). Thus, besides TFBMs we searched for TFBSs using available ChIP-seq datasets (Supplementary Table S1).

In *D. melanogaster*, there is ChIP-seq data available in non-stress conditions for four of the nine TFs studied (Table 2A). Based on these data, we retrieved a total of 58 244 TFBSs, of which 12 884 were located within TEs (Table 2A and Figure 1A). This is one order of magnitude less than the total number of predicted motifs with PWMs: 784 955. This suggests that most of the TFBMs predicted would not be bound by the TF, at least in the conditions and developmental stages in which the ChIP-seq experiments were performed. The number of TFBSs varies among TFs, which could be partly explained by the different number of experiments analyzed (Supplementary Table S1). While the num-

ber of TFBMs in TEs was overall not higher than expected if motifs were randomly distributed in the genome, when we looked at the ChIP-seq peaks, up to 37% of them occur in TEs (6.7-fold enrichment, *P*-value < 1 e-^{323}), with an average of 22% (4.1-fold enrichment, *P*-value = 2.2e^{-16}, Table 2A and Figure 1A).

In humans, there are ChIP-seq data available for all eight TFs studied (see 'Materials and Methods' section). Overall, the proportion of TFBSs occurring within TEs is smaller than expected for all TFs, except for NFE2L2 (1.9-fold enrichment, *P*-value = 1.42e^{-115}, Table 2B and Figure 1B). For HSF1, we also analyzed a ChIP-seq dataset obtained in stress conditions (60). In non-stress conditions, 22.94% (120 out of 523) of the peaks were located inside TEs, while in stress conditions 32.06% (680 out of 2121) of the peaks were inside TEs, suggesting that we might be underestimating the number of peaks in TEs by analyzing non-stress conditions. While more TFBS were identified in stress conditions compared to non-stress conditions, 71% (369 out of 523) of the peaks found in non-stress conditions are present in stress conditions, suggesting that most of the peaks in non-stress conditions were also present in stress conditions.

Overall, we found that TEs are enriched for caudal, dorsal, HSF and tango binding sites in *D. melanogaster* and for NFE2L2 binding sites in humans (Figure 1 and Table 2). Our results also suggest that we might be underestimating the number of binding sites in TEs since we analyzed ChIP-seq experiments performed in non-stress conditions.

TE families, superfamilies and classes are enriched for different TFBMs/TFBSs in *D. melanogaster* and humans

It has been described that particular TE families and TE classes are enriched for TFBSs (5,8). Thus, we measured the enrichment of TFBMs/TFBSs in TEs at the family, superfamily and class levels. In *D. melanogaster*, we found 14 families enriched both for TFBMs and TFBSs (Supplementary Table S5A and B). If we take into account the copy number of the families enriched for a certain TF, the three largest TE families (*1360, Cr1a* and *roo*) were enriched for tango (HIF1B) TFBSs, suggesting that these families could significantly contribute to the spreading of hypoxia response elements (HRE) in the *D. melanogaster* genome (Figure 2A). At the superfamily level, only TEs that belong to the *P* and *BEL* superfamilies were enriched both for TFBMs and TFBSs (Supplementary Table S5C and D). Finally, at the class level, LTRs, nonLTRs, and DNA transposons were enriched for TFBSs for at least one TF (Supplementary Table S5E). Note that LTRs are known to be enriched for TFBSs in human and mouse (5,8).

In humans, 214 families were enriched both for TFBMs and TFBSs (Supplementary Table S6A and B). The five families with the highest copy numbers were enriched for FOS, NFE2L2 and/or NFKB1 binding sites suggesting that these families could significantly contribute to the spreading of these three response elements (Figure 2B). At the superfamily level, six superfamilies were enriched both for TFBMs and TFBSs (Supplementary Table S6C and D). Finally, SINE were enriched both for TFBMs and TFBSs, while DNA transposons, LTRs, and LINEs were enriched for TFBSs (Supplementary Table S6E and F).

Overall, both in *D. melanogaster* and in humans, we found enrichment for different TFBMs/TFBSs at the family (Figure 2), superfamily and class levels (Supplementary Tables S5 and S6) suggesting that they could significantly contribute to the TFBMs/TFBSs repertoire in *D. melanogaster* and in humans.

The overlap between TFBMs and TFBSs predictions varies among stress-related TFs in *D. melanogaster* and humans

To identify the unique TFBMs/TFBSs in TEs, we checked the overlap among the predictions of the three methodologies used. The overlap between PWMs and TFFMs was in general low for all TFs in both species (Figure 3 and Supplementary Figure S4). If we consider the ChIP-seq peaks as true binding events (not necessarily functional), we observed that neither PWM nor TFFM predictions alone are able to predict all binding sites for a given TF (Figure 3 and Supplementary Figure S4). As mentioned above, only for some TFs, such as HSF in *D. melanogaster*, TFFMs outperformed PWMs at predicting motifs (Figure 3A). In humans, only FOS showed a high overlap between motif predictions and ChIP-seq peaks, while for other TFs the overlap was quite small (Figure 3B and Supplementary Figure S4B).

The fraction of ChIP-seq peaks for which we could not predict a motif with either PWMs or TFFMs might be explained by indirect binding through another TF, undiscovered minor motifs, or unspecific binding (5,61). Similarly, the fraction of motifs predicted by PWMs or TFFMs that did not overlap with a ChIP-seq peak could be explained because ChIP-seq data were obtained in non-stress conditions, and for a few developmental stages. Thus, because all three methods have limitations, to obtain the unique number of TFBMs/TFBSs identified, we merged the predictions for those TFs where we had multiple sources of motif and/or binding predictions (Table 2).

For the rest of this work, we focused on *D. melanogaster* TFBMs/TFBSs predictions as our ultimate goal was to test whether a subset of TFBMs/TFBSs were functional by using *in vivo* reporter gene assays.

TEs containing TFBMs/TFBSs are not globally enriched for enhancer and/or promoter distinctive features

To further investigate the potential role of TEs with predicted TFBMs/TFBSs as enhancers or promoters, we checked whether these TEs were enriched for several distinctive features associated with these regulatory regions: location in open chromatin, co-binding of CREB-binding protein (CBP), and presence of active histone marks. We also checked the genomic location of the identified TEs. We considered all the TEs with at least one predicted TFBMs/TFBSs (3593 TEs) and the TEs with three or more TFBMs/TFBSs (2183 TEs) as it has been shown that functional regulatory regions tend to be bound by multiple related TFs, usually three or more (58,62,63). Indeed, we found that the number of unique predicted TFBMs/TFBSs correlates very well with TE length: most TEs have at least one motif prediction if they have a minimum length of 220 bp (Supplementary Figure S2).

Active transcription has been linked to changes in nucleosome organization in regulatory elements due to TF

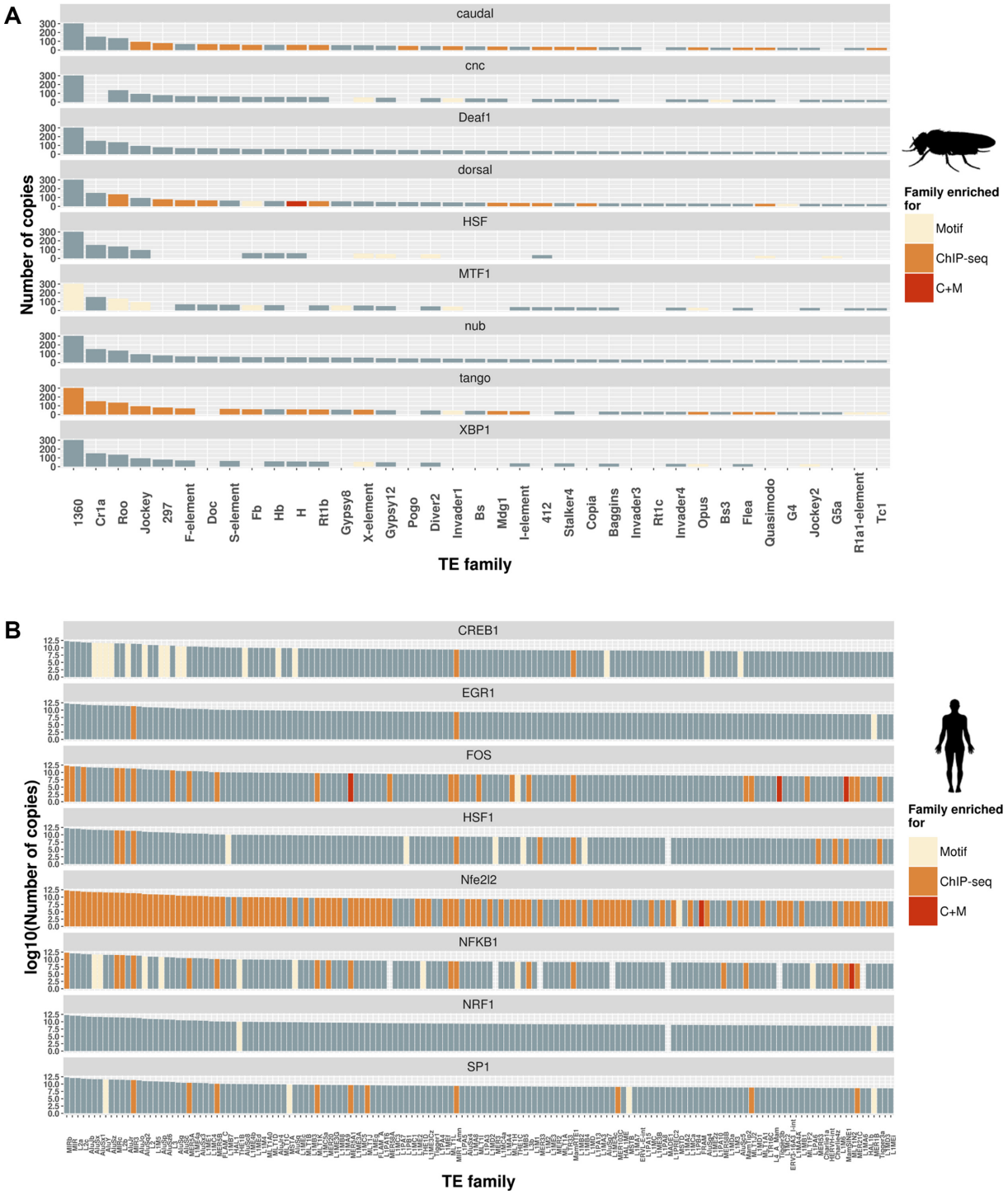


Figure 2. Several TE families are enriched for stress-related transcription factor motifs and binding sites. (**A**) The number of genomic copies for *D. melanogaster* TE families with at least 25 copies is represented. Families are painted depending on whether they are enriched for motifs, ChIP-seq peaks, or both (C+M). Absent columns for a particular TF indicate that the score could not be calculated due to lack of sufficient motifs or peaks. (**B**) Equivalent figure for humans. The number of copies is given in log scale due to the high number of copies of some families. Only families with more than 5000 copies are plotted.

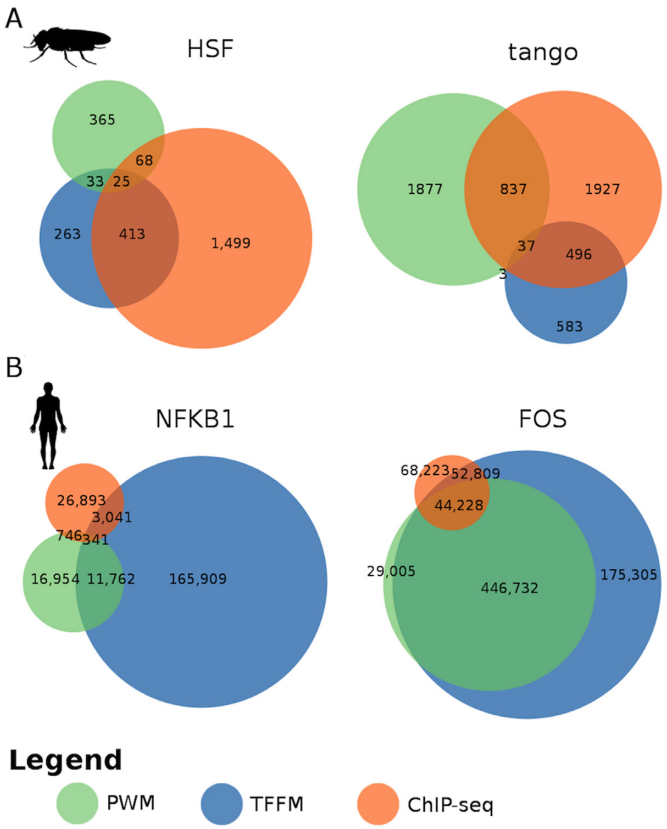


Figure 3. Overlap of TFBMs and TFBS predictions. Venn diagrams showing the overlap in the predictions across methods (PWM, TFMM, and ChIP-seq) within TEs for representative transcription factors in panel (**A**) *D. melanogaster* and panel (**B**) humans. A motif/peak is considered as shared if there is overlap in their coordinates. Note that a ChIP-seq peak can overlap with several motifs.

binding (36). Thus, identifying motifs located within open chromatin regions should be an effective strategy to identify functional binding sites (64). By combining, all ATAC-seq and FAIRE-seq experiments performed in Davie *et al.* (2015) and Koenecke *et al.* (2016), we obtained 36 507 distinct open chromatin regions (36,37). Only 637 open chromatin regions were detected inside TEs, corresponding to 489 unique TEs. This overlap is much smaller than expected by chance (permutation test, *P*-value = 0.0002, Supplementary Figure S5), suggesting that TEs in *D. melanogaster* do not tend to be located in open chromatin regions, as has been previously reported in humans (65). Overall, TEs containing one or more TFBMs/TFBSs were not preferentially located in open chromatin regions if we consider each one of the TFs independently (Figure 4 and Supplementary Table S7) or altogether (Supplementary Table S8). The same result was obtained for TEs containing three or more TFBMs/TFBSs (Supplementary Table S8). The only exceptions were TEs containing TFBMs/TFBSs for XBP1, which were slightly enriched in open chromatin regions (14.49% versus 9.04%, *P*-value = 0.04, Supplementary Table S7).

We also looked for evidence of co-binding of CBP, which has a role as an activator of several TFs, some of them related to different stress responses, such as CNC (66), HSF (67), HIF1A (68), MTF-1 (69), or immune response (70). We identified 815 TEs that contain a CBP-binding region. We did not find significantly more CBP interactions in TEs that have one or more TFBMs/TFBSs (Figure 4 and Supplementary Table S7), while we see a depletion of CBP peaks in TEs that contain one or more ChIP-seq peaks for a stress TF (Figure 4 and Supplementary Table S7). Overall, TEs with one or more TFBMs/TFBSs and TEs with three or more TFBMs/TFBSs are not enriched for CBP binding sites (Supplementary Table S8).

Binding of TFs to TEs has been found to be strongly associated with the epigenetic status of a TE (5,71). Indeed, TEs have been postulated as tissue-specific gene regulators through epigenetic modifications (72). We thus looked for the presence of two key histone modifications in TEs: H3K4me3 and H3K36me3, associated with promoters and transcriptional elongation, respectively (39, see 'Materials and Methods' section). We found that 286 TEs contained the H3K4me3 histone mark, and 584 TEs contained the H3K36me3 histone mark (Supplementary Table S9). We found that TEs containing TFBSs for HSF and dorsal were enriched for H3K4me3 and/or H3K36me3 (*P*-value = $1.11e^{-16}$ and $1.57e^{-18}$, respectively, Figure 4 and Supplementary Table S7). Note that histone marks are highly variable across cell types and strains; thus, the fraction of TEs with active epigenetic marks is an underestimation, and many more might be identified in other cell types or conditions.

Finally, we also tested whether TEs with TFBMs/TFBSs were located in proximal regulatory regions. We defined the proximal regulatory region of a gene as the 1000 bp upstream the TSS, the 5′UTR, and the first intron. Only TEs containing TFBSs for dorsal were slightly enriched in regulatory regions (*P*-value = $3.19e^{-4}$, Figure 4 and Supplementary Table S7).

Overall, TEs containing one or more, or three or more, TFBMs/TFBSs were not globally enriched for enhancer or promoter features (Figure 4 and Supplementary Table S8). Only TEs containing TFBMs/TFBSs for XBP1, HSF and dorsal were enriched in open chromatin regions, active histone marks and/or regulatory regions (Figure 4 and Supplementary Table S7).

TEs with three or more TFBMs/TFBSs were present at higher population frequencies

We expect TEs with functional TFBMs/TFBSs to be present at high frequencies or fixed in populations due to an increase in fitness of the individuals that carry them. We found that the proportion of TEs with one or more TFBMs/TFBSs present at high frequencies (≥10% to <95%) in populations is significantly higher than the proportion of all TEs present at high frequencies in the genome (16.2% versus 11%, *P*-value < $2.2e^{-16}$, Table 3). This percentage increased when we only considered TEs with three or more TFBMs/TFBSs (25.9%), and it was even higher in the subset of TEs that have ChIP-seq evidence for three or more TFBSs (42.1%, Table 3).

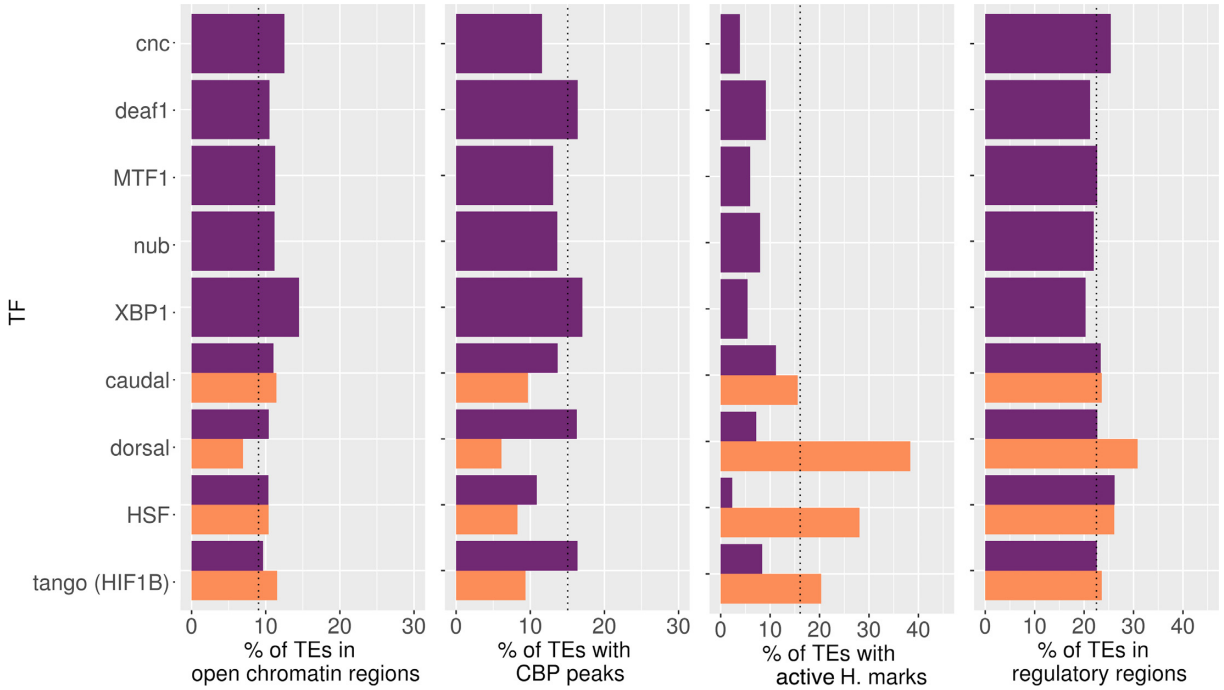


Figure 4. Enhancer/promoter genomic characteristics in TEs with predicted TFBMs/TFBSs in *D. melanogaster*. Percentage of TEs with at least one TFBMs/TFBSs for each one of the nine transcription factors studied overlapping with (**A**) open chromatin regions, (**B**) containing a CBP peak, (**C**) enriched for active histone marks or (**D**) located in a regulatory region. In purple, merged dataset of TFBMs/TFBSs and in orange dataset with evidence from ChIP-seq. The vertical dotted line showed the expected percentage for each feature.

Table 3. Number of TEs containing one or more, or three of more TFBMs/TFBSs present at high population frequencies or fixed

		High freq TEs			Fixed TEs (non-*INE-1*)			Fixed TEs (*INE-1*)		
Dataset	TE #	TE #	%	*P*-value	TE #	%	*P*-value	TE #	%	*P*-value
All TEs with frequency estimations	3894	424	11	NA	855	22	NA	2234	58	NA
≥1 TFBSs	2438	396	16.2	$<2.2e^{-16}$	621	25.5	$<2.2e^{-16}$	1086	44.5	$<2.2e^{-16}$
≥3 TFBSs	1314	340	25.9	$<2.2e^{-16}$	386	29.4	$<2.2e^{-16}$	275	20.9	$<2.2e^{-16}$
≥3 Chip-seq TFBSs	311	131	42.1	$<2.2e^{-16}$	12	3.9	$<2.2e^{-16}$	0	0	$<2.2e^{-16}$

For the fixed non-INE-1 TEs, we observed a small increase in the proportion of TEs with one or more TFBS present at high frequencies in populations (22% versus 25.5%, *P*-value $< 2.2e^{-16}$, Table 3). However, we found a significant decrease when we only considered ChIP-seq peaks (3.9%, *P*-value $< 2.2e^{-16}$, Table 3). This can be explained by the shorter length of fixed TEs that makes it more difficult to detect three relatively large non-overlapping ChIP-seq peaks in this dataset. Finally, the proportion of TEs from the *INE-1* family decreased in the datasets of TEs with TFBMs/TFBSs consistent with these TEs having reached fixation in populations through neutral processes rather than positive selection (Table 3).

Overall, these results suggest that the subset of TEs containing three or more TFBMs/TFBSs, and especially those TEs with evidence coming from ChIP-seq experiments, could be enriched for functional TFBMs/TFBSs, as the proportion of these TEs present at high frequencies in populations is higher compared to all TEs in the genome. On the other hand, INE-1 elements were depleted for TEs with three or more TFBMs/TFBSs.

TEs containing three or more TFBMs/TFBSs and present at high population frequencies were enriched nearby stress-associated genes

We tested whether TEs containing three or more TFBMs/TFBSs were enriched nearby stress-associated genes. Briefly, we considered as stress-associated genes those identified in GWAS, QTL, transcriptomics and/or protein–protein interaction analysis as described in Rech *et al.* (2019) (41). We first confirmed that the promoters of genes that have been reported as stress-associated are enriched for the corresponding stress-associated TFBMs/TFBSs compared with the promoters of nonstress-associated genes (Supplementary Figure S6 and Supplementary Table S10).

We observed that high frequency TEs were more often located nearby stress-response genes (28.81% versus 18.93, *P*-value $= 1e^{-6}$, Supplementary Table S11A). This association was also significant for TEs present at high frequencies and containing three or more TFBMs/TFBSs (30.31% versus 18.83, *P*-value $= 9.16e^{-7}$, Supplementary Table S11B). Thus, TEs containing three or more TFBMs/TFBSs and

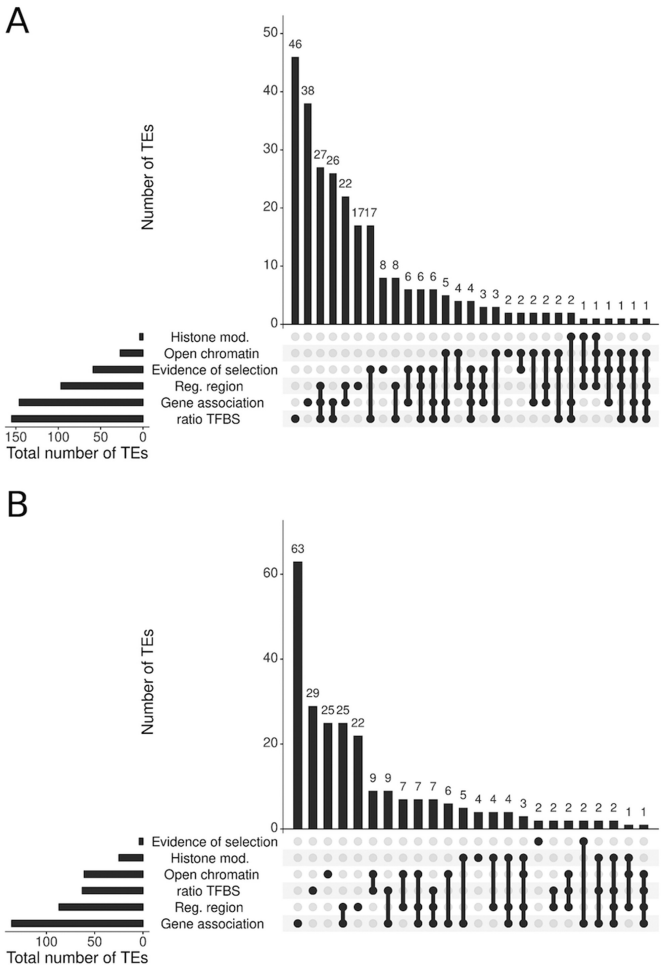


Figure 5. Characteristics of TEs containing three or more TFBMs/TFBS present at (**A**) high frequency or (**B**) fixed (non-INE-1). Histone mod: TE bears H3K4me3 or H3K36me3 marks associated with active chromatin. Open chromatin: TE is located in an open chromatin region. Evidence of selection: TEs with evidence of selection (41). Reg. region: TE is located in the proximal regulatory region of a gene (promoter, 5′UTR or first intron). Gene association: TEs located nearby stress-associated genes. Ratio TFBS: TE contains 20% more TFBS than expected given their length.

present at high population frequencies are enriched nearby stress-associated genes.

TEs containing TFBMs/TFBSs affect the expression of genes that were already part of a stress regulatory network

We summarized all the information suggesting that TEs could be adding functional TFBMs/TFBSs (Figure 5 and Supplementary Table S12). Based on our results, we focused on those TEs containing at least three TFBMs/TFBSs, and present at high population frequencies or fixed (non-*INE-1*). We considered TEs that were (i) enriched for active histone marks, (ii) located in open chromatin regions, (iii) located in regulatory regions, (iv) located nearby stress-associated genes and/or (v) have more TFBSs than expected given their total length (ratio TFBS, see 'Materials and Methods' section). In addition, we also considered whether there is evidence suggesting that the region flanking the TE insertion is under positive selection (41, see 'Materials and Methods' section). We found that 73 TEs containing at least three TFBMs/TFBSs, and present at high population frequencies or fixed (non-*INE-1*) showed signatures of selection in their flanking regions (Supplementary Tables S3 and S12).

We found that 82.5% (264 out of 320) of the TEs with three or more TFBMs/TFBSs and present at high population frequencies have at least one additional line of evidence suggesting that they might be functional. This percentage is significantly smaller for fixed *non-INE-1* TEs suggesting that fixed non-INE TEs are less likely to contain functional TFBMs/TFBSs (63%, 243 out of 386, chi-square P-value < 0.0001).

We chose 11 TEs with at least one additional line of evidence to perform *in vivo* enhancer reporter assays (Table 4). We also included in Table 4 three TEs, *tdn8* (*transpac*), *FBti0020057* (*BS*) and *FBti0018868* (*297*), which were previously tested in our laboratory (44). All these TEs contained three or more TFBMs/TFBSs, except *FBti0019453* (*jockey*) that contained two and *FBti0020057* (*BS*) that contained one (Table 4). The majority of them were present at high population frequencies, except *FBti0019197* (*Tc1*) and *FBti0061578* (*baggins*) that were fixed. Based on the TFBMs/TFBSs added by the TE and on the functional information available for the nearest gene, we tested the role of these TEs in four stress responses: heat-shock, oxidative, xenobiotic and immune (Table 4). Seven of the 14 TEs were tested in two stress conditions.

Six of the 14 tested TEs affected the expression of the reporter gene under stress conditions: three TEs up-regulated and three down-regulated the reporter gene (Table 4 and Figure 6; Supplementary Table S13). Four of the six validated TEs added TFBSs that were already present in the promoter region of the nearby gene (Table 4). For five of the six validated TEs, the intergenic region containing the TE was compared with the intergenic region cloned from a strain without the TE insertion, while in the other case only the TE was cloned and significance was determined by comparing with the empty vector (see 'Materials and Methods' section). On the other hand, only the TE or only the TE fragment containing the TFBMs/TFBSs was cloned for five of the eight TEs that were not validated (Supplementary Table S13). These results suggest that in most cases the TE sequence is not enough to drive the expression of nearby genes but rather modulates their level of expression.

Finally, for three insertions, we cloned the TE in sense and in antisense orientation. We did not find differences between the two constructs: *FBti0019985* (*roo*) affected the expression of the reporter gene regardless of the orientation while *FBti0019012* (*pogo*) and *FBti0019309* (*1360*) did not affect the expression in any of the two orientations (Table 4 and Supplementary Table S13).

DISCUSSION

In this work, we showed that transposable elements (TEs) contribute to stress-related transcription factor binding motifs/sites (TFBMs/TFBSs) in *D. melanogaster* and in humans. This contribution is transcription factor (TF) specific, ranging from 17% to 37% in *D. melanogaster* and from

Table 4. Results summary for the *in vivo* enhancer assays performed

	TFBS/TFBM					q-PCR result (*t*-test *P*-value)		
TE Family Class	TE	Reg. region	Additional evidence	Experimental design	Stress tested	Control	Treated	Reference
FBti0019386 *Invader4 LTR*	DEAF1: 1 **CAD: 1** **tango:1**	CAD: 2 NUB: 1 DEAF1: 1 dorsal: 1	Regulatory region CBP TFBS ratio Histone marks Selection evidence	Intergenic	IRE	No	Up-regulation (8.91E-05)	This work
FBti0019082 *Rt1b non-LTR*	CAD: 1 DEAF1: 3 MTF-1: 3 CNC: 3 dorsal: 2	NUB: 2 XBP1: 1	Regulatory region Open chromatin CBP Histone marks Selection evidence	Intergenic	IRE	No	Down-regulation (0.033)	This work
FBti0019985 *roo* LTR	DEAF1: 1 NUB: 1 MTF-1: 1 dorsal: 1	DEAF1: 1	Regulatory region Selection evidence TFBS ratio	TE/antisense	IRE	No	Up-regulation (0.0126)	This work
				TE/sense		No	Up-regulation	(44)
tdn8 *transpac* LTR	NUB: 2 DEAF1: 4	NA	Regulatory region	Intergenic	IRE	No	Up-regulation (0.046)	(44)
FBti0020057 *BS* non-LTR	NUB: 1	NUB: 3 CAD: 1 DEAF1:1	Regulatory region Open chromatin Gene: *Acbp6* Selection evidence	Intergenic	IRE	Down-regulation (0.0193)	Down-regulation (0.0161)	(44)
FBti0019453 *jockey* non-LTR	NUB: 1 CAD: 1	NUB: 2 DEAF1:1	Regulatory region Open chromatin Selection evidence	Intergenic	XRE	No	No	This work
					IRE	No	Down-regulation (0.007)	This work
FBti0019012 *Pogo* TIR	NUB: 4 **HSF:1 tango: 2 CAD: 2** dorsal: 1	NUB: 3 DEAF1: 1 XBP1: 1	Regulatory region Gene: *mir-31a* TFBS ratio	TFBS/sense	IRE	No	No	This work
					HSE	No	No	This work
				TFBS/antisense	HSE	No	No	This work
FBti0019309 *1360* TIR	DEAF1: 2 NUB: 3 MTF-1: 2 **tango: 1**	NUB: 3 CAD: 1 DEAF1:2 dorsal:1	Regulatory region TFBS ratio	TFBS/sense	IRE	No	No	This work
				TFBS/antisense	HSE	No	No	This work
FBti0018880 *Bari1* TIR	CNC: 1 DEAF1: 3 NUB: 2 MTF-1: 1 **HSF: 2** **tango: 1** **CAD: 2** **dorsal:** 3	MTF-1: 1 DEAF1: 2	Regulatory region TFBS ratio Gene: *Jheh2* Selection evidence	TFBS	ARE	No	No	This work
				Intergenic	ARE	No	No	This work
					IRE	No	No	This work
FBti0061428 *Hobo* TIR	dorsal: 3 DEAF1: 3 **tango: 2** **CAD:** 2 CNC: 1	NA	Open chromatin Histone marks Gene: *CG31809* TFBS ratio	TFBS	IRE	No	No	This work
					HSE	No	No	This work
FBti0019197* *Tc1* TIR	tango: 1 dorsal: 1 **CAD: 1**	MTF-1: 1 NUB: 1	Regulatory region Histone marks	TE	IRE	No	No	This work
					ARE	No	No	This work
FBti0019978 *1360* TIR	MTF-1: 2 **tango: 1** CAD: 1 **HSF:1** DEAF1: 1	MTF-1: 1 CAD: 1 DEAF1: 1	Regulatory region Open chromatin Histone marks	Intergenic	XRE	No	No	This work
FBti0061578* *baggins* non-LTR	DEAF1: 2 tango: 1	CAD: 1 DEAF1: 1 dorsal: 1	Regulatory region Histone Marks TFBS ratio Gene:*CG2217*	Intergenic	ARE	No	No	This work

Table 4. Continued

TE Family Class	TFBS/TFBM TE	TFBS/TFBM Reg. region	Additional evidence	Experimental design	Stress tested	q-PCR result (*t*-test *P*-value) Control	Treated	Reference
					IRE	No	No	This work
FBti0018868 *297* LTR	DEAF1: 1 NUB: 1 CAD: 1	NA	Regulatory region TFBS ratio Gene: *TM4SF*	TE	IRE	No	No	(44)

*Fixed TEs; In bold, TFs for which the evidence for the presence of TFBSs in that particular TE comes from ChIP-seq data. Experimental design indicates the region that was cloned in front of the reporter gene. We also included the data for three reporter assays performed previously in the lab (44).

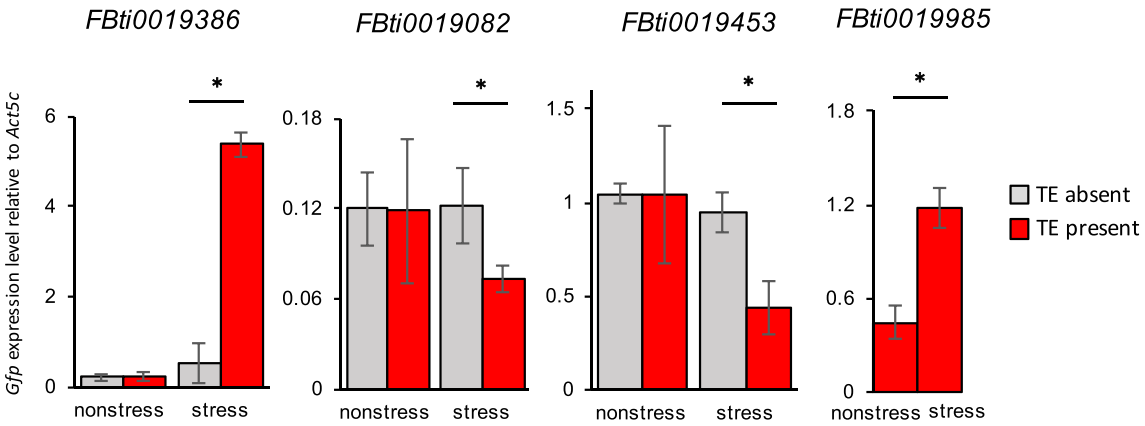


Figure 6. Four TE insertions analyzed in this work affect the expression of a reporter gene. qRT-PCR experiments comparing the expression of the *gfp* reporter gene in transgenic flies containing the genomic region under study without the TE insertion (gray) and with the TE insertions (red), in stress and non-stress conditions. The error bars represent the standard deviation of three biological replicates. Significant results are indicated with *.

6% to 86% in humans (Figure 1 and Table 2). This is consistent with previous reports in humans in which the contribution of TEs was also highly TF specific (5,6). Some of the families with the highest copy number, such as *1360* and *Cr1a* in *D. melanogaster* and *MIRb* and *L2c* in humans, were enriched for TFBSs suggesting that these families could be significantly contributing to the spreading of particular stress response elements (Figure 2). Indeed, *MIRs* have previously been shown to contribute to functional enhancers genome-wide in mammals (73).

We showed that while *D. melanogaster* TEs are not enriched in open chromatin regions, TEs containing binding sites for HSF and dorsal were enriched for active histone marks (Figure 4). Histone marks are often used to identify active regulatory regions at a genome-wide level (72,74,75). Interestingly, SINEs involved in neural gene activation were enriched for active histone marks in control conditions suggesting that these insertions were epigenetically primed prior to neural activation (75). Thus, histone mark enrichment in control conditions, as we have studied in this work, could be informative about the enhancer role of TEs in specific conditions.

We also found that TEs containing three or more TFBSs had a higher proportion of TEs present at high population frequencies (Table 3), and were enriched in the promoter regions of stress-related genes, suggesting that this subset of TEs is likely to be enriched for functional TFBSs. Our results are consistent with previous studies showing that TEs containing three or more TFBSs are more likely to be functional (58,62,63). Indeed, based on the integration of ChIP-seq data for enhancer histone marks and TFs, *ERVs* have been shown to disproportionally overlap with genomic regions showing combinatorial binding of several TFs (76).

While we could not confirm the functional role of the two TEs that were fixed in all the populations analyzed, six of the 12 TEs present at high population frequencies were validated (Table 4). Five of these six TEs affected the expression of the nearby gene only under stress conditions suggesting that their effect is stress-response specific (Table 4). Most of these TEs, four out of six, add TFBSs that were already present in the promoter region of the nearby gene. This result suggests that rather than recruiting new genes to stress-regulatory networks, these TEs affect the level of expression of genes that were already part of the cellular stress response (Table 4).

Interestingly, all six validated TEs were either LTR or LINE elements, while most of the non-validated TEs, six out of eight, were TIR elements (Table 4). Each validated TE belong to a different family: *Invader4*, *Rt1b*, *roo*, *transpac*, *BS* and *jockey*. These results suggest a different dynamics in *D. melanogaster* compared with humans or mouse in which often is a particular TE family or subfamily that contributes most of the TFBSs for a given TF (72–73,75,77). It is also noteworthy that five of the six TEs that were functionally validated showed signatures of selection in their flanking regions, suggesting that the changes in expres-

sion they induced could have an adaptive effect (Supplementary Table S3). Three of these TEs are associated with down-regulation of the reporter gene (Table 4). These three insertions contain TFBMs related with immune-response TFs, with some of them involved in the negative regulation of genes in response to an immune challenge (78). Finally, we cannot discard that mechanisms other than adding TFBMs/TFBSs could affect the changes in expression of the reporter gene described in this work as TEs have been shown to affect gene expression through a variety of mechanisms (1,2).

While it is possible that the non-validated TEs are false positives, that is, TEs containing non-functional TFBSs, it could also be that these TEs are false negatives. First, in order for some TFs to be able to bind the DNA, they could require genomic context that is missing in the genomic region where the transgene is inserted. For example, it has been reported that binding of HSF to the corresponding motif sequences required the presence of active chromatin marks (79). Moreover, instead of affecting expression of nearby genes, it has been argued that TEs containing TFBSs could provide a buffer of extra binding sites to trap TFs or could serve as a landing pad to allow TFs to scan the DNA (5). Thus, although we cannot discard that the non-validated TEs are indeed non-functional, there are other possible explanations for the lack of effect of these TEs on the expression of the reporter gene. If we extrapolate our validation rate to the subset of TEs with similar characteristics, we can speculate that at least 132 reference TE insertions in the *D. melanogaster* genome could be adding functional TFBSs to their nearby genes. This is likely an underestimation as we only analyzed binding peaks in non-stress conditions. Thus, our results suggest that TEs are likely to be important contributors to the regulation of stress-response genes in the *D. melanogaster* genome. Experimental data on binding sites and chromatin features, obtained both under control and stress conditions, should help further quantify their contribution.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Gabriel E. Rech for the analysis of the signatures of selection for reference TEs located in low recombination regions, and for the analysis of the population frequencies of the TEs. We thank the Drosophila Genomics Resource center, supported by NIH grant 2P40OD010949, for sharing flies.

FUNDING

Ministerio de Economia y Competitividad [BFU2014-57779-P]; Ministerio de Ciencia, Innovación y Universidades/AEI [BFU2017-82937-P]; European Commission [H2020-ERC-2014-CoG-647900]. Funding for open access charge: European Commission [H2020-ERC-2014-CoG-647900].
Conflict of interest statement. None declared.

REFERENCES

1. Elbarbary,R.A., Lucas,B.A. and Maquat,L.E. (2016) Retrotransposons as regulators of gene expression. *Science*, **351**, aac7247.
2. Chuong,E.B., Elde,N.C. and Feschotte,C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.*, **18**, 71–86.
3. Casacuberta,E. and González,J. (2013) The impact of transposable elements in environmental adaptation. *Mol. Ecol.*, **22**, 1503–1517.
4. Batut,P., Dobin,A., Plessy,C., Carninci,P. and Gingeras,T.R. (2013) High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.*, **23**, 169–180.
5. Sundaram,V., Cheng,Y., Ma,Z., Li,D., Xing,X., Edge,P., Snyder,M.P. and Wang,T. (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.*, **24**, 1963–1976.
6. Kunarso,G., Chia,N.Y., Jeyakani,J., Hwang,C., Lu,X., Chan,Y.S., Ng,H.H. and Bourque,G. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.*, **42**, 631–634.
7. Lynch,V.J., Leclerc,R.D., May,G. and Wagner,G.P. (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.*, **43**, 1154–1159.
8. Chuong,E.B., Elde,N.C. and Feschotte,C. (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, **351**, 1083–1087.
9. Dunn-Fletcher,C.E., Muglia,L.M., Pavlicev,M., Wolf,G., Sun,M.A., Hu,Y.C., Huffman,E., Tumukuntala,S., Thiele,K., Mukherjee,A. *et al.* Anthropoid primate-specific retroviral element THE1B controls expression (2018) of CRH in placenta and alters gestation length. *PLoS Biol.*, **16**, e2006337.
10. Kantorovitz,M.R., Kazemian,M., Kinston,S., Miranda-Saavedra,D., Zhu,Q., Robinson,G.E., Göttgens,B., Halfon,M.S. and Sinha,S. (2009) Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse. *Dev. Cell*, **17**, 568–579.
11. Mathelier,A. and Wasserman,W.W. (2013) The next generation of transcription factor binding site prediction. *PLoS Comput. Biol.*, **9**, e1003214.
12. Mathelier,A., Fornes,O., Arenillas,D.J., Chen,C.Y., Denay,G., Lee,J., Shi,W., Shyr,C., Tan,G., Worsley-Hunt,R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **44**, D110–D115.
13. Shlyueva,D., Stampfel,G. and Stark,A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.*, **15**, 272–286.
14. Venuto,D. and Bourque,G. (2018) Identifying co-opted transposable elements using comparative epigenomics. *Dev. Growth Differ.*, **60**, 53–62.
15. Straalen,N.MV. and Roelofs,D. (2012) *An Introduction to Ecological Genomics*. Oxford University Press, UK.
16. Fan,S., Hansen,M.E., Lo,Y. and Tishkoff,S.A. (2016) Going global by adapting local: A review of recent human adaptation. *Science*, **354**, 54–59.
17. Mackay,T.FC. and Huang,W. (2018) Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. *Wiley Interdiscip Rev. Dev. Biol.*, **7**, e289.
18. Clos,J., Rabindran,S., Wisniewski,J. and Wu,C. (1993) Induction temperature of human heat shock factor is reprogrammed in a Drosophila cell environment. *Nature*, **364**, 252–255.
19. Lemaitre,B. and Hoffmann,J. (2007) The host defense of Drosophila melanogaster. *Annu. Rev. Immunol.*, **25**, 697–743.
20. Espinosa-Diez,C., Miguel,V., Mennerich,D., Kietzmann,T., Sánchez-Pérez,P., Cadenas,S. and Lamas,S. (2015) Antioxidant responses and cellular adjustments to oxidative stress. *Redox Biol.*, **6**, 183–197.
21. Zhou,D. and Haddad,G.G. (2013) Genetic analysis of hypoxia tolerance and susceptibility in Drosophila and humans. *Annu. Rev. Genomics Hum. Genet.*, **14**, 25–43.
22. Thiel,G. and Cibelli,G. (2002) Regulation of life and death by the zinc finger transcription factor Egr-1. *J. Cell Physiol.*, **193**, 287–292.

23. Sims,H.I., Chirn,G.W. and Marr,M.T. (2012) Single nucleotide in the MTF-1 binding site can determine metal-specific transcription activation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16516–16521.
24. Reed,D.E., Huang,X.M., Wohlschlegel,J.A., Levine,M.S. and Senger,K. (2008) DEAF-1 regulates immunity gene expression in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 8351–8356.
25. Hu,F., Yu,X., Wang,H., Zuo,D., Guo,C., Yi,H., Tirosh,B., Subjeck,J.R., Qiu,X. and Wang,X.Y. (2011) ER stress and its regulator X-box-binding protein-1 enhance polyIC-induced innate immune response in dendritic cells. *E. J. Immunol.*, **41**, 1086–1097.
26. Tan,G. and Lenhard,B. (2016) TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics*, **32**, 1555–1556.
27. Wagih,O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645–3647.
28. Thurmond,J., Goodman,J.L., Strelets,V.B., Attrill,H., Gramates,L.S., Marygold,S.J., Matthews,B.B., Millburn,G., Antonazzo,G., Trovisco,V. *et al.* (2019) FlyBase 2.0: the next generation. *Nucleic Acids Res.*, **47**, D759–D765.
29. Ma,W., Noble,W.S. and Bailey,T.L. (2014) Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.*, **9**, 1428–1450.
30. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
31. Worsley,H.R., Mathelier,A., Del Peso,L. and Wasserman,W.W. (2014) Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC Genomics*, **15**, 472.
32. Chung,D., Kuan,P.F., Li,B., Sanalkumar,R., Liang,K., Bresnick,E.H., Dewey,C. and Keleş,S. (2011) Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS Comput. Biol.*, **7**, e1002111.
33. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
34. Feng,J., Liu,T., Qin,B., Zhang,Y. and Liu,X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.*, **7**, 1728–1740.
35. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
36. Davie,K., Jacobs,J., Atkins,M., Potier,D., Christiaens,V., Halder,G. and Aerts,S. (2015) Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLos Genet.*, **11**, e1004994.
37. Koenecke,N., Johnston,J., Gaertner,B., Natarajan,M. and Zeitlinger,J. (2016) Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.*, **17**, 196.
38. Gel,B., Diez-Villanueva,A., Serra,E., Buschbeck,M., Peinado,M.A. and Malinverni,R. (2016) regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*, **32**, 289–291.
39. Jung,Y.L., Luquette,L.J., Ho,J.W., Ferrari,F., Tolstorukov,M., Minoda,A., Issner,R., Epstein,C.B., Karpen,G.H., Kuroda,M.I. *et al.* (2014) Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res.*, **42**, e74.
40. Fiston-Lavier,A.S., Barron,M.G., Petrov,D.A. and Gonzalez,J. (2015) T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Res.*, **43**, e22.
41. Rech,G.E., Bogaerts-Márquez,M., Barrón,M.G., Merenciano,M., Villanueva-Cañas,J.L., Horváth,V., Fiston-Lavier,A.S., Luyten,I., Venkataram,S., Quesneville,H. *et al.* (2019) Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila. *PLos Genet.*, **15**, e1007900.
42. Huang,W., Massouras,A., Inoue,Y., Peiffer,J., Ràmia,M., Tarone,A.M., Turlapati,L., Zichner,T., Zhu,D., Lyman,R.F. *et al.* (2014) Natural variation in genome architecture among 205 drosophila melanogaster genetic reference panel lines. *Genome Res.*, **24**, 1193–1208.
43. Ramia,M., Librado,P., Casillas,S., Rozas,J. and Barbadilla,A. (2012) PopDrowser: the population drosophila browser. *Bioinformatics*, **28**, 595–596.
44. Ullastres,A., Merenciano,M. and González,J. (2019) Natural transposable element insertions drive expression changes in genes underlying *Drosophila melanogaster* immune response. bioRxiv doi: https://doi.org/10.1101/655225, 31 May 2019, preprint: not peer reviewed.
45. Tian,S., Haney,R.A. and Feder,M.E. (2010) Phylogeny disambiguates the evolution of heat-shock cis-regulatory elements in Drosophila. *PLoS One*, **5**, e10669.
46. Sykiotis,G.P. and Bohmann,D. (2008) Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. *Dev. Cell*, **14**, 76–85.
47. Brandt,A., Scharf,M., Pedra,J.H., Holmes,G., Dean,A., Kreitman,M. and Pittendrigh,B.R. (2002) Differential expression and induction of two Drosophila cytochrome P450 genes near the Rst(2)DDT locus. *Insect Mol. Biol.*, **11**, 337–341.
48. Davies,T.G., Field,L.M., Usherwood,P.N. and Williamson,M.S. (2007) DDT., pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life*, **59**, 151–162.
49. Vodovar,N., Vinals,M., Liehl,P., Basset,A., Degrouard,J., Spellman,P., Boccard,F. and Lemaitre,B. (2005) Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 11414–11419.
50. Neyen,C., Bretscher,A.J., Binggeli,O. and Lemaitre,B. (2014) Methods to study Drosophila immunity. *Methods*, **68**, 116–128.
51. Kristensen,T.N., Sorensen,P., Pedersen,K.S., Kruhoffer,M. and Loeschcke,V. (2006) Inbreeding by environmental interactions affect gene expression in Drosophila melanogaster. *Genetics*, **173**, 1329–1336.
52. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
53. Balamurugan,K., Egli,D., Selvaraj,A., Zhang,B., Georgiev,O. and Schaffner,W. (2004) Metal-responsive transcription factor (MTF-1) and heavy metal stress response in Drosophila and mammalian cells: a functional comparison. *Biol. Chem.*, **385**, 597–603.
54. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
55. Kapitonov,V.V. and Jurka,J. (2003) Molecular paleontology of transposable elements in the Drosophila melanogaster genome. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 6569–6574.
56. Singh,N.D. and Petrov,D.A. (2004) Rapid sequence turnover at an intergenic locus in Drosophila. *Mol. Biol. Evol.*, **21**, 670–680.
57. Yang,H.P. and Barbash,D.A. (2008) Abundant and species-specific DINE-1 transposable elements in 12 Drosophila genomes. *Genome Biol.*, **9**, R39.
58. Spitz,F. and Furlong,E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
59. Spivakov,M. (2014) Spurious transcription factor binding: non-functional or genetically redundant? *Bioessays*, **36**, 798–806.
60. Vihervaara,A., Sergelius,C., Vasara,J., Blom,M.A., Elsing,A.N., Roos-Mattjus,P. and Sistonen,L. (2013) Transcriptional response to stress in the dynamic chromatin environment of cycling and mitotic cells. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E3388–E3397.
61. White,M.A., Myers,C.A., Corbo,J.C. and Cohen,B.A. (2013) Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 11952–11957.
62. Xie,D., Chen,C.C., Ptaszek,L.M., Xiao,S., Cao,X., Fang,F., Ng,H.H., Lewin,H.A., Cowan,C. and Zhong,S. (2010) Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.*, **20**, 804–815.
63. Paris,M., Kaplan,T., Li,X.Y., Villalta,J.E., Lott,S.E. and Eisen,M.B. (2013) Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. *PLos Genet.*, **9**, e1003748.
64. Neph,S., Vierstra,J., Stergachis,A.B., Reynolds,A.P., Haugen,E., Vernot,B., Thurman,R.E., John,S., Sandstrom,R., Johnson,A.K.

et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, **489**, 83–90.
65. Vrljicak,P., Lucas,E.S., Lansdowne,L., Lucciola,R., Muter,J., Dyer,N.P., Brosens,J.J. and Ott,S. (2018) Analysis of chromatin accessibility in decidualizing human endometrial stromal cells. *FASEB J.*, **32**, 2467–2477.
66. Katoh,Y., Itoh,K., Yoshida,E., Miyagishi,M., Fukamizu,A. and Yamamoto,M. (2001) Two domains of Nrf2 cooperatively bind CBP., a CREB binding protein, and synergistically activate transcription. *Genes Cells*, **6**, 857–868.
67. Takii,R., Fujimoto,M., Tan,K., Takaki,E., Hayashida,N., Nakato,R., Shirahige,K. and Nakai,A. (2015) ATF1 modulates the heat shock response by regulating the stress-inducible heat shock factor 1 transcription complex. *Mol. Cell Biol.*, **35**, 11–25.
68. Bhattacharya,S., Michels,C.L., Leung,M.K., Arany,Z.P., Kung,A.L. and Livingston,D.M. (1999) Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. *Genes Dev.*, **13**, 64–75.
69. Li,Y., Kimura,T., Huyck,R.W., Laity,J.H. and Andrews,G.K. (2008) Zinc-induced formation of a coactivator complex containing the zinc-sensing transcription factor MTF-1, p300/CBP., and Sp1. *Mol. Cell Biol.*, **28**, 4275–4284.
70. Revilla,Y. and Granja,A. (2009) Viral mechanisms involved in the transcriptional CBP/p300 regulation of inflammatory and immune responses. *Crit. Rev. Immunol.*, **29**, 131–154.
71. Du,J., Leung,A., Trac,C., Lee,M., Parks,B.W., Lusis,A.J., Natarajan,R. and Schones,D.E. (2016) Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics Chromatin*, **9**, 28.
72. Xie,M., Hong,C., Zhang,B., Lowdon,R.F., Xing,X., Li,D., Zhou,X., Lee,H.J., Maire,C.L., Ligon,K.L. *et al.* (2013) DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.*, **45**, 836–841.
73. Jjingo,D., Conley,A.B., Wang,J., Marino-Ramirez,L., Lunyak,V.V. and Jordan,I.K. (2014) Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob. DNA.*, **5**, 14.
74. Calo,E. and Wysocka,J. (2013) Modification of enhancer chromatin: what, how, and why? *Mol. Cell*, **49**, 825–837.
75. Policarpi,C., Crepaldi,L., Brookes,E., Nitarska,J., French,S.M., Coatti,A. and Riccio,A. (2017) Enhancer SINEs link pol III to pol II transcription in neurons. *Cell Rep.*, **21**, 2879–2894.
76. Teng,L., He,B., Gao,P., Gao,L. and Tan,K. (2014) Discover context-specific combinatorial transcription factor interactions by integrating diverse ChIP-Seq data sets. *Nucleic Acids Res.*, **42**, e24.
77. Wang,J., Xie,G., Singh,M., Ghanbarian,A.T., Rasko,T., Szvetnik,A., Cai,H., Besser,D., Prigione,A., Fuchs,N.V. *et al.* (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, **516**, 405–409.
78. Dantoft,W., Davis,M.M., Lindvall,J.M., Tang,X., Uvell,H., Junell,A., Beskow,A. and Engström,Y. (2013) The Oct1 homolog Nubbin is a repressor of NF-κB-dependent immune gene expression that increases the tolerance to gut microbiota. *BMC Biol.*, **11**, 99.
79. Lelli,K.M., Slattery,M. and Mann,R.S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.*, **46**, 43–68.

3.4. The transcriptomic and physiological basis of desiccation tolerance in natural European *Drosophila melanogaster* populations

Vivien Horváth, Sara Guirao-Rico, Judit Salces-Ortiz, Gabriel E. Rech, Eugenio Aprea , Mirco Rodeghiero , Gianfranco Anfora , and Josefa González

The transcriptomic and physiological basis of desiccation stress response in natural European *Drosophila melanogaster* populations

Vivien Horváth[1], Sara Guirao-Rico[1], Judit Salces-Ortiz[1], Gabriel E. Rech[1], Eugenio Aprea [2], Mirco Rodeghiero [2], Gianfranco Anfora [2], and Josefa González[1]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain
[2] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trentino, Italy

ABSTRACT
Climate change is one of the main factors shaping the distribution and biodiversity of organisms, among others by greatly altering water availability, thus exposing species and ecosystems to harsh desiccation conditions. Insects are especially threatened by these challenging dry environments, because of their small size and thus large surface area to volume ratio. *Drosophila melanogaster* is a great model to study the response of populations to rapidly changing conditions, because of its southern African origin and recent and fast worldwide colonization. Desiccation stress response is a complex and extensively studied trait, however the natural variation in tolerance, and the underlying transcriptomic and physiological mechanisms are still not clear. Here we subjected to desiccation stress 74 natural *D. melanogaster* European strains, belonging to five different climate zones. We found that the strains from cold semi-arid climates are more tolerant compared with the ones from hot summer mediterranian climate zones. Moreover, the variance in the tolerance of the strains correlates with the interaction of altitude and evaporation. We found that the tolerant strains had a lower level of initial water content and lose less water during desiccation stress. The reduction in the water loss is probably due to the decrease in the respiration rate in desiccation stress conditions, and to the cuticular hydrocarbon composition found in tolerant strains. Moreover, we found that the genes related to response to stimulus and environmental sensing are up-regulated only in the tolerant strains. We also identified transposable element insertions possibly affecting the expression of genes relevant in desiccation tolerance. However, except for four insertions, there is no clear association between the presence of the TE insertions and the tolerance level of the strains. Overall, our study for the first time described the physiological and transcriptomic changes underlying the desiccation tolerance of natural European *D. melanogaster* strains, and identifies new genes that are likely to play a role in this ecologically relevant phenotype.

KEY WORDS
Climate change, transposable elements, water content, water loss, respiration rate, cuticular hydrocarbons

INTRODUCTION

Global climate changes such as increased temperature, elevated CO_2 levels, changes in UV radiation levels, and unpredictable changes in precipitation, pose a severe and widespread impact on organisms, from human health and crop production, to species distribution and biodiversity (Parmesan 2006, Stott 2016, Wheeler and Watts 2018, Waldvogel et al. 2020). Among the natural disasters caused by climate change, droughts are one of the costliest ones (Grillakis 2019). The unpredictable patterns of precipitation are causing an increase in aridity and the expansion of drylands in many regions (Schlaepfer et al. 2017).

Water related challenges are threatening a lot of species, but insects are particularly vulnerable due to their small size and thus large surface area to volume ratio (Edney 1977, Gibbs et al. 1997, Gibbs and Rajpurohit 2010). In a recent study, almost 50% decline in insect abundance has been reported, partly due to climate change (Sánchez-Bayo and Wyckhuys 2019, Møller 2020). Because insects represent most of the animal diversity, and include some economically and ecologically extremely

relevant species, such as bees, mosquitos and moths, understanding the adaptive responses of insects to climate change is crucial (Forister et al. 2019, Kellermann and Heerwaarden 2019). Most of the insect-related climate change studies so far have focused on the effect of increased temperature (Harvell et al. 2002, Jaramillo et al. 2009, Laws and Belovsky 2010, Robinet and Roques 2010, Chown et al. 2011). However, patterns of rainfall, humidity, and water availability, which can cause severe desiccation conditions, are getting more attention lately (Chown et al. 2011, Kellermann et al. 2018, Kellermann and Heerwaarden 2019). Despite the increasing body of knowledge related to desiccation stress response, the adaptive mechanisms underlying tolerance to this trait are not fully understood yet (Telonis-Scott et al. 2006, Telonis-Scott et al. 2012, Rajpurohit et al. 2018).

Drosophila species are good models for the study of physiological and genetic adaptation to dry environments, as species of this genus have adapted to diverse climatic conditions, including arid regions, during their evolutionary history (Coyne et al. 1983, Parsons 1983, Lemeunier et al. 1986, Gibbs and Matzkin 2001). Indeed, geographical variation for desiccation tolerance among Drosophila populations has been found, which suggests that it is an important trait in these species (Hoffmann et al. 2001, Hoffmann et al. 2003, Rouault et al. 2004, Matzkin et al. 2007, Parkash et al. 2008, Parkash and Aggarwal 2012, Rajpurohit and Nedved 2013, Rajpurohit et al. 2013a, Rajpurohit et al. 2013b, Rajpurohit et al. 2017, Rajpurohit et al. 2018).

Three main physiological mechanisms have been related to desiccation tolerance in Drosophila: water loss reduction, water loss tolerance, and increased bulk water content (Hadley 1994, Hoffmann and Harshman 1999, Chown 2002, Gibbs et al. 2003, Chown et al. 2011). Reduced water loss rate appears to be the most common mechanism to survive desiccation (Hoffmann and Harshman 1999, Gibbs and Matzkin 2001, Gibbs 2002, Gibbs et al. 2003, Telonis-Scott and Hoffmann 2003, Telonis-Scott et al. 2006). Water loss happens mostly by two routes, the first occurs through the spiracles during the open phase in respiration (Gibbs et al. 2003, Lehmann and Schützner 2010). The second is related to the cuticular hydrocarbons (CHCs), which are the most prominent fatty acid-derived lipids on the insect body surface (Jallon et al. 1997, Chown 2002). The variation in water loss through the cuticle has been related to the amount, chain length, and saturation of CHCs, notably a negative correlation between the length of the hydrocarbon chain and rates of water loss has been described (Gibbs et al. 1997, Gibbs 1998, Chown 2002, Gibbs 2002). In some studies, water loss (dehydration) tolerance has also been shown to be important in desiccation stress conditions, however, this desiccation tolerance mechanism appears to be the least common in Drosophila (Telonis-Scott et al. 2006, Telonis-Scott et al. 2016). Finally, the role of increased bulk water content in desiccation tolerance is still not clear. While flies more tolerant to desiccation stress were found to have higher bulk water content (Gibbs et al. 1997, Folk et al. 2001, Gefen et al. 2006, Telonis-Scott et al. 2006, Parkash and Aggarwal 2012), in other studies either no significant differences were described (Hoffmann and Parsons 1993) or higher water content was associated with lower desiccation tolerance (Ferveur et al. 2018). Some of the previously mentioned studies were performed in xeric Drosophila species, and most of the ones which studied *D. melanogaster* used strains selected for desiccation tolerance (Gibbs et al. 1997, Hoffmann and Harshman 1999, Folk et al. 2001, Gibbs 2002, Gibbs et al. 2003, Telonis-Scott and Hoffmann 2003, Gefen et al. 2006, Telonis-Scott et al. 2006, Ferveur et al. 2018). Thus, even though the physiological traits relevant for desiccation tolerance has been extensively studied, a comprehensive picture in natural *D. melanogaster* populations is still not available (Rouault et al. 2004, Parkash et al. 2008, Foley and Telonis-Scott 2011, Parkash and Aggarwal 2012, Rajpurohit et al. 2018).

There are several genome wide studies investigating the underlying genetic architecture of desiccation tolerance, however, the knowledge on the genome wide transcriptomic variation is still limited (Telonis-Scott et al. 2012, Kang et al. 2016, Telonis-Scott et al. 2016, Griffin et al. 2017, Rajpurohit et al. 2018). The great majority of

studies investigating the expression of genes responsible for desiccation tolerance in *D. melanogaster* focus on a few candidate genes. Moreover, most of these studies are done on laboratory selected lines (Sinclair et al. 2007, Sørensen et al. 2007, Terhzaz et al. 2012, Cannell et al. 2016, Clemson et al. 2018, Sun et al. 2018). Thus, to our knowledge, a comprehensive study investigating the genome wide transcriptomic changes in natural populations and physiological traits underlying desiccation tolerance in the same natural populations is not available yet.

While the role of single nucleotide polymorphisms (SNPs) and inversions in *D. melanogaster* desiccation stress tolerance has been investigated (Telonis-Scott et al. 2012, Telonis-Scott et al. 2016), the potential role of transposable elements (TEs) in this stress response has never been studied. TEs are mobile genetic sequences, and thanks to their dynamic nature, they are very powerful mutagens (Elbarbary et al. 2016, Chuong et al. 2017, Schrader and Schmitz 2019). They can have an effect on gene expression, for instance by adding transcription factor binding sites (TFBSs) or alternative transcription start sites (Casacuberta and González 2013, Chuong et al. 2017). The adaptive role of TE insertions has been shown in several environmental conditions and organisms, including *D. melanogaster* (Guio et al. 2014, Merenciano et al. 2016, Van't Hof et al. 2016).

In this work, we assessed the desiccation tolerance of natural European *D. melanogaster* strains belonging to five different climate zones, and identified the underlying physiological mechanisms and transcriptomic profiles. Moreover, we analysed the TE content in sensitive and tolerant strains and identified four candidate insertions possibly playing a role in desiccation stress resistance. To our knowledge, this is the first comprehensive study on European natural populations combining genome wide transcriptomic analysis with the characterization of physiological traits relevant in desiccation stress, and considering the potential role of transposable elements in the desiccation stress response.

MATERIALS AND METHODS

Fly husbandry

Flies were collected in 2015 from nine different European locations by members of the *DrosEU* consortium (Figure 1 and Table S1). Using the *Köppen-Geiger* climate classification system applied in Merenciano et al (2016) the nine locations belong to five different climate zones: Subarctic (*Dfc*), Oceanic (*Cfb*), Cool Summer Mediterranean (*Csb*), Cold Semi-Arid (*Bsk*) and Hot Summer Mediterranean (*Csa*) (Table S1 (Rubel and Kottek 2010, Merenciano et al. 2016)). In total, 74 inbred strains were generated from the aforementioned natural populations. Flies were inbred for 20 generations, except for two strains which were inbred for 21 generations and 11 strains that were too weak to continue with the inbreeding process and were thus stopped before reaching 20 generations (Table S1). Fly stocks were kept at 25 °C and 65% relative humidity with 12h day and night cycles.

Desiccation experiments

Desiccation survival assays were performed for the 74 strains used in this study. For each strain, three replicates of 15 individuals of four to eight days old females were used. Both, in treated and control conditions the vials were closed with cotton and sealed with parafilm in order to stop the airflow. In treated conditions, flies were placed in empty vials and between the cotton and the parafilm, three grams of silica gel (Merck) were placed, so they were starved and desiccated. Control vials were prepared similarly, except that it contained 1mL of 1% agar on the bottom of the vial to prevent desiccation. Agar in the tubes provides hydration but not food source for the control flies (Sinclair et al 2007). For 15 out of the 74 strains, >10% mortality was observed in control conditions. Thus, these strains were eliminated and the subsequent analysis were performed using 59 strains (Table S2A).
Temperature and humidity were continuously monitored using three *iBotton*s (Mouser electronics) (Table S2). Fly survival was monitored every four hours until hour 12, and at shorter intervals afterwards (1 to 3.5 hours). Flies that died before the first survival check

were considered to have been injured during the experiment setup and were not included in the analysis (Table S2).

Analysis of the LT100 values and their correlation with environmental and geographical variables

The average LT100 of the three biological replicates and the standard deviation was calculated, except for some exceptions where only two replicates were available (Table S2A). Additionally, five out of the 59 strains were removed from the analysis due to the absence of LT100 information (Table S2A). To test if the LT100 data follows a normal distribution we used the Shapiro-Wilk normality test. As the data was normally distributed we performed an ANOVA analysis to test if there were differences in the average of the LT100 values among different climate zones, and a Tukey Test to test which climates differ from each other. We then tested if the LT100 correlates with geographical or environmental variables. The geographical variables used were longitude, latitude and altitude (Table S1). For environmental data we used two different sources: (i) *World Clim* (www.worldclim.org; (Fick and Hijmans 2017) and (ii) Copernicus (*ERA5)* (Table S1) (Copernicus and (C3s) 2017). Environmental variables related to temperature and precipitation have been shown to explain the greatest variability in desiccation resistance in Drosophila (Kellermann et al. 2018), so from WorldClim, we used the 19 bioclimatic variables, which are derived from the monthly temperature and rainfall values between 1970-2000, and the yearly maximum and minimum temperature. We used the R package raster for downloading this data. We also used evaporation and solar radiation data from the year previous to the collection date obtained from ERA5 database from Copernicus. Evaporation is known to cause desiccation stress, and solar radiation has been showed to affect mortality and development combined with desiccation stress in gastropods (Strauch et al. 2004, Przeslawski 2005)

Multicollinearity is very common when working with geographical/environmental variables, so we calculated the variance inflation factor (VIF) for the geographical/environmental variables used in this study in order to remove variables that correlate with each other (Rogerson 2001). Environmental and geographical variables were sequentially removed based on the highest VIF, until the VIF number was lower than five. A linear regression analysis, with the three variables with VIF < 5 (altitude, longitude and evaporation, the last one based on ERA5) was performed (Table S1).

Desiccation-related phenotypic experiments

In the experiments detailed below, 10 tolerant and 10 sensitive strains were used from the tail of the phenotypic distribution of the lethal time 50 (LT50, the time when 50% of the flies are dead) values, unless it is stated differently. In all the experiments, we used four to seven-day old female flies.

Initial water content. The initial water content measurements were done as described in Gibbs and Matzkin (2001) with small modifications (Gibbs and Matzkin 2001). Briefly, from each strain 10 replicates of 10 females were anesthetized with CO_2 and placed into microcentrifuge tubes, put at -80 °C for a few seconds, and then measured. The tubes were placed at 55 °C for 72 hours and the dry body weight was measured again. The initial water content was estimated as the difference between wet and dry body mass (Rajpurohit et al. 2013b).

Water loss analysis. Five replicates per strain, of five flies each, were anesthetized on ice and their weight was measured. Flies were then transferred to vials containing 3 grams of silica for 6 hours and their weight was measured again. The silica reduces the humidity to < 20% in about three hours, so the flies were exposed to low (<20%) humidity conditions for three hours. Water loss was calculated as the difference of the initial and final weight after desiccation stress. All the initial water content and water loss measurements were done in a *Mettler Toledo AJ100* microbalance (000.1-gram accuracy).

Respiration rate measurements. The experiment was done in three replicates on three tolerant (GIM-012, GIM-024, and COR-

023) and three sensitive strains (TOM-08, LUN-07, and MUN-013) from the phenotypic extremes of the LT50 distribution of the desiccation experiment. Insect CO_2 exchange rate was measured with a portable photosynthesis system (Li- 6400XT, Li-Cor Biosciences, Lincoln, Nebraska, USA). We used a 7.5 cm diameter clear conifer chamber (LI-6400-05; 220 cm^3 approx. volume) to measure the respiration rate (μg CO_2 g $Insect^{-1}$ min^{-1}). The system was calibrated daily for zero water and CO_2 concentrations, moreover infrared gas analyzers (IRGAs) were matched before introducing the insects in the chamber (as suggested by 6400-89 Insect Respiration Chamber Manual). Groups of five females were placed in a net container placed inside the measuring chamber. The flow rate of the air was set to 150 mol s^{-1}, whereas the CO_2 concentration inside the chamber was fixed to 400 ppm, which is the concentration found in nature. Measurements were collected every 60 seconds. We measured for 3 hours in normal humidity conditions (65 ± 5 %) and then we reduced the humidity to 20 ± 5% and measured for 3 more hours. In both conditions, only the measurements of the second hour were used for statistical analysis, because flies need at least one hour to stabilize the respiration after the introduction in the chamber (Rajpurohit et al. 2018), and once in the chamber the rates did not change between the second and the third hour. The chamber was covered with a 2x2 cm of dark paper to keep the flies in a less active state. The effect of the treatment and strain on the respiration rate were checked using ANOVA. The difference in the amount by which the strains changed their respiration rate was analyzed using Wilcoxon signed-rank test, since the data does not follow a normal distribution.

Extraction and analysis of cuticular hydrocarbons. To extract the cuticular hydrocarbons **(**CHCs) 10 replicates of five flies for each strain were used. Flies were plunged in 200 μl of *hexane* (Sigma Aldrich) containing 20 μl of an internal standard (tridecane, 10ng/μl) and soaked for 9 minutes. Samples were vortexed gently for one minute and then the extract was removed and placed in a conical glass insert. The samples were stored at -20 °C until analysis.

Gas Chromatography Mass-Spectrometry (GS-MS) analysis was performed using a gas chromatograph (GC Agilent 7890B) coupled with a quadrupole mass spectrometer (MS Agilent 5977A MSD) operating in electron ionization mode (internal ionization source; 70 eV). 2 μl of sample were injected in the GC injection port held at 260 °C using a split ratio of 1:5. A DB-5 ms fused silica capillary column (30m x 0.250 mm; film thickness of 0.25 μm) was used for separation using helium as the carrier gas at a constant flow rate of 1.5 mL/min. The temperature program was as follows: 40°C (1 min), then it was increased with a rate of 5°C min^{-1} until 110°C (0 min), followed by 10°C min^{-1} to 190°C (0 min), then it was increased with a rate of 10°C min^{-1} to 300°C (held 2 min). The mass spectra were recorded from m/z 33 to 450. A C7-C30 n-alkane series (Supelco, Bellefonte, PA) under the same chromatographic conditions was injected in order to calculate the linear retention indices (LRIs). Tentative identification of compounds was based on mass spectra matching the NIST-2014/Wiley 7.0 libraries and comparing the calculated LRIs with those available from the literature (Everaerts et al. 2010, Dembeck et al. 2015, Stinziano et al. 2015, Flaven-Pouchon et al. 2016, Rajpurohit et al. 2017, Ferveur et al. 2018). The amount (ng/insect) of each component was calculated relative to the internal standard. The absolute and relative (%) amount of each component was calculated.

Descriptive statistical analysis was performed using R version 4.0.0. and SPSS software (v26). PCA analysis was performed with the log transformed data. Relative % and balanced ratios of desaturated (D) and saturated (S) compounds were calculated as in Rouault et al (2004): (D-S)/(D+S). Because the data was not normally distributed (Shapiro-Wilk test p-value=9.229 e-11), we used nonparametric tests. The comparison of balanced rations, and amounts of specific compounds were calculated using a Wilcoxon-signed rank test (Stinziano et al. 2015, Ferveur et al. 2018). The correlation between the hydrocarbons and the survival of the flies

(LT100) was calculated with Spearman's correlation test.

Genotyping of the *Desat2* promoter region and the *In(3R)K* inversion

We determined the presence of the 16-bp deletion in the promotor region of the *Desat2* gene (Dallerac et al. 2000) by mapping short read sequences to the *Drosophila melanogaster* reference genome release 6 (Hoskins et al. 2015). Briefly, Illumina short sequences for strains COR-023, GIM-024, LUN-007, MUN-013, TOM-008, GIM-012 from this study and for three additional strains (JUT-011, MUN-009 and MUN-015) were obtained from the BioProject PRJNA559813. Two of the three additional strains (JUT-011, MUN-09) are among the ten most tolerant and one (MUN-015) is among the ten most sensitive strains based on the LT50 values of this study (Table S2A). Reads were trimmed using *cutadapt* v1.18 (Martin 2011), mapped to the *D. melanogaster* reference genome release 6 using *bwa mem* (v0.7.15) (Li and Durbin 2010) and sorted and indexed using *samtools* (v.1.6) (Li et al. 2009). Once alignments were generated, we used IGV (v2.8.2) (Robinson et al. 2011) to manually inspect the alignments in the *Desat2* promoter region according to FlyBase gene annotation r6.26.

We also investigated In(3R)K segregating autosomal inversion (Huang et al. 2014) by mapping long read sequences available for the same nine strains from the BioProject PRJNA559813 to the *Drosophila melanogaster* reference genome release 6 (Hoskins et al. 2015). Sequences were mapped using the *ngmlr* (v.0.2.7) long read mapper (Sedlazeck et al. 2018) followed by sorting and indexing using *samtools* (v.1.6) (Li et al. 2009). Mapped sequences were then analysed using *svim* (v1.4.0) (Heller and Vingron 2019) in the *alignment* mode and with the parameters *--min_sv_size 2000000 --max_sv_size 16000000 --types INV* since this covers the characteristics of the inversion. Moreover, we consider only inversions with at least 2 supporting reads and whose breakpoint coordinates overlapped at least 90% of the In(3R)K autosomal inversion described by (Huang et al. 2014). Note that with this methodology we were able to reproduce the results obtained by cytological genotyping by (Huang et al. 2014), but using long read sequences for strain RAL-426.

RNA-seq experiments

Fly strains. From the results of the desiccation phenotyping with the 59 strains, and from the phenotyping with the 10 most sensitive and 10 most tolerant strains, LT50 values were calculated using *Probit* analysis (Finney 1971, Johnson et al 2013) (Table S2A, S2B, Figure S1, Figure S2B). As expected, tolerant strains have a higher LT50 compared with sensitive strains. Although the LT50 range was different, it was consistent with differences in humidity between the two experiments (12 to 30 hours *vs* 16 to 31 hours and 13.65% vs 19.47% humidity; Table S2A, S2B, Figure S1, S2A, S2B). Three tolerant, ES_GIM_15_12 (GIM-012), ES_GIM_15_24 (GIM-024), and ES_COR_15_23 (COR-023), and three sensitive ES_TOM_15_8 (TOM-08), LUN_15_7 (LUN-07) and MUN_15_13 (MUN-013) strains were chosen from the five strains from the tail of the LT50 distribution for RNA sequencing (Table S2C). Flies were subjected to desiccation stress as described above, until the second most sensitive strain reached 50% mortality (18h and 45 minutes). For the most sensitive strain this time was 15h and 45 minutes (Table S2C).

RNA extraction for qRT-PCR and RNA-sequencing. RNA was extracted using *GenElute*™ Mammalian Total RNA Miniprep Kit from 30 whole female flies, four to six days old, per replicate. RNA samples were treated with DNAse I (Thermo Fisher Scientific) following manufacturer's instructions. RNA concentration was checked with *NanoDrop* spectrophotometer (NanoDrop Technologies) and quality was assessed with Bioanalyzer.

qRT-PCR. We checked whether the desiccation stress triggered molecular changes in the flies and confirmed that the replicates gave similar results by measuring the expression of *frost* gene which has been reported to be up-regulated in desiccation conditions (Sinclair et al. 2007). Primers used were: *frost* forward (F) (5'-

CGATTCTTCAGCGGTCTAGG-3') and *frost* reverse (R) (5'-CTCGGAAAC GCCAAATTTTA-3'). qRT-PCR data were normalized with *Actin 5* (*Act5*) expression and mRNA abundance of each gene was compared to that in control samples using the 2(-Delta Delta C(T)) method and Student's *t*-test (Livak and Schmittgen 2001) (Table S3).

RNA-seq. 1.5 µg of total RNA from each sample was used for library preparation and sequencing. Library preparation was performed using the Truseq Stranded mRNA Sample Prep kit from Illumina following the manufacturer's instructions. Libraries were sequenced using Illumina 125 bp paired-end reads.

Analysis of RNA-seq data

Overall, we obtained 1.202 million pair-end sequences for all samples in the experiment. Fastq sequence quality was first assessed using *F*ast*QC* (v.0.11.8) (www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapter and quality trimming was performed using *C*utadapt (v. 1.18) (Martin 2011) with the parameters *--quality-cutoff* 20, *-a* AGATCGGAAGAGC and the *paired-end* option. Trimmed reads were then mapped to the *D. melanogaster* genome r6.15 using *STAR* (v.2.6) (Dobin et al. 2013). On average, 96.3% of the reads mapped to the reference genome. Technical duplications were explored using dup*R*adar (Sayols et al. 2016). Overall, we found no bias towards high number of duplicates at low read counts, so we did not remove duplicates from the alignments. We then used feature*C*ounts (v.1.6.2) (Liao et al. 2014) for counting the number of reads mapping to genes (*reverse-stranded* parameter). Overall, 91.81% of the aligned reads were uniquely assigned to a gene feature. We used *RSeQC* (v.2.6.4) (http://rseqc.sourceforge.net/) for determining junction saturation and we found all samples saturated the number of splice junctions, meaning that the sequencing depth used in the analysis was sufficient. Raw sequencing data and matrix of raw counts per gene have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE153850.

***Transcriptogramer* analysis**

Transcriptogramer R package v. 1.4.1 was used, to perform topological analysis, differential expression (DE) and gene ontology (GO) enrichment analysis (Morais et al. 2019). *Transcriptogramer* identifies expression profiles and analyzes GO enrichment of entire genetic systems instead of individual genes. We normalized and filtered raw counts of RNA-seq reads using the functions, fread (), *calcNormFactors()* and *filterByExpr()* (count per million values (CPM) higher than 0.5) in the *R data.table v. 1.12.2* (https://github.com/Rdatatable/data.table/wiki) and the *edgeR package* v *3.24.3* (Robinson et al. 2010, Mccarthy et al. 2012). The filtering step was performed to remove those genes that are lowly expressed and thus, would not be retained in the posterior statistical analysis. Then, we analyzed the processed data using the *Transcriptogramer* pipeline to identify the differential expression of functionally associated genes (hereafter clusters). The workflow of *Transcriptogramer* requires (i) an edge list with the gene connections, which was downloaded from *STRINGdb* 11.0 with a combined score greater or equal to 800; (ii) an ordered gene list, where genes are sorted by the probability if their products interact with each other, which was obtained using the *Transcriptogramer* v 1.0 for Windows (https://lief.if.ufrgs.br/pub/biosoftwares/transcriptogramer/); (iii) expression data, which in this case was the processed reads (described above) of our RNA-seq analysis; and (iv) a dictionary, for mapping proteins to gene identifiers used as expression data row-names. The name of the genes in our data were converted to *Ensembl* Peptide IDs using the *biomaRt R/Bioconductor* package v 2.38.0 to build a dictionary, to map the *Ensembl* Peptide IDs to *Ensembl* Gene IDs. First, the program assigns expression values (obtained from the expression data) to each respective gene in the ordered gene list. Then, the average expression of neighbor genes gets assigned to each gene in the ordered gene list. In order to measure the average expression of functionally associated genes, represented by neighbor genes in the

ordered gene list we have to define a sliding window centered on a given gene with a fixed radius. We initially specified three different radius (50, 80 and 125) and finally choose 125, because this gives us the highest number of statistically significant windows per cluster (as authors recommended). The p-value threshold for FDR for the differential expression was set to 0.01.

We then checked if the clusters of genes that were differentially expressed were enriched for specific GO terms representing specific pathways. The p-value threshold for FDR for the gene ontology analysis was set to 0.005 and we focused on the first 10 GOs with the highest adjusted p-value for each the cluster when interpreting the results.

We run *Transcriptogramer* with i) the six strains comparing treated and control conditions ("All DEGs") ii) the three tolerant strains comparing treated and control conditions ("Tolerant DEGs") iii) the three sensitive strains comparing treated and control conditions ("Sensitive DEGs") iv) the six strains used in the RNA-seq analysis comparing tolerant and sensitive strains in basal conditions ("Basal DEGs").

Protein-protein interaction (PPI) network analysis

Even though *Transcriptogramer* identifies the DEGs based on the network connections, it does not identify the hub genes with statistics in which we were interested in. So, in order to find the differentially expressed hub genes, which are likely to have a bigger biological impact through being connected to more genes, we did PPI network analysis and calculated network properties (Szklarczyk et al. 2019). Analysis were performed using *STRING version 11*, on the four previously mentioned four groups ("All DEGs", "Tolerant DEGs", "Sensitive DEGs", "Basal DEGs"). We only considered the results with a minimum required interaction score of 0.8 as recommended by Zhang et al (2019). As interaction sources we used *experiments* and *co-expression* data. The hub genes were determined using a *Cytoscape* plugin (Cytoscape version 3.7.1), *cytoHubba*, which calculates 11 properties of PPI networks. Among these 11 properties, it calculates Maximal Clique Centrality (MCC) which is one of the most efficient ones to find hub genes (Chin et al. 2014). We ranked the genes by MCC and considered as hub genes and candidates for being involved in desiccation stress response the 30% of the genes with the highest MCC values.

Differentially expressed gene location analysis

Using Drosophila gene expression tool (DGET), we checked the previously reported location and the level of expression of the DEGs in this study (Hu et al. 2017). DGET uses modENCODE and RNA-seq experiment data and offers information in several life stages and tissues of *Drosophila melanogaster*. Since we were working with four to six days old females, we used the information only for adult, mated four-day old females. Expression data was available for head, digestive system, carcass and ovary. We considered as highly expressed genes the ones with RPKM > 51. The enrichment of DEGs in tissues was checked using a hypergeometric test, and the significant p-value after a Bonferroni correction was 0.001.

Mutant and RNAi knockdown strains

Three of the hub genes among the ones with the highest MCC values were selected for experimental validation (*nclb, Nsun2*, and *Dbp73D*). In order to determine if the candidate hub genes have an effect on desiccation tolerance, RNAi transgenic lines and transposable element insertion mutants were analyzed (Table S4A). The effect of each gene was tested in two different backgrounds, when possible. In the case of *nclb* gene, the flies generated with the ubiquitous GAL4 driver were not viable, so we crossed the RNAi lines with the 6g1HR-GAL4-6c (HiKone) driver line, which only affects the expression of the gene of interest in the midgut, Malphigian tubulus and fat body, where the *nclb* gene is mostly expressed (Chung et al. 2007) (Table S4A). To generate the *Nsun2* mutant strain, we crossed strains carrying the RNAi controlled by an *UAS* promoter with flies carrying an ubiquitous GAL4 driver to silence the gene (Table S4A). In case of the *Dbp73D* gene, to overcome the lethality of the

pupae caused by the ubiquitous GAL4 driver, we used an Act5c-GAL4 strain regulated by the temperature sensitive repressor GAL80 (P tubP-GAL80ts), which allows us to time the activation of the driver. For these transgenic flies we transferred flies from 25ºC to 29ºC before emerging, to activate the driver which causes the mutation. Since not all offspring of each previously mentioned cross would inherit the UAS-RNAi construct, we separated the flies with the construct from the ones which do not have it in the F1 generation based on the phenotypic markers. In all cases, we did reciprocal crosses of the transgenic lines and the driver strain and experiments were carried out in the F1 generation. As controls in the experiment, reciprocal crosses with the wild-type strain of each RNAi line and the corresponding drivers were generated (Table S4A).

Besides the four RNAi lines, we analyzed mutant strains generated with a P-element transposable element insertion for *nclb* and *Nsun2*. We used the wild type strain in which the mutant was generated as control in the experiment (Table S4A).

We checked if the expression of the genes was different in transgenic and mutant strains compared to control strains by performing qRT-PCR analysis (Table S4B). For *Dbp73D* gene, when we crossed the #108310 female with P tubP-GAL80ts driver male we found no difference in the expression in one of the reciprocal crosses, and slightly higher expression in the other reciprocal cross (Table S4B). Thus, no further experiments were performed with this RNAi line.

All the desiccation experiments were done with four to seven days old females, and at least three replicates per strain were performed. We used the desiccation phenotyping protocol as described above. Survival curves were analyzed with log-rank test using SPSS statistical software (v26). Only in one case, we found differences in the survival of the strains in control conditions. One of the reciprocal crosses for *Dbp73D* (RNAi #36131 F x Gal80ts M) showed mortality in control condition (Table S12). Still, we found that the mortality of these flies in desiccation condition was higher than in control condition (long rank test p-value: <0.0001; Table S12).

Transposable elements analysis

We analyzed the TE annotations of three tolerant (GIM-024, GIM-012, COR-023) and three sensitive (LUN-07, TOM-08, MUN-013) strains (G. Rech, personal communication). Genomes of these strains were sequenced using Oxford Nanopore Technologies and TEs were *de novo* annotated using REPET package v.2.5 (Quesneville et al. 2005, Flutre et al. 2011, Hoede et al. 2014). We did not consider TEs smaller than 120bp as they are known to have a high false positives/negatives rates (G. Rech, personal communication).

TE family expression analysis. We analysed the expression of TE families by adapting the *TEtools* pipeline (Lerat et al. 2017) to the newly annotated TE copies in the *de novo* assembled genomes. Briefly, we used the first module of *TEtools* (TEcount) with customized *rosette* and *TE_fasta* files specific for each genome under study. These files were obtained based on the specific *de novo* TE annotation on each genome (G. Rech, personal communication). For each of the six genomes we run *TEcount* for mapping the RNA-Seq trimmed reads for each sample using *Bowtie2* (Langmead and Salzberg 2012). Once the counts were obtained for each sample, we normalized the counts to Reads per Kilobase per Million (RPKM) based on the size (in base pairs) of each TE family at each genome (Table S12A). Average kurtosis per sample was 55.38, so we performed a log normalization to avoid the effects of extreme RPKM values (Table S14A). Normalized values were then used to perform differential expression analysis between different set of samples using the *limma* package (Ritchie et al. 2015). The differentially expression (DE) of the TE families was done in the same comparisons as in the case of RNA-seq analysis ("All DEGs", "Tolerant DEGs", "Sensitive DEGs", "Basal DEGs"). In the case of the "All DEGs" group we controlled for the effect of strain since we observed strains represent a mayor batch in the analysis (Figure S3). We consider as significantly DE families the ones showing an adjusted p-value FDR < 0.05 and a Fold Change ≥ |1.5|.

Candidate transposable element insertions. We analyzed the TE presence/absence nearby the DEG in the "All DEGs", "Tolerant DEGs" and "Sensitive DEGs", "Basal DEGs" groups in the six genomes. We focused on the TEs located inside genes or < 1 kb distance to the closest gene. For five TEs which were present in more than one tolerant strain we did manual curation (FBti0064269, 2R_14230537_14230539_Copia, 2L_13842323_13842344_jockey, 3R_20696637_20696644_BS, X_7887128_7887141_297). We found that FBti0064269 which was annotated in three tolerant and one sensitive strain is fixed. For the other four TEs the manual curation matched the annotations.

Statistical analysis

All the statistical analysis were done using R version 3.5.2 for Mac unless stated differently.

RESULTS

Altitude and evaporation correlate with desiccation tolerance of European natural *D. melanogaster* strains

To test whether European *D. melanogaster* populations differ in their level of desiccation tolerance, we exposed 74 inbred strains from nine locations to low humidity conditions (< 20% humidity) (Figure 1 and Table S1). These nine populations belong to five different climate zones including subarctic, oceanic, cool summer mediterranean, cold semi-arid, and hot summer mediterranean climates. We found variability in the desiccation tolerance among the nine populations (Table S2A, Figure S1): the LT50 values, representing the time when half the flies were dead, ranged from 12.5 to 25.6 hours, which is wider than the one found in North American strains (LT50 = 17.9 to 20 hours; (Rajpurohit et al. 2018). The LT100 values, representing the time when all the flies were dead, varied between 16 and 32 hours (Table S2A, Figure 2A, Figure S1). Similar differences in desiccation tolerance were found in Indian populations from different latitudes, where the more tolerant strains survived about twice as much as the sensitive ones (LT100= 13.6 vs 28.6; (Parkash et al. 2008), while in Australian populations LT100 varied between 14.2 and 17.5 hours (Clemson et al. 2018). Thus, European populations showed similar or wider ranges of variation in survival times to desiccation stress compared with strains from other continents.

Flies from temperate climates have been shown to be more tolerant to desiccation stress than flies from tropical climates (Hoffmann and Harshman 1999, Kennington et al. 2001, Clemson et al. 2018, Rajpurohit et al. 2018). We thus tested whether there were significant differences in survival among flies from the five climates in our dataset (Figure 1), and we found significant differences in the average LT100 values (ANOVA, p-value: 0.023; Figure 2B and Table S2A). Pairwise comparisons using Tukey Test showed significant differences between strains from the cold semi-arid (*Bsk*) and hot summer mediterranean (*Csa)* climate zones (Tukey comparison, p-adj=0.01): the cold semi-arid strains were more tolerant (Figure 2B; Table S2A).

Finally, desiccation tolerance has been correlated with altitude, latitude, and with environmental variables such as annual precipitation and minimum temperature (Parkash et al. 2008, Kellermann et al. 2012, Kellermann et al. 2018, Rajpurohit et al. 2018). To test whether geographical and environmental variables were associated with desiccation tolerance in European strains, we did a multiple regression analysis. First, we checked if latitude, longitude or altitude explain variability in the desiccation tolerance, however none of the variables correlated with the desiccation tolerance of the strains (linear model, p-value = 0.648 (altitude), p-value = 0.853 (latitude), p-value = 0.686 (longitude); Table S2A).

Next, besides geographical variables we also considered the 19 bioclimatic variables related with temperature and rainfall available at *WorldClim*, and evaporation and solar radiation available from *ERA5* (see Material and Methods). After applying variance inflation factor (VIF) calculations, we performed the regression analysis with the non-collinear geographical/environmental variables (altitude, longitude, evaporation) and their interactions. We found that the LT100 values significantly correlated with the interaction of altitude and evaporation (p-value = 0.0005, adjusted R-squared: 0.135; Table S2A).

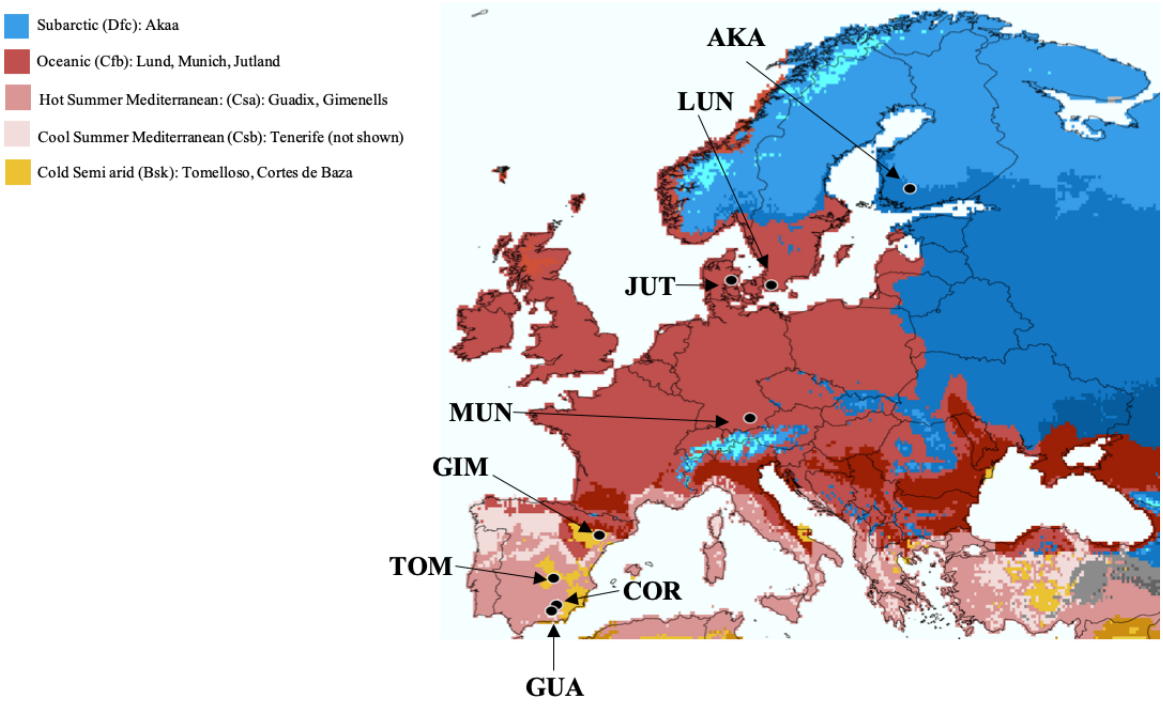


Figure 1. Geographical origin of the nine populations used in this study.
The location of the populations is indicated with arrows in a map of Europe coloured by the Köppen-Geiger climate zones, except for Tenerife which is not shown in the map. The five climate zones can be classified in three main groups: Continental climates (Subarctic); Temperate climates (Oceanic, Hot Summer Mediterranean, Cool Summer Mediterranean); Dry climates (Cold Semi arid).

Tolerant strains have lower bulk water content and lose less water during desiccation stress compared with sensitive strains

Differences in water content and in the rate of water loss, which can be influenced by the respiration rate and cuticular transpiration, have been associated with the level of desiccation tolerance in flies (Gibbs 2002, Parkash et al. 2008, Parkash and Aggarwal 2012, Ferveur et al. 2018). In order to test the differences in these physiological traits, we checked the bulk water content and the water loss in the 10 most tolerant and 10 most sensitive strains of the LT50 distribution (Figure S1; see Material and Methods). We found that the tolerant strains had significantly lower initial water content, compared to the sensitive ones (45% vs 50% bulk water content in tolerant vs sensitive, respectively, Wilcoxon, p-value: < 0.0001, Table S5A, Figure 3A), which is consistent with some previous studies (Ferveur et al. 2018).

Another way for insects to protect themselves against desiccation stress is by reducing the rate of water loss, which is a key

mechanism for desiccation survival (Gibbs and Matzkin 2001, Chown 2002, Telonis-Scott et al. 2006, Parkash et al. 2008). To quantify the amount of water loss, we measured the weight of the flies before and after three hours of desiccation stress. We found that the sensitive strains lose around 15% while the tolerant strains lose 10% of their water content (Wilcoxon, p-value: 2.88E-04) (Table S5B, Figure 3B). Our results are in agreement with previous studies performed both in populations selected for desiccation stress tolerance and natural populations (Telonis-Scott et al. 2006, Parkash et al. 2008). Overall, we showed that the tolerant strains have a lower amount of bulk water content and tend to lose less water during desiccation stress.

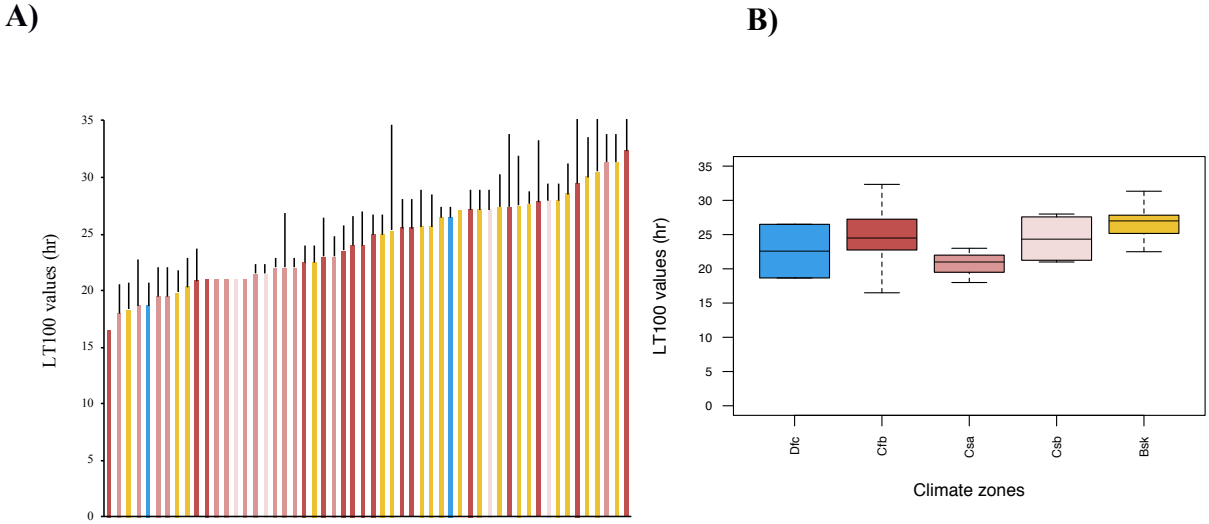


Figure 2. Desiccation survival of European natural popul
A) Bar graph showing the LT100 values for the 59 inbred strains used in this work. Y axis shows the average hour when the flies in all the replicates were dead. X axis represents the individual strains coloured by the climate zone in which they were collected. **B)** Boxplot of the distribution of the LT100 values of the strains, grouped by climate zones.

Tolerant strains decrease more their respiration rate in desiccation stress conditions compared to the sensitive ones

The primary routes for water loss in insects are respiration and cuticular transpiration, thus we tested whether differences in respiration rate or in cuticular hydrocarbons were associated with the higher capacity of water retention found in the tolerant strains (Chown et al. 2011). We measured the respiration rate of tolerant and sensitive strains in control and in desiccation stress conditions, and we found that there is an effect of the desiccation treatment (ANOVA, p-value: < 2e-16) and also of the interaction of the treatment and the strain (ANOVA, p-value: <2e-16) (Table S6A-G, Figure 3C). We found that the tolerant strains have a higher respiration rate in control conditions (Wilcoxon, p-value: 0.006). However, after desiccation stress the sensitive strains lower their respiration rate in average by 20% while the tolerant ones by 33% and this difference is statistically significant (Wilcoxon, p-value: 0.0044).

Tolerant strains have higher relative amount of desaturated hydrocarbons

The level of cuticular transpiration, which is another influential factor in water loss, depends on the composition of the cuticle, thus we next analyzed the cuticular hydrocarbon (CHC) composition of the 10 most tolerant and the 10 most sensitive strains (Figure S1, see Material and Methods). We identified 13 main hydrocarbons with chain lengths varying between 23 and 29 carbons, including three saturated (n-alkanes) and ten desaturated compounds (alkene, alkadiene) (Table 1, Table S7A). We performed principal component analysis (PCA) to explore the variability of the

strains in terms of CHC composition. We found that neither PC1 nor PC2 clearly separate the 10 tolerant from the 10 sensitive strains when considering the total CHC composition (Figure S4A), although tolerant and sensitive strains differed in the relative amounts of individual hydrocarbons (see below). When considering the subset of strains used for the RNA-seq analysis (see below), we found that PC1 does separated the tolerant from the sensitive strains, and explains 63.72% of the variation (Figure S4B). Note that these six strains are on the extremes of the phenotypic distribution for water loss measurements (Table S5B).

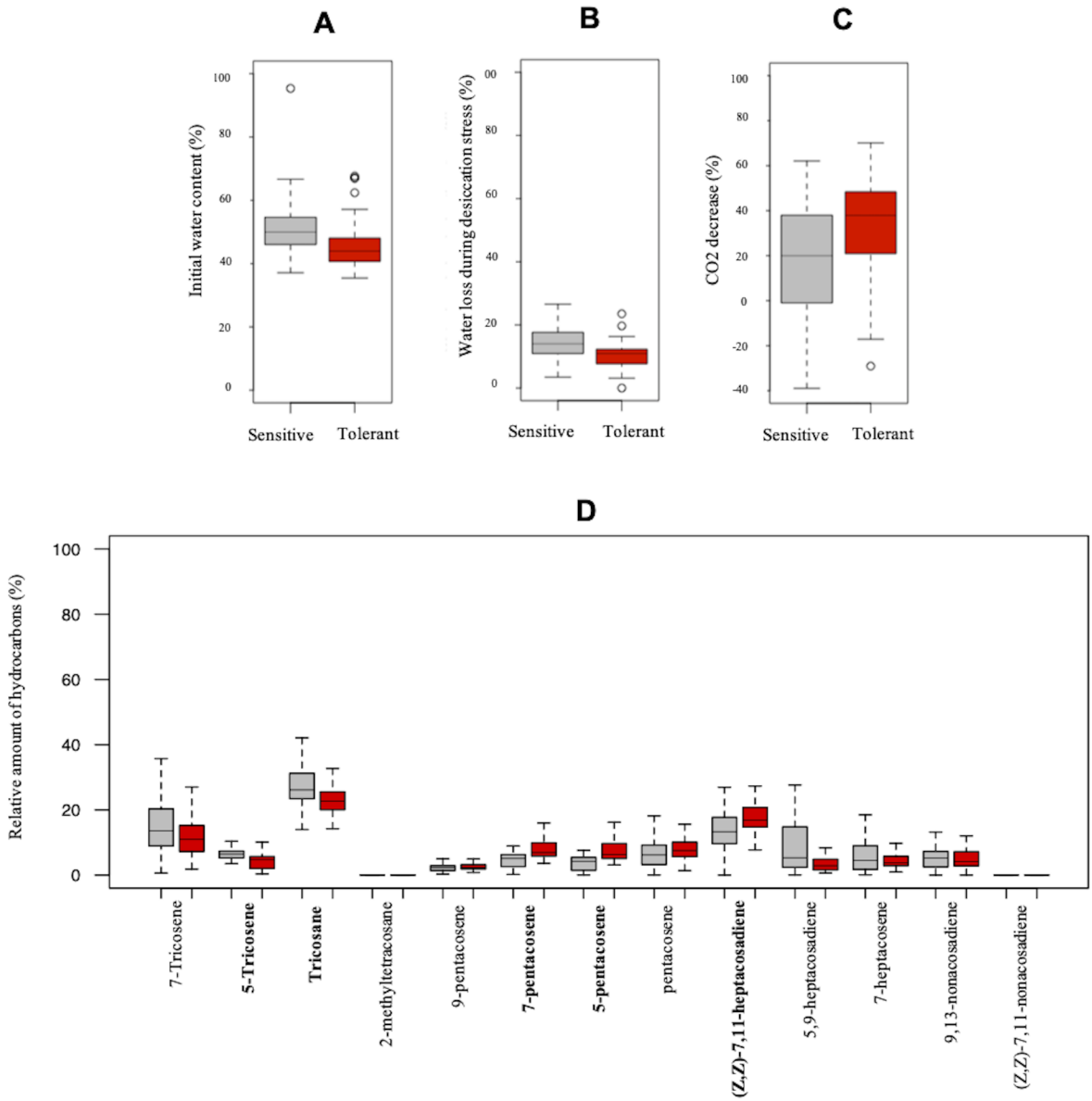


Figure 3. **Desiccation-related physiological traits and cuticular hydrocarbon variation in natural European populations.**
Results of **A)** initial water content; **B)** percentage of water loss during desiccation stress; **C)** percentage of CO_2 decrease (respiration); and **D)** relative amount of hydrocarbons in sensitive (grey) and tolerant (red) strains. Hydrocarbons that showed significant differences between sensitive and tolerant strains are depicted in bold.

We found that the tolerant strains had a higher relative amount of desaturated hydrocarbons (Wilcoxon, p-value: 0.004), and a higher desaturated:saturated balanced ratio, compared with sensitive strains, as previously reported (Wilcoxon, p-value = 0.004191;

Table S7A; (Ferveur et al. 2018). However, the percentage of 7,11:Cn alkadienes, which was previously reported to be negatively correlated with desiccation tolerance, was found to be positively correlated in our strains (Spearman's correlation = 0.203, p-value = 0.02; (Foley and Telonis-Scott 2011)). The relative percentage of longer chain hydrocarbons (> 27C) was not correlated with desiccation tolerance (Spearman's correlation = -0.19, p-value = 0.828; Table S7A, (Foley and Telonis-Scott 2011)). However, in Foley & Telonis-Scott (2011) compounds with 27 carbons or more represented approximately half of the total CHC composition, while in our European strains compounds with 25 carbons or more represented more than half of the total CHC composition. Indeed, we found a higher percentage of >25C compounds in the tolerant strains (Wilcoxon, p-value: 6.265e-10). Moreover, the percentage of >25C compounds positively correlated with the LT100 (Spearman's correlation = 0.178, p-value = 0.041) (Table S7A). In line with these results, we also found that tolerant strains had a higher relative percentage of 5-pentacosene, 7-pentacosene and 7,11-heptacosadiene, which are compounds with 25C (Wilcoxon, p-value:<0.0001 in all cases; Fig 3D), while sensitive strains had a higher relative percentage of 5-tricosene and tricosene (Wilcoxon, p-value:<0.0001 in all cases), which are compounds with 23C (Figure 3D).

Overall, the most tolerant and most sensitive strains to desiccation stress did not differed in the global CHC composition, but they differed in the relative percentage of individual CHCs. While we did not find some of the previously described correlations between particular CHC and desiccation tolerance, this is most likely explained by the different CHC composition of the European strains in which CHC with 25 or more carbons, instead of CHC with 27C or more, represent approximately half of the total CHCs.

One of the European strains showed the African cuticular hydrocarbon composition phenotype

While the primary role of CHCs is in desiccation tolerance, some CHCs have been co-opted to function as chemical signals relevant for social interactions. Differences in the levels of 5,9-C27:2 and 7,11-C27:2, which is the primary female sex pheromone, are associated with the adaptive divergence of African and out-of-Africa populations (Dembeck et al. 2015). The characteristic high level of 5,9-C27:2 and low level of 7,11-C27:2 found in African flies have been associated with the intact promoter region (no 16bp deletion) of the *Desat2* gene, that is located inside the inversion *In(3R)K*. However, some North American strains did not show a clear association between the *Desat2* locus and the CHC phenotype (Dembeck et al. 2015). Thus, we tested whether European strains also fail to show this association. We analyzed nine strains for which we have both the CHC profile and the genome sequence (see Material and Methods). Eight strains had the 16bp-deletion characteristic of the out-of-Africa strains, however only five of them showed the corresponding CHC phenotype (Table S7B). The other European strain although showing the African CHC phenotype, besides the 16bp deletion it also had two small deletions in the coding region of the *Desat2* gene. Our results thus confirmed that the relationship between the presence/absence of an intact *Desat2* gene and the CHC phenotype is not as clear as previously suggested (Dallerac et al. 2000, Takahashi et al. 2001, Dembeck et al. 2015).

Most of the differentially expressed genes are down-regulated after desiccation stress in tolerant strains while sensitive strains showed a limited coordinated response

We choose three tolerant and three sensitive strains from the extremes of the LT50 distribution to investigate the transcriptional response to desiccation stress (Figure S2A, S2B, Table S2B, see Material and Methods). We identified the differentially expressed genes (DEGs) using *Transcriptogramer* (Morais et al. 2019), and the differentially expressed hub genes using protein-protein interaction networks (see Material and Methods). We investigated the transcriptomic changes in response to desiccation stress of the six strains analyzed ("All DEGs"), and we also investigated whether the tolerant ("Tolerant DEGs") and the sensitive strains ("Sensitive DEGs") differed in this response.

We found that 90% of the "All DEGs" category, including all the hub genes, were down-regulated (Figure 4A, Table S8A, Table S9A). The same pattern was found in the "Tolerant DEGs": 83% of the genes were down-regulated, including all the hub genes (Figure 4A, Table S8B, Table S9B). This result is not unexpected as 86% of the "Tolerant DEGs" overlap with the "All DEGs" . On the other hand, in the sensitive strains we found very few up-regulated genes and no down-regulated ones in response to desiccation stress (Figure 4A, Table S8C, Table S9C). This low number of genes suggested that sensitive strains have a limited coordinated desiccation stress response.

Among the 1,524 genes which changed expression during desiccation stress conditions in our three analysis ("All DEGs", "Tolerant DEGs", and "Sensitive DEGs"), we found 379 genes (25 %) which have been previously related to desiccation stress response (Sørensen et al. 2007, Foley and Telonis-Scott 2011, Telonis-Scott et al. 2012, Dembeck et al. 2015, Kang et al. 2016, Telonis-Scott et al. 2016, Griffin et al. 2017) (Table S10A). There was a significant overlap between our candidate genes and genes showing evidence of selection in desiccation selected lines (261 genes, Fisher's exact test, p-value = 0.037, Table S10A; Kang et al. 2016). Four of our DEGs, *CG18609, Pxd, CG8814,* and *CG9801* have been previously shown to affect the cuticular composition of *D. melanogaster*, using RNAi transgenic flies (Dembeck et al. 2015) (Table S8A, B). Moreover, 10 of our DEGs overlap with candidate genes associated with different mechanisms against desiccation stress: stress sensing (malpighian tubules and stress responsive pathways), water balance, and primary hemolymph sugar protectant (Table S10B; Telonis-Scott et al. 2016).

Overall, most of the DEGs after desiccation stress response were down-regulated, and we found a significant overlap among our DEGs and genes previously related to desiccation stress response.

Genes related to metabolic process are down-regulated and genes related to response to stimulus are up-regulated after desiccation stress

Next, we explored which genetic pathways have a role in desiccation stress response and whether the tolerant and sensitive strains differed in these pathways. When analyzing the "All DEGs", we found five clusters of genes: four down- and one up-regulated (Table S8A, Figure 4B). The genes in the two biggest down-regulated clusters were mostly related to ncRNA, rRNA processing and metabolic process, and gene expression. In the other two down-regulated clusters, there were genes related to egg coat formation, chitin metabolic process, and very long chain fatty acid biosynthetic process. Note that genes related to gene expression, RNA metabolism, and chitin metabolism have been previously associated with desiccation tolerance (Telonis-Scott et al. 2012, Rajpurohit et al. 2013b, Telonis-Scott et al. 2016). On the other hand, up-regulated genes were related to signal transduction, cell communication, and response to stimulus (Table S8A). Genes belonging to response to stimulus and environmental sensing have also been previously associated to desiccation stress response in *D. melanogaster* and *D. mojavensis* (Rajpurohit et al. 2013b, Telonis-Scott et al. 2016) (Table S8A, Figure 4B).

In the tolerant strains we found six clusters of genes that were differentially expressed: four down-regulated and two up-regulated. Down-regulated clusters were enriched for similar GO terms as the ones found in the "All DEGs" analysis: RNA metabolic processes and gene expression. Also, the results for the up-regulated clusters were very similar to those found in the "All DEGs" analysis, but among the tolerant strains we also found genes related to localization and transport (Table S8B, Figure 4B). In the "Sensitive DEGs" category, we found only one up-regulated cluster, which has genes related to nucleotide metabolic and catabolic processes (Table S8C, Figure 4B).

Overall, we found that the desiccation stress response in European *D. melanogaster* strains consists of the up-regulation of genes related to response to stimulus and the down-regulation of genes related to gene expression

and metabolic processes. Interestingly, the genes related to metabolic processes are down-regulated in the tolerant but up-regulated in the sensitive strains.

Tolerant strains have a higher level of basal expression of genes related to response to stimulus, chitin metabolic process, and fatty acid elongation

Besides analyzing the transcriptional response to desiccation stress, we also compared the gene expression in tolerant *versus* sensitive strains in basal conditions. We found 3,456 DEGs including 851 genes (25%) which have been previously related to desiccation stress, including eight genes that affect the cuticular composition of *D. melanogaster* (Sinclair et al. 2007, Sørensen et al. 2007, Foley and Telonis-Scott 2011, Telonis-Scott et al. 2012, Rajpurohit et al. 2013b, Dembeck et al. 2015, Kang et al. 2016, Telonis-Scott et al. 2016, Griffin et al. 2017, Sun et al. 2018) (Table S8D, Table S10A). We found significant overlap among our gene candidates and the genes reported in Rajpurohit et al (2018), Kang et al (2016) and Telonis-Scott et al (2012) (Fisher's exact test, p-value = 0.013; p-value < 0.0001; and p-value < 0.006, respectively; Table S8D, S10A). Moreover, 15 of the DEGs in basal conditions overlapped with candidate genes reported in previous studies to be involved in pathways related to desiccation resistance: stress sensing (malphigian tubulus, and stress responsive pathways) primary hemolymph sugar protectant, and hygrosensing (Table S8D, Table S10B; Telonis-Scott et al 2016).

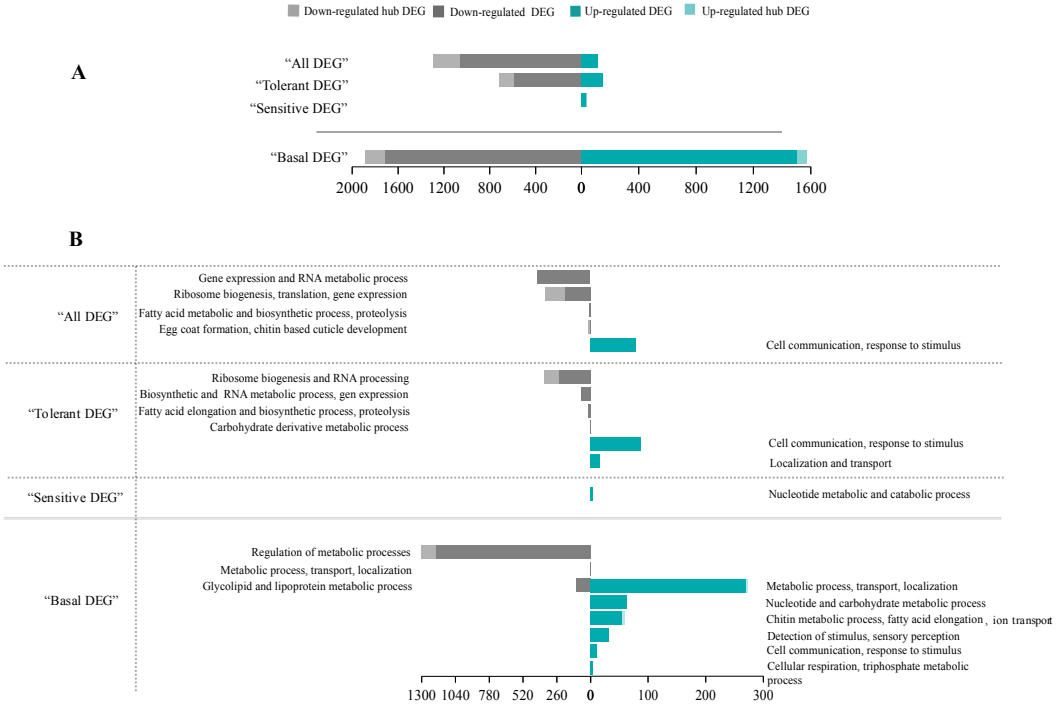


Figure 4. **Transcriptomic profile of desiccation stress sensitive and tolerant strains.**
(A) Number of differentially expressed genes in the four categories; **(B)** clusters of differentially expressed genes and enriched GO terms according to *Transcriptogramer*. The x axis represents the number of genes. The grey colour scale represents the down-regulated and the blue the up-regulated genes both in figure A and B

Next, we assessed whether DEGs between the tolerant and sensitive strains in basal conditions were significantly enriched for any specific pathways. We found eight significantly enriched clusters (Figure 4B, Table S8D): down-regulated gene clusters were mostly enriched in metabolic processes such as nucleic acid, RNA, macromolecule, and liposaccharide metabolic processes (Figure 4B, Table S8D), while up-regulated clusters contained genes related to response to stimulus, ion transport, sensory perception, cell communication, metabolic processes such as chitin metabolic processes, fatty acid

elongation and very long chain fatty acid biosynthetic process (Figure 4B, Table S8D).
Overall, our results showed that a significant proportion of DEGs between tolerant and sensitive strains in basal conditions have been previously identified as desiccation stress candidates. Moreover, tolerant strains have a higher level of basal expression of genes related to sensory perception, response to stimulus, and some metabolic processes including chitin and fatty acid elongation, while genes related to other metabolic processes were down-regulated (Table S8D, Figure 4B).

Desiccation responsive genes are enriched for highly-expressed genes in the ovary

Desiccation stress response genes are enriched for highly expressed genes in the ovary (Griffin et al. 2017). We thus used the Drosophila Gene Expression Tool (DGET) to check whether DEGs after desiccation stress or DEGs in basal conditions were enriched for genes highly expressed in the ovary, or in the other three tissues (head, digestive system, and carcass) with available data in the same age-range as the flies in which we performed our experiments (Table S11).
We found that DEGs ("All DEGs", "Tolerant DEGs", and "Sensitive DEGs") after desiccation stress, and also hub genes, were enriched for highly-expressed genes in the ovary (Hypergeometric test, p-value < 0.0001, for all comparisons; Table S11A-C). While basal DEGs were enriched for highly-expressed genes in all four tissues (Hypergeometric test, p-value < 0.0001, for all comparisons), basal hub genes were enriched for highly expressed genes in the ovary and digestive system (Hypergeometric test, p-value < 0.0001, for all comparisons, Table S11D).

Overall, similar to the previously reported results, we found that desiccation stress response genes are enriched in genes highly expressed in the ovary further suggesting that maternal effects are relevant for desiccation tolerance (Griffin et al. 2017).

Linear retention index	Component	Formula (acronym)	Hydrocarbon type	Saturated/ desaturated
2279	7-Tricosene	C23H46 (7-C23:1)	Alkene	Desaturated
2290	**5-Tricosene**	C23H46 (5-C23:1)	Alkene	Desaturated
2298	**Tricosane**	C23H48 (n-C23)	Alkane	Saturated
2464	2-Methyltetracosane	C25H52 (2-Me-C25)	Alkane	Saturated
2473	9-pentacosene	C25H50 (9-C25:1)	Alkene	Desaturated
2480	**7-pentacosene**	C25H50 (7-C25:1)	Alkene	Desaturated
2480	**5-pentacosene**	C25H50 (5-C25:1)	Alkene	Desaturated
2498	pentacosane	C25H52 (n-C25)	Alkane	Saturated
2657	**(Z,Z)-7,11 heptacosadiene**	C27H52 (7,11-C27:2)	Alkadiene	Desaturated
2670	5,9-heptacosadiene	C27H52 (5,9-C27:2)	Alkadiene	Desaturated
2693	7-heptacosene	C27H54 (7-C27:1)	Alkene	Desaturated
2841	9,13-nonacosadiene	C29H56 (9,13-C29:2)	Alkadiene	Desaturated
2872	(Z,Z)-7,11-nonacosadiene	C29H56 (7,11-C29:2)	Alkadiene	Desaturated

Table 1. Cuticular hydrocarbons (CHCs) identified in the 10 most tolerant and 10 most sensitive strains. CHCs with a different amount between the tolerant and sensitive strains are marked in bold.

nclb*, *Nsun2*, and *DBp73D* genes affect desiccation tolerance in *D. melanogaster

Besides detecting genes previously known to play a role in desiccation tolerance, our transcriptomic analysis also identified new candidate genes (Figure 4, Table S8). We choose three hub genes among the ones with the highest MCC values, *nclb, Nsun2*, *Dbp73D*, to perform functional validation experiments (Table S9). These genes were (i) mostly expressed in the ovary and digestive system; (ii) related to gene expression and

RNA methylation (Table 2); and (iii) were down-regulated in the "All DEGs" and "Tolerant DEGs" groups (Table S8).

In case of *nclb*, the insertional mutant and the two reciprocal crosses of the RNAi transgenic line analyzed showed a lower level of expression and higher survival in desiccation stress conditions when compared with the wild-type strains with similar genetic backgrounds (Log rank test, p-value: <0.0001 in all three comparisons; Table 2, Table S4 and Table S12). These results are consistent with the observed down-regulation of the *nclb* gene in our tolerant strains in desiccation stress conditions (Table S8).

For *Nsun2*, we found that while the expression of the gene in the insertional mutant and the two reciprocal crosses of the RNAi transgenic line was lower, the survival to desiccation stress was lower for the insertional mutant but higher for the RNAi reciprocal crosses (Log rank test, p-value: <0.0001 for the three comparisons; Table 2 and Table S4 and Table S12). These results are consistent with a role of *Nsun2* in desiccation tolerance and suggest that the effect of this gene is background dependent (Table 2, Table S12).

Finally, for *Dbp73D*, the two reciprocal crosses performed with the RNAi line showed lower gene expression and lower survival under desiccation stress conditions (Log rank test, p-value: <0.0001, p-value: <0.0001; Table 2 and Table S4 and Table S12). This result suggests that the effect of *Dbp73D* on desiccation tolerance is also background dependent as we found this gene to be down-regulated in tolerant strains (Table S8).

Overall, we found that all three genes affect the desiccation survival of the flies, however in some cases this effect depends on the genetic background.

Gene name/ Flybase ID	**Function (BP)**	**Mutant/RNAi (stock number)**	**qRT-PCR results (p-value)**	**Survival of the mutant /RNAi strain**	**Log-rank test p-value**
nclb/ FBgn0263510	Regulation of gene expression	*P-element* insertion in first intron (#21138)	Lower expression in the mutants (3.80E-03)	Higher survival	<0.0001
		RNAi (#41826) F x Gal4-6c M	Lower expression in the mutants (0.03)	Higher survival	<0.0001
		RNAi (#41826) M x Gal4-6c F	Marginally lower expression in the mutants (0.067)	Higher survival	<0.0001
Nsun2/ FBgn0026079	RNA methylation, tRNA methylation	*P-element* insertion in first intron (#33452)	Lower expression in the mutant (0.024)	Lower survival	<0.0001
		RNAi (#62495) F x Gal4 M	Lower expression in the mutant (5.07E-04)	Higher survival	<0.0001
		RNAi (#62495) M x Gal4 F	Lower expression in the mutant (0.03)	Higher survival	<0.0001
Dbp73D/ FBgn0004556	RNA binding	RNAi (#36131) F x Gal80ts M	Lower expression in the mutant (0.007)	Lower survival	<0.0001
		RNAi (#36131) M x Gal80ts F	Lower expression in the mutant (0.01)	Lower survival	<0.0001

Table 2. Summary of the results obtained with mutants and RNAi strains. The survival of mutant and RNAi strains is relative to the control strains, and results for each reciprocal cross are shown. F= female; M= male; NS=not significant results

Multiple transposable element families are differentially expressed after desiccation treatment

Transposable elements (TEs) can be activated and/or repressed in response to stress (for a review see (Horváth et al. 2017)). We thus checked whether expression of TE families present in the six genomes analyzed was affected by desiccation stress (Table S13). Overall, we found 25 differentially expressed families (16 down and nine up-regulated) (Table S14B, Figure 5). The sensitive strains had the highest number of differentially expressed TE families, 21, and 16 of them

were down-regulated. Interestingly, even though the tolerant strains had only eight differentially expressed TE families, six were up-regulated (Table S14B, Figure 5). Eleven of the 25 differentially expressed families found in this study were also differentially expressed in response to the insecticide malathion (Table S14C; (Salces-Ortiz et al. 2020). Similarly, down-regulated families were enriched for long-terminal repeat retrotransposons (LTRs) (Chi-square, p-value: 0.009; Table S14; (Salces-Ortiz et al. 2020).

Besides checking whether expression of TE families were affected by desiccation stress, we also checked whether the expression of genes known to be involved in TE regulation were affected (Fabian et al. 2019). We found that 34% (33/96) of the TE regulating genes were differentially expressed, 31 down and two up-regulated when sensitive and tolerant strains were analyzed together. Moreover, we found three genes differentially expressed only in the tolerant strains (two up- and one down-regulated), and none in the sensitive. Note that the number of DEG in the sensitive strains was very low (33 genes, Figure 4A, Table S8C).

Finally, we also checked whether tolerant and sensitive strains differed in the expression of any TE family or TE regulating genes in basal conditions. We found that one of the TE families, *Quasimodo2*, and 70% (six up and 59 are down-regulated) of the genes which have a role in TE regulation were differentially expressed in tolerant versus sensitive strains (Table S14B).

Overall, our results showed that sensitive strains have the highest number of differentially expressed TE families and that most of the TE families were down-regulated, while TE families differentially expressed in the tolerant strains were up-regulated. Tolerant and sensitive strains also differed in the expression of genes involved in TE regulation in basal conditions.

Candidate transposable element insertions

To identify individual TE insertions that might play a role in the change of expression of genes after desiccation stress, we look for TEs present in the coding region or in the 1kb upstream/downstream region of the identified DEGs (Table S13). We found that 138 of the "All DEGs" have TEs inside or nearby (183 TEs in total), although overall they were depleted in TEs (Chi-square, p-value: >0.001). 55 of these TEs were present in at least one tolerant and zero sensitive strains, and six of them were nearby hub DEGs. Nine of the genes unique for the "Tolerant DEGs" category have TEs inside or nearby (9 TEs in total). All nine TEs were present in one tolerant and zero sensitive strains (Table S15A and S15B). Note that 13 genes of the "All DEGs" and five genes of the "Tolerant DEGs" that have TEs nearby were previously found to be desiccation stress candidates (Sørensen et al. 2007, Foley and Telonis-Scott 2011, Telonis-Scott et al. 2012, Dembeck et al. 2015, Kang et al. 2016, Telonis-Scott et al. 2016).

Finally, we checked whether TEs could play a role in the differential expression of genes between tolerant and sensitive strains in basal conditions. We found 510 TE insertions nearby DEGs in basal conditions. 167 of them were present in at least one tolerant and zero sensitive strains, and four of them were present in two tolerant and zero sensitive strain (Table S15C). 77 of these 167 TEs are nearby genes previously shown to play a role in desiccation resistance (Sørensen et al. 2007, Foley and Telonis-Scott 2011, Telonis-Scott et al. 2012, Dembeck et al. 2015, Kang et al. 2016, Telonis-Scott et al. 2016, Griffin et al. 2017, Rajpurohit et al. 2018).

Overall, we did not find a clear correlation between the presence/absence of TEs and the tolerance/sensitivity of the strains. However, we found four TEs present in two tolerant and zero sensitive strains that were located inside or nearby DEGs in basal conditions.

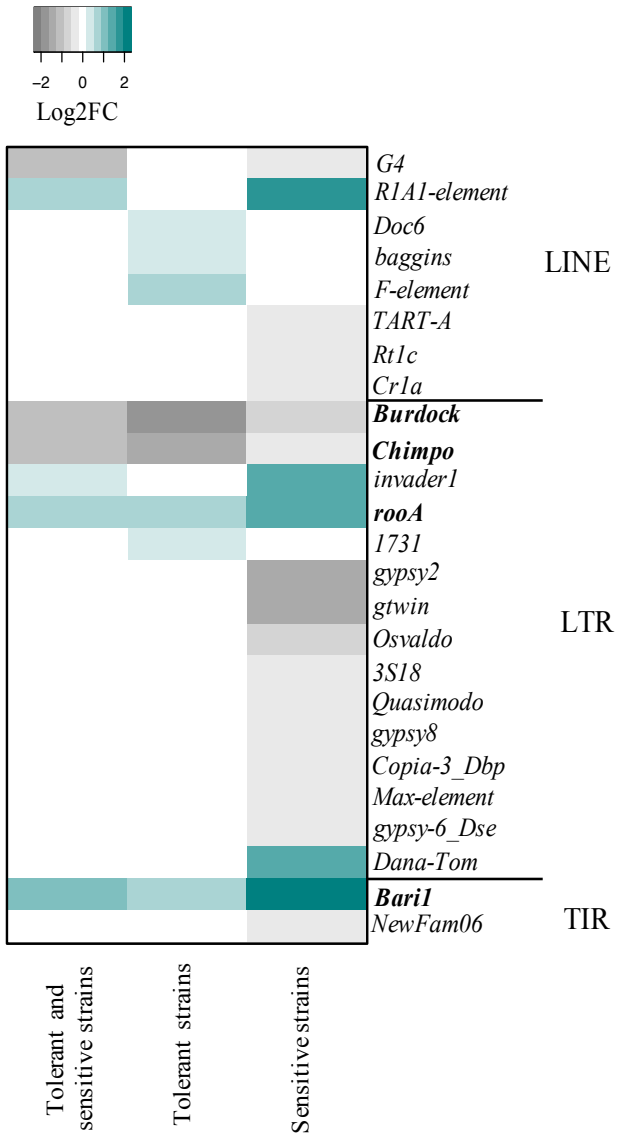


Figure 5. **Differentially expressed transposable element families.** The families common in the three categories are marked in bold.

DISCUSSION

In this study, we analysed the desiccation tolerance of European *D. melanogaster* strains, which belong to five different climate zones, including temperate, continental, and arid regions (Figure 1). We found that strains from the cold-semi arid climate (Bsk) were more tolerant compared to the temperate strains from the hot-summer mediterranean (Csa) climate (Figure 2). While in previous studies in Australia (Hoffmann and Harshman 1999, Clemson et al. 2018), South America (Kennington et al. 2001), and North America (Rajpurohit et al. 2018) temperate strains were found to be more tolerant compared to tropical ones, here we showed that arid European strains are more tolerant to desiccation compared to the temperate ones. We also found that variation in desiccation resistance in European strains can be partly explained by altitude and evaporation. The importance of altitude was previously shown in Indian populations of *D. melanogaster*, where flies from highlands were more tolerant compared to flies from lowlands (Parkash et al. 2008).

Besides describing the natural variation in desiccation stress tolerance, we sought to uncover the physiological traits which influence this variation and the coordinated response of genes which orchestrated it. In control conditions, the tolerant strains showed a higher level of respiration (Table S6).

Consistent with this result, the genes related to respiration i.e. respiratory electron transport chain, and cellular respiration, were up-regulated in tolerant strains in control conditions (Figure 4 and Table S8D). Genes related to ion transport were also up-regulated in the tolerant strains compared to the sensitive ones in basal conditions. Ion homeostasis genes have been suggested to be involved in water retention by the Malpighian tubules (MT) and cells throughout the body (Telonis-Scott et al. 2012, Rajpurohit et al. 2018), and have been related to desiccation survival before by regulating water retention in the flies (Gáliková et al. 2018).

Although tolerant strains have a higher level of respiration rate in control conditions, following desiccation stress they lower more their respiration compared to the sensitive strains, and were consistently found to lose less water (Figure 3). Reduced water loss after desiccation stress has previously been found in desiccation tolerant *D. melanogaster* strains and xeric Drosophila species (Hoffmann and Parsons 1993, Gibbs et al. 1997, Gibbs and Matzkin 2001, Telonis-Scott et al. 2006, Parkash et al. 2008, Parkash and Aggarwal 2012). Moreover, we found that genes related to metabolic processes were down-regulated after desiccation stress in the tolerant strains, thus probably causing a lowered metabolism. Reduction in the metabolic rate is known to reduce the need to open the spiracles, which is consistent with tolerant flies losing less water (Lighton 1996, Zachariassen 1996, Addo-Bediako et al. 2001).

The analysis of the CHC composition revealed that desiccation tolerant strains contained a higher proportion of CHC with 25 carbons or more, and their proportion correlated with desiccation tolerance. This is in contrast with previous studies that found CHCs with ≥27 carbons to be the most abundant, and to correlate with desiccation tolerance (Foley and Telonis-Scott 2011). However, it has been described that in populations from higher latitudes the CHCs with 23C- 25C tend to be over-represented, while in the populations from lower latitudes the hydrocarbons with 27C- 29C tend to be more abundant. While populations in the Foley and Telonis-Scott (2011) analysis were collected in latitude 38.5, our strains were collected in latitudes from 37.33 to 61 (except for one strain collected in Tenerife at latitude 28.5). Thus, the difference in the most abundant CHC hydrocarbons in European strains could be explained by the latitude at which the populations were collected. While we found that genes related to very long chain fatty acid elongation (bigger than 20C;(Swanson et al. 1995) and chitin metabolic process were up-regulated in the tolerant strains in basal conditions, after desiccation stress these genes were down-regulated. This result suggests that tolerant strains do not improve the water retaining properties of the cuticle during desiccation stress by increasing the production of cuticular proteins as has been suggested in *D. mojavensis,* where genes involved in chitin metabolism and cuticle constituents were found to be up-regulated after desiccation stress (Rajpurohit et al. 2013b). Thus, it seems that the tolerant strains have a different, more favourable CHC composition in control conditions compared with sensitive strains which might be related with their capacity to better survive low humidity conditions.

Tolerant strains showed an increased expression of genes related to response to stimulus, signaling, localization and transport after desiccation stress. Furthermore, tolerant strains also up-regulated genes related to sensory perception and detection of chemical stimulus in basal conditions compared with sensitive strains. These results suggested that the response to desiccation has an important environmental sensing component, as has been previously suggested in *D. melanogaster* and in *D. mojavensis* (Telonis-Scott et al. 2012, Rajpurohit et al. 2013b, Telonis-Scott et al. 2016). Indeed *Pkd2*, *pyrexia*, and *painless* genes involved in hygrorecepcion, a sense that allows the flies to detect changing levels of moisture in the air, were found to be differentially expressed in tolerant strains in basal conditions (Telonis-Scott et al. 2012, Telonis-Scott et al. 2016). Moreover, 27% (14 out of 52) of odorant binding proteins (*Obp*), which are also associated with hygroreception (Sun et al. 2018), were up-regulated in tolerant strains in basal conditions, which also underpins the importance of environmental sensing in control conditions.

Finally, we also investigated the potential role of transposable element (TE)-induced mutations in desiccation stress tolerance, as TEs have previously been shown to play an important role in stress response in Drosophila (Horváth et al. 2017, Rech et al. 2019, Salces-Ortiz et al. 2020). Although several of the DEGs in response to stress had TEs in their coding or regulatory regions, we did not find a clear correlation between the presence/absence of the TEs and the tolerance/sensitivity of the strains. In the case of genes differentially expressed in basal conditions between tolerant and sensitive strains, we found four genes which had TE insertions in two out of the three tolerant strains analyzed, while these TEs were absent in the sensitive strains. These four insertions are promising candidates to play a role in the tolerance of these strains to desiccation stress.

Overall our results suggest, that desiccation stress response is a complex trait, as reported before, and it is influenced by variation in several physiological traits. Moreover, differences in gene expression between sensitive and tolerant strains are relevant both in basal and upon desiccation stress conditions. We propose that the composition of the cuticle in control conditions is also relevant, and upon desiccation stress the tolerant strains lower their metabolism which probably results in lowering their respiration rate.

ACKNOWLEDGMENTS

We thank Maria Bogaerts-Márquez and Santiago Radío for technical help. We also thank Omar Rota-Stabelli for his help during the stay at the *Fondazione Edmund Mach*. We also thank Diego Morais and Rodrigo Dalmolin for their advice related to the *Transcriptogramer* package.

DATA AVAILABILITY

Supplementary tables are figures are available in Figshare**:**
https://figshare.com/s/79b11a350c9a803de0a3

REFERENCES

1. Addo-Bediako, A., S.L. Chown and K.J. Gaston (2001). Revisiting water loss in insects: a large scale view.J Insect Physiol(12): 1377-1388.10.1016/s0022-1910(01)00128-7
2. Cannell, E., A.J. Dornan, K.A. Halberg, S. Terhzaz, J.a.T. Dow, et al. (2016). The corticotropin-releasing factor-like diuretic hormone 44 (DH44) and kinin neuropeptides modulate desiccation and starvation tolerance in Drosophila melanogaster.Peptides(96-107.10.1016/j.peptides.2016.02.004
3. Casacuberta, E. and J. González (2013). The impact of transposable elements in environmental adaptation.Mol Ecol(6): 1503-1517.10.1111/mec.12170
4. Chin, C.H., S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, et al. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome.BMC Syst Biol(S11.10.1186/1752-0509-8-S4-S11
5. Chown, S.L. (2002). Respiratory water loss in insects.Comp Biochem Physiol A Mol Integr Physiol(3): 791-804.10.1016/s1095-6433(02)00200-3
6. Chown, S.L., J.G. Sørensen and J.S. Terblanche (2011). Water loss in insects: an environmental change perspective.J Insect Physiol(8): 1070-1084.10.1016/j.jinsphys.2011.05.004
7. Chung, H., M.R. Bogwitz, C. Mccart, A. Andrianopoulos, R.H. Ffrench-Constant, et al. (2007). Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1.Genetics(3): 1071-1077.10.1534/genetics.106.066597
8. Chuong, E.B., N.C. Elde and C. Feschotte (2017). Regulatory activities of transposable elements: from conflicts to benefits.Nat Rev Genet(2): 71-86.10.1038/nrg.2016.139
9. Clemson, A.S., C.M. Sgrò and M. Telonis-Scott (2018). Transcriptional profiles of plasticity for desiccation stress in Drosophila.Comp Biochem Physiol B Biochem Mol Biol(1-9.10.1016/j.cbpb.2017.11.003

10. Copernicus, C.C.S. and (C3s) (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS).
11. Coyne, J.A., J. Bundgaard and T. Prout (1983). Geographic variation of tolerance to environmental stress in Drosophila pseudoobscura. . The American Naturalist. **122:** 474-488.
12. Dallerac, R., C. Labeur, J.M. Jallon, D.C. Knipple, W.L. Roelofs, et al. (2000). A delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in Drosophila melanogaster.Proc Natl Acad Sci U S A(17): 9449-9454.10.1073/pnas.150243997
13. Dembeck, L.M., K. Böröczky, W. Huang, C. Schal, R.R. Anholt, et al. (2015). Genetic architecture of natural variation in cuticular hydrocarbon composition in Drosophila melanogaster.Elife(10.7554/eLife.09861
14. Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, et al. (2013). STAR: ultrafast universal RNA-seq aligner.Bioinformatics (Oxford, England)(1): 15-21.10.1093/bioinformatics/bts635
15. Edgar, R., M. Domrachev and A.E. Lash (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.Nucleic Acids Res(1): 207-210.10.1093/nar/30.1.207
16. Edney, E.B. (1977). Water Balance in Land Arthropods. Springer, Berlin.
17. Elbarbary, R.A., B.A. Lucas and L.E. Maquat (2016). Retrotransposons as regulators of gene expression.Science(6274): aac7247.10.1126/science.aac7247
18. Everaerts, C., J.P. Farine, M. Cobb and J.F. Ferveur (2010). Drosophila cuticular hydrocarbons revisited: mating status alters cuticular profiles.PLoS One(3): e9607.10.1371/journal.pone.0009607
19. Fabian, D.K., H.M. Dönertaş, M. Fuentealba, L. Partridge and J.M. Thornton (2019). Transposable element landscape in Drosophila populations selected for longevity.bioRxiv(867838.10.1101/867838
20. Ferveur, J.F., J. Cortot, K. Rihani, M. Cobb and C. Everaerts (2018). Desiccation resistance: effect of cuticular hydrocarbons and water content in.PeerJ(e4318.10.7717/peerj.4318
21. Fick, S.E. and R.J. Hijmans (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology. **37:** 4302-4315.
22. Flaven-Pouchon, J., J.P. Farine, J. Ewer and J.F. Ferveur (2016). Regulation of cuticular hydrocarbon profile maturation by Drosophila tanning hormone, bursicon, and its interaction with desaturase activity.Insect Biochem Mol Biol(87-96.10.1016/j.ibmb.2016.10.007
23. Flutre, T., E. Duprat, C. Feuillet and H. Quesneville (2011). Considering transposable element diversification in de novo annotation approaches.PLoS One(1): e16526.10.1371/journal.pone.0016526
24. Foley, B.R. and M. Telonis-Scott (2011). Quantitative genetic analysis suggests causal association between cuticular hydrocarbon composition and desiccation survival in Drosophila melanogaster.Heredity (Edinb)(1): 68-77.10.1038/hdy.2010.40
25. Folk, D.G., C. Han and T.J. Bradley (2001). Water acquisition and partitioning in Drosophila melanogaster: effects of selection for desiccation-resistance.J Exp Biol(Pt 19): 3323-3331
26. Forister, M.L., E.M. Pelton and S.H. Black (2019). Declines in insect abundance and diversity: We know enough

27. to act now. Conservation Science and Practice. **1:e80**.
28. Gáliková, M., H. Dircksen and D.R. Nässel (2018). The thirsty fly: Ion transport peptide (ITP) is a novel endocrine regulator of water homeostasis in Drosophila.PLoS Genet(8): e1007618.10.1371/journal.pgen.1007618
29. Gefen, E., A.J. Marlon and A.G. Gibbs (2006). Selection for desiccation resistance in adult Drosophila melanogaster affects larval development and metabolite accumulation.J Exp Biol(Pt 17): 3293-3300.10.1242/jeb.02397
30. Gibbs, A.G. (1998). Water-Proofing Properties of Cuticular Lipids. American Zoologist, Oxford University Press. **38:** 471-482.

31. Gibbs, A.G. (2002). Water balance in desert Drosophila: lessons from non-charismatic microfauna.Comp Biochem Physiol A Mol Integr Physiol(3): 781-789.10.1016/s1095-6433(02)00208-8
32. Gibbs, A.G., A.K. Chippindale and M.R. Rose (1997). Physiological mechanisms of evolved desiccation resistance in Drosophila melanogaster.J Exp Biol(Pt 12): 1821-1832
33. Gibbs, A.G., F. Fukuzato and L.M. Matzkin (2003). Evolution of water conservation mechanisms in Drosophila.J Exp Biol(Pt 7): 1183-1192.10.1242/jeb.00233
34. Gibbs, A.G. and L.M. Matzkin (2001). Evolution of water balance in the genus Drosophila.J Exp Biol(Pt 13): 2331-2338
35. Gibbs, A.G. and S. Rajpurohit (2010). Cuticular lipids and water balance. In G. Blomquist & A. Bagnères (Eds.), Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology, Cambridge University Press**:** 100-120.
36. Griffin, P.C., S.B. Hangartner, A. Fournier-Level and A.A. Hoffmann (2017). Genomic Trajectories to Desiccation Resistance: Convergence and Divergence Among Replicate Selected Drosophila Lines.Genetics(2): 871-890.10.1534/genetics.116.187104
37. Grillakis, M.G. (2019). Increase in severe and extreme soil moisture droughts for Europe under climate change.Sci Total Environ(1245-1255.10.1016/j.scitotenv.2019.01.001
38. Guio, L., M.G. Barrón and J. González (2014). The transposable element Bari-Jheh mediates oxidative stress response in Drosophila.Mol Ecol(8): 2020-2030.10.1111/mec.12711
39. Hadley, N.F. (1994). Water Relations of Terrestrial Arthropods. Academic Press, San Diego. .
40. Harvell, C.D., C.E. Mitchell, J.R. Ward, S. Altizer, A.P. Dobson, et al. (2002). Climate warming and disease risks for terrestrial and marine biota.Science(5576): 2158-2162.10.1126/science.1063699
41. Heller, D. and M. Vingron (2019). SVIM: structural variant identification using mapped long reads.Bioinformatics(17): 2907-2915.10.1093/bioinformatics/btz041
42. Hoede, C., S. Arnoux, M. Moisset, T. Chaumier, O. Inizan, et al. (2014). PASTEC: an automatic transposable element classification tool.PLoS One(5): e91929.10.1371/journal.pone.0091929
43. Hoffmann, A.A., R. Hallas, C. Sinclair and P. Mitrovski (2001). Levels of variation in stress resistance in drosophila among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold resistance, and associated traits.Evolution(8): 1621-1630.10.1111/j.0014-3820.2001.tb00681.x
44. Hoffmann, A.A., R.J. Hallas, J.A. Dean and M. Schiffer (2003). Low potential for climatic stress adaptation in a rainforest Drosophila species.Science(5629): 100-102.10.1126/science.1084296
45. Hoffmann, A.A. and L.G. Harshman (1999). Desiccation and starvation resistance in Drosophila: patterns of variation at the species, population and intrapopulation levels.Heredity (Edinb)(637-643.10.1046/j.1365-2540.1999.00649.x
46. Hoffmann, A.A. and P.A. Parsons (1993). Direct and correlated responses to selection for desiccation resistance: a comparison of *Drosophila melanogaster* and *D. simulans*. Journal of Evolutionary Biology. **6:** 643-657.
47. Horváth, V., M. Merenciano and J. González (2017). Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response.Trends Genet(11): 832-841.10.1016/j.tig.2017.08.007
48. Hoskins, R.A., J.W. Carlson, K.H. Wan, S. Park, I. Mendez, et al. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome.Genome Res(3): 445-458.10.1101/gr.185579.114
49. Hu, Y., A. Comjean, N. Perrimon and S.E. Mohr (2017). The Drosophila Gene Expression Tool (DGET) for expression analyses.BMC Bioinformatics(1): 98.10.1186/s12859-017-1509-z
50. Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Ramia, et al. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines.Genome Res(7): 1193-1208.10.1101/gr.171546.113
51. Jallon, J., G. Kunesch, L. Bricard and M. Pennanec'h (1997). Incorporation of fatty acids into cuticular hydrocarbons of male and female Drosophila melanogaster.J Insect

Physiol(12): 1111-1116.10.1016/s0022-1910(97)00082-6
52. Jaramillo, J., A. Chabi-Olaye, C. Kamonjo, A. Jaramillo, F.E. Vega, et al. (2009). Thermal tolerance of the coffee berry borer Hypothenemus hampei: predictions of climate change impact on a tropical insect pest.PLoS One(8): e6487.10.1371/journal.pone.0006487
53. Kang, L., D.D. Aggarwal, E. Rashkovetsky, A.B. Korol and P. Michalak (2016). Rapid genomic changes in Drosophila melanogaster adapting to desiccation stress in an experimental evolution system.BMC Genomics(233.10.1186/s12864-016-2556-y
54. Kellermann, V. and B.V. Heerwaarden (2019). Terrestrial insects and climate change: adaptive responses in key traits. Physiological Entomology. **44:** 99-115.
55. Kellermann, V., A.A. Hoffmann, J. Overgaard, V. Loeschcke and C.M. Sgrò (2018). Plasticity for desiccation tolerance across Drosophila species is affected by phylogeny and climate in complex ways.Proc Biol Sci(1874).10.1098/rspb.2018.0048
56. Kellermann, V., V. Loeschcke, A.A. Hoffmann, T.N. Kristensen, C. Fløjgaard, et al. (2012). Phylogenetic constraints in key functional traits behind species' climate niches: patterns of desiccation and cold resistance across 95 Drosophila species.Evolution(11): 3377-3389.10.1111/j.1558-5646.2012.01685.x
57. Kennington, W.J., A.S. Gilchrist, D.B. Goldstein and L. Partridge (2001). The genetic bases of divergence in desiccation and starvation resistance among tropical and temperate populations of Drosophila melanogaster.Heredity (Edinb)(Pt 3): 363-372.10.1046/j.1365-2540.2001.00925.x
58. Langmead, B. and S.L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2.Nat Methods(4): 357-359.10.1038/nmeth.1923
59. Laws, A.N. and G.E. Belovsky (2010). How will species respond to climate change? Examining the effects of temperature and population density on an herbivorous insect.Environ Entomol(2): 312-319.10.1603/en09294
60. Lehmann, F.O. and P. Schützner (2010). The respiratory basis of locomotion in Drosophila.J Insect Physiol(5): 543-550.10.1016/j.jinsphys.2009.04.019
61. Lemeunier, F., J.R. David, L. Tsacas and M. Ashburner (1986). The melanogaster species group In M. Ashburner, H. L. Carson, & J. Thompson (Eds.), The genetics and biology of Drosophila, , London and Orlando: Academic Press. **e3:** 147-256.
62. Lerat, E., M. Fablet, L. Modolo, H. Lopez-Maestre and C. Vieira (2017). TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes.Nucleic Acids Research(4): e17-e17.10.1093/nar/gkw953
63. Li, H. and R. Durbin (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform.Bioinformatics(5): 589-595.10.1093/bioinformatics/btp698
64. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, et al. (2009). The Sequence Alignment/Map format and SAMtools.Bioinformatics(16): 2078-2079.10.1093/bioinformatics/btp352
65. Liao, Y., G.K. Smyth and W. Shi (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.Bioinformatics(7): 923-930.10.1093/bioinformatics/btt656
66. Lighton, J.R. (1996). Discontinuous gas exchange in insects.Annu Rev Entomol(309-324.10.1146/annurev.en.41.010196.001521
67. Livak, K.J. and T.D. Schmittgen (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.Methods(4): 402-408.10.1006/meth.2001.1262
68. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.2011(1): 3.10.14806/ej.17.1.200
69. Matzkin, L.M., T.D. Watts and T.A. Markow (2007). Desiccation resistance in four Drosophila species: sex and population effects.Fly (Austin)(5): 268-273.10.4161/fly.5293
70. Mccarthy, D.J., Y. Chen and G.K. Smyth (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.Nucleic Acids Res(10): 4288-4297.10.1093/nar/gks042

71. Merenciano, M., A. Ullastres, M.A. De Cara, M.G. Barrón and J. González (2016). Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in Drosophila.PLoS Genet(8): e1006249.10.1371/journal.pgen.1006249
72. Møller, A.P. (2020). Quantifying rapidly declining abundance of insects in Europe using a paired experimental design.Ecol Evol(5): 2446-2451.10.1002/ece3.6070
73. Morais, D.a.A., R.M.C. Almeida and R.J.S. Dalmolin (2019). Transcriptogramer: an R/Bioconductor package for transcriptional analysis based on protein-protein interaction.Bioinformatics(16): 2875-2876.10.1093/bioinformatics/btz007
74. Parkash, R. and D.D. Aggarwal (2012). Trade-off of energy metabolites as well as body color phenotypes for starvation and desiccation resistance in montane populations of Drosophila melanogaster.Comp Biochem Physiol A Mol Integr Physiol(2): 102-113.10.1016/j.cbpa.2011.09.010
75. Parkash, R., S. Rajpurohit and S. Ramniwas (2008). Changes in body melanisation and desiccation resistance in highland vs. lowland populations of D. melanogaster.J Insect Physiol(6): 1050-1056.10.1016/j.jinsphys.2008.04.008
76. Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change.Annual Review of Ecology, Evolution, and Systematics(1): 637-669.10.1146/annurev.ecolsys.37.091305.110100
77. Parsons, P.A. (1983). The evolutionary biology of colonizing species. Cambridge: Cambridge University Press. .
78. Przeslawski, R. (2005). Combined effects of solar radiation and desiccation on the mortality and development of encapsulated embryos of rocky shore gastropods Marine Ecology Progress Series. **298:** 169-177.
79. Quesneville, H., C.M. Bergman, O. Andrieu, D. Autard, D. Nouaud, et al. (2005). Combined evidence annotation of transposable elements in genome sequences.PLoS Comput Biol(2): 166-175.10.1371/journal.pcbi.0010022
80. Rajpurohit, S., E. Gefen, A.O. Bergland, D.A. Petrov, A.G. Gibbs, et al. (2018). Spatiotemporal dynamics and genome-wide association genome-wide association analysis of desiccation tolerance in Drosophila melanogaster.Mol Ecol(17): 3525-3540.10.1111/mec.14814
81. Rajpurohit, S. and O. Nedved (2013). Clinal variation in fitness related traits in tropical drosophilids of the Indian subcontinent. Journal of Thermal Biology. **38:** 345-354.
82. Rajpurohit, S., O. Nedved and A.G. Gibbs (2013a). Meta-analysis of geographical clines in desiccation tolerance of Indian drosophilids.Comp Biochem Physiol A Mol Integr Physiol(2): 391-398.10.1016/j.cbpa.2012.11.013
83. Rajpurohit, S., C.C. Oliveira, W.J. Etges and A.G. Gibbs (2013b). Functional genomic and phenotypic responses to desiccation in natural populations of a desert drosophilid.Mol Ecol(10): 2698-2715.10.1111/mec.12289
84. Rajpurohit, S., X. Zhao and P.S. Schmidt (2017). A resource on latitudinal and altitudinal clines of ecologically relevant phenotypes of the Indian Drosophila.Sci Data(170066.10.1038/sdata.2017.66
85. Rech, G.E., M. Bogaerts-Márquez, M.G. Barrón, M. Merenciano, J.L. Villanueva-Cañas, et al. (2019). Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila.PLoS Genet(2): e1007900.10.1371/journal.pgen.1007900
86. Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies.Nucleic Acids Research(7): e47-e47.10.1093/nar/gkv007
87. Robinet, C. and A. Roques (2010). Direct impacts of recent climate warming on insect populations.Integr Zool(2): 132-142.10.1111/j.1749-4877.2010.00196.x
88. Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, et al. (2011). Integrative genomics viewer.Nature Biotechnology(1): 24-26.10.1038/nbt.1754

89. Robinson, M.D., D.J. Mccarthy and G.K. Smyth (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.Bioinformatics(1): 139-140.10.1093/bioinformatics/btp616
90. Rogerson, P.A. (2001). Statistical methods for geography. UK: Sage, London.
91. Rouault, J.D., C. Marican, C. Wicker-Thomas and J.M. Jallon (2004). Relations between cuticular hydrocarbon (HC) polymorphism, resistance against desiccation and breeding temperature; a model for HC evolution in D. melanogaster and D. simulans.Genetica(1-3): 195-212.10.1023/b:gene.0000017641.75820.49
92. Rubel, F. and M. Kottek (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the K ö ppen-Geiger climate classif cation Meteorologische Zeitschrift, Gebr Ü der Borntraeger. **19:** 135-141.
93. Salces-Ortiz, J., C. Vargas-Chavez, L. Guio, G.E. Rech and J. González (2020). Transposable elements contribute to the genomic response to insecticides in.Philos Trans R Soc Lond B Biol Sci(1795): 20190341.10.1098/rstb.2019.0341
94. Sánchez-Bayo, F. and K.a.G. Wyckhuys (2019). Worldwide decline of the entomofauna: A review of its drivers. Biological Conservation. **232:** 8-27.
95. Sayols, S., D. Scherzinger and H. Klein (2016). dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data.BMC Bioinformatics(1): 428.10.1186/s12859-016-1276-2
96. Schlaepfer, D.R., J.B. Bradford, W.K. Lauenroth, S.M. Munson, B. Tietjen, et al. (2017). Climate change reduces extent of temperate drylands and intensifies drought in deep soils.Nat Commun(14196.10.1038/ncomms14196
97. Schrader, L. and J. Schmitz (2019). The impact of transposable elements in adaptive evolution.Mol Ecol(6): 1537-1549.10.1111/mec.14794
98. Sedlazeck, F.J., P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing.Nat Methods(6): 461-468.10.1038/s41592-018-0001-7
99. Sinclair, B.J., A.G. Gibbs and S.P. Roberts (2007). Gene transcription during exposure to, and recovery from, cold and desiccation stress in Drosophila melanogaster.Insect Mol Biol(4): 435-443.10.1111/j.1365-2583.2007.00739.x
100. Sørensen, J.G., M.M. Nielsen and V. Loeschcke (2007). Gene expression profile analysis of Drosophila melanogaster selected for resistance to environmental stressors.J Evol Biol(4): 1624-1636.10.1111/j.1420-9101.2007.01326.x
101. Stinziano, J.R., R.J. Sové, H.D. Rundle and B.J. Sinclair (2015). Rapid desiccation hardening changes the cuticular hydrocarbon profile of Drosophila melanogaster.Comp Biochem Physiol A Mol Integr Physiol(38-42.10.1016/j.cbpa.2014.11.004
102. Stott, P. (2016). CLIMATE CHANGE. How climate change affects extreme weather events.Science(6293): 1517-1518.10.1126/science.aaf7271
103. Strauch, O., J. Oestergraard, S. Hollmer and R.-U. Ehlers (2004). Genetic improvement of the desiccation tolerance of the entomopathogenic nematode *Heterorhabditis bacteriophora* through selective breeding Biological Control. **31**.
104. Sun, J.S., N.K. Larter, J.S. Chahda, D. Rioux, A. Gumaste, et al. (2018). Humidity response depends on the small soluble protein Obp59a in.Elife(10.7554/eLife.39249
105. Swanson, R.L., C.A. Baumgardner and B.W. Geer (1995). Very Long-Chain Fatty Acids Change the Ethanol Tolerance of Drosophila melanogaster Larvae.The Journal of Nutrition(3): 553-564.10.1093/jn/125.3.553
106. Szklarczyk, D., A.L. Gable, D. Lyon, A. Junge, S. Wyder, et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.Nucleic Acids Res(D1): D607-D613.10.1093/nar/gky1131
107. Takahashi, A., S.C. Tsaur, J.A. Coyne and C.I. Wu (2001). The nucleotide changes

governing cuticular hydrocarbon variation and their evolution in Drosophila melanogaster.Proc Natl Acad Sci U S A(7): 3920-3925.10.1073/pnas.061465098

108. Telonis-Scott, M., M. Gane, S. Degaris, C.M. Sgrò and A.A. Hoffmann (2012). High resolution mapping of candidate alleles for desiccation resistance in Drosophila melanogaster under selection.Mol Biol Evol(5): 1335-1351.10.1093/molbev/msr294
109. Telonis-Scott, M., K.M. Guthridge and A.A. Hoffmann (2006). A new set of laboratory-selected Drosophila melanogaster lines for the analysis of desiccation resistance: response to selection, physiology and correlated responses.J Exp Biol(Pt 10): 1837-1847.10.1242/jeb.02201
110. Telonis-Scott, M. and A.A. Hoffmann (2003). Isolation of a Drosophila melanogaster desiccation resistant mutant.J Insect Physiol(11): 1013-1020.10.1016/s0022-1910(03)00184-7
111. Telonis-Scott, M., C.M. Sgrò, A.A. Hoffmann and P.C. Griffin (2016). Cross-Study Comparison Reveals Common Genomic, Network, and Functional Signatures of Desiccation Resistance in Drosophila melanogaster.Mol Biol Evol(4): 1053-1067.10.1093/molbev/msv349
112. Terhzaz, S., P. Cabrero, J.H. Robben, J.C. Radford, B.D. Hudson, et al. (2012). Mechanism and function of Drosophila capa GPCR: a desiccation stress-responsive receptor with functional homology to human neuromedinU receptor.PLoS One(1): e29897.10.1371/journal.pone.0029897
113. Van't Hof, A.E., P. Campagne, D.J. Rigden, C.J. Yung, J. Lingley, et al. (2016). The industrial melanism mutation in British peppered moths is a transposable element.Nature(7605): 102-105.10.1038/nature17951
114. Waldvogel, A.M., B. Feldmeyer, G. Rolshausen, M. Exposito-Alonso, C. Rellstab, et al. (2020). Evolutionary genomics can improve prediction of species' responses to climate change.Evol Lett(1): 4-18.10.1002/evl3.154
115. Wheeler, N. and N. Watts (2018). Climate Change: From Science to Practice.Curr Environ Health Rep(1): 170-178.10.1007/s40572-018-0187-y
116. Zachariassen, K.E. (1996). The water conserving physiological compromise of desert insects European Journal of Entomology. **93:** 359-367.

SECTION 4

DISCUSSION

4. DISCUSSION

In the first section of the results (Section 3.3.), we aimed at identifying candidate transposons, that are possibly responsible for stress response due to a specific mechanism in six different stress conditions. By using a genome-wide approach, we described the contribution of TEs to the transcription factor binding site (TFBS) repertoire of different stress-regulatory networks in humans and *D. melanogaster*. We proved that TEs can contain functional TFBSs related to immune stress response, and we found that instead of introducing new TFBSs, they are adding extra binding sites to the already existing *cis-regulatory* networks (Figure 4.1).

In the second section (Section 3.4), we choose a stress specific approach, and studied the physiological and transcriptomic basis of the desiccation tolerance in European natural *D. melanogaster* strains. We contributed to the better understanding of the underlying changes in gene expression and the physiological mechanisms which make some strains more tolerant compared to others. We also investigated the role of TEs in desiccation stress response, and generated a list of TEs putatively affecting the expression of desiccation stress response genes (Figure 4.1).

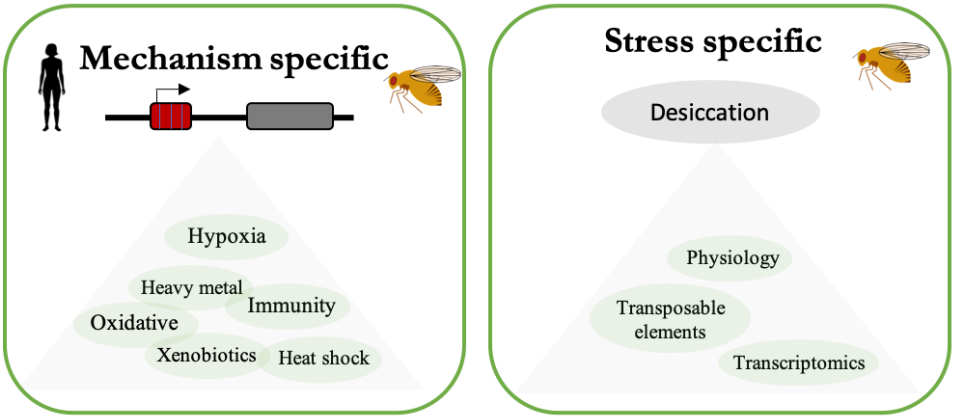


Figure 4.1. Strategies followed in the present thesis to describe the role of transposable elements in the eukaryotic stress response.

4.1. TEs can harbour TFBSs in different contexts and organisms

The role of TEs as gene regulators has been proposed in the 1950s and since then growing number of evidences strengthen this idea (Mcclintock 1956, Feschotte 2008, Bourque et al. 2018). There are several ways for TEs to act as gene regulators, such as by adding transcription factor binding sites (Feschotte 2008, Chuong et al. 2017). Even though in the first part of this thesis we focused on stress related TFBSs embedded within TEs, the body of knowledge related to the role of TEs

as enhancer repertoires in other biological processes is greatly increasing and cannot be neglected (Table 4.1).

4.1.1. TEs adding stress related transcription factor binding sites

In the first chapter of this thesis, we found that TEs are enriched among others for NF-κB transcription factor binding motifs, which is an important immune regulator in humans (Lowe et al. 2014). Similar to our results, a recent study in immune stressed human macrophages found TEs harbouring NF-κB binding motifs (Bogdan et al. 2020). Consistent with our findings, they described that repeats of the *THE1* family (*THE1B*, *THE1C*) are enriched in these binding motifs (Bogdan et al. 2020). However, we also found that apart from *THE1*, some families with the highest copy number such as repeats belonging to *MIR*, *L2* and *Alu* groups are enriched in NF-κB binding sites. The members of these families are recurrently linked to human enhancers, reflecting their regulatory potential in different contexts, including immune stress response (Lynch et al. 2011, Lynch et al. 2015, Chuong et al. 2016, Cao et al. 2019). The growing body of knowledge associating TEs with enhancers in immune stress conditions in humans suggests that TEs are important regulators in this process (Bourque et al. 2008, Sundaram et al. 2014, Chuong et al. 2016, Trizzino et al. 2017, Nikitin et al. 2018, Bogdan et al. 2020, Macchietto et al. 2020, Ye et al. 2020). However, to our knowledge, there are no studies tackling the role of TEs as enhancers in human oxidative and hypoxia stress response. Thus, for the first time we described the enrichment of transposable elements for *NFEL2L* binding sites and *CREB1* binding motifs in humans, which are important gene regulators in oxidative stress and hypoxia stress conditions (Section 3.3, Villanueva-Cañas, Horvath et al 2019) (Espinosa-Diez et al. 2015).

Apart from humans, we also looked for TE-derived TFBSs in *D. melanogaster* and found enrichment of TEs in *HSF*, *tango*, *caudal*, and *dorsal* binding sites (Section 3.3, Villanueva-Cañas, Horvath et al 2019). These are binding sites for transcription factors involved in the regulation of heat, oxidative, hypoxia and immune stress response genes. These findings suggest, that TEs can contribute with a representative fraction of binding sites to these stress regulatory networks on a genome-wide scale. Interestingly, the role of TEs in immune stress response has already been shown in *D. melanogaster* (Magwire et al. 2011, Ullastres et al. 2019). Indeed, Ullastres and co-workers showed that individual TEs can add TFBSs and thus regulate the expression of immune stress response genes, however the contribution of TEs to the genome wide immune related TFBSs repertoire was not uncovered. In another recent study, they used the data generated in our work to check the contribution of TEs to xenobiotic stress response. They focused on cap-n-collar (*cnc)* binding sites inside TEs, which is the main transcriptional regulator of xenobiotic detoxification. They found

that several insertions nearby xenobiotic responsive genes contain binding sites for *cnc*, and suggest that TEs could be responsible for the changes in gene expression in these conditions (Salces-Ortiz et al. 2020). Apart from these examples, in *D. melanogaster* the contribution of TEs to specific stress response networks by adding TFBSs has not been studied, therefore our study is unprecedented and adds a very important body of knowledge to the topic (Section 3.3, Villanueva-Cañas, Horvath et al 2019).

There are also several examples of TEs adding stress related TFBSs in other organisms such as mice, nematodes and plants (Table 4.1) (Makarevitch et al. 2015, Ito et al. 2017, Garrigues et al. 2019, Ye et al. 2020). In a recent example in *C. elegans* they found that the majority of heat shock elements (HSEs) are harboured by *Helitron* TEs and that these TEs have an effect on the expression of nearby genes in heat stress conditions (Garrigues et al. 2019).

4.1.2. Are TEs more prone to be involved in immune gene regulation compared to other stresses?

As mentioned before, there are several studies in different species and biological conditions uncovering the co-option of TEs to regulatory regions, however, it is still not clear whether different regulatory networks are similarly affected by TEs (Ivancevic and Chuong 2020). For instance, in case of mammals it is known that TE-orchestrated regulatory activity can be find in nearly all cell types and tissues. However, the relative contribution of TEs to the genome-wide regulatory landscape of each tissue is highly different (Sundaram et al. 2014, Jang et al. 2019, Pehrsson et al. 2019). Some evidence suggests that certain cell types and processes are particularly more prone to TE-mediated gene regulation, such as embryonic stem cells, placental cells and immune cells (Wang et al. 2015, Chuong et al. 2016, Römer et al. 2017, Dunn-Fletcher et al. 2018, Fuentes et al. 2018, Bogdan et al. 2020).

It is known that the adaptation of organisms to immune related stresses is very relevant, and immune response has previously been reported to be important for local adaptation not only in mammals, including humans, but also in *D. melanogaster* (Gobert et al. 2003, Fumagalli et al. 2011, Kolaczkowski et al. 2011, Magwire et al. 2011, Fabian et al. 2012, Fan et al. 2016, Quach and Quintana-Murci 2017). The mechanisms by which TEs add TFBSs could be particularly relevant fot the evolution of the immune system, which needs to adapt rapidly in order to be able to respond to a lot of different stimuli. There are several examples of TEs adding immune related TFBSs in mammals (Chuong et al. 2016, Trizzino et al. 2017, Bogdan et al. 2020, Ye et al. 2020). In a recent example Ye et al (2020) showed that in mice TEs are more likely to be involved in gene regulation in immune related cells, compared to other cell types (Ye et al. 2020). They showed that

endogenous retroviruses (ERVs) bare immune-related TFBSs and that enhancers specific to immune tissues are more prone to a putative TE co-option, compared to enhancers specific to other tissues. They found that a higher proportion of immune enhancers are TE derived, consistent with the idea that immune tissues are more prone to co-opting TEs as regulatory elements (Ye et al. 2020). Protein coding immune genes are among the most rapidly evolving genes in the genome, reflecting the constant need to adapt against new and evolving pathogens (Obbard et al. 2009, Daugherty and Malik 2012). Thus, Ye and co-workers suggest that since active or recently active TEs are a major source of genetic polymorphism, they may facilitate rapid adaptive evolution of immune responses at the gene regulatory level (Ye et al. 2020). The biased TE representation in immune-associated genomic regions the authors explain by proposing a model in which a high density of TEs in immune-associated genomic regions favored selection and regulatory cooption of functional TE sequences. This may have accelerated evolution of immune enhancers and acquisition of robustness of immune functions. The existing examples suggest, that the role of TEs in mammalian immune stress response is bigger than in other stress types, however, we need to take this with caution. For instance, in our study in section 3.3., we found that TEs in humans are not only enriched in TFBSs related to immune stress response, but also in TFBSs related to oxidative and hypoxia stress. This might suggest, that the existing studies in mammals are biased towards immune stress response, and further studies are necessary to systematically analyse the role of TEs in different stress regulatory landscapes.

Apart from human and mice it has also been shown in *D. melanogaster*, that the contribution of TEs to immune related regulatory networks is important (Ullastres et al. 2019). However, apart from immunity stress, the role of TEs seems to be important in heat, oxidative, xenobiotic and hypoxia stress response, some of which were described in our study (Section 3.3., Villanueva-Cañas, Horvath et al 2019) and also in others (Salces-Ortiz et al. 2020). These results suggest, that unlike in mammals, in Drosophila the contribution of TEs to stress regulatory networks has been shown in a great variety of stress responses. However, it is important to emphasize, that in the above-mentioned studies (including ours), the authors only validated experimentally the TEs which bear immune related binding sites. Thus, in order to be able to say that transposons have a biological role in other stresses by adding TFBSs, further experimental validation is necessary (see Section 4.2).

Overall, the existing data suggests, that in Drosophila and other non mammal species TEs contribute to different stress regulatory networks. However, we have to take into account that in mammals the contribution of TEs might be biased towards immune-related stress response, nonetheless, further studies are needed in order to be able to say this with confidence.

Condition		Organism	Reference
Stress	**Immunity**	Human	Bogdan et al 2020 Nikitin et al 2018
		D.melanogaster	Ullastres et al 2019
		Mice	Ye et al 2020
	Heat-shock	C. *elegans*	Carrigues et al 2019
	Xenobiotics	*D. melanogaster*	Sales-Ortiz et al 2020
Development		Human	Pontis et al 2019 Nishihara 2019
		D. simulans	Loreto et al 2018
		Arabidopsis thaliana	Batista et al 2019
		Mice	Todd et al 2019 Sakashita et al 2020 Thybert et al 2018
Metabolic pathway		Human	Kellner et al 2019
Disease		Human	Jang et al 2019 Jiang & Upton 2019

Table 4.1. Studies investigating the role of TEs in different gene regulatory networks. This is a non-exhaustive list of studies, since it contains the studies from 2018 onwards.

4.1.3. TEs adding development related TFBSs

Apart from being important in regulating stress response genes, TEs have been proved to play a key role during development by spreading TFBSs in several organisms, including humans (Kunarso et al. 2010, Lynch et al. 2011, Schmidt et al. 2012, Batut et al. 2013, Pontis et al. 2019). In a recent study, Nishihara (2019) found that TEs can have a role in the development of mammary glands. He identified that thousands of TFBSs for estrogen receptor α (ER α) and three related pioneer factors (FoxA1, GATA3 and AP2 γ) that are essential regulators of mammary gland development, arose from a spreading of the binding motifs by retrotransposons (Nishihara 2019). These TE-derived functional elements primarily serve as distal enhancers and are enriched around genes that are associated with mammalian gland morphogenesis (Nishihara 2019). Apart from humans, the role of TEs as key regulators of developmental genes have been also described in *Drosophila simulans* (Loreto et al. 2018). Loreto and co-workers identified a *hobo* element which adds binding sites for some master developmental genes and is able to produce remarkable changes in development (Loreto et al. 2018). The phenomenon of TEs adding binding sites to developmental genes has also been described in other organisms such as plants and mice (Table 4.1) (Thybert et al. 2018, Batista et al. 2019, Todd et al. 2019, Sakashita et al. 2020). A study in mouse identified a small fraction of TEs that seem to act as enhancers in early development in embryonic and trophoblast stem cells (Todd et al. 2019).

4.1.4. TEs adding metabolic pathway and disease related TFBSs

Several studies have found TFBSs related to different metabolic pathways and diseases embedded within TEs (Table 4.2) (Jang et al. 2019, Jiang and Upton 2019). For instance, in one of them they looked for TE-derived TFBSs in the promoter of genes transcribed by polymerase II in the human genome (Kellner and Makałowski 2019). It was found that more than 6% of active TFBSs in these regions are located in TE-originated sequences. Moreover, they found that the genes which have TFBSs spread in their promoter region by TEs are mostly enriched in metabolic and disease pathways such as Huntington's disease and different type of cancers (Kellner and Makałowski 2019). TEs have also been associated with driving the expression of oncogenes (Jang et al. 2019, Jiang and Upton 2019). In a recent work genome-wide binding sites for TFs with oncogenic capacity in cancer (C/EBPβ, E2F1 and MYC) were mapped, using MCF7 breast cancer cells. They found that up to 55% of the identified TFBSs are overlapping TEs, and that there are 268 TE subfamilies enriched in these TFBSs, 30% of which belong to the LTR class (Jiang and Upton 2019).

Overall, the contribution of TEs to regulatory networks is widespread and relevant in a great variety of conditions such as stress response, development or cancer, and has been described in different organisms from nematodes to humans. As it was fittingly proposed in the early years, even though in several cases TEs are identified as non-functional, in a representative number of examples TEs form a functional part of gene regulatory networks, even though it is not always beneficial for the host.

4.2. The importance of functional experiments when evaluating the role of TEs as gene regulatory elements

In Section 3.3. we performed enhancer reporter assays in *D. melanogaster* with 11 TE insertions, and found that four TEs are acting as enhancers in immune stress conditions (*FBti0019386*, *FBti0019985, FBti0019082*, *FBti0019453*). Reporter assays are an effective tool as a first step in validating TE insertions which contain TFBSs. However, TEs have been shown to affect gene expression through a variety of mechanisms, which we cannot uncover by doing enhancer assays (Elbarbary et al. 2016, Chuong et al. 2017). Moreover, it would be interesting to mutate the TFBSs inside the TEs and generate mutants flies with this construct. These future experiments would allow us to identify if the TEs which acted as enhancers in the previous studies are still having an effect on the expression of the reporter gene with the binding sites mutated. Even if enhancer assays give negative results, we have to consider that it might not mean that the region of interest is not acting as an enhancer. For instance, it is known that the genetic background plays a very important role in the regulation of gene activity, thus it is possible that the negative results are due

to the lack of the surrounding regions in the original genome (Lelli et al. 2012). In order to be able to investigate the role of a TE insertion in the genome where it was described, we would need to do molecular manipulations such as CRISPR/Cas9. Even though up to date there are very few studies that use this technique to excise TEs, there are some examples in several organisms such as Drosophila, plants, human cells or mice (Chuong et al. 2016, Ding et al. 2016, Fuentes et al. 2018, Jang et al. 2019, Saika et al. 2019, Todd et al. 2019). Note that only some of the examples are aiming to validate the role of TEs containing TFBSs as enhancers (Chuong et al. 2016, Jang et al. 2019, Todd et al. 2019).

In *Drosophila simulans*, the deletion of a *Shedller* element using CRISPR/Cas system in natural strains showed that the absence of this TE has an effect on the phenotype of the courtship song of the flies (Ding et al. 2016). Also, using this system, Saika and co-workers reported for the first time the deletion of a transposable element *(Tos17)* in rice (Saika et al. 2019). The authors suggest, that this strategy could be applied to other plant species also, as a rapid alternative breeding method (Saika et al. 2019).

In mammals, there are recent examples of successfully deleting the members of a TE class and also of deleting individual insertions (Fuentes et al. 2018, Todd et al. 2019). Fuentes and co-workers inactivated the vast majority of *LTR5Hs* elements by CRISPRi technique and they found that a large proportion of these elements play significant roles in the regulation of nearly 300 genes in a human embryonal carcinoma cell line (Fuentes et al. 2018). The method applied by Fuentes et al could gain a lot of significance nowadays, since there are a lot of genome wide studies looking at the gene regulatory role of TEs, as the work previously exposed in section 3.3. and many others (Jiao and Deng 2007, Oki et al. 2008, González et al. 2010, Khalkhali-Evrigh et al. 2019, Rech et al. 2019, Mascagni et al. 2020, Salces-Ortiz et al. 2020). For instance, if we could inactivate most of the members of a specific TE family, we could have a comprehensive view of the role of TEs in adaptation and stress response at the TE family level, instead of just identifying unique TE insertions with adaptive roles.

In a recent study in mice Todd and co-workers performed CRISPR-mediated genetic excision of individual TEs and they identified that only a small set of these TEs influence the regulation of gene expression in embryonic and trophoblast stem cells (Todd et al. 2019). The evidence that only a small portion of TEs has been validated in Todd et al (2019) and also in our study, underpins the fact that some of the binding events will not be functional, thus it is crucial to validate them experimentally in order to confirm enhancer function (Todd et al. 2019). However, after validating the role of TEs as gene regulators using the CRISPR/Cas9 system, the question if these TE insertions have an impact on the phenotype and fitness of the host still remains open, thus

conducting further phenotypic experiments with the generated mutants is essential. Even though the CRIPR/Cas9 method is one of the best ways to validate the role of TEs in gene regulation and also their effect on the phenotype of the host, it also has its limitations such as off-target events which can generate unexpected mutations in the genomes.

All the previously mentioned examples for validating TEs as gene regulators are limited to enhancers in the near proximity of the genes, thus the distant enhancers stay hidden. In order to be able to find TEs acting as distal enhancers, we should investigate the chromatin conformation using methods such as the so-called 4C methodologies. TEs which act as gene regulators may also act as enhancers from further away, and they are suggested to have a role in shaping the chromatin (Glinsky 2018, Cao et al. 2019, Choudhary et al. 2020, Diehl et al. 2020). Thus, being able to find TEs which are acting as distal enhancers would broaden the amount of candidate TEs putatively acting as enhancers. Even though our knowledge of biochemical activity of the regulatory landscape has increased in the last decade, the insight about the functional roles of these activities remain largely uncovered (Todd et al. 2019). Thus, one of the biggest challenges in the future is finding the association between the biochemical changes and their regulatory function in the genome and its phenotypic consequences.

4.3. Novel gene regulators: tRNA derived small RNAs

Apart from transposable elements, other non-coding sequences, such as small RNAs (sncRNA) has been described to regulate gene expression. Some groups of non-coding RNAs (ncRNA) such as microRNAs (miRNAs), small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs) have been intensively studied for decades and their mechanisms are well described (Ghildiyal and Zamore 2009, Jacquier 2009). A relatively new group of small ncRNAs, derived from transfer RNAs (tRNA), the so called tRNA derived small RNAs (tsRNA) were described in the 1970s, however back then they attracted little attention. Two different classes of these molecules have been reported. One class are the tRNA halves, which are between 30–36 nucleotides (nt) long. The other class consists of tRNA-derived RNA fragments (tRFs), which are between 18–20 nt long (Dhahbi 2015, Martinez et al. 2017). Note that for the rest of this section, I will refer to tRNA halves and tRFs as tsRNAs. tsRNAs have been identified in various species ranging from prokaryotes to eukaryotes, considerably expanding the small RNA repertoire (Cole et al. 2009, Hsieh et al. 2009, Garcia-Silva et al. 2010, Couvillion et al. 2012, Gebetsberger et al. 2012, Peng et al. 2012, Kumar et al. 2014, Shigematsu et al. 2014, Keam and Hutvagner 2015, Kumar et al. 2015, Kumar et al. 2016, Li et al. 2018). Moreover, they have been found to be very conserved across the tree of life (Martinez et al. 2017, Dou et al. 2019).

Until recently tsRNAs were thought to be the non-functional by-products of tRNA synthesis. However, with the advent of next-generation sequencing techniques it is becoming clear that these tsRNAs are important gene regulators. Some tdRNAs have been shown to silence genes through an Argonaut (AGO) independent manner, by interacting with the general translational machinery to inhibit translation (Gebetsberger et al. 2012, Ivanov et al. 2014, Keam and Hutvagner 2015, Gebetsberger et al. 2017, Lyons et al. 2017), while others has been described to bind to AGO proteins in order to silence genes in a sequence specific manner (Yeung et al. 2009, Couvillion et al. 2010, Couvillion et al. 2012, Loss-Morais et al. 2013, Maute et al. 2013, Nie et al. 2013, Hirano et al. 2014, Karaiskos et al. 2015). tsRNAs have been showed to play regulatory roles in several biological processes including transposable element silencing and different stress responses in several organisms from nematodes to humans (Thompson et al. 2008, Hsieh et al. 2009, Thompson and Parker 2009, Yamasaki et al. 2009, Emara et al. 2010, Peng et al. 2012, Maute et al. 2013, Nie et al. 2013, Goodarzi et al. 2015, Pekarsky et al. 2016, Sharma et al. 2016, Martinez et al. 2017, Schorn et al. 2017, Liu et al. 2018, Luo et al. 2018, Fricker et al. 2019, Angelova et al. 2020).

In a recent work, Liu and co-workers used human alveolar type II-like epithelial cells (A549), and checked the change of global sncRNAs profiles. They found that after heavy metal (arsenite) treatment, the amount of tsRNAs increased by 1.6-fold. They also described a specific tsRNA (tRF5-AlaCGC), which leads to enhanced secretion of an inflammatory cytokine IL-8 by activating the p65 transcription factor (Liu et al. 2018). These results suggest the potential role of tsRNAs in heavy metal stress response in humans.

Another recent example can be found in Drosophila, where Luo and co-workers investigated the role of tsRNAs in modulating Drosophila gene expression in control and starvation conditions in S2 cells. They show that tsRNAs inhibit the translational efficiency of key components of the general translation machinery, and that this repression depends on AGO2. These findings suggest that in Drosophila tsRNAs inhibit specific targets in an RNAi-like manner. Moreover, the authors describe that tsRNAs participate in cellular starvation response by regulating translation of specific and general mRNAs (Luo et al. 2018).

In Section 3.4. of this thesis we did not find obvious evidence for the role of TEs in desiccation stress response. Thus, we must look for an additional explanation that is behind the regulation of desiccation stress responsive genes. Based on the existing examples in Drosophila related to the role of tsRNAs in stress response, it is right to hypothesize that they might be important players in desiccation stress response (Luo et al. 2018, Angelova et al. 2020). In Section 3.4., when looking for genes which differ in expression after desiccation stress, we found that most of the genes which responded were down-regulated. This suggests that the flies in order to survive apart

from utilizing other physiological mechanisms, lower down the transcriptional machinery in general. Indeed, repression of global protein synthesis has been shown to be important in cell survival in stress conditions (Racle et al. 2013, Liu and Qian 2014). Moreover, it has also been shown that tsRNAs are able to lower down the abundance of translation products and as a consequence, slowing down protein synthesis. Through this mechanism tsRNAs could be important players in desiccation stress conditions, by cleaving the translation products of genes. Overall, the newly emerged and constantly growing field of tsRNAs harbors a great potential for future studies aiming at understanding the stress regulation of different organisms including humans and *D. melanogaster*.

4.4. The role of plastic responses in insect survival due to climate change

As discussed in Section 3.4., climate change is dramatically altering environmental traits such as thermal extremes and water availability, causing severe stress conditions (Sgrò et al. 2016, Kellermann and Heerwaarden 2019). In order to survive, organisms can either migrate to suitable habitats, or go through adaptive evolution or phenotypic plasticity (Fox et al. 2019). Phenotypic plasticity is the ability of one genotype to generate multiple phenotypes, induced by rapidly changing environmental conditions (Sgrò et al. 2016, Waldvogel et al. 2020). Merilä and Hendry (2014) draws the attention to an important point, that temporal changes in climate-related traits of organisms were until recently often interpreted as adaptive evolution without actually proving that phenotypic plasticity could be excluded (Merilä and Hendry 2014). While adaptive evolution is relevant for long term survival, it is argued that since plastic responses occur rapidly, they might help organisms to cope with rapid environmental changes, including global climate change, and thus be beneficial in the short term (Chevin et al. 2010, Bonamour et al. 2019, Kellermann and Heerwaarden 2019). An increasing number of studies suggests that the role of plasticity in shaping phenotypic responses to climate change is larger than previously thought (Charmantier et al. 2008, Gienapp et al. 2008, Hendry et al. 2008, Merilä and Hendry 2014, Chevin and Hoffmann 2017). Interestingly, evidence shows that the level of plasticity might be more correlated with environmental variability than with the overall level of stress, or exposure to stressful environments. For instance, since changes in temperature can happen rapidly and in a short period of time, it is more likely that species will respond plastically to these changes (Stone et al. 2020). Indeed, plastic responses related to either low or high temperatures has been reported in several insect species (Czajka and Lee 1990, Hoffmann et al. 2003, Ayrinhac et al. 2004, Nyamukondiwa et al. 2010, Overgaard et al. 2011, Gunderson and Stillman 2015, Stone et al. 2020, Xue and Ma 2020). Even though plastic changes in heat tolerance have been shown, it seems that the capacity

of many species to shift their critical thermal maximum (CT_{Max}) via plasticity is small (Overgaard et al. 2011, Gunderson and Stillman 2015, Sørensen et al. 2016, Kellermann and Sgrò 2018). For instance, when subjecting Drosophila to heat acclimation (long term) and hardening (short term) treatments, that are known to induce plasticity, only a slight increase (0.6–1.0 °C) in heat resistance has been described (Van Heerwaarden et al. 2016). The authors also found, that basal CT_{Max} increased, while hardening capacity decreased. They argue that the limited potential for phenotypic plasticity to increase CT_{Max} could be explained by the tolerance-plasticity trade-off hypothesis, which predicts that individuals already adapted to high temperatures have limited potential to improve their heat tolerance via phenotypic plasticity (Van Heerwaarden et al. 2016, Van Heerwaarden and Kellermann 2020). In another study the authors exposed Drosophila flies for short term cold conditions to induce phenotypic plasticity. They found that cold hardening is a highly plastic trait that does not exhibit classic signatures of adaptive tracking (Stone et al. 2020).

Apart from changes in temperature, alterations in humidity are also threatening terrestrial ectotherms and the response to these changes have also been shown to be plastic (Kellermann et al. 2012, Siepielski et al. 2017, Kellermann et al. 2018). For instance, rapid desiccation hardening in adult *D. melanogaster*, as well as developing larvae at lower relative humidity, increases resistance to desiccation (Hoffmann 1990, Hoffmann 1991, Bubliy et al. 2012, Parkash et al. 2012, Aggarwal et al. 2013, Bubliy et al. 2013). Similar to heat stress, for some Drosophila species a trade-off between basal desiccation resistance and plasticity has been observed, where species with the highest basal tolerance also have the lowest capacity to respond plastically to the environment (Kellermann et al. 2018).

Kellermann et al (2020) using a subset of the dataset published in Kellermann et al (2018) studied the role of plasticity in shaping species distribution. They found, that even though in the widespread Drosophila species such as *D. melanogaster* the role of plasticity might be higher than in the restricted species, overall, plasticity for desiccation resistance will have a limited impact on distribution and survival of Drosophila species under climate change (Kellermann et al 2020).

Overall, while it is unlikely that the small plastic responses in heat or desiccation resistance will contribute much to climate change responses across Drosophila (Schiffer et al. 2013, Johnstone et al. 2017, Kellermann et al. 2018, Kellermann and Sgrò 2018), the contribution of plastic responses for cold resistance can be substantial (Ayrinhac et al. 2004, Colinet and Hoffmann 2012, Stone et al. 2020). This underpins the importance of understanding the mechanisms behind the differences in basal resistance to heat and desiccation among individuals and populations.

SECTION 5

CONCLUSIONS

5. CONCLUSIONS

The main conclusions of this thesis are the following:

1. In humans, the contribution of TEs is relevant for immunity, hypoxia and oxidative stress regulatory networks, since ~60% of *NF-κB and CREB1* transcription factor binding motifs and 86% of *NFE2L2* binding sites are located inside TEs.

2. In *D. melanogaster*, the contribution of TEs to heavy-metal, immunity, oxidative, hypoxia and heat shock stress regulatory networks is substantial. We found that TEs are enriched for *MTF-1* binding motifs (7.42%) and that 16.58-36.57% of *caudal*, *dorsal*, *HSF* and *tango* binding sites reside within TEs.

3. TEs containing three or more transcription factor binding sites are more likely to be functional. We found, that TEs which have three or more transcription factor binding sites and are present at high frequencies and are enriched nearby stress related genes.

4. In *D. melanogaster*, TEs affect the level of expression of genes that were already part of the cellular stress response, rather than recruiting new genes to stress regulatory networks. We found that four out of the six TEs with confirmed functional role added TFBSs that were already present in the promoter region of the nearby gene.

5. In *D. melanogaster* diverse families contribute to the transcription factor binding sites repertoire of stress response genes, since several families were enriched for TFBSs and each one of the validated TE belongs to a different family.

6. The variation in desiccation tolerance of natural European strains is similar to those of other continents. Desiccation tolerance of European strains correlates with the interaction of altitude and evaporation. Moreover, the strains belonging to Cold semi-arid climates are more tolerant compared to strains from Hot summer mediterranean climates.

7. The combination of water content, rate of water loss and the composition of the cuticle is relevant for *D. melanogaster* survival in desiccation stress conditions. Tolerant strains have a lower water content, and lower more their respiration rate during desiccation stress compared to the sensitive ones. Moreover, the tolerant strains have a more favorable cuticle composition, which suggests that they also loose less water through the cuticle.

8. Lowering down the metabolism and sensing the stress is relevant for survival in desiccation stress conditions. We found that the genes related to metabolic processes are down-regulated, while the genes related to response to stimulus and environmental sensing are up-regulated after desiccation stress conditions only in the tolerant strains.

9. TE families are more active in the sensitive strains, however most of the differentially expressed TE families are down-regulated. Moreover, TEs might affect the expression of desiccation stress responsive genes in a strain specific manner. We suggest that TEs might not be the main drivers of desiccation stress response, however their contribution might be relevant prior to stress.

SECTION 6

REFERENCES

References

Acuña-Galindo, M.A., R.E. Mason, N.K. Subramanian and D.B. Hays (2015). Meta-Analysis of Wheat QTL Regions Associated with Adaptation to Drought and Heat Stress. Crop Sci. **55:** 477-492.

Aggarwal, D.D., P. Ranga, B. Kalra, R. Parkash, E. Rashkovetsky, et al. (2013). Rapid effects of humidity acclimation on stress resistance in Drosophila melanogaster.Comp Biochem Physiol A Mol Integr Physiol(1): 81-90.10.1016/j.cbpa.2013.05.012

Agrawal, A., Q.M. Eastman and D.G. Schatz (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.Nature(6695): 744-751.10.1038/29457

Ambrozová, K., T. Mandáková, P. Bures, P. Neumann, I.J. Leitch, et al. (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of Fritillaria lilies.Ann Bot(2): 255-268.10.1093/aob/mcq235

Angelova, M.T., D.G. Dimitrova, B. Da Silva, V. Marchand, C. Jacquier, et al. (2020). tRNA 2'-O-methylation by a duo of TRM7/FTSJ1 proteins modulates small RNA silencing in Drosophila.Nucleic Acids Res(4): 2050-2072.10.1093/nar/gkaa002

Arguello, J.R., S. Laurent and A.G. Clark (2019). Demographic History of the Human Commensal Drosophila melanogaster.Genome Biol Evol(3): 844-854.10.1093/gbe/evz022

Arkhipova, I.R. (2017). Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories.Mob DNA(19.10.1186/s13100-017-0103-2

Ayrinhac, A., V. Debat, P. Gibert, A.-G. Kister, H. Legout, et al. (2004). Cold adaptation in geographical populations of Drosophila melanogaster: phenotypic plasticity is more important than genetic variability.Functional Ecology(5): 700-706.10.1111/j.0269-8463.2004.00904.x

Bao, W., K.K. Kojima and O. Kohany (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes.Mob DNA(11.10.1186/s13100-015-0041-9

Batista, R.A., J. Moreno-Romero, Y. Qiu, J. Van Boven, J. Santos-González, et al. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons.Elife(10.7554/eLife.50541

Batut, P., A. Dobin, C. Plessy, P. Carninci and T.R. Gingeras (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression.Genome Res(1): 169-180.10.1101/gr.139618.112

Baucom, R.S. (2019). Evolutionary and ecological insights from herbicide-resistant weeds: what have we learned about plant adaptation, and what is left to uncover?New Phytol(1): 68-82.10.1111/nph.15723

Beall, C.M. (2006). Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia.Integr Comp Biol(1): 18-24.10.1093/icb/icj004

Bejerano, G., C.B. Lowe, N. Ahituv, B. King, A. Siepel, et al. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon.Nature(7089): 87-90.10.1038/nature04696

Bennetzen, J.L. and H. Wang (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes.Annu Rev Plant Biol(505-530.10.1146/annurev-arplant-050213-035811

Benoit, M., H.G. Drost, M. Catoni, Q. Gouil, S. Lopez-Gomollon, et al. (2019). Environmental and epigenetic regulation of Rider retrotransposons in tomato.PLoS Genet(9): e1008370.10.1371/journal.pgen.1008370

Bickler, P.E. and L.T. Buck (2007). Hypoxia tolerance in reptiles, amphibians, and fishes: life with variable oxygen availability.Annu Rev Physiol(145-170.10.1146/annurev.physiol.69.031905.162529

Bijlsma, K. and V. Loeschcke (1997). Environmental Stress, Adaptation and Evolution, Birkhauser Verlag, Basel, CH.

Bingham, P.M., M.G. Kidwell and G.M. Rubin (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family.Cell(3): 995-1004.10.1016/0092-8674(82)90463-9

Boeke, J.D., D.J. Garfinkel, C.A. Styles and G.R. Fink (1985). Ty elements transpose through an RNA intermediate.Cell(3): 491-500.10.1016/0092-8674(85)90197-7

Bogdan, L., L. Barreiro and G. Bourque (2020). Transposable elements have contributed human regulatory regions that are activated upon bacterial infection.Philos Trans R Soc Lond B Biol Sci(1795): 20190332.10.1098/rstb.2019.0332

Bonamour, S., L.M. Chevin, A. Charmantier and C. Teplitsky (2019). Phenotypic plasticity in response to climate change: the importance of cue variation.Philos Trans R Soc Lond B Biol Sci(1768): 20180178.10.1098/rstb.2018.0178

Bourque, G., K.H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, et al. (2018). Ten things you should know about transposable elements.Genome Biol(1): 199.10.1186/s13059-018-1577-z

Bourque, G., B. Leong, V.B. Vega, X. Chen, Y.L. Lee, et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements.Genome Res(11): 1752-1762.10.1101/gr.080663.108

Brandt, J., S. Schrauth, A.M. Veith, A. Froschauer, T. Haneke, et al. (2005). Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals.Gene(1): 101-111.10.1016/j.gene.2004.11.022

Britten, R.J. (1996). DNA sequence insertion and evolutionary variation in gene regulation.Proc Natl Acad Sci U S A(18): 9374-9377.10.1073/pnas.93.18.9374

Britten, R.J. and E.H. Davidson (1969). Gene regulation for higher cells: a theory.Science(3891): 349-357.10.1126/science.165.3891.349

Brönner, G., H. Taubert and H. Jäckle (1995). Mesoderm-specific B104 expression in the Drosophila embryo is mediated by internal cis-acting elements of the transposon.Chromosoma(10): 669-675.10.1007/BF00344227

Brosius, J. (1991). Retroposons--seeds of evolution.Science(4995): 753.10.1126/science.1990437

Bryan-Walker, K., T.L.F. Leung and R. Poulin (2007). Local adaptation of immunity against a trematode parasite in marine amphipod populations. Marine Biology, Springer-Verlag. **152:** 687-695.

Bubliy, O.A., T.N. Kristensen, V. Kellermann and V. Loeschcke (2012). Plastic responses to four environmental stresses and cross-resistance in a laboratory population of Drosophila melanogaster.Functional Ecology(1): 245-253.10.1111/j.1365-2435.2011.01928.x

Bubliy, O.A., T.N. Kristensen and V. Loeschcke (2013). Stress-induced plastic responses in Drosophila simulans following exposure to combinations of temperature and humidity levels.J Exp Biol(Pt 24): 4601-4607.10.1242/jeb.092502

Bubliy, O.A. and V. Loeschcke (2005). Correlated responses to selection for stress resistance and longevity in a laboratory population of Drosophila melanogaster.J Evol Biol(4): 789-803.10.1111/j.1420-9101.2005.00928.x

Buffry, A.D., C.C. Mendes and A.P. Mcgregor (2016). The Functionality and Evolution of Eukaryotic Transcriptional Enhancers.Adv Genet(143-206.10.1016/bs.adgen.2016.08.004

Cai, H.N. and M. Levine (1997). The gypsy insulator can function as a promoter-specific silencer in the Drosophila embryo.EMBO J(7): 1732-1741.10.1093/emboj/16.7.1732

Cao, Y., G. Chen, G. Wu, X. Zhang, J. Mcdermott, et al. (2019). Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions.Genome Res(1): 40-52.10.1101/gr.235747.118

Cappucci, U., F. Noro, A.M. Casale, L. Fanti, M. Berloco, et al. (2019). The Hsp70 chaperone is a major player in stress-induced transposable element activation.Proc Natl Acad Sci U S A(36): 17943-17950.10.1073/pnas.1903936116

Capy, P., G. Gasperi, C. Biémont and C. Bazin (2000). Stress and transposable elements: co-evolution or useful parasites?Heredity (Edinb)(101-106.10.1046/j.1365-2540.2000.00751.x

Carroll, S.B., J.K. Greiner and S.D. Weatherbee (2001). From DNA to diversity: molecular genetics and the evolution of animal design Malden, Mass, Blackwell Science.

Casacuberta, E. and J. González (2013). The impact of transposable elements in environmental adaptation.Mol Ecol(6): 1503-1517.10.1111/mec.12170

Cavalcanti Luna, M.A., E. Rodrigues Vieira, K. Okada, G.M. Campos-Takaki and A.E. Do Nascimento (2015). Copper-induced adaptation, oxidative stress and its tolerance in Aspergillus niger UCP1261.Electronic Journal of Biotechnology(418-427

Chakraborty, M., N.W. Vankuren, R. Zhao, X. Zhang, S. Kalsow, et al. (2018). Hidden genetic variation shapes the structure of functional elements in Drosophila.Nat Genet(1): 20-25.10.1038/s41588-017-0010-y

Charmantier, A., R.H. Mccleery, L.R. Cole, C. Perrins, L.E. Kruuk, et al. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population.Science(5877): 800-803.10.1126/science.1157174

Chénais, B., A. Caruso, S. Hiard and N. Casse (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments.Gene(1): 7-15.10.1016/j.gene.2012.07.042

Chevin, L.M. and A.A. Hoffmann (2017). Evolution of phenotypic plasticity in extreme environments.Philos Trans R Soc Lond B Biol Sci(1723).10.1098/rstb.2016.0138

Chevin, L.M., R. Lande and G.M. Mace (2010). Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory.PLoS Biol(4): e1000357.10.1371/journal.pbio.1000357

Choudhary, M.N., R.Z. Friedman, J.T. Wang, H.S. Jang, X. Zhuo, et al. (2020). Co-opted transposons help perpetuate conserved higher-order chromosomal structures.Genome Biol(1): 16.10.1186/s13059-019-1916-8

Chown, S.L. (2002). Respiratory water loss in insects.Comp Biochem Physiol A Mol Integr Physiol(3): 791-804.10.1016/s1095-6433(02)00200-3

Chown, S.L., J.G. Sørensen and J.S. Terblanche (2011). Water loss in insects: an environmental change perspective.J Insect Physiol(8): 1070-1084.10.1016/j.jinsphys.2011.05.004

Chuong, E.B., N.C. Elde and C. Feschotte (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses.Science(6277): 1083-1087.10.1126/science.aad5497

Chuong, E.B., N.C. Elde and C. Feschotte (2017). Regulatory activities of transposable elements: from conflicts to benefits.Nat Rev Genet(2): 71-86.10.1038/nrg.2016.139

Clemmons, A.W. and S.A. Wasserman (2013). Combinatorial effects of transposable elements on gene expression and phenotypic robustness in Drosophila melanogaster development.G3 (Bethesda)(9): 1531-1538.10.1534/g3.113.006791

Cole, C., A. Sobala, C. Lu, S.R. Thatcher, A. Bowman, et al. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs.RNA(12): 2147-2160.10.1261/rna.1738409

Colinet, H. and A.A. Hoffmann (2012). Comparing phenotypic effects and molecular correlates of developmental, gradual and rapid cold acclimation responses in Drosophila melanogaster.Functional Ecology(1): 84-93.10.1111/j.1365-2435.2011.01898.x

Conte, C., B. Dastugue and C. Vaury (2002). Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes.Mol Cell Biol(6): 1767-1777.10.1128/mcb.22.6.1767-1777.2002

Couvillion, M.T., G. Bounova, E. Purdom, T.P. Speed and K. Collins (2012). A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus.Mol Cell(4): 509-520.10.1016/j.molcel.2012.09.010

Couvillion, M.T., R. Sachidanandam and K. Collins (2010). A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo.Genes Dev(24): 2742-2747.10.1101/gad.1996210

Cowley, M. and R.J. Oakey (2013). Transposable elements re-wire and fine-tune the transcriptome.PLoS Genet(1): e1003234.10.1371/journal.pgen.1003234

Czajka, M.C. and R.E. Lee, Jr. (1990). A rapid cold-hardening response protecting against cold shock injury in Drosophila melanogaster.J Exp Biol(245-254

Daborn, P.J., J.L. Yen, M.R. Bogwitz, G. Le Goff, E. Feil, et al. (2002). A single p450 allele associated with insecticide resistance in Drosophila.Science(5590): 2253-2256.10.1126/science.1074170

Daugherty, M.D. and H.S. Malik (2012). Rules of engagement: molecular insights from host-virus arms races.Annu Rev Genet(677-700.10.1146/annurev-genet-110711-155522

David, J.R. and P. Capy (1988). Genetic variation of Drosophila melanogaster natural populations.Trends Genet(4): 106-111.10.1016/0168-9525(88)90098-4

De Cecco, M., T. Ito, A.P. Petrashen, A.E. Elias, N.J. Skvir, et al. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation.Nature(7742): 73-78.10.1038/s41586-018-0784-9

De Koning, A.P., W. Gu, T.A. Castoe, M.A. Batzer and D.D. Pollock (2011). Repetitive elements may comprise over two-thirds of the human genome.PLoS Genet(12): e1002384.10.1371/journal.pgen.1002384

Deprá, M., V.L. Valente, R. Margis and E.L. Loreto (2009). The hobo transposon and hobo-related elements are expressed as developmental genes in Drosophila.Gene(1): 57-63.10.1016/j.gene.2009.08.012

Dhahbi, J.M. (2015). 5' tRNA Halves: The Next Generation of Immune Signaling Molecules.Front Immunol(74.10.3389/fimmu.2015.00074

Diehl, A.G., N. Ouyang and A.P. Boyle (2020). Transposable elements contribute to cell and species specific chromatin looping and gene regulation in mammalian genomes. Nature Communications. **11:1796**.

Ding, Y., A. Berrocal, T. Morita, K.D. Longden and D.L. Stern (2016). Natural courtship song variation caused by an intronic retroelement in an ion channel gene.Nature(7616): 329-332.10.1038/nature19093

Dixit, G., A.P. Singh, A. Kumar, S. Dwivedi, F. Deeba, et al. (2015). Sulfur alleviates arsenic toxicity by reducing its accumulation and modulating proteome, amino acids and thiol metabolism in rice leaves.Sci Rep(16205.10.1038/srep16205

Doolittle, W.F. and C. Sapienza (1980). Selfish genes, the phenotype paradigm and genome evolution.Nature(5757): 601-603.10.1038/284601a0

Dou, S., Y. Wang and J. Lu (2019). Metazoan tsRNAs: Biogenesis, Evolution and Regulatory Functions.Noncoding RNA(1).10.3390/ncrna5010018

Dunn-Fletcher, C.E., L.M. Muglia, M. Pavlicev, G. Wolf, M.A. Sun, et al. (2018). Anthropoid primate-specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length.PLoS Biol(9): e2006337.10.1371/journal.pbio.2006337

Edney, E.B. (1977). Water Balance in Land Arthropods. Springer, Berlin.

Elbarbary, R.A., B.A. Lucas and L.E. Maquat (2016). Retrotransposons as regulators of gene expression.Science(6274): aac7247.10.1126/science.aac7247

Ellison, C.E., M.S. Kagda and W. Cao (2020). The Drosophila TART transposon manipulates the piRNA pathway as a counter-defense strategy to limit host silencing.bioRxiv(2020.2002.2020.957324.10.1101/2020.02.20.957324

Emara, M.M., P. Ivanov, T. Hickman, N. Dawra, S. Tisdale, et al. (2010). Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly.J Biol Chem(14): 10959-10968.10.1074/jbc.M109.077560

Engels, W.R. (1979). Extrachromosomal control of mutability in Drosophila melanogaster.Proc Natl Acad Sci U S A(8): 4011-4015.10.1073/pnas.76.8.4011

Esnault, C., M. Lee, C. Ham and H.L. Levin (2019). Transposable element insertions in fission yeast drive adaptation to environmental stress.Genome Res(1): 85-95.10.1101/gr.239699.118

Espinosa-Diez, C., V. Miguel, D. Mennerich, T. Kietzmann, P. Sánchez-Pérez, et al. (2015). Antioxidant responses and cellular adjustments to oxidative stress.Redox Biol(183-197.10.1016/j.redox.2015.07.008

Fabian, D.K., M. Kapun, V. Nolte, R. Kofler, P.S. Schmidt, et al. (2012). Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America.Mol Ecol(19): 4748-4769.10.1111/j.1365-294X.2012.05731.x

Fablet, M. and C. Vieira (2011). Evolvability, epigenetics and transposable elements.Biomol Concepts(5): 333-341.10.1515/BMC.2011.035

Fan, S., M.E. Hansen, Y. Lo and S.A. Tishkoff (2016). Going global by adapting local: A review of recent human adaptation.Science(6308): 54-59.10.1126/science.aaf5098

Fedoroff, N.V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution.Science(6108): 758-767.10.1126/science.338.6108.758

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks.Nat Rev Genet(5): 397-405.10.1038/nrg2337

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution.Trends Genet(4): 103-107.10.1016/0168-9525(89)90039-5

Flutre, T., E. Duprat, C. Feuillet and H. Quesneville (2011). Considering transposable element diversification in de novo annotation approaches.PLoS One(1): e16526.10.1371/journal.pone.0016526

Fox, R.J., J.M. Donelson, C. Schunter, T. Ravasi and J.D. Gaitán-Espitia (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change.Philos Trans R Soc Lond B Biol Sci(1768): 20180174.10.1098/rstb.2018.0174

Fricker, R., R. Brogli, H. Luidalepp, L. Wyss, M. Fasnacht, et al. (2019). A tRNA half modulates translation as stress response in Trypanosoma brucei.Nat Commun(1): 118.10.1038/s41467-018-07949-6

Fuentes, D.R., T. Swigut and J. Wysocka (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation.Elife(10.7554/eLife.35989

Fumagalli, M., M. Sironi, U. Pozzoli, A. Ferrer-Admetlla, A. Ferrer-Admettla, et al. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution.PLoS Genet(11): e1002355.10.1371/journal.pgen.1002355

Gadd, G.M. (2010). Metals, minerals and microbes: geomicrobiology and bioremediation.Microbiology(Pt 3): 609-643.10.1099/mic.0.037143-0

Garcia-Silva, M.R., M. Frugier, J.P. Tosar, A. Correa-Dominguez, L. Ronalte-Alves, et al. (2010). A population of tRNA-derived small RNAs is actively produced in Trypanosoma cruzi and recruited to specific cytoplasmic granules.Mol Biochem Parasitol(2): 64-73.10.1016/j.molbiopara.2010.02.003

Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum.Nature(6906): 498-511.10.1038/nature01097

Garrigues, J.M., B.V. Tsu, M.D. Daugherty and A.E. Pasquinelli (2019). Diversification of the.Elife(10.7554/eLife.51139

Gebetsberger, J., L. Wyss, A.M. Mleczko, J. Reuther and N. Polacek (2017). A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress.RNA Biol(10): 1364-1373.10.1080/15476286.2016.1257470

Gebetsberger, J., M. Zywicki, A. Künzi and N. Polacek (2012). tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii.Archaea(260909.10.1155/2012/260909

Gerhart, J. and M. Kirschner (1997). Cells, embryos and Evolution, Wiley, Hoboken.

Ghildiyal, M. and P.D. Zamore (2009). Small silencing RNAs: an expanding universe.Nat Rev Genet(2): 94-108.10.1038/nrg2504

Gibbs, A.G. (2002). Water balance in desert Drosophila: lessons from non-charismatic microfauna.Comp Biochem Physiol A Mol Integr Physiol(3): 781-789.10.1016/s1095-6433(02)00208-8

Gibbs, A.G., F. Fukuzato and L.M. Matzkin (2003). Evolution of water conservation mechanisms in Drosophila.J Exp Biol(Pt 7): 1183-1192.10.1242/jeb.00233

Gienapp, P., C. Teplitsky, J.S. Alho, J.A. Mills and J. Merilä (2008). Climate change and evolution: disentangling environmental and genetic responses.Mol Ecol(1): 167-178.10.1111/j.1365-294X.2007.03413.x

Glinsky, G.V. (2018). Contribution of transposable elements and distal enhancers to evolution of human-specific features of interphase chromatin architecture in embryonic stem cells.Chromosome Res(1-2): 61-84.10.1007/s10577-018-9571-6

Gobert, V., M. Gottar, A.A. Matskevich, S. Rutschmann, J. Royet, et al. (2003). Dual activation of the Drosophila toll pathway by two pattern recognition receptors.Science(5653): 2126-2130.10.1126/science.1085432

Gompel, N., B. Prud'homme, P.J. Wittkopp, V.A. Kassner and S.B. Carroll (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.Nature(7025): 481-487.10.1038/nature03235

González, J., T.L. Karasov, P.W. Messer and D.A. Petrov (2010). Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila.PLoS Genet(4): e1000905.10.1371/journal.pgen.1000905

Goodarzi, H., X. Liu, H.C. Nguyen, S. Zhang, L. Fish, et al. (2015). Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement.Cell(4): 790-802.10.1016/j.cell.2015.02.053

Guerreiro, I., A. Nunes, J.M. Woltering, A. Casaca, A. Nóvoa, et al. (2013). Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine.Proc Natl Acad Sci U S A(26): 10682-10686.10.1073/pnas.1300592110

Guio, L. and J. González (2019). New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans.Methods Mol Biol(505-530.10.1007/978-1-4939-9074-0_16

Gunderson, A.R. and J.H. Stillman (2015). Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming.Proc Biol Sci(1808): 20150401.10.1098/rspb.2015.0401

Heap, I. (1997). The occurrence of herbicide-resistant weeds world wide. Pesticide Science. **51:** 235-243.

Hendry, A.P., T.J. Farrugia and M.T. Kinnison (2008). Human influences on rates of phenotypic change in wild animal populations.Mol Ecol(1): 20-29.10.1111/j.1365-294X.2007.03428.x

Hirano, T., Y.W. Iwasaki, Z.Y. Lin, M. Imamura, N.M. Seki, et al. (2014). Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate.RNA(8): 1223-1237.10.1261/rna.045310.114

Hoffmann, A.A. (1990). Acclimation for desiccation resistance in D*rosophila melanogaster* and the association between acclimation responses and genetic variation. Journal of Insect Physiology. **36:** 885-891.

Hoffmann, A.A. (1991). Acclimation for desiccation resistance in D*rosophila:* Species and population comparisons. Journal of Insect Physiology. **37:** 757-762.

Hoffmann, A.A., R.J. Hallas, J.A. Dean and M. Schiffer (2003). Low potential for climatic stress adaptation in a rainforest Drosophila species.Science(5629): 100-102.10.1126/science.1084296

Hoffmann, A.A. and P.A. Parsons (1989). Selection for increased desiccation resistance in Drosophila melanogaster: additive genetic control and correlated responses for other stresses.Genetics(4): 837-845

Hoffmann, A.A. and P.A. Parsons (1991). Evolutionary Genetics and Environmental Stress Oxford University Press, Oxford, UK.

Hoostal, M.J., M.G. Bidart-Bouzat and J.L. Bouzat (2008). Local adaptation of microbial communities to heavy metal stress in polluted sediments of Lake Erie.FEMS Microbiol Ecol(1): 156-168.10.1111/j.1574-6941.2008.00522.x

Horváth, V., M. Merenciano and J. González (2017). Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response.Trends Genet(11): 832-841.10.1016/j.tig.2017.08.007

Hsieh, L.C., S.I. Lin, A.C. Shih, J.W. Chen, W.Y. Lin, et al. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing.Plant Physiol(4): 2120-2132.10.1104/pp.109.147280

Hua-Van, A., A. Le Rouzic, T.S. Boutin, J. Filée and P. Capy (2011). The struggle for life of the genome's selfish architects.Biol Direct(19.10.1186/1745-6150-6-19

Huang, S., X. Tao, S. Yuan, Y. Zhang, P. Li, et al. (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination.Cell(1): 102-114.10.1016/j.cell.2016.05.032

Hurst, G.D. and J.H. Werren (2001). The role of selfish genetic elements in eukaryotic evolution.Nat Rev Genet(8): 597-606.10.1038/35084545

Ikeda, Y., T. Pélissier, P. Bourguet, C. Becker, M.N. Pouch-Pélissier, et al. (2017). Arabidopsis proteins with a transposon-related domain act in gene silencing.Nat Commun(15122.10.1038/ncomms15122

Ito, J., R. Sugimoto, H. Nakaoka, S. Yamada, T. Kimura, et al. (2017). Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses.PLoS Genet(7): e1006883.10.1371/journal.pgen.1006883

Ivancevic, A. and E.B. Chuong (2020). Transposable elements teach T cells new tricks.Proc Natl Acad Sci U S A(17): 9145-9147.10.1073/pnas.2004493117

Ivanov, P., E. O'day, M.M. Emara, G. Wagner, J. Lieberman, et al. (2014). G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments.Proc Natl Acad Sci U S A(51): 18201-18206.10.1073/pnas.1407361111

Jacques, P., J. Jeyakani and G. Bourque (2013). The majority of primate-specific regulatory sequences are derived from transposable elements.PLoS Genet(5): e1003504.10.1371/journal.pgen.1003504

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs.Nat Rev Genet(12): 833-844.10.1038/nrg2683

Jang, H.S., N.M. Shah, A.Y. Du, Z.Z. Dailey, E.C. Pehrsson, et al. (2019). Transposable elements drive widespread expression of oncogenes in human cancers.Nat Genet(4): 611-617.10.1038/s41588-019-0373-3

Jangam, D., C. Feschotte and E. Betrán (2017). Transposable Element Domestication As an Adaptation to Evolutionary Conflicts.Trends Genet(11): 817-831.10.1016/j.tig.2017.07.011

Jiang, J.C. and K.R. Upton (2019). Human transposons are an abundant supply of transcription factor binding sites and promoter activities in breast cancer cell lines.Mob DNA(16.10.1186/s13100-019-0158-3

Jiao, Y. and X.W. Deng (2007). A genome-wide transcriptional activity survey of rice transposable element-related genes.Genome Biol(2): R28.10.1186/gb-2007-8-2-r28

Johnstone, M., M. Schiffer and A.A. Hoffmann (2017). Separating multiple sources of variation on heat resistance in Drosophila hydei.J Insect Physiol(122-127.10.1016/j.jinsphys.2016.11.001

Juneja, P., A. Quinn and F.M. Jiggins (2016). Latitudinal clines in gene expression and cis-regulatory element variation in Drosophila melanogaster.BMC Genomics(1): 981.10.1186/s12864-016-3333-7

Kapitonov, V.V. and J. Jurka (2008). A universal classification of eukaryotic transposable elements implemented in Repbase.Nat Rev Genet(5): 411-412; author reply 414.10.1038/nrg2165-c1

Kapusta, A., Z. Kronenberg, V.J. Lynch, X. Zhuo, L. Ramsay, et al. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs.PLoS Genet(4): e1003470.10.1371/journal.pgen.1003470

Karaiskos, S., A.S. Naqvi, K.E. Swanson and A. Grigoriev (2015). Age-driven modulation of tRNA-derived fragments in Drosophila and their potential targets.Biol Direct(51.10.1186/s13062-015-0081-6

Kazazian, H.H., C. Wong, H. Youssoufian, A.F. Scott, D.G. Phillips, et al. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man.Nature(6160): 164-166.10.1038/332164a0

Keam, S.P. and G. Hutvagner (2015). tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression.Life (Basel)(4): 1638-1651.10.3390/life5041638

Kellermann, V. and B.V. Heerwaarden (2019). Terrestrial insects and climate change: adaptive responses in key traits. Physiological Entomology. **44:** 99-115.

Kellermann, V., A.A. Hoffmann, J. Overgaard, V. Loeschcke and C.M. Sgrò (2018). Plasticity for desiccation tolerance across Drosophila species is affected by phylogeny and climate in complex ways.Proc Biol Sci(1874).10.1098/rspb.2018.0048

Kellermann, V., V. Loeschcke, A.A. Hoffmann, T.N. Kristensen, C. Fløjgaard, et al. (2012). Phylogenetic constraints in key functional traits behind species' climate niches: patterns of desiccation and cold resistance across 95 Drosophila species.Evolution(11): 3377-3389.10.1111/j.1558-5646.2012.01685.x

Kellermann, V. and C.M. Sgrò (2018). Evidence for lower plasticity in CT(MAX) at warmer developmental temperatures.J Evol Biol(9): 1300-1312.10.1111/jeb.13303

Kellner, M. and W. Makałowski (2019). Transposable elements significantly contributed to the core promoters in the human genome.Sci China Life Sci(4): 489-497.10.1007/s11427-018-9449-0

Khalkhali-Evrigh, R., N. Hedayat-Evrigh, S.H. Hafezian, A. Farhadi and M.R. Bakhtiarizadeh (2019). Genome-Wide Identification of Microsatellites and Transposable Elements in the Dromedary Camel Genome Using Whole-Genome Sequencing Data.Front Genet(692.10.3389/fgene.2019.00692

Kim, J.H., J.A. Moreau, J.M. Zina, L. Mazgaeen, K.S. Yoon, et al. (2018). Identification and interaction of multiple genes resulting in DDT resistance in the 91-R strain of Drosophila melanogaster by RNAi approaches.Pestic Biochem Physiol(90-99.10.1016/j.pestbp.2018.03.003

Koehn, R.K. and B.L. Bayne (1989). Towards a physiological and genetical understanding of the energetics of the stress response Biological Journal of the Linnean Society. **37 (1-2):** 157-171.

Kolaczkowski, B., A.D. Kern, A.K. Holloway and D.J. Begun (2011). Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster.Genetics(1): 245-260.10.1534/genetics.110.123059

Kültz, D. (2005). Molecular and evolutionary basis of the cellular stress response.Annu Rev Physiol(225-257.10.1146/annurev.physiol.67.040403.103635

Kültz, D. (2020). Defining biological stress and stress responses based on principles of physics.J Exp Zool A Ecol Integr Physiol(10.1002/jez.2340

Kumar, P., J. Anaya, S.B. Mudunuri and A. Dutta (2014). Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets.BMC Biol(78.10.1186/s12915-014-0078-0

Kumar, P., C. Kuscu and A. Dutta (2016). Biogenesis and Function of Transfer RNA-Related Fragments (tRFs).Trends Biochem Sci(8): 679-689.10.1016/j.tibs.2016.05.004

Kumar, P., S.B. Mudunuri, J. Anaya and A. Dutta (2015). tRFdb: a database for transfer RNA fragments.Nucleic Acids Res(Database issue): D141-145.10.1093/nar/gku1138

Kunarso, G., N.Y. Chia, J. Jeyakani, C. Hwang, X. Lu, et al. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells.Nat Genet(7): 631-634.10.1038/ng.600

Kwiatkowski, D.P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria.Am J Hum Genet(2): 171-192.10.1086/432519

Lachance, J. and S.A. Tishkoff (2013). Population Genomics of Human Adaptation.Annu Rev Ecol Evol Syst(123-143.10.1146/annurev-ecolsys-110512-135833

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, et al. (2001). Initial sequencing and analysis of the human genome.Nature(6822): 860-921.10.1038/35057062

Lazzaro, B.P., H.A. Flores, J.G. Lorigan and C.P. Yourth (2008). Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in Drosophila melanogaster.PLoS Pathog(3): e1000025.10.1371/journal.ppat.1000025

Lelli, K.M., M. Slattery and R.S. Mann (2012). Disentangling the many layers of eukaryotic transcriptional regulation.Annu Rev Genet(43-68.10.1146/annurev-genet-110711-155437

Lerman, D.N. and M.E. Feder (2005). Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster.Mol Biol Evol(3): 776-783.10.1093/molbev/msi063

Li, H. and W. Stephan (2006). Inferring the demographic history and rate of adaptive substitution in Drosophila.PLoS Genet(10): e166.10.1371/journal.pgen.0020166

Li, S., Z. Xu and J. Sheng (2018). tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA.Genes (Basel)(5).10.3390/genes9050246

Lindgren, B. and A. Laurila (2005). Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of Rana temporaria.J Evol Biol(4): 820-828.10.1111/j.1420-9101.2004.00875.x

Lippman, Z., A.V. Gendrel, M. Black, M.W. Vaughn, N. Dedhia, et al. (2004). Role of transposable elements in heterochromatin and epigenetic control.Nature(6998): 471-476.10.1038/nature02651

Liu, B. and S.B. Qian (2014). Translational reprogramming in cellular stress response.Wiley Interdiscip Rev RNA(3): 301-315.10.1002/wrna.1212

Liu, S., Y. Chen, Y. Ren, J. Zhou, J. Ren, et al. (2018). A tRNA-derived RNA Fragment Plays an Important Role in the Mechanism of Arsenite -induced Cellular Responses.Sci Rep(1): 16838.10.1038/s41598-018-34899-2

Loeschcke, V., J.G. Sørensen and T.N. Kristensen (2004). Ecologically relevant stress resistance: from microarrays and quantitative trait loci to candidate genes - a research plan and preliminary results using Drosophila as a model organism and climatic and genetic stress as model stresses.J Biosci(4): 503-511.10.1007/bf02712122

Long, H.K., S.L. Prescott and J. Wysocka (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.Cell(5): 1170-1187.10.1016/j.cell.2016.09.018

Loreto, E.L.S., M. Deprá, J.F. Diesel, Y. Panzera and V.L.S. Valente-Gaiesky (2018). Drosophila relics hobo and hobo-MITEs transposons as raw material for new regulatory networks.Genet Mol Biol(1 suppl 1): 198-205.10.1590/1678-4685-GMB-2017-0068

Loss-Morais, G., P.M. Waterhouse and R. Margis (2013). Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets.Biol Direct(6.10.1186/1745-6150-8-6

Lowe, J.M., D. Menendez, P.R. Bushel, M. Shatz, E.L. Kirk, et al. (2014). p53 and NF-κB coregulate proinflammatory gene responses in human macrophages.Cancer Res(8): 2182-2192.10.1158/0008-5472.can-13-1070

Lu, X., F. Sachs, L. Ramsay, P. Jacques, J. Göke, et al. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity.Nat Struct Mol Biol(4): 423-425.10.1038/nsmb.2799

Lunyak, V.V., G.G. Prefontaine, E. Núñez, T. Cramer, B.G. Ju, et al. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis.Science(5835): 248-251.10.1126/science.1140871

Luo, S., F. He, J. Luo, S. Dou, Y. Wang, et al. (2018). Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response.Nucleic Acids Res(10): 5250-5268.10.1093/nar/gky189

Lynch, V.J., R.D. Leclerc, G. May and G.P. Wagner (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals.Nat Genet(11): 1154-1159.10.1038/ng.917

Lynch, V.J., M.C. Nnamani, A. Kapusta, K. Brayer, S.L. Plaza, et al. (2015). Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy.Cell Rep(4): 551-561.10.1016/j.celrep.2014.12.052

Lyons, S.M., D. Gudanis, S.M. Coyne, Z. Gdaniec and P. Ivanov (2017). Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs.Nat Commun(1): 1127.10.1038/s41467-017-01278-w

Macchietto, M.G., R.A. Langlois and S.S. Shen (2020). Virus-induced transposable element expression up-regulation in human and mouse host cells.Life Sci Alliance(2).10.26508/lsa.201900536

Maggert, K.A. (2019). Stress: An evolutionary mutagen.Proc Natl Acad Sci U S A(36): 17616-17618.10.1073/pnas.1912725116

Magwire, M.M., F. Bayer, C.L. Webster, C. Cao and F.M. Jiggins (2011). Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication.PLoS Genet(10): e1002337.10.1371/journal.pgen.1002337

Main, B.J., A. Everitt, A.J. Cornel, F. Hormozdiari and G.C. Lanzaro (2018). Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii.Parasit Vectors(1): 225.10.1186/s13071-018-2817-5

Makarevitch, I., A.J. Waters, P.T. West, M. Stitzer, C.N. Hirsch, et al. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress.PLoS Genet(1): e1004915.10.1371/journal.pgen.1004915

Marsano, R.M., R. Caizzi, R. Moschetti and N. Junakovic (2005). Evidence for a functional interaction between the Bari1 transposable element and the cytochrome P450 cyp12a4 gene in Drosophila melanogaster.Gene(2): 122-128.10.1016/j.gene.2005.06.005

Martinez, G., S.G. Choudury and R.K. Slotkin (2017). tRNA-derived small RNAs target transposable element transcripts.Nucleic Acids Res(9): 5142-5152.10.1093/nar/gkx103

Mascagni, F., A. Vangelisti, G. Usai, T. Giordani, A. Cavallini, et al. (2020). A computational genome-wide analysis of long terminal repeats retrotransposon expression in sunflower roots (Helianthus annuus L.).Genetica(1): 13-23.10.1007/s10709-020-00085-4

Mateo, L., A. Ullastres and J. González (2014). A transposable element insertion confers xenobiotic resistance in Drosophila.PLoS Genet(8): e1004560.10.1371/journal.pgen.1004560

Maute, R.L., C. Schneider, P. Sumazin, A. Holmes, A. Califano, et al. (2013). tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma.Proc Natl Acad Sci U S A(4): 1404-1409.10.1073/pnas.1206761110

Mcclintock, B. (1950). The origin and behavior of mutable loci in maize.Proc Natl Acad Sci U S A(6): 344-355.10.1073/pnas.36.6.344

Mcclintock, B. (1953). Induction of Instability at Selected Loci in Maize.Genetics(6): 579-599

Mcclintock, B. (1956). Controlling elements and the gene.Cold Spring Harb Symp Quant Biol(197-216.10.1101/sqb.1956.021.01.017

Mccue, A.D. and R.K. Slotkin (2012). Transposable element small RNAs as regulators of gene expression.Trends Genet(12): 616-623.10.1016/j.tig.2012.09.001

Mckenzie, J.A., S.W. Mckechnie and P. Batterham (1994). Perturbation of gene frequencies in a natural population of Drosophila melanogaster: evidence for selection at the Adh locus.Genetica(3): 187-196.10.1007/bf00132537

Mclaughlin, J.F., J.J. Hellmann, C.L. Boggs and P.R. Ehrlich (2002). Climate change hastens population extinctions.Proc Natl Acad Sci U S A(9): 6070-6074.10.1073/pnas.052131199

Mclaughlin, R.N., J.M. Young, L. Yang, R. Neme, H.A. Wichman, et al. (2014). Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency.PLoS Genet(9): e1004531.10.1371/journal.pgen.1004531

Meharg, A.A., Q.J. Cumbes and M.R. Macnair (1993). PRE-ADAPTATION OF YORKSHIRE FOG, HOLCUS LANATUS L. (POACEAE) TO ARSENATE TOLERANCE.Evolution(1): 313-316.10.1111/j.1558-5646.1993.tb01220.x

Menees, T.M. and S.B. Sandmeyer (1996). Cellular stress inhibits transposition of the yeast retrovirus-like element Ty3 by a ubiquitin-dependent block of virus-like particle formation.Proc Natl Acad Sci U S A(11): 5629-5634.10.1073/pnas.93.11.5629

Merenciano, M., A. Ullastres, M.A. De Cara, M.G. Barrón and J. González (2016). Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in Drosophila.PLoS Genet(8): e1006249.10.1371/journal.pgen.1006249

Merilä, J. and A.P. Hendry (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence.Evol Appl(1): 1-14.10.1111/eva.12137

Miki, Y., I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.Cancer Res(3): 643-645

Miller, W.J., J.F. Mcdonald and W. Pinsker (1997). Molecular domestication of mobile elements.Genetica(1-3): 261-270

Moschetti, R., A. Palazzo, P. Lorusso, L. Viggiano and R.M. Marsano (2020). "What You Need, Baby, I Got It": Transposable Elements as Suppliers of Cis-Operating Sequences in Drosophila.Biology (Basel)(2).10.3390/biology9020025

Negi, P., A.N. Rai and P. Suprasanna (2016). : Emerging Regulatory Roles for Transposons in Plant Stress Response.Front Plant Sci(1448.10.3389/fpls.2016.01448

Nie, Z., F. Zhou, D. Li, Z. Lv, J. Chen, et al. (2013). RIP-seq of BmAgo2-associated small RNAs reveal various types of small non-coding RNAs in the silkworm, Bombyx mori.BMC Genomics(661.10.1186/1471-2164-14-661

Nikitin, D., D. Penzar, A. Garazha, M. Sorokin, V. Tkachev, et al. (2018). Profiling of Human Molecular Pathways Affected by Retrotransposons at the Level of Regulation by Transcription Factor Proteins.Front Immunol(30.10.3389/fimmu.2018.00030

Nishihara, H. (2019). Retrotransposons spread potential cis-regulatory elements during mammary gland evolution.Nucleic Acids Res(22): 11551-11562.10.1093/nar/gkz1003

Notwell, J.H., T. Chung, W. Heavner and G. Bejerano (2015). A family of transposable elements co-opted into developmental enhancers in the mouse neocortex.Nat Commun(6644.10.1038/ncomms7644

Nyamukondiwa, C., E. Kleynhans and J.S. Terblanche (2010). Phenotypic plasticity of thermal tolerance contributes to the invasion potential of Mediterranean fruit flies (Ceratitis capitata).Ecological Entomology(5): 565-575.10.1111/j.1365-2311.2010.01215.x

Obbard, D.J., K.H. Gordon, A.H. Buck and F.M. Jiggins (2009). The evolution of RNAi as a defence against viruses and transposable elements.Philos Trans R Soc Lond B Biol Sci(1513): 99-115.10.1098/rstb.2008.0168

Oerke, E.C. (2005). Crop losses to pests. Journal of Agricultural Science. **144:** 31-43.

Ohno, S. (1972). So much "junk" DNA in our genome.Brookhaven Symp Biol(366-370

Oki, N., K. Yano, Y. Okumoto, T. Tsukiyama, M. Teraishi, et al. (2008). A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp. japonica.Genes Genet Syst(4): 321-329.10.1266/ggs.83.321

Oliver, K.R. and W.K. Greene (2009). Transposable elements: powerful facilitators of evolution.Bioessays(7): 703-714.10.1002/bies.200800219

Ong-Abdullah, M., J.M. Ordway, N. Jiang, S.E. Ooi, S.Y. Kok, et al. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm.Nature(7570): 533-537.10.1038/nature15365

Orgel, L.E., F.H. Crick and C. Sapienza (1980). Selfish DNA.Nature(5792): 645-646.10.1038/288645a0

Overgaard, J., T.N. Kristensen, K.A. Mitchell and A.A. Hoffmann (2011). Thermal tolerance in widespread and tropical Drosophila species: does phenotypic plasticity increase with latitude?Am Nat(S80-96.10.1086/661780

Palumbi, S.R. (2001). Humans as the world's greatest evolutionary force.Science(5536): 1786-1790.10.1126/science.293.5536.1786

Pardue, M.L. and P.G. Debaryshe (2003). Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres.Annu Rev Genet(485-511.10.1146/annurev.genet.38.072902.093115

Parkash, R., S. Ramniwas, B. Kajla and D.D. Aggarwal (2012). Divergence of desiccation-related traits in two Drosophila species of the takahashii subgroup from the western Himalayas.J Exp Biol(Pt 13): 2181-2191.10.1242/jeb.065730

Pehrsson, E.C., M.N.K. Choudhary, V. Sundaram and T. Wang (2019). The epigenomic landscape of transposable elements across normal human development and anatomy.Nat Commun(1): 5640.10.1038/s41467-019-13555-x

Pekarsky, Y., V. Balatti, A. Palamarchuk, L. Rizzotto, D. Veneziano, et al. (2016). Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer.Proc Natl Acad Sci U S A(18): 5071-5076.10.1073/pnas.1604266113

Peng, H., J. Shi, Y. Zhang, H. Zhang, S. Liao, et al. (2012). A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm.Cell Res(11): 1609-1612.10.1038/cr.2012.141

Percharde, M., T. Sultana and M. Ramalho-Santos (2020). What Doesn't Kill You Makes You Stronger: Transposons as Dual Players in Chromatin Regulation and Genomic Variation.Bioessays(e1900232.10.1002/bies.201900232

Petri, R., P.L. Brattås, Y. Sharma, M.E. Jönsson, K. Pircs, et al. (2019). LINE-2 transposable elements are a source of functional human microRNAs and target sites.PLoS Genet(3): e1008036.10.1371/journal.pgen.1008036

Piégu, B., S. Bire, P. Arensburger and Y. Bigot (2015). A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity.Mol Phylogenet Evol(90-109.10.1016/j.ympev.2015.03.009

Piriyapongsa, J. and I.K. Jordan (2007). A family of human microRNA genes from miniature inverted-repeat transposable elements.PLoS One(2): e203.10.1371/journal.pone.0000203

Pontis, J., E. Planet, S. Offner, P. Turelli, J. Duc, et al. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs.Cell Stem Cell(5): 724-735.e725.10.1016/j.stem.2019.03.012

Potter, S.S., W.J. Brorein, P. Dunsmuir and G.M. Rubin (1979). Transposition of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila.Cell(2): 415-427.10.1016/0092-8674(79)90168-5

Quach, H. and L. Quintana-Murci (2017). Living in an adaptive world: Genomic dissection of the genus.J Exp Med(4): 877-894.10.1084/jem.20161942

Racle, J., F. Picard, L. Girbal, M. Cocaign-Bousquet and V. Hatzimanikatis (2013). A genome-scale integration and analysis of Lactococcus lactis translation data.PLoS Comput Biol(10): e1003240.10.1371/journal.pcbi.1003240

Rajpurohit, S., E. Gefen, A.O. Bergland, D.A. Petrov, A.G. Gibbs, et al. (2018). Spatiotemporal dynamics and genome-wide association genome-wide association analysis of desiccation tolerance in Drosophila melanogaster.Mol Ecol(17): 3525-3540.10.1111/mec.14814

Rasmuson, B., I. Montell, A. Rasmuson, H. Svahlin and B.M. Westerberg (1980). Genetic instability in Drosophila melanogaster: evidence for regulation, excision and transposition at the white locus.Mol Gen Genet(4): 567-570.10.1007/bf00272664

Rebollo, R., M.M. Karimi, M. Bilenky, L. Gagnier, K. Miceli-Royer, et al. (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms.PLoS Genet(9): e1002301.10.1371/journal.pgen.1002301

Rech, G.E., M. Bogaerts-Márquez, M.G. Barrón, M. Merenciano, J.L. Villanueva-Cañas, et al. (2019). Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila.PLoS Genet(2): e1007900.10.1371/journal.pgen.1007900

Relyea, R.A. (2005). The lethal impacts of Roundup and predatory stress on six species of North American tadpoles.Arch Environ Contam Toxicol(3): 351-357.10.1007/s00244-004-0086-0

Rey, O., E. Danchin, M. Mirouze, C. Loot and S. Blanchet (2016). Adaptation to Global Change: A Transposable Element-Epigenetics Perspective.Trends Ecol Evol(7): 514-526.10.1016/j.tree.2016.03.013

Rius, N., Y. Guillén, A. Delprat, A. Kapusta, C. Feschotte, et al. (2016). Exploration of the Drosophila buzzatii transposable element content suggests underestimation of repeats in Drosophila genomes.BMC Genomics(344.10.1186/s12864-016-2648-8

Römer, C., M. Singh, L.D. Hurst and Z. Izsvák (2017). How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency.Curr Opin Virol(49-58.10.1016/j.coviro.2017.07.001

Roseman, R.R., J.M. Swan and P.K. Geyer (1995). A Drosophila insulator protein facilitates dosage compensation of the X chromosome min-white gene located at autosomal insertion sites.Development(11): 3573-3582

Rubin, G.M., M.G. Kidwell and P.M. Bingham (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations.Cell(3): 987-994.10.1016/0092-8674(82)90462-7

Rubin, G.M. and A.C. Spradling (1982). Genetic transformation of Drosophila with transposable element vectors.Science(4570): 348-353.10.1126/science.6289436

Saika, H., A. Mori, M. Endo and S. Toki (2019). Targeted deletion of rice retrotransposon Tos17 via CRISPR/Cas9.Plant Cell Rep(4): 455-458.10.1007/s00299-018-2357-7

Sakashita, A., S. Maezawa, K.G. Alavattam, M. Yukawa, A. Barski, et al. (2020). Endogenous retroviruses drive species-specific germline transcriptomes in mammals.bioRxiv(2020.2003.2011.987230.10.1101/2020.03.11.987230

Salces-Ortiz, J., C. Vargas-Chavez, L. Guio, G.E. Rech and J. González (2020). Transposable elements contribute to the genomic response to insecticides in.Philos Trans R Soc Lond B Biol Sci(1795): 20190341.10.1098/rstb.2019.0341

Sarkar, A., C. Sim, Y.S. Hong, J.R. Hogan, M.J. Fraser, et al. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences.Mol Genet Genomics(2): 173-180.10.1007/s00438-003-0909-0

Scharsack, J.P., M. Kalbe, C. Harrod and G. Rauch (2007). Habitat-specific adaptation of immune responses of stickleback (Gasterosteus aculeatus) lake and river ecotypes.Proc Biol Sci(1617): 1523-1532.10.1098/rspb.2007.0210

Schiffer, M., S. Hangartner and A.A. Hoffmann (2013). Assessing the relative importance of environmental effects, carry-over effects and species differences in thermal stress resistance: a comparison of Drosophilids across field and laboratory generations.J Exp Biol(Pt 20): 3790-3798.10.1242/jeb.085126

Schlebusch, C.M., L.M. Gattepaille, K. Engström, M. Vahter, M. Jakobsson, et al. (2015). Human adaptation to arsenic-rich environments.Mol Biol Evol(6): 1544-1555.10.1093/molbev/msv046

Schmidt, D., P.C. Schwalie, M.D. Wilson, B. Ballester, A. Gonçalves, et al. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages.Cell(1-2): 335-348.10.1016/j.cell.2011.11.058

Schorn, A.J., M.J. Gutbrod, C. Leblanc and R. Martienssen (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs.Cell(1): 61-71.e11.10.1016/j.cell.2017.06.013

Schrader, L. and J. Schmitz (2019). The impact of transposable elements in adaptive evolution.Mol Ecol(6): 1537-1549.10.1111/mec.14794

Seebacher, F., V. Ducret, A.G. Little and B. Adriaenssens (2015). Generalist-specialist trade-off during thermal acclimation.R Soc Open Sci(1): 140251.10.1098/rsos.140251

Sessegolo, C., N. Burlet and A. Haudry (2016). Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies.Biol Lett(8).10.1098/rsbl.2016.0407

Sgrò, C.M., J.S. Terblanche and A.A. Hoffmann (2016). What Can Plasticity Contribute to Insect Responses to Climate Change?Annu Rev Entomol(433-451.10.1146/annurev-ento-010715-023859

Sharma, U., C.C. Conine, J.M. Shea, A. Boskovic, A.G. Derr, et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals.Science(6271): 391-396.10.1126/science.aad6780

Shen, S., L. Lin, J.J. Cai, P. Jiang, E.J. Kenkel, et al. (2011). Widespread establishment and regulatory impact of Alu exons in human genes.Proc Natl Acad Sci U S A(7): 2837-2842.10.1073/pnas.1012834108

Shigematsu, M., S. Honda and Y. Kirino (2014). Transfer RNA as a source of small functional RNA.J Mol Biol Mol Imaging(2)

Siepielski, A.M., M.B. Morrissey, M. Buoro, S.M. Carlson, C.M. Caruso, et al. (2017). Precipitation drives global variation in natural selection.Science(6328): 959-962.10.1126/science.aag2773

Sørensen, J.G., T.N. Kristensen and J. Overgaard (2016). Evolutionary and ecological patterns of thermal acclimation capacity in Drosophila: is it important for keeping up with climate change?Curr Opin Insect Sci(98-104.10.1016/j.cois.2016.08.003

Sørensen, J.G., M.M. Nielsen, M. Kruhøffer, J. Justesen and V. Loeschcke (2005). Full genome gene expression analysis of the heat stress response in Drosophila melanogaster.Cell Stress Chaperones(4): 312-328.10.1379/csc-128r1.1

Spana, C., D.A. Harrison and V.G. Corces (1988). The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon.Genes Dev(11): 1414-1423.10.1101/gad.2.11.1414

Spitz, F. and E.E. Furlong (2012). Transcription factors: from enhancer binding to developmental control.Nat Rev Genet(9): 613-626.10.1038/nrg3207

Spradling, A.C. and G.M. Rubin (1982). Transposition of cloned P elements into Drosophila germ line chromosomes.Science(4570): 341-347.10.1126/science.6289435

Sprengelmeyer, Q.D., S. Mansourian, J.D. Lange, D.R. Matute, B.S. Cooper, et al. (2020). Recurrent Collection of Drosophila melanogaster from Wild African Environments and Genomic Insights into Species History.Mol Biol Evol(3): 627-638.10.1093/molbev/msz271

Stern, D.L. (2000). Evolutionary biology. The problem of variation.Nature(6812): 529, 531.10.1038/35046183

Stone, H.M., P.A. Erickson and A.O. Bergland (2020). Phenotypic plasticity, but not adaptive tracking, underlies seasonal variation in post-cold hardening freeze tolerance of Drosophila melanogaster.Ecol Evol(1): 217-231.10.1002/ece3.5887

Straalen, N.M.V. and D. Roelofs (2012). An introduction to Ecological Genomics. United States, Oxford University Press Inc.

Strobel, H.W. and J.D. Dignam (1978). Purification and properties of NADPH-cytochrome P-450 reductase.Methods Enzymol(89-96.10.1016/s0076-6879(78)52009-0

Sun, X., X. Wang, Z. Tang, M. Grivainis, D. Kahler, et al. (2018). Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression.Proc Natl Acad Sci U S A(24): E5526-E5535.10.1073/pnas.1722565115

Sundaram, V., Y. Cheng, Z. Ma, D. Li, X. Xing, et al. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks.Genome Res(12): 1963-1976.10.1101/gr.168872.113

Tanda, S. and V.G. Corces (1991). Retrotransposon-induced overexpression of a homeobox gene causes defects in eye morphogenesis in Drosophila.EMBO J(2): 407-417

Telonis-Scott, M., M. Gane, S. Degaris, C.M. Sgrò and A.A. Hoffmann (2012). High resolution mapping of candidate alleles for desiccation resistance in Drosophila melanogaster under selection.Mol Biol Evol(5): 1335-1351.10.1093/molbev/msr294

Thompson, D.M., C. Lu, P.J. Green and R. Parker (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes.RNA(10): 2095-2103.10.1261/rna.1232808

Thompson, D.M. and R. Parker (2009). The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae.J Cell Biol(1): 43-50.10.1083/jcb.200811119

Thompson, P.J., T.S. Macfarlan and M.C. Lorincz (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire.Mol Cell(5): 766-776.10.1016/j.molcel.2016.03.029

Thybert, D., M. Roller, F.C.P. Navarro, I. Fiddes, I. Streeter, et al. (2018). Repeat associated mechanisms of genome evolution and function revealed by the Mus caroli and Mus pahari genomes.Genome Res(4): 448-459.10.1101/gr.234096.117

Tinsley, M.C. and M.E. Majerus (2006). A new male-killing parasitism: Spiroplasma bacteria infect the ladybird beetle Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae).Parasitology(Pt 6): 757-765.10.1017/S0031182005009789

Todd, C.D., Ö. Deniz, D. Taylor and M.R. Branco (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells.Elife(10.7554/eLife.44344

Touchon, M. and E.P. Rocha (2007). Causes of insertion sequences abundance in prokaryotic genomes.Mol Biol Evol(4): 969-981.10.1093/molbev/msm014

Travis, J., J. Leips and F.H. Rodd (2013). Evolution in population parameters: density-dependent selection or density-dependent fitness?Am Nat(S9-S20.10.1086/669970

Trivedi, M., J. Shah, N. Hodgson, H.M. Byun and R. Deth (2014). Morphine induces redox-based changes in global DNA methylation and retrotransposon transcription by inhibition of excitatory amino acid transporter type 3-mediated cysteine uptake.Mol Pharmacol(5): 747-757.10.1124/mol.114.091728

Trizzino, M., Y. Park, M. Holsbach-Beltrame, K. Aracena, K. Mika, et al. (2017). Transposable elements are the primary source of novelty in primate gene regulation.Genome Res(10): 1623-1633.10.1101/gr.218149.116

Tuğrul, M., T. Paixão, N.H. Barton and G. Tkačik (2015). Dynamics of Transcription Factor Binding Site Evolution.PLoS Genet(11): e1005639.10.1371/journal.pgen.1005639

Ullastres, A., M. Merenciano and J. González (2019). Natural transposable element insertions drive expression changes in genes underlying *Drosophila melanogaster* immune response. bioRxiv.

Van Heerwaarden, B. and V. Kellermann (2020). Does Plasticity Trade Off With Basal Heat Tolerance?Trends Ecol Evol(10.1016/j.tree.2020.05.006

Van Heerwaarden, B., V. Kellermann and C.M. Sgrò (2016). Limited scope for plasticity to increase upper thermal limits.Functional Ecology(12): 1947-1956.10.1111/1365-2435.12687

Van Meter, M., M. Kashyap, S. Rezazadeh, A.J. Geneva, T.D. Morello, et al. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age.Nat Commun(5011.10.1038/ncomms6011

Van't Hof, A.E., P. Campagne, D.J. Rigden, C.J. Yung, J. Lingley, et al. (2016). The industrial melanism mutation in British peppered moths is a transposable element.Nature(7605): 102-105.10.1038/nature17951

Volff, J.N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes.Bioessays(9): 913-922.10.1002/bies.20452

Waldvogel, A.M., B. Feldmeyer, G. Rolshausen, M. Exposito-Alonso, C. Rellstab, et al. (2020). Evolutionary genomics can improve prediction of species' responses to climate change.Evol Lett(1): 4-18.10.1002/evl3.154

Wang, J., C. Vicente-García, D. Seruggia, E. Moltó, A. Fernandez-Miñán, et al. (2015). MIR retrotransposon sequences provide insulators to the human genome.Proc Natl Acad Sci U S A(32): E4428-4437.10.1073/pnas.1507253112

Wang, J., G. Xie, M. Singh, A.T. Ghanbarian, T. Raskó, et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells.Nature(7531): 405-409.10.1038/nature13804

Wang, T., J. Zeng, C.B. Lowe, R.G. Sellers, S.R. Salama, et al. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53.Proc Natl Acad Sci U S A(47): 18613-18618.10.1073/pnas.0703637104

Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, et al. (2007). A unified classification system for eukaryotic transposable elements.Nat Rev Genet(12): 973-982.10.1038/nrg2165

Wilkins, A.S. (2002). The evolution of developmental pathways. Sinauer Associates, Sunderland, Mass.

Wilson, S., L.V. Matyunina and J.F. Mcdonald (1998). An enhancer region within the copia untranslated leader contains binding sites for Drosophila regulatory proteins.Gene(1-2): 239-246.10.1016/s0378-1119(98)00048-1

Wray, G.A. (2003). Transcriptional regulation and the evolution of development.Int J Dev Biol(7-8): 675-684

Xue, Q. and C.S. Ma (2020). Aged virgin adults respond to extreme heat events with phenotypic plasticity in an invasive species, Drosophila suzukii.J Insect Physiol(104016.10.1016/j.jinsphys.2020.104016

Yamasaki, S., P. Ivanov, G.F. Hu and P. Anderson (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression.J Cell Biol(1): 35-42.10.1083/jcb.200811106

Yao, P., P. Lin, A. Gokoolparsadh, A. Assareh, M.W. Thang, et al. (2015). Coexpression networks identify brain region-specific enhancer RNAs in the human brain.Nat Neurosci(8): 1168-1174.10.1038/nn.4063

Ye, M., C. Goudot, T. Hoyler, B. Lemoine, S. Amigorena, et al. (2020). Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers.Proc Natl Acad Sci U S A(14): 7905-7916.10.1073/pnas.1912008117

Yeung, M.L., Y. Bennasser, K. Watashi, S.Y. Le, L. Houzet, et al. (2009). Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid.Nucleic Acids Res(19): 6575-6586.10.1093/nar/gkp707

Zatsepina, O.G., V.V. Velikodvorskaia, V.B. Molodtsov, D. Garbuz, D.N. Lerman, et al. (2001). A Drosophila melanogaster strain from sub-equatorial Africa has exceptional thermotolerance but decreased Hsp70 expression.J Exp Biol(Pt 11): 1869-1881

Zhao, L.L., J.L. Sun, J. Liang, Q. Liu, J. Luo, et al. (2020). Enhancing lipid metabolism and inducing antioxidant and immune responses to adapt to acute hypoxic stress in *Schizothora prenanti*. Agriculture. **519:** 734933.

SECTION 7

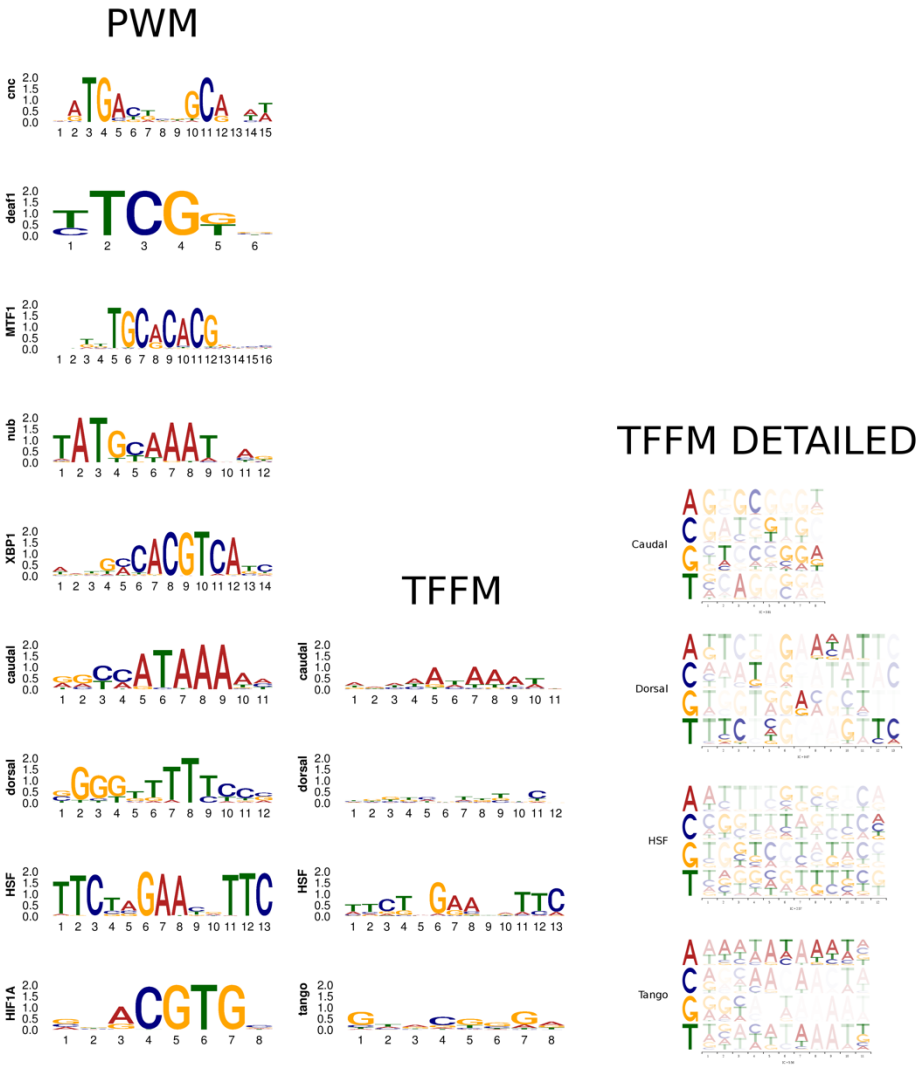
ANNEXES

7. ANNEXES

7.1 Supplementary material: Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*

Figure S1. Sequence logos for all the PWMs and TFFMs used in this work.
(A) *D. melanogaster*, and (B) humans. Notice that for the TFFMs we are plotting both the simplified version and the detailed version with the emission probabilities.

A) *D. melanogaster*



A) Humans

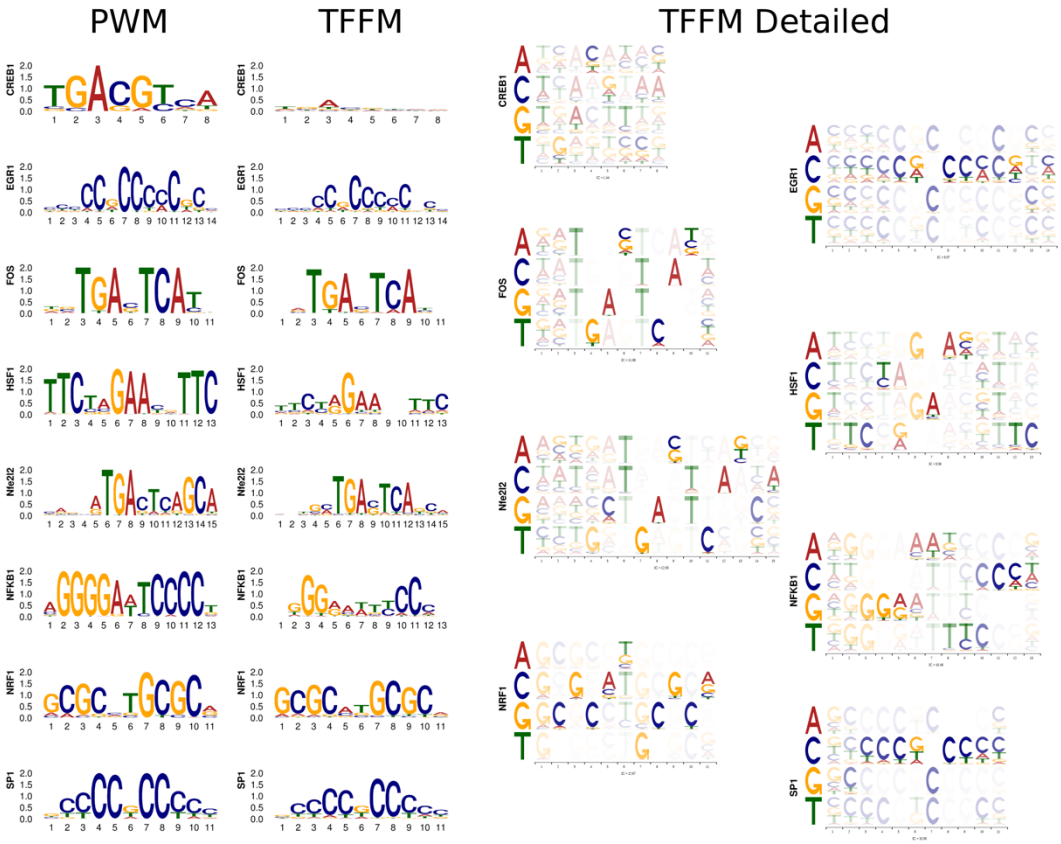

PWM
TFFM
TFFM Detailed
CREB1
EGR1
FOS
HSF1
Nfe2l2
NFKB1
NRF1
SP1

Figure S2. General linear model between TE length (≤ 10,000 bp) and the number of unique TFBSs. Spearman's rank correlation, p-value < 2.2e-16, rho= 0.882.

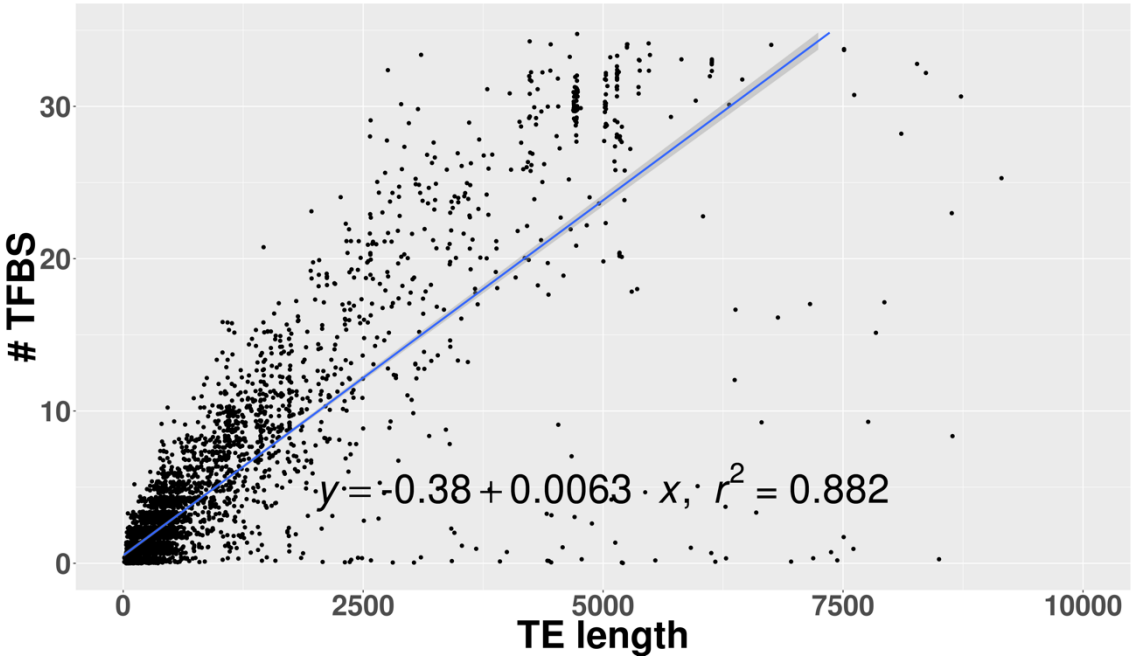


Figure S3. Percentage of TFBMs and TFBSs located in non-INE1 TEs in *D. melanogaster*. In green, motif predictions using PWMs. The vertical dotted line depicts the expected percentage of motifs falling in non-INE1 TEs (5.05%). In blue, ratio of number of motifs predicted in TEs and number of motifs predicted in background sequences with the same properties than TEs using TFFMs. Expected ratio is 1 (vertical dotted line). In orange, percentage of TFBSs predicted using ChIP-seq peaks located in TEs. Expected percentage of TFBS falling in non-INE1 TEs (vertical dotted line, 5.05%).

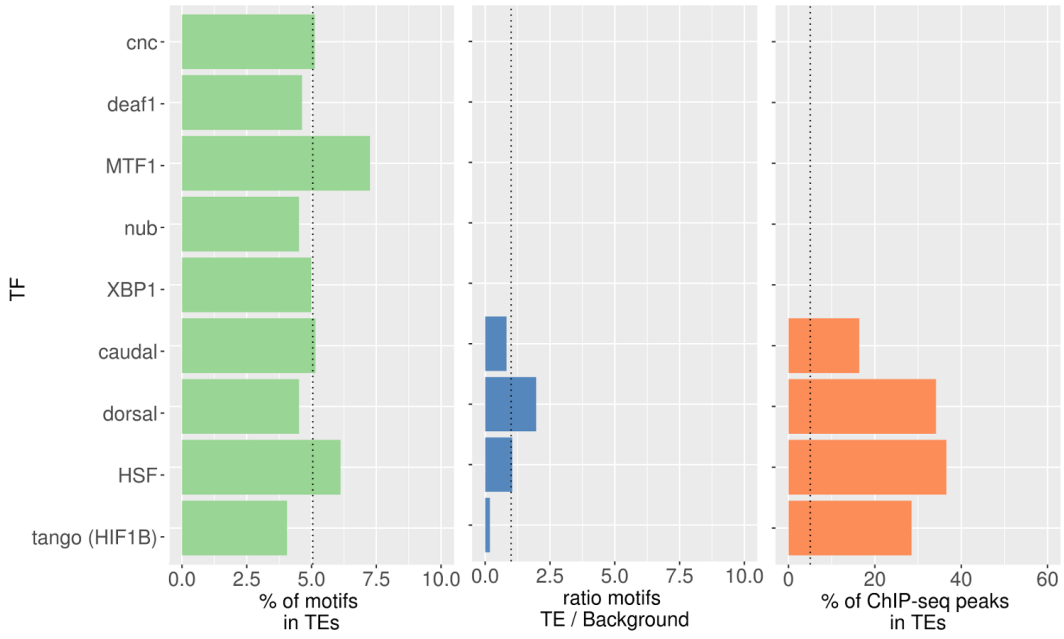
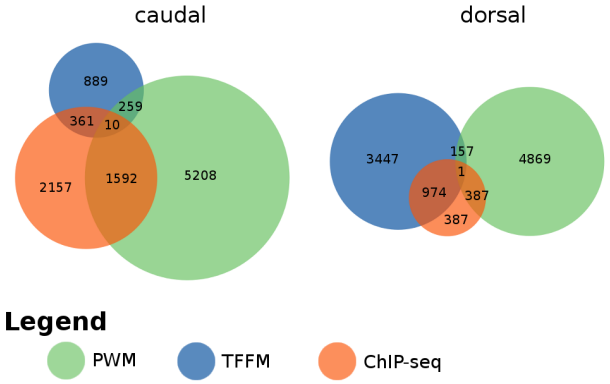


Figure S4. Overlap of motif and TFBS prediction.

Venn diagrams showing the overlap across methods (PWM, TFMM, and ChIP-seq) in the prediction of TFBM/TFBS within TEs in Drosophila (**A)** and humans (**B)**. A motif/peak is considered shared if there is overlap in their coordinates. Notice that a ChIP-seq peak can overlap with several motifs.

A)



B)

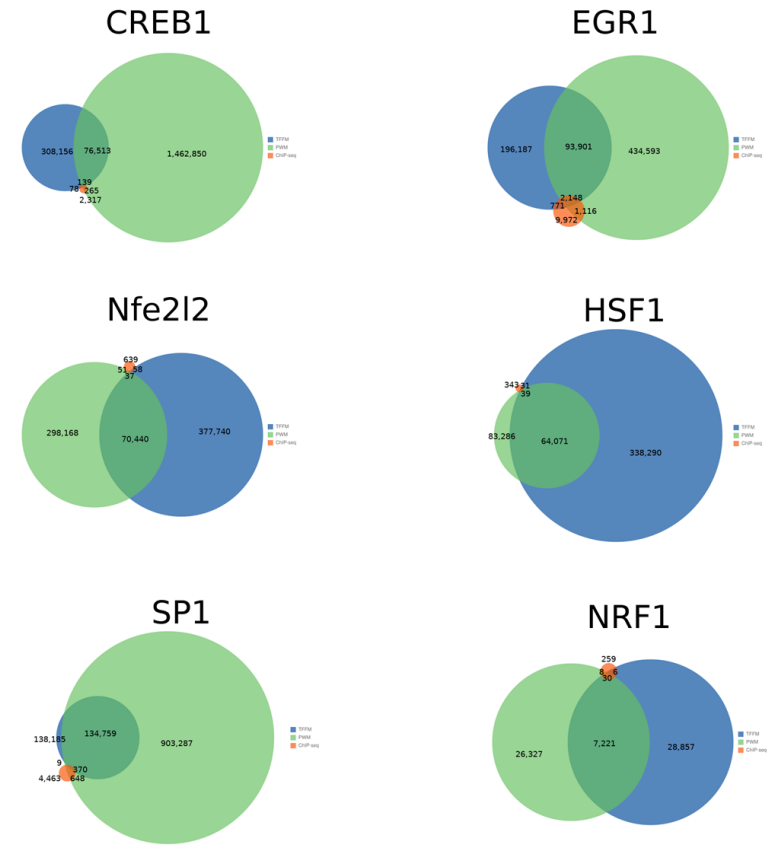


Figure S5. Distribution of the expected overlap between TEs and open chromatin regions. Random coordinates without overlap were generated for TEs and then we calculated the overlap with open chromatin regions. In green the actual overlap between TEs and open chromatin regions.

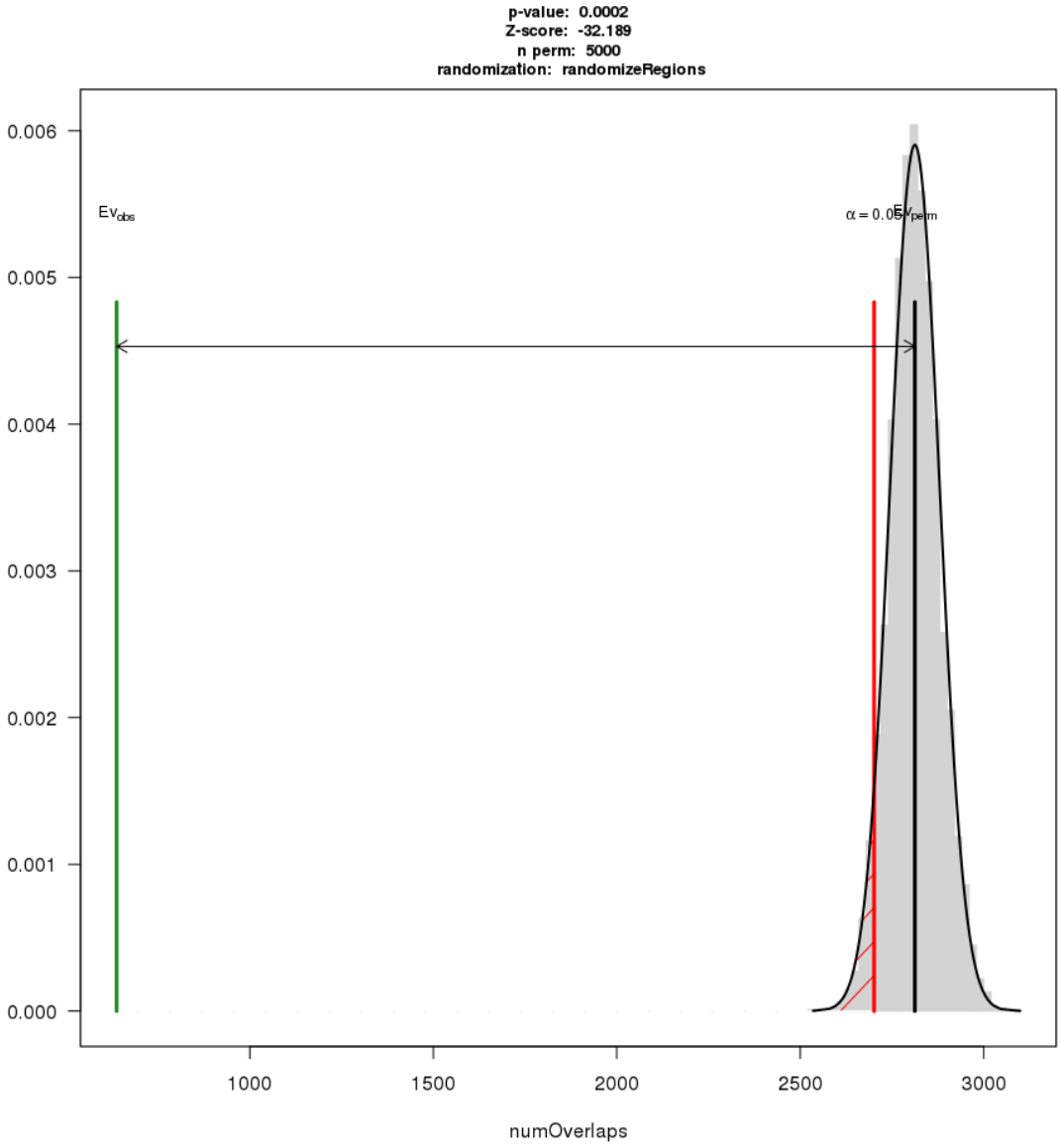


Figure S6. Stress associated genes are enriched in stress TFBMs/TFBSs.

Proportion of genes with 3 or more TFBMs/TFBSs in their promoter region (green) for each of the stresses analysed in this work. The first column shows the data for nonstress-associated genes and the second column for stress-associated genes.

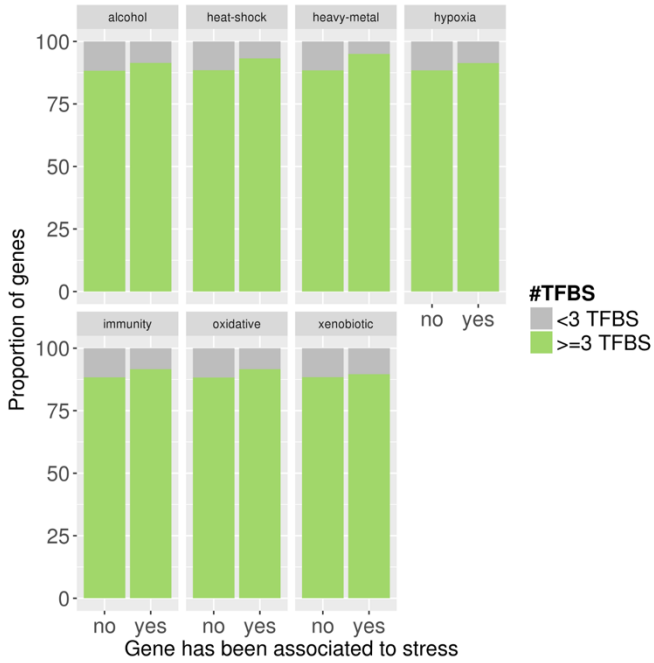


Table S1. Accession codes for the *D. melangoaster* ChIP-seq data used in this work.

TF or Histone mark	Sample name	SRR/modENCODE identifier
HSF	GSM470839: S2_20HS_HSF_IP	SRR039093
HSF	GSM470839: S2_20HS_HSF_IP	SRR039094
HSF	GSM470839: S2_20HS_HSF_IP	SRR039095
HSF_control	GSM470838: S2_20HS_PREIMMUNE	SRR039090
HSF_control	GSM470838: S2_20HS_PREIMMUNE	SRR039091
HSF_control	GSM470838: S2_20HS_PREIMMUNE	SRR039092
caudal	E0-4h_7T_ChIPSeq_1 and 2	SRR034714
caudal	E0-4h_7T_ChIPSeq_1 and 2	SRR034715
caudal_control	E0-4h_7T_INPUT_1	SRR034716
caudal_control	E0-4h_7T_INPUT_2	SRR034717
caudal	3dFemale_ChIPSeq_1 and 2	SRR063875
caudal	3dFemale_ChIPSeq_1 and 2	SRR063876
caudal_control	3dFemale_INPUT_2	SRR063877
caudal	7T_W3L_GFP_ChIPSeq_1, 2 and 3	SRR073946
caudal_control	7T_W3L_GFP_INPUT_3	SRR073949
cnc_control	CNC_16-24-seq_Input_Rep1	SRR1198694
cnc_control	CNC_16-24-seq_Input_Rep2	SRR1198695
cnc	CNC_16-24-seq_ChIP_Rep1	SRR1198696
cnc	CNC_16-24-seq_ChIP_Rep2	SRR1198697
cnc_control	CNC_AdultFemale-seq_Input_Rep1	SRR1198698
cnc_control	CNC_AdultFemale-seq_Input_Rep2	SRR1198699
cnc	CNC_AdultFemale-seq_ChIP_Rep1	SRR1198700
cnc	CNC_AdultFemale-seq_ChIP_Rep2	SRR1198701
dorsal	Dorsal ChIP-seq	SRR086224
dorsal_control	Dorsal ChIP-seq input control	SRR086225
tango	White-prepupae-WPP:ChIP-seq:Rep-1:ChIP	modENCODE ID=5025
tango	White-prepupae-WPP:ChIP-seq:Rep-2:ChIP	modENCODE ID=5025
tango	White-prepupae-WPP:ChIP-seq:Rep-3:ChIP	modENCODE ID=5025
tango_control	White-prepupae-WPP:ChIP-seq:Rep-1:input	modENCODE ID=5025
tango	Larvae-L3-stage:ChIP-seq:Rep-1:ChIP	modENCODE ID=5024
tango	Larvae-L3-stage:ChIP-seq:Rep-2:ChIP	modENCODE ID=5024
tango	Larvae-L3-stage:ChIP-seq:Rep-3:ChIP	modENCODE ID=5024
tango_control	Larvae-L3-stage:ChIP-seq:Rep-1:input	modENCODE ID=5024
H3K4me3	14-16 hr late embryo	SRR1170749
Input (H3K4me3)	15-16 hr late embryo	SRR1170751
H3K36me3	16-16 hr late embryo	SRR1170752
Input (H3K36me3)	17-16 hr late embryo	SRR1170753
H3K27me3	18-16 hr late embryo	SRR1170755
Input (H3K27me3)	19-16 hr late embryo	SRR1170756
H3K9me2	20-16 hr late embryo	SRR1170757
Input (H3K9me2)	21-16 hr late embryo	SRR1170759

Table S2. *MTF1* binding motifs identified in heavy metal response genes. A) The gene Flybase ID of heavy metal response genes and the number of MTF1 binding motifs that each one of these genes contained. Out of 195 heavy metal response genes, 143 containing MTF1 binding sites including 27 of the 29 genes that have been experimentally validated (in bold). **B)** References for the publications reporting that genes in S11A are heavy metal response genes

https://academic.oup.com/nar/article/47/13/6842/5512985#supplementary-data

Table S3. List of polymorphic and fixed TEs with evidence of selection. The recombination rate, the type of statistics and the work where it was done are given.

	Recombination rate			
TE	**Comeron_cM_Mb**	**Fiston-Lavier_midpoint_rate_cM_Mb**	**Evidence of selection**	**Reference**
FBti0020172	0	0	Fst	This work
FBti0020146	1.8	0.62	Fst	Rech et al 2019
FBti0020116	0.76	1.88	H12	Rech et al 2019
FBti0019079	1.27	2.46	H12	Rech et al 2019
FBti0019679	1.72	0.66	nSL, H12, Fst	This work
FBti0019677	0.72	0.81	H12	Rech et al 2019
FBti0019732	0.18	0.25	H12	This work
FBti0020125	2.15	1.35	Allele age	Blumenstiel et al 2014
FBti0019065	2.35	3.5	Fst, nSL / fTE / CSTV	Rech et al 2019/ Gonzalez et al 2008/ Lerat et al 2019
FBti0019613	4.16	4.23	H12	Rech et al 2019
FBti0019234	0	0	H12	Rech et al 2019
FBti0019279	0.76	1.89	H12	Rech et al 2019
FBti0019632	2.08	4.13	H12	Rech et al 2019
FBti0019617	2.26	4.24	iHS	Rech et al 2019
FBti0019318	0	0	H12	This work
FBti0018984	1.19	0	H12	This work
FBti0019601	1.81	4.04	H12	Rech et al 2019
FBti0019611	4.25	4.22	CSTV	Lerat et al 2019
FBti0018880	8.89	3.76	H12, nSL / iHS / Phenotypic	Rech et al 2019/ Gonzalez et al 2008 Gonzalez et al 2009/ Guio et al 2014
FBti0019657	2.8	1.26	iHS	Rech et al 2019
FBti0020393	0.65	0	iHS	Rech et al 2019
FBti0060442	0	0	H12	This work
FBti0019386	1.31	2.43	CL test, TajimaD, Phenotypic	Ullastres et al 2015
FBti0020091	4.08	2.91	iHS	Rech et al 2019
FBti0020086	1.04	3.12	Fst, iHS / Allele age	Rech et al 2019/ Blumenstiel et al 2014
FBti0020243	0.07	0	H12	This work
FBti0019430	1.2	3.21	H12 / TajimaD / iHS, fTE / Alllele age / Phenotypic	Rech et al 2019/ Kofler et al 2012/ Gonzalez et al 2008/ Blumenstiel et al 2014/ Aminetzach et al 2005 Schmidt et al 2010
FBti0020114	0.35	1.93	iHS, nSL	Rech et al 2019
FBti0019344	0.98	0.45	Fst	Mateo et al 2018
FBti0019360	0.65	1.38	Fst	Mateo et al 2018
FBti0060381	0	0.19	H12, Fst	This work
FBti0019294	0	0	H12	This work
FBti0019082	2.53	2.18	TajimaD	Kofler et al 2012
FBti0019771	1.17	0.42	Fst	Rech et al 2019
FBti0019081	1.18	2.29	nSL	Rech et al 2019
FBti0018916	3.15	3.74	iHS	Rech et al 2019
FBti0019747	0.09	0.08	H12	This work
FBti0019975	0.22	0.61	iHS	Rech et al 2019
FBti0020036	2.22	3.36	iHS	Rech et al 2019
FBti0019354	0.98	0.99	iHS / Allele age	Rech et al 2019/ Blumenstiel et al 2014
FBti0020149	1.32	0.47	H12, nSL / Allele age	Rech et al 2019/ Blumenstiel et al 2014
FBti0019276	0.54	0.29	CSTV	Lerat et al 2019
FBti0018937	3.36	2.77	iHS	Rech et al 2019
FBti0020207	0	0	H12	This work
FBti0019737	0.27	0.17	H12	This work
FBti0020322	1.09	0	iHS	This work
FBti0020155	1.04	0.31	Phenotypic	Zhu et al 2014
FBti0019372	7.51	1.94	H12	Rech et al 2019
FBti0020046	0.9	3.4	Allele age	Blumenstiel et al 2014
FBti0020305	0	0	H12	This work
FBti0020306	0	0	H12, Fst	This work
FBti0020301	0	0	H12	This work
FBti0019200	0.23	0	Allele age	Blumenstiel et al 2014
FBti0019985	1.08	1.84	TajimaD, iHS, H12, Phenotypic	Merenciano et al 2016
FBti0019457	1.63	2.82	FST/nSL	Rech et al 2019
FBti0019112	0.94	3.25	iHS, H12, nSL	Rech et al 2019
FBti0020224	0	0	H12	This work
FBti0020006	0	0	H12	This work
FBti0020096	1.66	2.79	iHS/ nSL	Rech et al 2019
FBti0019655	2.71	1.34	TajimaD	Kofler et al 2012
FBti0020147	0.14	0.49	TajimaD	Rech et al 2019
FBti0062854	2.5	1.55	TajimaD	Rech et al 2019
FBti0020013	2.7	2.87	Young&Long insertion	Rech et al 2019
FBti0059674	8.77	4.17	Young&Long insertion	Rech et al 2019
FBti0059794	1.88	3.75	Young&Long insertion	Rech et al 2019
FBti0019149	2.35	3.72	Young&Long insertion	Rech et al 2019
FBti0020098	2.63	2.54	Young&Long insertion	Rech et al 2019
FBti0019153	1.99	3.53	Young&Long insertion	Rech et al 2019
FBti0019191	0.12	0.13	Young&Long insertion	Rech et al 2019
FBti0020015	1.66	2.91	Young&Long insertion	Rech et al 2019
FBti0019590	12.75	3.77	Young&Long insertion	Rech et al 2019
FBti0020101	1.45	2.36	Young&Long insertion	Rech et al 2019
FBti0019199	0.23	0	Young&Long insertion / Allele age	Rech et al 2019/ Blumenstiel et al 2014

Table S4. Primers used to amplify the fragments being tested (A). Primers used to check if the cloned fragment does not contain any mutations due to the amplification step (B). Primers used to do the qRT-PCR

A)

	Primer name	Primer sequence	Enzymes	Vector	Flybase v 6 coordinates of the cloned fragments
TFBS	FBti0019012 outer F	TGATGATGAAATCGTGCGCT		PlacZattB	2R: 17,783,104..17,783,524
	FBti0019012 outer R	AGCTGCCTCGAGTACTTTGC			
	FBti0019012 inner F	sense: *AAGCTT*TCAAGATGCAAAAACCAGCGA antisense: *TCTAGA*TCAAGATGCAAAAACCAGCGA	sense: *HindIII*, antisense: *XbaI*		
	FBti0019012 inner R	sense: *TCTAGA*TGCCTCGATGCAGGTAACTT antisense: *AAGCTT*TGCCTCGATGCAGGTAACTT	sense: *XbaI*, antisense: *HindIII*		
	FBti0019309 outer F	CAACACCAACGGATTCACAA		PlacZattB	3R: 5,822,535..5,823,518
	FBti0019309 outer R	ACCCGTTCCCAAAATAGCTT			
	FBti0019309 inner F	sense: *CTCGAG*CTCCCGCTTAGAGCGTAAGA, antisense: *GGTACC*CTCCCGCTTAGAGCGTAAGA	sense: *XhoI*, antisense: *KpnI*		
	FBti0019309 inner R	sense: *GGTACC*GCGTAACGCCATACGATTTT, antisense: *CTCGAG*GCGTAACGCCATACGATTTT	sense: *KpnI*, antisense: *XhoI*		
	FBti0061428 outer F	CCGTGACAGTTCAACGACGA		PlacZattB	2L:16,857,131..16,861,438
	FBTi0061428 outer R	AGTCGCGAATGTAAATAGTAGAGAA			
	FBti0061428 inner F	antisense: *TCTAGA*ATTAACAGCCGACACACTCG	*XbaI*		
	FBti0061428 inner R	antisense: *AAGCTT*TGCCACTCACCACTCACATC	*HindIII*		
TE	FBti0019985 F	*GGTACC*CGACGTTTCTCTGCGGACTA	*KpnI*	*pGreenRabbit*	2R: 9871040..9871567
	FBti0019985 R	*GCGGCCGC*ACGAGAAGACAGCGTAGATCG	*NotI*		
	FBti0019197 inner F	sense: *CTCGAG*ACAGCTGCGGTTAAAATAATAGC antisnse: *TCTAGA*ACAGCTGCGGTTAAAATAATAGC	sense: *XhoI*, antisense: *XbaI,*	PlacZattB	2L: 18,113,230..18,114,199
	FBti0019197 inner R	sense: *TCTAGA*CAGTTTCCCCGCTTTCAGC antisense: *CTCGAG*CAGTTTCCCCGCTTTCAGC	sense: *XbaI*, antisense: *XhoI*		
Intergenic	FBti0019386 F	*GGTACC*AAGCAGACGTAACCGAAAGCA	*KpnI*	*pGreenRabbit*	3R:16188900..16190392
	FBti0019386 R	*GCGGCCGC*TTTCCTTGGCCTTTTGCAGTC	*NotI*		
	FBti0019082 F	*GGTACC*CGCGCTTCCGCTCTATAATCT	*KpnI*	*pGreenRabbit*	X:18783330..18786241
	FBti0019082 R	*GCGGCCGC*CAGCAGTCCCACTATAGACACAA	*NotI*		
	FBti0019082 part1 F	*GGTACC*CGCGCTTCCGCTCTATAATCT	*KpnI*		
	FBti0019082 part1 R	TGTTTTTGTTGTTGCGCAGGT			
	FBti0019082 part2 F	tgatacctgcgcaacaacaaaaacaCAACCAGATGGGTGACGAGA			
	FBti0019082 part2 R	*GCGGCCGC*CAGCAGTCCCACTATAGACACAA	*NotI*		
	FBti0019453 F	*GCGGCCGC*CCTCTTCGCCTGCTTTGGTA	*NotI*	*pGreenRabbit*	3R:29319602..29320694
	FBti0019453 R	*TCTAGA*ACTACTTGGCTCAAAATAGTCAGA	*XbaI*		
	FBti0019978 F	*GGTACC*CTTTCCGGAGGTGAAAACGG	*KpnI*	*pGreenRabbit*	2R:8565520..8567485
	FBti0019978 R	*GCGGCCGC*AGGCAAACTTGTTTGTTGATTCG	*NotI*		
	FBti0061578 F	*GCGGCCGC*TCCGCAGCTAAGCCGATTCT	*NotI*	*pGreenRabbit*	3R:30422306..30423491
	FBti0061578 R	*TCTAGA*GCAAAAGGACCCAAAAGGGAC	*XbaI*		
	FBti0061578 F part1	*GCGGCCGC*TCCGCAGCTAAGCCGATTCT	*NotI*		
	FBti0061578 R part1	CAGACAGCGTCGAGGGATG			
	FBti0061578 part2 F	catccctcgacgctgtctgtTGTAGCTCCGAATGGTACGC			
	FBti0061578 part 2 R	*TCTAGA*GCAAAAGGACCCAAAAGGGAC	*XbaI*		
TFBS and intergenic	FBti0018880 inner F	sense: *TCTAGA*TGTTGTTGTCCGATTTGCCTGA	*XbaI*	*PlacZattB*	2R: 18,858,124..18,860,103
	FBi0018880 inner R	sense: *CTCGAG*AAGTGCATTTTTGTGCATGGGT	*XhoI*		
	FBti0018880 F	*GAATTC*TTTATTTACCTCTGGCTTTACGCTGATG	*EcoRI*		2R:18,857,644..18,860,622
	FBti0018880 R	*GAATTC*CTCCAGCTGGATCGTGTGCAGTCCC			
Positive controls	ARE F	GGGTAAGCCCAAGCTATGGTA	*XhoI*	*PlacZattB*	3R:12,369,100..12,372,099
	ARE R	TACCTTTTGGGCCGGCTTTT	*HindIII*		
	HSE F	TGGGTTATTACCTACCCGGC	*HindIII*	*PlacZattB*	3L:9381300..9381999
	HSE R	AGCTGTTATCGCTTTGGCTTT	*XhoI*		

Table S4. Continued

B)

Primer name	Primers
PlacZattB F	GGTGGGCATAATAGTGTTGTTTAT
PlacZattB R	CGACGTGTTCACTTTGCTTGT
pGreenRabbit F	CGACGTGTTCACTTTGCTTGT
pGreenRabbit R	GTACTTCAAATACCCTTGGATCG

C)

Primer name	Primers
lacZ F	CCTGCTGATGAAGCAGAACAACT
lacZ R	CACCACATACAGGCCGTAGC
gfp F	ATGATCAGCGAGTTGCACGCC
gfp R	GACGGAAACATCCTCGGCCACA
actine F	GCGCCCTTACTCTTTCACCA
actine R	ATGTCACGGACGATTTCACG

Table S5. TE family and superfamily enrichment analysis in D. melanogaster. Nested TEs were not considered (3,768 analyzed TEs). lor_score threshold was 1.5 and only families with at least 20 copies were analyzed. A) Significant TE families enriched based on PWMs predictions. B) Significant TE families enriched based on ChIP-seq data. Families that are enriched for both PWM and ChiP-seq peaks are in bold. C) Significant TE superfamilies enriched based on PWMs predictions. D) Significant TE superfamilies enriched based on ChIP-seq data. Superfamilies that are enriched for both PWM and ChiP-seq peaks are in bold. E) Significant TE classes based on ChIP-seq data. Note that there is no significant TE class enrichement based on PWMs predictions.

https://academic.oup.com/nar/article/47/13/6842/5512985#supplementary-data

Table S6. TE family, superfamily, and class enrichment analysis in the human genome. lor_score threshold was 1.5 and only families with at least 50 copies were analyzed. A) Significant TE families enriched based on PWMs predictions. B) Significant TE families enriched based on ChIP-seq data. C) Significant TE superfamilies enriched based on PWMs predictions. D) Significant TE superfamilies enriched based on ChIP-seq data. E) Significant TE classes based on PWMs prediction. F) Significant TE classes based on ChIP-seq data.

https://academic.oup.com/nar/article/47/13/6842/5512985#supplementary-data

Table S7. For each TF, number of TEs containing at least one TFBM or 1 ChIP-seq peak located in open chromatin, containing CBP peaks, enriched for enhancer histone marks, located in proximal regulatory regions, present at high frequencies in populations and fixed in populations. We only considered significant those enriched x1.3 (in bold)

				Open chromatin			CBP peaks			Active histone marks			Proximal regulatory regions		
Number of TFBMs	TF	Dataset	Number of TEs	TEs in open chromatin	%	p-value	TEs with CBP peaks	%	p-value	TEs with active histone mark	%	p-value	Number of TEs	%	p-value
	ALL	control	5412	489	9.04	1.00	815	15.06	1	870	16.08	1	1218	22.51	1
≥ 1 TFBSs/TFBMs	Cnc	merged	846	106	12.53	0.05	98	11.58	1	33	3.90	0.01	215	25.41	0.612482071
	Caudal	merged	1912	211	11.04	0.04	262	13.70	1	214	11.19	0.04	447	23.38	1
	HSF	merged	386	40	10.36	1.00	42	10.88	0.31	9	**2.33**	**1.11E-16**	101	26.17	1
	Dorsal	merged	1431	149	10.41	1.00	233	16.28	1	103	7.20	1	325	22.71	1
	tango (HIF1B)	merged	1142	110	9.63	1.00	187	16.37	1	96	8.41	0.05	258	22.59	1
	MTF1	merged	887	100	11.27	0.31	116	13.08	1	53	5.98	1	201	22.66	1
	XBP1	merged	276	40	**14.49**	**0.04**	47	17.03	1	15	5.43	1	56	20.29	1
	Deaf1	merged	2188	230	10.51	0.26	359	16.41	1	200	9.14	1	464	21.21	1
	nub	merged	2022	226	11.18	0.02	276	13.65	1	161	7.96	1	444	21.96	1
≥ 1 ChIP-seq peaks	Caudal	chip	1188	136	11.45	0.1	115	**9.68**	**7.01118E-07**	185	**15.57**	**4.50E-10**	280	23.57	1.00E+00
	HSF	chip	712	74	10.39	1.00	59	**8.29**	**1.07517E-06**	200	**28.09**	**1.11E-16**	186	26.12	3.10E-01
	Dorsal	chip	490	34	6.94	1.00	30	**6.12**	**1.84704E-08**	188	**38.37**	**1.57E-18**	151	**30.82**	**3.19E-04**
	tango (HIF1B)	chip	1145	132	11.53	0.06	107	**9.34**	**1.85159E-07**	233	**20.35**	**9.01E-15**	270	23.58	1.00E+00

Table S8. Summary of TEs with one or more and three or more TFs. For TEs with one or more and three or more TFs, the number of TEs located in open chromatin, containing CBP peaks, enriched for enhancer histone marks, located in proximal regulatory regions.

Dataset	Total number of TEs	TEs with ≥1 TFBM/TFBS			TEs with ≥3 TFBM/TFBS		
		Number	%	p-value	Number	%	p-value
All TEs	5416	3593	66.34		2183	60.76	
TEs in open chromatin	489	354	72.39	0.0914	234	66.10	0.00716
TEs containing CBP	815	572	70.18	0.1482	362	63.29	0.04826
TEs containing active histone marks	870	441	50.69	**1.645E-10**	196	44.44	**<2.2e-16**
TEs in regulatory regions	1218	801	65.76	0.7951	463	57.80	0.1513

Table S9. Number of TEs with different chromatin marks, classified according to the number of TFBS they contain.

	TEs	no data	H3K4me3	H3K36me3
All TEs	5412	2618	286	584
≥ 1 TFBMs/TFBSs	3593	2038	171	352
≥ 3 TFBMs/TFBSs	2183	1563	70	161

Table S10. Number of TFBSs in promoter regions of genes associated and not associated with the stress responses analysed in this work

Dataset	Stress-associated genes	Number of TFBS in promoter	Number of genes	% of genes	p-value (bonferroni correction)
alcohol	no	<3	5017	31.74	2.79E-04
alcohol	no	>=3	10791	68.26	
alcohol	yes	<3	370	27.90	
alcohol	yes	>=3	956	72.10	
heat-shock	no	>=3	11603	68.49	4.11E-02
heat-shock	no	<3	5339	31.51	
heat-shock	yes	>=3	144	75.00	
heat-shock	yes	<3	48	25.00	
heavymetal	no	<3	5348	31.54	3.43E-03
heavymetal	no	>=3	11606	68.46	
heavymetal	yes	<3	39	21.67	
heavymetal	yes	>=3	141	78.33	
hypoxia	no	<3	5196	31.56	1.83E-02
hypoxia	no	>=3	11270	68.44	
hypoxia	yes	<3	191	28.59	
hypoxia	yes	>=3	477	71.41	
immunity	no	>=3	11067	68.49	1.17E-03
immunity	no	<3	5091	31.51	
immunity	yes	>=3	680	69.67	
immunity	yes	<3	296	30.33	
oxidative	no	>=3	10811	68.41	1.03E-04
oxidative	no	<3	4993	31.59	
oxidative	yes	>=3	936	70.38	
oxidative	yes	<3	394	29.62	
xenobiotic	no	<3	5004	31.44	2.27E-01
xenobiotic	no	>=3	10911	68.56	
xenobiotic	yes	<3	383	31.42	
xenobiotic	yes	>=3	836	68.58	

Table S11. Number and percentage of TEs located nearby any gene in the genome, and nearby stress-associated genes, for all TEs in the genome **(A)** and for TEs with ≥ 3TFBSs **(B)**

A)

Dataset	Total number of TEs	TEs nearby any gene	% of TEs nearby any genes	p-value	TEs nearby stress-associated genes	% of TEs nearby stress associated	p-value
All TEs	5416	3578	66.06	NA	1025	18.93	NA
High Freq	420	310	73.81	6.63E-04	121	**28.81**	**1.00E-06**
Fixed (INE1)	2234	1713	76.68	2.20E-16	465	20.81	2.50E-02
Fixed (non-INE1)	855	601	70.29	9.28E-03	178	20.82	1.62E-01

B)

Dataset	Total number of TEs	TEs nearby any gene	% of TEs nearby any genes	p-value	TEs nearby stress-associated genes	% of TEs nearby stress associated genes	p-value
All TEs	2183	1368	62.67	NA	411	18.83	NA
High Freq ≥ 3TFBS	320	241	75.31	1.88E-06	97	**30.31**	**9.16E-07**
Fixed (INE1) ≥ 3TFBS	275	209	76.00	3.42E-06	45	16.36	3.17E-01
Fixed (non-INE1)≥ 3TFBS	386	272	70.47	1.56E-03	68	17.62	6.02E-01

Table S12. Summary of all lines of evidence suggesting that TEs containing three or more TFBM/TFBS might be functional. A) TEs present at high population frequencies. B) Fixed non-INE1 TEs. *0 indicates that the TE does not have that line of evidence, and 1 indicates that the TE has that line of evidence.

https://academic.oup.com/nar/article/47/13/6842/5512985#supplementary-data

Table S13. A) delta Ct calculations for the qRT-PCR experiments **B)** Raw data of the qRT-PCR experiments

https://academic.oup.com/nar/article/47/13/6842/5512985#supplementary-data

7.2. Supplementary material: Transcriptomic and physiological basis of desiccation tolerance in natural European Drosophila melanogaster populations

Figure S1. Bar graph showing the LT50 values for each of the strains. Y axis shows the average lethal time 50 (LT50) hour when 50% of the flies in all the replicates are dead. X axis represents the individual strains. The strains marked by star (black and red) were used for water content, water loss and CHC analysis. The strains marked with red stars were used to repeat the phenotyping to assure that the results are repeatable. The strains with red circles were used for RNA-sequencing and respirometry analysis. The sensitive strain ES_Gua_15_9 was not involved in further analysis because the stock died.

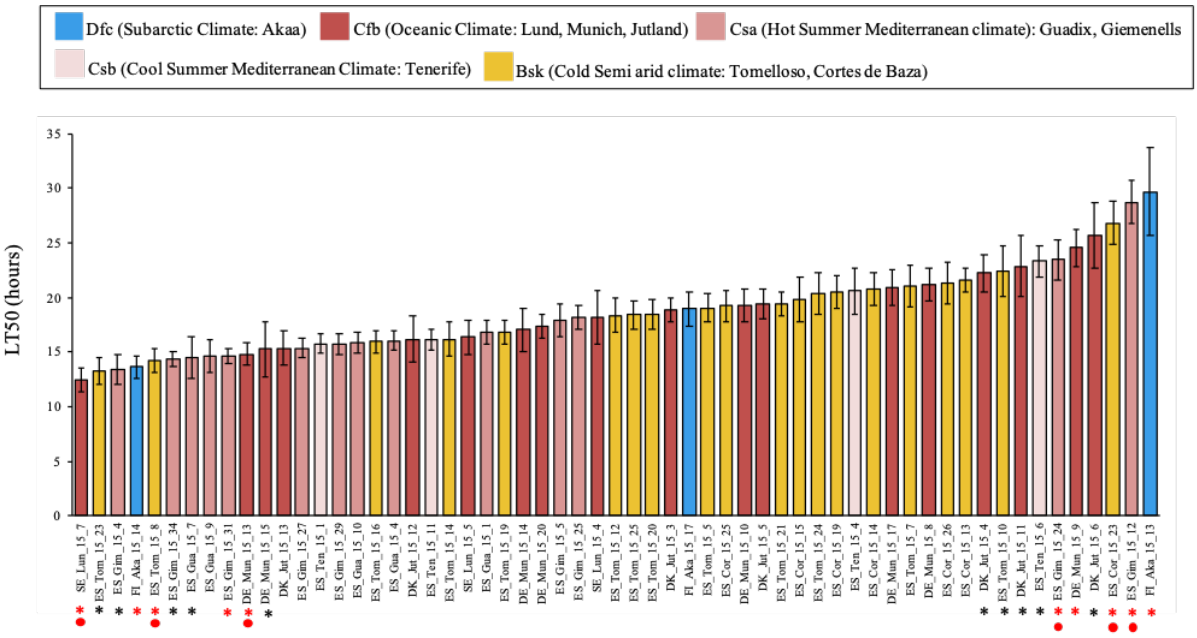


Figure S2. LT100 and LT50 values of the desiccation phenotyping experiments with five strains from the extremes of the phenotypic distribution. A) LT100 values. **B)** LT50 values. The Y axes show the hours (LT100, LT50). The strains chosen for RNA-sequencing are marked with a dot.

A) **B)**

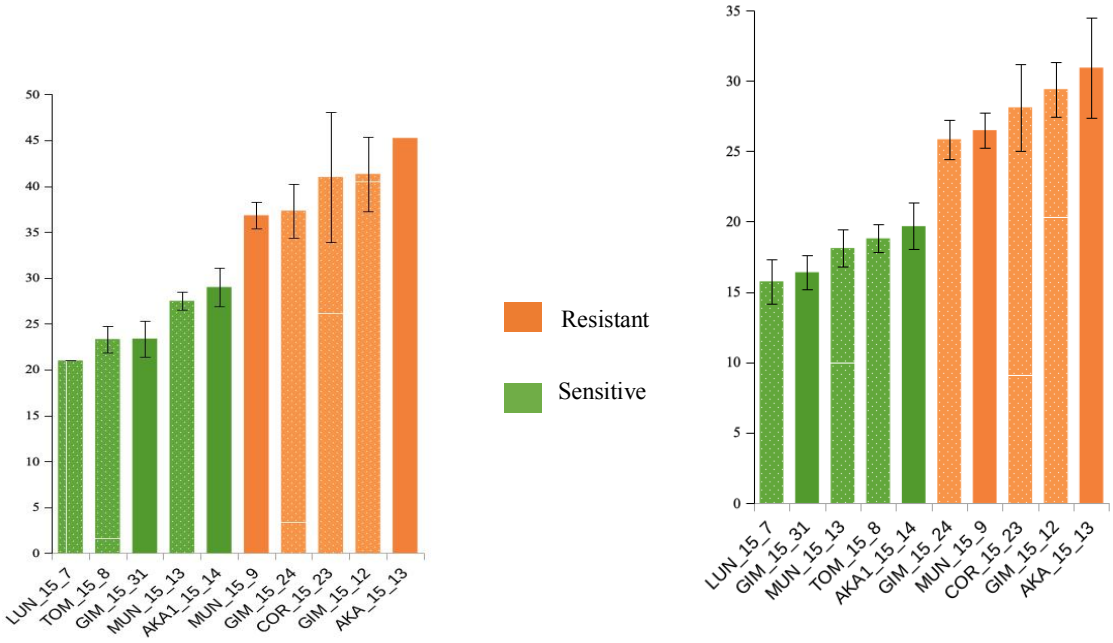


Figure S3. Clustering analysis performed using the Log normalized RPMK counts. Euclidean distance and the "Average" linkage methods were applied. CO=COR-023, G2=GIM-024, G3=GIM-012, MU=MUN-013, SW=LUN-07, TO=TOM-08

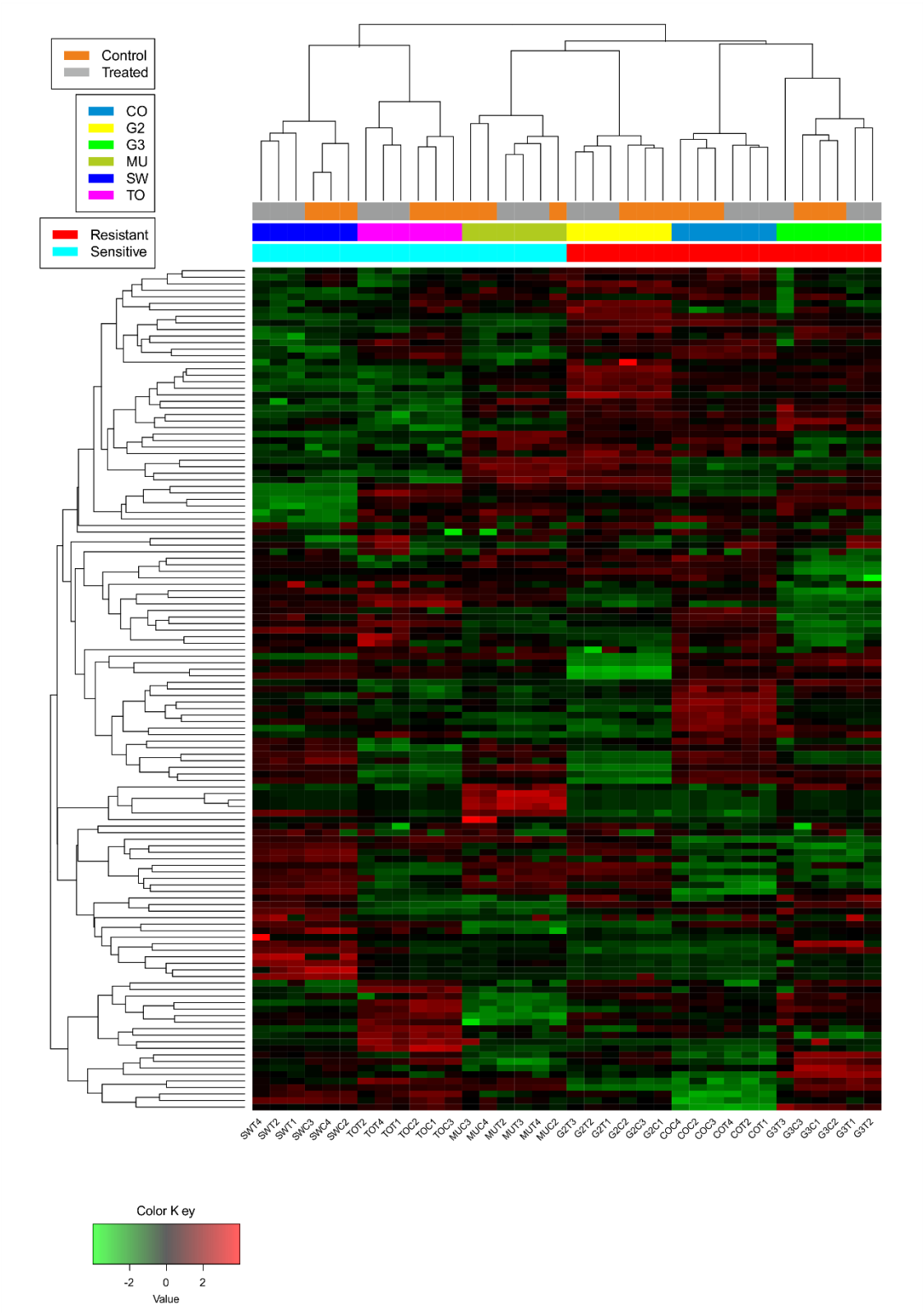
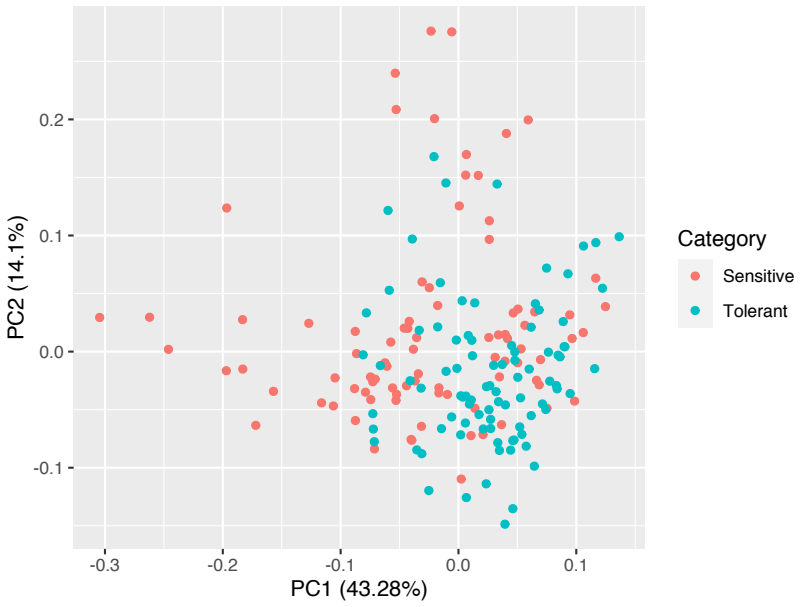


Figure S4. PCA of 10+10 strains from the phenotypic extremes and for the 6 strains selected for subsequent RNA sequencing. **A)** PC1 and PC2 of the CHC variability of the 20 strains from the phenotypic extremes comparing tolerant and sensitive strains **B)** PCA of the 6 strains used for RNA-sequencing based on the CHC.

A)



B)

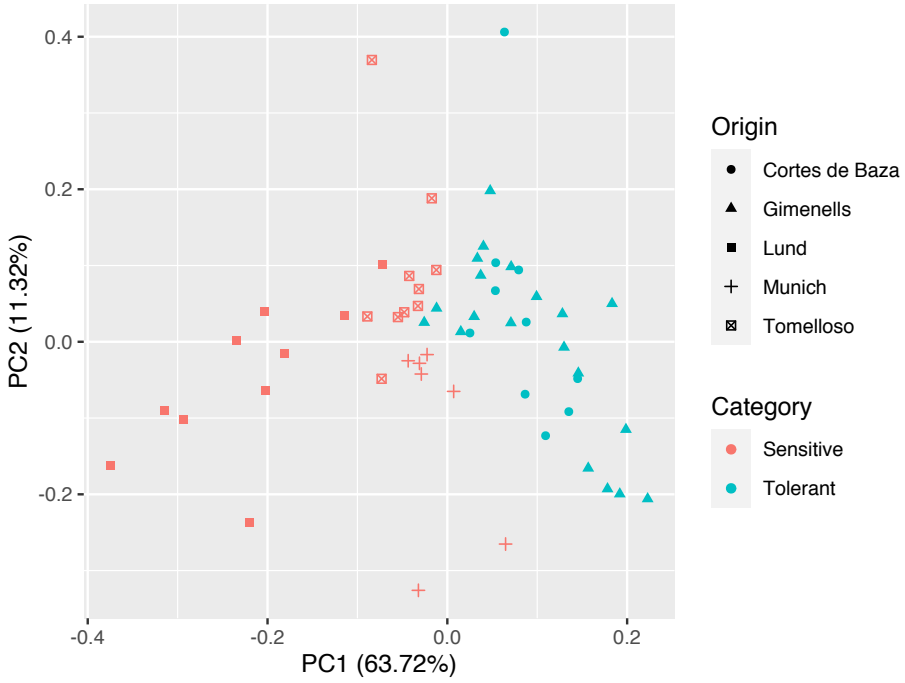


Table S1. Inbred strains used in this work and environmental and geographical variables of the populations in which they were collected. In the collection date column, the numbers in brackets represent the collection dates in summer (1) and fall (2), corresponding to collection dates before and after September 1st, respectively (Kapun et al 2020).

https://figshare.com/s/79b11a350c9a803de0a3

Tale S2. Desiccation survival assays including the LT100 and LT50 values. The temperature and humidity measurements corresponding to each experiment are also given. The numbers corresponding to each time point are the number of dead flies. A) 74 inbred strains. The 15 strains excluded from the analysis with higher than 10% control mortality are depicted in red. The results of the statistical analysis comparing the LT100 values of strains from different climate zones and the correlation between geographical and environmental variables is also shown. B) Desiccation survival assays for the 10 strains from the phenotypic extremes of the LT50 distribution. C) Desiccation survival assays for the six strains used in the RNA-seq analysis.

https://figshare.com/s/79b11a350c9a803de0a3

Table S3. Raw qRT-PCR results of the frost gene for the six strains used in the RNA-seq experiment. At least three replicates were performed for each qPCR indicated with 1, 2, 3, 4 A) LUN15-07. B) TOM15-08. C) MUN15-013 D) COR15-023. E) GIM15-012. F) GIM15-024.

https://figshare.com/s/79b11a350c9a803de0a3

Table S4. Genotypic information of the mutant and RNAi stocks used in this study (A) and the results of qRT-PCR analysis cecking the expression level of the mutated gene (B). In the bar charts, the y axis represents the gene expression level relative to Act5c.

A) Information of the mutant and RNAi stocks used in this study.

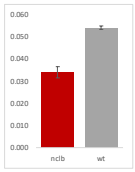
						qRT-PCR primers	
Gene	**Stock number**	**Genotype**	**Wild-type stock number**	**WT genotype**	**Driver strain used**	**Forward**	**Reverse**
Nsun2	33452	w[*] P{w[+mC]=EP}Nsun2[G493]	w1118	w1118	-	CAATTGCACTGATCCCACGC	GTAAACCGCTGGTCCACGTA
	62495	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMJ24019}attP40/CyO	36304	y[1] v[1]; P{y[+t7.7]=CaryP}attP40	w;act GAL4/TM6+tb		
Dbp73D	v36131	w1118; P{GD14287}v36131/TM3	W1118	w1118	P{tubP-GAL80ts}	GAGAAGCTGGTGGAGCAATACA	GAACCCTTTCGTGGCGTGC
	v108310	P{KK102538}VIE-260B	60100	y,w[1118];P{attP,y[+],w[3`]}	P{tubP-GAL80ts}		
Nclb	21138	y[1] w[67c23]; P{w[+mC] y[+mDint2]=EPgy2}nclb[EY15483]/CyO	6599	y[1] w[67c23]	-	GTGGAAAGTTGACGGTGAAGTG	GATTTCCTCGTTGTGGGCCT
	41826	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL01254}attP2	36303	y[1] v[1]; P{y[+t7.7]=CaryP}attP2	6g1HR-GAL4-6c		

Table S4. Continued.

B) The results of qRT-PCR analysis checking the expression level of the gene of interest

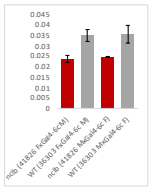
nclb #21138

	Ct *Act5c*	Ct *nclb*	delta Ct (*Act5c* - nclb)	2^deltaCt	Mean	SD	SEM
nclb repl. 1	23.00	27.73	-4.73	0.04		0.004	0.003
nclb repl. 2	21.73	26.83	-5.10	0.03	0.034		
nclb repl. 3	20.93	25.78	-4.84	0.03			
WT repl. 1	21.99	26.24	-4.25	0.05		0.001	0.001
WT rep.2	22.81	26.99	-4.18	0.06	0.054		
WT rep.3	22.46	26.67	-4.21	0.05			
					T.TEST		3.80E-03



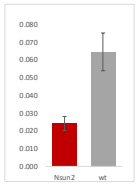
nclb #41826

	Ct *Act5c*	Ct *nclb*	delta Ct (*Act5c* - *nclb*)	2^deltaCt	Mean	SD	SEM
nclb (41826 FxGal4-6c M) repl. 1	23.37	28.75	-5.38	0.024			
nclb (41826 FxGal4-6c M) repl. 2	23.18	28.42	-5.23	0.027	0.024	0.0030	0.0017
nclb (41826 FxGal5-6c M) repl. 3	24.11	29.70	-5.60	0.021			
nclb (41826 MxGal4-6c F) repl. 1	22.85	28.18	-5.34	0.025			
nclb (41826 MxGal4-6c F) repl. 2	22.67	27.99	-5.32	0.025	0.025	0.0003	0.0002
nclb (41826 MxGal4-6c F) repl. 3	23.56	28.92	-5.36	0.024			
WT (36303 FxGal4-6c M) repl. 1	22.81	27.46	-4.64	0.040			
WT (36303 FxGal4-6c M) repl. 2	23.28	28.13	-4.84	0.035	0.035	0.0049	0.0029
WT (36303 FxGal4-6c M) repl. 3	22.92	27.98	-5.05	0.030			
WT (36303 MxGal4-6c F) repl. 2	22.79	27.83	-5.04	0.030	0.036	0.0073	0.0042
WT (36303 MxGal4-6c F) repl. 3	22.73	27.35	-4.61	0.041			
		T.TEST (41826FxGal4-6c M with 36303FxGal4-6c M)					0.028
		T.TEST(41826MxGal4-6c F with 36303MxGal4-6c F)					0.068



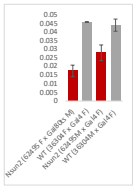
NSUN2 #33452

	Ct *Act5c*	Ct *Nsun2*	delta Ct (Act5c - Nsun2)	2^deltaCt	Mean	SD	SEM
Nsun2 repl.1	24.65	30.35	-5.70	0.02		0.007	0.004
Nsun2 repl.2	21.14	26.68	-5.54	0.02	0.024		
Nsun2 repl.3	21.41	26.37	-4.96	0.03			
WT repl.1	21.78	25.34	-3.57	0.08		0.018	0.011
WT repl.2	21.84	26.23	-4.39	0.05	0.065		
WT repl.3	22.30	26.32	-4.02	0.06			
					T.TEST		0.024



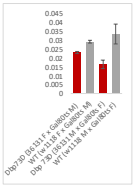
NSUN2 #62495

	Ct *Act5c*	Ct *Nsun2*	delta Ct (Act5c - Nsun2)	2^deltaCt	Mean	SD	SEM
Nsun2 (62495 F x Gal4 M) repl. 1	21.92	27.89	-5.97	0.02			
Nsun2 (62495 F x Gal4 M) repl. 2	23.39	28.87	-5.47	0.02			
Nsun2 (62495 F x Gal4 M) repl. 3	23.56	29.79	-6.23	0.01	0.017	0.005	0.003
Nsun2 (62495M x Gal4 F) repl. 1	23.65	28.49	-4.83	0.04			
Nsun2 (62495M x Gal4 F) repl. 2	22.30	27.69	-5.39	0.02			
Nsun2 (62495M x Gal4 F) repl. 3	23.01	28.37	-5.36	0.02	0.028	0.006	0.005
WT (36304F x Gal4s F) repl. 1	22.52	26.99	-4.47	0.05			
WT (36304F x Gal4 F) repl. 2	22.40	26.86	-4.46	0.05			
WT (36304F x Gal4 F) repl. 3	22.50	26.95	-4.45	0.05	0.045	0.000	0.000
WT (36304M x Gal4 F) repl. 1	22.65	27.25	-4.60	0.04			
WT (36304M x Gal4 F) repl. 2	22.76	27.10	-4.34	0.05			
WT (36304M x Gal4 F) repl. 3	22.98	27.61	-4.63	0.04	0.044	0.005	0.003
		T.TEST(62495 F x Gal80ts M with 36304F x Gal80ts M)					5.07E-04
		T.TEST(62495M x Gal80ts F with 36304M x Gal80ts F)					0.03



Dbp73D #36131

	Ct Act5c	Ct *Dbp73D*	delta Ct (*Act5c - nclb*)	2^deltaCt	Mean	SD	SEM
Dbp73D (36131 F x Gal80ts M) repl. 1	22.90	28.31	-5.40	0.0236			
Dbp73D (36131 F x Gal80ts M) repl. 2	21.95	27.43	-5.48	0.0224	0.0230	0.0009	0.0006
Dbp 73D (36131 M x Gal80ts F) repl. 1	23.20	29.22	-6.03	0.0153			
Dbp 73D (36131 M x Gal80ts F) repl. 2	21.81	27.65	-5.85	0.0174			
Dbp 73D (36131 M x Gal80ts F) repl. 3	21.33	26.84	-5.51	0.0219	0.0164	0.0034	0.0024
WT (w1118 F x Gal80ts M) repl. 1	25.29	30.34	-5.06	0.0301			
WT (w1118 F x Gal80ts M) repl. 2	22.49	27.58	-5.09	0.0293			
WT (w1118 F x Gal80ts M) repl. 3	23.89	29.06	-5.16	0.0279	0.0291	0.0011	0.0008
WT (w1118 M x Gal80ts F) repl. 1	22.31	27.22	-4.92	0.0331			
WT (w1118 M x Gal80ts F) repl. 2	22.79	27.68	-4.89	0.0337	0.0334	0.0004	0.0003
		T.TEST (36131 F x Gal80ts M with w1118 F x Gal80ts M)					0.0075
		T.TEST(36131 M x Gal80ts F with w1118 M x Gal80ts F)					0.01



Dbp73D #108310

	Ct *Act5c*	Ct *Dbp73D*	delta Ct (*Act5c - Dbp73D*)	2^deltaCt	Mean	SD	SEM
Dbp73D (108310 FxGal80ts M) repl. 1	22.892	27.834	-4.943	0.033			
Dbp73D (108310 FxGal80ts M) repl. 2	24.326	29.520	-5.194	0.027			
Dbp73D (108310 FxGal80ts M) repl. 3	23.061	28.190	-5.129	0.029	0.0295	0.0027	0.0019
Dbp73D (108310 MxGal80ts F) repl. 1	21.501	26.876	-5.374	0.024			
Dbp73D (108310 MxGal80ts F) repl. 2	23.241	28.254	-5.013	0.031	0.0278	0.0035	0.0020
Dbp73D (108310 MxGal80ts F) repl. 3	21.952	27.087	-5.135	0.028			
WT (60100 Fx Gal80ts M) repl. 1	24.390	29.676	-5.286	0.026			
WT (60100 Fx Gal80ts M) repl. 2	23.586	29.112	-5.526	0.022	0.0255	0.0037	0.0021
WT (60100 Fx Gal80ts M) repl. 2	22.572	27.678	-5.106	0.029			
WT (60100 Mx Gal80ts F) repl. 1	22.971	28.425	-5.454	0.023			
WT (60100 Mx Gal80ts F) repl. 2	24.913	30.705	-5.792	0.018	0.0208	0.0024	0.0014
WT (60100 Mx Gal80ts F) repl. 3	23.593	29.139	-5.546	0.021			
		T.TEST (108310FxGal80ts M with 60100Fx Gal80ts M)					0.202
		T.TEST (108310MxGal80ts F with 60100Mx Gal80tsF)					0.044

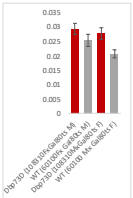


Table S5. Water content and water loss measurements with the 10 most sensitive and 10 most tolerant strains according to the LT50 distribution. A) Bulk water content B) Water loss rates

A) Bulk water content measurements.

Strain	Replicates	Body weight before drying (gr)	Body weight after drying (gr) (72h, 56 C)	%of water content (after 72h)	Average % of water content(after 72h)	Standard deviation
Sensitive strains						
SE_LUN_15_7	Replicate 1	0.01140	0.00390	65.79	45.75	7.36
	Replicate 2	0.01100	0.00620	43.64		
	Replicate 3	0.01150	0.00630	45.22		
	Replicate 4	0.01170	0.00670	42.74		
	Replicate 5	0.01130	0.00610	46.02		
	Replicate 6	0.01150	0.00680	40.87		
	Replicate 7	0.01130	0.00650	42.48		
	Replicate 8	0.01010	0.00530	47.52		
	Replicate 9	0.01130	0.00660	41.59		
	Replicate 10	0.01080	0.00630	41.67		
ES_TOM_15_23	Replicate 1	0.00910	0.00420	53.85	55.94	3.12
	Replicate 2	0.00910	0.00420	53.85		
	Replicate 3	0.00930	0.00420	54.84		
	Replicate 4	0.00910	0.00410	54.95		
	Replicate 5	0.00830	0.00300	63.86		
	Replicate 6	0.00900	0.00410	54.44		
	Replicate 7	0.00840	0.00360	57.14		
	Repicate 8	0.00900	0.00400	55.56		
	Replicate 9	0.00850	0.00360	57.65		
	Replicate 10	0.00920	0.00430	53.26		
ES_GIM_15_4	Replicate 1	0.00990	0.00510	48.48	47.80	3.28
	Replicate 2	0.01020	0.00560	45.10		
	Replicate 3	0.00970	0.00440	54.64		
	Replicate 4	0.00950	0.00490	48.42		
	Replicate 5	0.01000	0.00480	52.00		
	Replicate 6	0.01050	0.00570	45.71		
	Replicate 7	0.01110	0.00610	45.05		
	Replicate 8	0.01100	0.00610	44.55		
	Replicate 9	0.01030	0.00540	47.57		
	Replicate 10	0.00990	0.00530	46.46		
FI_AKA_15_14	Replicate 1	0.01140	0.00550	51.75	52.24	3.43
	Replicate 2	0.01010	0.00490	51.49		
	Replicate 3	0.01100	0.00540	50.91		
	Replicate 4	0.01180	0.00610	48.31		
	Replicate 5	0.01110	0.00530	52.25		
	Replicate 6	0.01100	0.00490	55.45		
	Replicate 7	0.01150	0.00590	48.70		
	Replicate 8	0.01220	0.00490	59.84		
	Replicate 9	0.01120	0.00560	50.00		
	Replicate 10	0.00950	0.00440	53.68		
ES_TOM_15_8	Replicate 1	0.00900	0.00430	52.22	54.18	1.03
	Replicate 2	0.00870	0.00420	51.72		
	Replicate 3	0.00850	0.00390	54.12		
	Replicate 4	0.00940	0.00440	53.19		
	Replicate 5	0.00840	0.00380	54.76		
	Replicate 6	0.00870	0.00410	52.87		
	Replicate 7	0.00900	0.00430	52.22		
	Replicate 8	0.00860	0.00400	53.49		
	Replicate 9	0.00880	0.00370	57.95		
	Replicate 10	0.00810	0.00330	59.26		
ES_GIM_15_34	Replicate 1	0.01150	0.00600	47.83	43.53	3.37
	Replicate 2	0.01230	0.00710	42.28		
	Replicate 3	0.01210	0.00750	38.02		
	Replicate 4	0.01140	0.00650	42.98		
	Replicate 5	0.01160	0.00630	45.69		
	Replicate 6	0.01120	0.00600	46.43		
	Replicate 7	0.01230	0.00720	41.46		
ES_GUA_15_7	Replicate 1	0.00950	0.00450	52.63	54.40	6.23
	Replicate 2	0.00970	0.00490	49.48		
	Replicate 3	0.00780	0.00320	58.97		
	Replicate 4	0.00870	0.00330	62.07		
	Replicate 5	0.00840	0.00340	59.52		
	Replicate 6	0.00790	0.00300	62.03		
	Replicate 7	0.00860	0.00370	56.98		
	Replicate 8	0.00780	0.00420	46.15		
	Replicate 9	0.00710	0.00360	49.30		
	Replicate 10	0.00790	0.00420	46.84		
ES_GIM_15_31	Replicate 1	0.00960	0.00420	56.25	50.31	4.31
	Replicate 2	0.01060	0.00540	49.06		
	Replicate 3	0.01030	0.00510	50.49		
	Replicate 4	0.01120	0.00600	46.43		
	Replicate 5	0.01020	0.00480	52.94		
	Replicate 6	0.01060	0.00570	46.23		
	Replicate 7	0.01080	0.00520	51.85		
	Replicate 8	0.01100	0.00610	44.55		
	Replicate 9	0.01010	0.00430	57.43		
	Replicate 10	0.00960	0.00500	47.92		
DE_MUN_15_13	Replicate 1	0.01070	0.00660	38.32	46.46	10.95
	Replicate 2	0.01090	0.00650	40.37		
	Replicate 3	0.01100	0.00610	44.55		
	Replicate 4	0.01190	0.00410	65.55		
	Replicate 5	0.01150	0.00700	39.13		
	Replicate 6	0.01120	0.00670	40.18		
	Replicate 7	0.00990	0.00330	66.67		
	Replicate 8	0.01060	0.00600	43.40		
	Replicate 9	0.00830	0.00420	49.40		
	Replicate 10	0.01240	0.00780	37.10		
DE_MUN_15_15	Replicate 1	0.01070	0.00570	46.73	51.45	4.08
	Replicate 2	0.01100	0.00450	59.09		
	Replicate 3	0.01020	0.00500	50.98		
	Replicate 4	0.01080	0.00580	46.30		
	Replicate 5	0.01020	0.00510	50.00		
	Replicate 6	0.01020	0.00510	50.00		
	Replicate 7	0.01050	0.00540	48.57		
	Replicate 8	0.00970	0.00450	53.61		
	Replicate 9	0.01000	0.00470	53.00		
	Replicate 10	0.00960	0.00420	56.25		

Strain	Replicates	Body weight before drying (gr)	Body weight after drying (gr) (72h, 56 C)	%of water content (after 72h)	Average % of water content(after 72h)	Standard deviation
Tolerant strains						
DK_JUT_15_4	Replicate 1	0.01020	0.00550	46.08	49.45	7.77
	Replicate 2	0.01000	0.00510	49.00		
	Replicate 3	0.01030	0.00530	48.54		
	Replicate 4	0.00980	0.00470	52.04		
	Replicate 5	0.01110	0.00580	47.75		
	Replicate 6	0.01020	0.00330	67.65		
	Replicate 7	0.01060	0.00350	66.98		
	Replicate 8	0.01010	0.00510	49.50		
	Replicate 9	0.00930	0.00440	52.69		
	Replicat e10	0.01100	0.00550	50.00		
ES_TOM_15_10	Replicate 1	0.01230	0.00740	39.84	43.47	3.54
	Replicate 2	0.01220	0.00630	48.36		
	Replicate 3	0.01200	0.00700	41.67		
	Replicate 4	0.01160	0.00690	40.52		
	Replicate 5	0.01180	0.00700	40.68		
	Replicate 6	0.01240	0.00710	42.74		
	Replicat 7	0.01250	0.00740	40.80		
	Replicate 8	0.01140	0.00570	50.00		
	Replicate 9	0.01190	0.00660	44.54		
	Replicat e10	0.01230	0.00670	45.53		
DK_JUT_15_11	Replicate 1	0.01210	0.00730	39.67	43.29	4.24
	Replicate 2	0.01100	0.00650	40.91		
	Replicate 3	0.01120	0.00630	43.75		
	Replicate 4	0.01120	0.00620	44.64		
	Replicate 5	0.01220	0.00760	37.70		
	Replicate 6	0.01180	0.00720	38.98		
	Replicate 7	0.01110	0.00550	50.45		
	Replicate 8	0.01030	0.00550	46.60		
	Replicate 9	0.01030	0.00600	41.75		
	Replicate 10	0.00970	0.00500	48.45		
TEN_15_6	Replicate 1	0.01000	0.00510	49.00	51.06	6.68
	Replicate 2	0.01050	0.00550	47.62		
	Replicate 3	0.01030	0.00550	46.60		
	Replicate 4	0.00960	0.00480	50.00		
	Replicate 5	0.00970	0.00530	45.36		
	Replicate 6	0.01000	0.00540	46.00		
	Replicate 7	0.01030	0.00340	66.99		
	Replicate 8	0.00980	0.00420	57.14		
	Replicate 9	0.00930	0.00430	53.76		
	Replicate 10	0.01060	0.00550	48.11		
ES_GIM_15_24	Replicate 1	0.00980	0.00520	46.94	44.28	2.58
	Replicate 2	0.00960	0.00540	43.75		
	Replicate 3	0.01080	0.00600	44.44		
	Replicate 4	0.01080	0.00650	39.81		
	Replicate 5	0.01130	0.00660	41.59		
	Replicate 6	0.00990	0.00520	47.47		
	Replicate 7	0.01000	0.00550	45.00		
	Replicate 8	0.01130	0.00650	42.48		
	Replicate 9	0.01030	0.00540	47.57		
	Replicate 10	0.01030	0.00580	43.69		
DE_MUN_15_9	Replicate 1	0.01310	0.00840	35.88	38.99	1.99
	Replicate 2	0.01330	0.00820	38.35		
	Replicate 3	0.01300	0.00810	37.69		
	Replicate 4	0.01290	0.00760	41.09		
	Replicate 5	0.01260	0.00730	42.06		
	Replicate 6	0.01230	0.00750	39.02		
	Replicate 7	0.01270	0.00760	40.16		
	Replicate 8	0.01320	0.00810	38.64		
	Replicate 9	0.01260	0.00750	40.48		
	Replicate 10	0.01260	0.00800	36.51		
DK_JUT_15_6	Replicate 1	0.01000	0.00520	48.00	44.85	6.65
	Replicate 2	0.01050	0.00610	41.90		
	Replicate 3	0.01130	0.00630	44.25		
	Replicate 4	0.01120	0.00650	41.96		
	Replicate 5	0.01100	0.00620	43.64		
	Replicate 6	0.01150	0.00700	39.13		
	Replicate 7	0.01180	0.00710	39.83		
	Replicate 8	0.01070	0.00610	42.99		
	Replicate 9	0.01080	0.00600	44.44		
	Replicate 10	0.01090	0.00410	62.39		
ES_COR_15_23	Replicate 1	0.01120	0.00640	42.86	41.18	2.48
	Replicate 2	0.01110	0.00660	40.54		
	Replicate 3	0.01090	0.00630	42.20		
	Replicate 4	0.01100	0.00710	35.45		
	Replicate 5	0.01050	0.00580	44.76		
	Replicate 6	0.01130	0.00680	39.82		
	Replicate 7	0.01130	0.00660	41.59		
	Replicate 8	0.01150	0.00680	40.87		
	Replicate 9	0.01070	0.00610	42.99		
	Replicate 10	0.01080	0.00640	40.74		
ES_GIM_15_12	Replicate 1	0.01240	0.00720	41.94	40.47	3.11
	Replicate 2	0.01270	0.00790	37.80		
	Replicate 3	0.01220	0.00780	36.07		
	Replicate 4	0.01270	0.00820	35.43		
	Replicate 5	0.01230	0.00730	40.65		
	Replicate 6	0.01120	0.00650	41.96		
	Replicate 7	0.01110	0.00620	44.14		
	Replicate 8	0.01180	0.00680	42.37		
	Replicate 9	0.01220	0.00730	40.16		
	Replicate 10	0.01110	0.00620	44.14		
FI_AKA_15_13	Replicate 1	0.01010	0.00530	47.52	49.30	2.35
	Replicate 2	0.01000	0.00490	51.00		
	Replicate 3	0.01050	0.00510	51.43		
	Replicate 4	0.00980	0.00480	51.02		
	Replicate 5	0.01020	0.00530	48.04		
	Replicate 6	0.01080	0.00570	47.22		
	Replicate 7	0.01060	0.00580	45.28		
	Replicate 8	0.01020	0.00480	52.94		
	Replicate 9	0.01050	0.00540	48.57		
	Replicate 10	0.01040	0.00520	50.00		

B) Water loss measurements.

Strain	Replicate number	Body weight before the experiment	Weight after 3 h in desiccation condiions	% of water loss	Average % of water loss	Standard deviation
Sensitive strains						
SE_LUN_15_7	Replicate 1	0.00630	0.0049	22.22	17.69	8.96
	Replicate 2	0.00580	0.0045	22.41		
	Replicate 3	0.00550	0.0047	14.55		
	Replicate 4	0.00620	0.0046	25.81		
	Replicate 5	0.00580	0.0056	3.45		
ES_TOM_15_23	Replicate 1	0.00470	0.0040	14.89	14.34	0.81
	Replicate 2	0.00460	0.0040	13.04		
	Replicate 3	0.00470	0.0040	14.89		
	Replicate 4	0.00570	0.0049	14.04		
	Replicate 5	0.00540	0.0046	14.81		
ES_GIM_15_4	Replicate 1	0.00610	0.0053	13.11	14.20	6.92
	Replicate 2	0.00650	0.0060	7.69		
	Replicate 3	0.00650	0.0058	10.77		
	Replicate 4	0.00590	0.0051	13.56		
	Replicate 5	0.00580	0.0043	25.86		
FI_AKA_15_14	Replicate 1	0.00470	0.0041	12.77	13.60	3.74
	Replicate 2	0.00530	0.0047	11.32		
	Replicate 3	0.00590	0.0048	18.64		
	Replicate 4	0.00500	0.0042	16.00		
	Replicate 5	0.00540	0.0049	9.26		
ES_TOM_15_8	Replicate 1	0.00530	0.0046	13.21	18.59	6.72
	Replicate 2	0.00590	0.0053	10.17		
	Replicate 3	0.00660	0.0052	21.21		
	Replicate 4	0.00550	0.0043	21.82		
	Replicate 5	0.00640	0.0047	26.56		
ES_GIM_15_34	Replicate 1	0.00460	0.0043	6.52	10.40	5.49
	Replicate 2	0.00560	0.0048	14.29		
ES_GUA_15_7	Replicate 1	0.00440	0.0035	20.45	13.84	4.83
	Replicate 2	0.00450	0.0040	11.11		
	Replicate 3	0.00420	0.0036	14.29		
	Replicate 4	0.00420	0.0038	9.52		
ES_GIM_15_31	Replicate 1	0.00600	0.0052	13.33	10.60	2.15
	Replicate 2	0.00370	0.0034	8.11		
	Replicate 3	0.00470	0.0042	10.64		
	Replicate 4	0.00500	0.0044	12.00		
	Replicate 5	0.00560	0.0051	8.93		
DE_MUN_15_13	Replicate 1	0.00550	0.0043	21.82	16.39	3.35
	Replicate 2	0.00560	0.0047	16.07		
	Replicate 3	0.00560	0.0048	14.29		
	Replicate 4	0.00610	0.0053	13.11		
	Replicate 5	0.00540	0.0045	16.67		
DE_MUN_15_15	Replicate 1	0.00490	0.0045	8.16	11.97	6.47
	Replicate 2	0.00460	0.0044	4.35		
	Replicate 3	0.00540	0.0048	11.11		
	Replicate 4	0.00520	0.0041	21.15		
	Replicate 5	0.00530	0.0045	15.09		
Tolerant strains						
DK_JUT_15_4	Replicate 1	0.00600	0.0056	6.67	10.50	3.83
	Replicate 2	0.00580	0.0051	12.07		
	Replicate 3	0.00590	0.0050	15.25		
	Replicate 4	0.00620	0.0058	6.45		
	Replicate 5	0.00580	0.0051	12.07		
ES_TOM_15_10	Replicate 1	0.00550	0.0047	14.55	9.48	4.07
	Replicate 2	0.00570	0.0052	8.77		
	Replicate 3	0.00520	0.0049	5.77		
	Replicate 4	0.00540	0.0051	5.56		
	Replicate 5	0.00470	0.0041	12.77		
DK_JUT_15_11	Replicate 1	0.00570	0.0050	12.28	10.74	4.56
	Replicate 2	0.00510	0.0049	3.92		
	Replicate 3	0.00550	0.0050	9.09		
	Replicate 4	0.00570	0.0050	12.28		
	Replicate 5	0.00620	0.0052	16.13		
TEN_15_6	Replicate 1	0.00550	0.0050	9.09	11.43	7.07
	Replicate 2	0.00510	0.0039	23.53		
	Replicate 3	0.00370	0.0035	5.41		
	Replicate 4	0.00500	0.0046	8.00		
	Replicate 5	0.00540	0.0048	11.11		
ES_GIM_15_24	Replicate 1	0.00510	0.0047	7.84	8.36	4.96
	Replicate 2	0.00570	0.0051	10.53		
	Replicate 3	0.00550	0.0049	10.91		
	Replicate 4	0.00580	0.0058	0.00		
	Replicate 5	0.00560	0.0049	12.50		
DE_MUN_15_9	Replicate 1	0.00600	0.0056	6.67	12.79	5.33
	Replicate 2	0.00560	0.0049	12.50		
	Replicate 3	0.00650	0.0057	12.31		
	Replicate 4	0.00610	0.0049	19.67		
DK_JUT_15_6	Replicate 1	0.00600	0.0053	11.67	10.77	1.41
	Replicate 2	0.00520	0.0047	9.62		
	Replicate 3	0.00630	0.0057	9.52		
	Replicate 4	0.00570	0.0050	12.28		
ES_COR_15_23	Replicate 1	0.00540	0.0048	11.11	13.59	2.62
	Replicate 2	0.00450	0.0039	13.33		
	Replicate 3	0.00490	0.0041	16.33		
ES_GIM_15_12	Replicate 1	0.00650	0.0057	12.31	7.30	3.32
	Replicate 2	0.00600	0.0056	6.67		
	Replicate 3	0.00640	0.0062	3.12		
	Replicate 4	0.00630	0.0058	7.94		
	Replicate 5	0.00620	0.0058	6.45		
FI_AKA_15_13	Replicate 1	0.00510	0.0046	9.80	11.38	3.09
	Replicate 2	0.00520	0.0048	7.69		
	Replicate 3	0.00540	0.0048	11.11		
	Replicate 4	0.00570	0.0050	12.28		
	Replicate 5	0.00500	0.0042	16.00		

Table S6. Results of the ANOVA and respirometry data analysis in control and desiccation stress conditions. A) Results of the ANOVA analysis. B) TOM-08. C) LUN-07. D) MUN-013. E) GIM-024. F) COR-023. G) GIM-012. The strains from B-D are sensitive and from E-G are tolerant. The insect mass is given in gramms. The tables from B-G are organized horizontally and the control and desiccation conditions vertically.

https://figshare.com/s/79b11a350c9a803de0a3

Table S7. Results of the cuticular hydrocarbon analysis. A) Hydrocarbons identified and their peak area in the 10 most sensitive and 10 most tolerant strains. B) CHC phenotype, Desat2 allele status, and In(3R)K inversion status of the six strains sequenced in this study, and three additional strains whose genome sequence is also available. NS= non-significant difference among the CHC ompounds, A= ancestral, D= deletion

https://figshare.com/s/79b11a350c9a803de0a3

Table S8. Differentialy expressed genes and enriched gene ontology (GO) categories in the different clusters calculated by transcriptogramer.A) All DEGs: DEGs when comparing treated versus control conditions in the six strains analyzed. B) Tolerant DEGs: DEGs in tolerant strains when comparing treated versus control conditions C) Sensitive DEGs: DEGs in sensitive strains when comparing treated versus control conditions. D) Basal DEGs: DEGs when comparing tolerant vs sensitive strains in control conditions. The tables from A-D are organized horizontally.

https://figshare.com/s/79b11a350c9a803de0a3

Table S9. Results of the PPI network analysis with STRING and MCC calculation with cytohubba. A) All DEGs. B) Tolerant DEGs C) Sensitive DEGs D) Basal DEGs. Genes considered as hubs are in bold. Tables from A-D are organized horizontally.

https://figshare.com/s/79b11a350c9a803de0a3

Table S10. Overlap among differentially expressed genes in this study and in previous studies. A) Summary of previous studies that generated a desiccation stresss response candidate gene list and the overlap with the differentially expressed genes in this study. B) Summary of differentially expressed genes of this study overlapping with the genes summarized in Table 1 of Telonis-Scott et al (2016).

A)

References	Continent	Population	Source	Significant Genes	Overlapping genes in treated conditions	P-value	Overlapping genes in control conditions	P-value
Rajpurohit et al 2018	North America	North Carolina	SNP	21	0	NA	9	**0.0131**
Clemson et al 2018	Australia	Melbbourne, Innisfail	Expression	12	0	NA	0	NA
Griffin et al 2017	Australia	Melbourne	SNP	40	2	NA	9	0.69
Canell et al 2016	-	Canton S, mutant lines	Expression (Validated)	1	0	NA	0	NA
Dembeck et al 2015	North America, Africa	North Carolina, Zambia	SNP	243	25	0.2	58	0.1
Sinclair et al 2007	North America	New Yersey	Expression	5	0	NA	1	NA
Sorensen et al 2007	Europe	Hov, Hvidovre	Expression	261	26	0.127	61	0.134
Telonis-Scott et al 2012	Australia	Victoria	SNP	423	49	1.6	106	**0.006**
Telonis-Scott et al 2016 (*)	Australia	Romsey	SNP	369	40	0.2	96	**0.0029**
Terhzaz et al 2012	-	Canton S, mutant lines	Expression (Validated)	1	0	NA	0	NA
Kang et al 2016	India	Jabalpur	SNP (sweeps, selection)	2227	261	**0.0373**	571	**< 0.00001**
			SNP (selection)	344	29	**0.007**	71	0.6
Sun et al 2018	-	Mutant strains	Expression (Validated)	1	0	NA	1	NA
Foley and Telonis-Scott 2011	America	Winters	SNP	55	6	0.8	15	0.1

* We identified the genes where the significant SNPs were located using bedtools (v. 2.27.1) and the *D. melanogaster* reference genome annotation v6.12

B)

	Differentialy expressed genes in the treated groups	**Differentialy expressed genes in the control group**
Hygrosensing (Sensory moisture receptors)	-	*Pkd2* (FBgn0041195)
		pyrexia (FBgn0035113)
		painless (FBgn0060296)
Stress sensing (Malphigian tubulus)	*Relish* (FBgn0014018)	*Nplp2* (FBgn0287423)
	dunce (FBgn0000479)	*Pkg21D* (FBgn0000442)
	Pde11 FBgn0085370	*Nplp4* (FBgn0040717)
		nanos (FBgn0002962)
Stress sensing (Stress responsive pathway)	*eiger* (FBgn0033483)	*Src64B* (FBgn0262733)
	puckered (FBgn0243512)	*happyhour* (FBgn0263395)
	basket (FBgn0000229)	*cka* (FBgn0044323)
	Jra (FBgn0001291)	*anterior open* (FBgn0000097)
	kayak (FBgn0001297)	*canoe* (FBgn0259212)
Metabolic homeostasis and water balance	*chico* (FBgn0024248)	-
Resistance mechanism: Water loss barriers	-	-
Primary hemolymph sugar/tissue-protectant	*cryptocephal* (FBgn0000370)	*Tps1* (FBgn0027560)
		cryptocephal (FBgn0000370)
		Pgm1 (FBgn0003076)

Table S11. Output of the DGET tool in each DEG group tested. A) All strains comparing control and treated conditions ("All DEGs") B) Tolerant strains comparing control and treated conditions ("Tolerant DEGs") C) Sensitive strains comparing control and treated conditions ("Sensitive DEGs") D) Comparing tolerant and sensitive strains in control conditions ("Basal DEGs"). The hub genes are marked in bold.

https://figshare.com/s/79b11a350c9a803de0a3

Table S12. Results of survival assays for the mutant and RNAi stocks used in this work. Discontinuos lines show the survival of the two compared stocks in control conditions. Solid lines show the survival of the two stocks in desiccation stress conditions. The numbers corresponding to each time point are the number of dead flies.

https://figshare.com/s/79b11a350c9a803de0a3

Table S13. De-novo TE annotations of the six genomes from this study. All the information refers to the individual genomes, not the reference.

https://figshare.com/s/79b11a350c9a803de0a3

Table S14. Normalized read counts to RPKM and differentially expressed TE families. A) Normalized read counts to RPKM for each strains and replicates. The normalization was done based on the size of each TE family. B) Differentially expressed TE families in this study. The significantly dfferentially expressed TE families are marked in bold. C1-3 and T1-3 represents the three replicates in control and treated conditions for each strain, respectively. C) Overlap of differentially expressed TE families of this study with Salces-Ortiz et al 2020.

https://figshare.com/s/79b11a350c9a803de0a3

Table S15. TEs nearby DEGs in three categories in this study. A) TEs nearby differentially expressed genes in the "all DEG" group. The overlap between the different categories is shown. B) TEs nearby DEG unique for the tolerant category. C) TEs nearby DEG unique for the tolerant group. DEGs in three categories and TEs nearby. In brackets the distance to the closest gene and the size of TE is shown respectively. YES and NO indicate the presence and absence of a TE insertion in each of the strains of this study.

https://figshare.com/s/79b11a350c9a803de0a3