UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Effects of Environment and Genetic Background on Transposable Element Activity in *Drosophila melanogaster*

Mestrado em Biologia Evolutiva e do Desenvolvimento

Ana Teresa Mendes Eugénio

Dissertação orientada por:

Élio Sucena
(Faculdade de Ciências da Universidade de Lisboa)

Patrícia Beldade
(Instituto Gulbenkian de Ciência)

2015

Resumo em Português

Os transposões são sequência de ADN repetitivas que se replicam de forma independente e que são capazes de alterar a sua posição dentro do genoma hospedeiro. Estão presentes em todos os eucariotas estudados até à data e em cerca de 80% dos procariotas (Touchon *et al.* 2007), constituindo uma porção variável dos genomas, desde 1%, como no fungo *Fusarium graminearum* (Cuomo *et al.* 2007), até cerca de 85%, como nas espécies de milho *Zea mays* e *Zea luxurians* (Tenaillon *et al.* 2011, SanMiguel *et al.* 1996). Em humanos, os transposões constituem cerca 50% do genoma (Mills *et al.* 2007) e em *Drosophila melanogaster* cerca de 20% (Mukamel *et al*. 2013, Barr *et al.* 2014).

Os transposões são geralmente classificados de acordo com o seu mecanismo de transposição, ou seja, como se movem dentro do genoma hospedeiro. Wicker *et al.* (2007) propuseram um sistema de classificação que se tornou consensual e no qual os transposões são divididos em duas classes. A Classe I compreende os retrotransposões, semelhantes a retrovírus e que transpõem através de um intermediário de ARN, são reconvertidos em ADN pela enzima Transcriptase Reversa e são reintegrados novamente no genoma, num mecanismo ao qual se designou transposição de *copiar-e-colar*. Os membros desta classe dividem-se em duas ordens – os LTRs, que possuem longas repetições terminais (*long terminal repeats*) e os não-LTRs, que não possuem essas sequências. A Classe II inclui os transposões de ADN, que utilizam um mecanismo de *cortar-e-colar*, não passando por um intermediário de ARN, e que possuem, na sua maioria, repetições terminais invertidas (TIRs, *terminal inverted repeats*)(Pray 2008).

Quando se movem dentro do genoma, os transposões podem causar mutações que contribuem para a criação de novos variantes genéticos que podem servir de matéria prima para a evolução (Wagner *et al.* 2005). A transposição pode afetar a regulação de genes fornecendo, por exemplo, promotores novos, locais de *splicing* e sinais de poliadenilação (Cowley *et al.* 2013). Há dados que indicam um papel dos transposões na adaptação em populações naturais (González *et al*. 2008) e de laboratório (Sousa *et al.* 2013) e na evolução de caracteres novos (Bourque *et al.* 2008).

Apesar de serem uma fonte de variação genética e inovação, a amplificação e mobilização de transposões tem geralmente efeitos prejudiciais ao hospedeiro, ao se inserirem em genes que codificam proteínas, alterando redes de regulação de transcrição e causando quebra de cromossomas e rearranjos genómicos a grande escala (McClintock 1951, Hedges *et al*. 2007). Para evitar estes efeitos, os organismos evoluíram mecanismos para reprimir e silenciar transposões, como os piARNs na linha germinal (Klattenhoff *et al.* 2007) e mecanismos epigenéticos como metilação e modificação da cromatina nas células somáticas (Slotkin *et al.* 2007).

Drosophila melanogaster, também conhecida por mosca da fruta ou mosca do vinagre, é um sistema laboratorial modelo para estudos de genética e é também muito usada em investigação na área da biologia evolutiva e do desenvolvimento (Arbuthnoot *et al.* 2014, Campos *et al.* 2014, Tiwari *et al.* 2015).

Existem várias ferramentas genéticas disponíveis para este organismo, inclusivamente para o estudo dos transposões que se encontram inseridos no seu genoma. Um desses recursos é o *Drosophila melanogaster Genetic Reference Panel* (DGRP) (Mackay *et al.* 2012, Huang *et al*. 2014), que consiste num conjunto de cerca de 200 linhas isogénicas completamente sequenciadas, provenientes de uma única população natural. Através de anotação *in silico* das sequências genómicas, conhece-se a posição e identidade dos transposões inseridos nessas linhas.

Apesar dos grandes avanços que têm sido feitos no sentido de caracterizar transposões e de se descobrirem mecanismos que os silenciem e/ou controlem a sua expressão, tanto nas células da linha germinal como em tecidos unicamente somáticos, ainda não se sabe ao certo o que aciona e altera a transposição dos transposões. No entanto, é reconhecido que, como em muitos outros processos biológicos, a ativação destes elementos pode ser afetada por fatores genéticos e ambientais (Capi *et al.* 2000).

A temperatura é um fator ambiental importante que afeta o *fitness* do organismo a nível de fecundidade, viabilidade e sobrevivência (Allen *et al.* 2011, Stoks *et al.* 2011, Ciota *et al.* 2014). Em plantas, sabe-se que o aumento da temperatura pode levar a um aumento da expressão dos transposões (Grandbastien *et al.* 2005). Por outro lado, em *Drosophila melanogaster*, os estudos têm sido inconclusivos, havendo experiências onde o aumento de temperatura afeta e amplifica a expressão dos transposões (Zabanov *et al.* 1990, Vasilyeva *et al.* 1999, Bubenshchikova *et al.* 2002) e outras onde não se vê qualquer efeito (Arnault *et al.* 1997, Alonso-González *et al.* 2006, Vázquez *et al.* 2007).

Um outro fator ambiental potencialmente relevante é a presença da bactéria *Wolbachia*, maternalmente transmitida e presente em muitas espécies de insetos, crustáceos e nematodes filamentosos (Stouthamer *et al.* 1999). Este simbionte confere resistência a vírus em *Drosophila melanogaster* (Teixeira *et al.* 2008, Hedges *et al.* 2008) e, tendo em conta que muitos transposões possuem características semelhantes às dos vírus (Schaack *et al.* 2010, Xiong and Eickbush, 1988), é possível que este simbionte confira também alguma proteção contra transposões.

Nesta dissertação, utilizámos linhas DGRP para estudar potenciais efeitos do genótipo e dos fatores ambientais acima descritos na atividade dos transposões de *Drosophila melanogaster*. Focámos o estudo na expressão de transposões nos ovários, pois é na linha germinal que alterações na atividade destes elementos poderão ser herdadas pela geração seguinte e contribuir para a variação genética da população.

O nosso primeiro objetivo foi validar as previsões feitas *in silico* para as linhas do DGRP relativamente à localização e identidade das inserções de transposões. Para isso, sequenciámos várias inserções do painel e confirmámos a posição e identidade para a maioria dessas inserções. Isto nunca tinha sido testado e os nossos resultados conferem alguma robustez às previsões *in silico* feitas para o painel para que possam ser usadas com maior confiança em estudos futuros.

De seguida, procurámos verificar se haveria alguma associação entre a expressão dos transposões e o número de cópias desses elementos, dentro de várias linhas do DGRP, questionando se o número de cópias poderia explicar níveis de expressão.

Observámos diferenças evidentes entre genótipos na expressão de alguns mas não todos os transposões. Também observámos que as diferenças entre os genótipos para um mesmo transposão não podiam ser explicadas pelo número de inserções.

Finalmente, testámos os efeitos da temperatura e *Wolbachia*, em várias linhas DGRP, na expressão de transposões e concluímos que ambos os fatores ambientais afetam a expressão destes elementos, embora a forma como são afetados varie de acordo com o transposão e com o genótipo.

A temperatura poderá estar a afetar transposões que se encontrem sob o efeito de *enhancers* ou fatores de transcrição sensíveis a temperatura ou a afetar os mecanismos de repressão dos transposões, como os piARNs. Tal podia ser testado procurando-se *enhancers* e fatores de transcrição cujas sequências estejam anotadas em bases de dados. Podia-se também sequenciar piARNs de uma linha do DGRP após ter sido sujeita a diferentes condições de temperatura e verificar se quando há efeito da temperatura também há alteração nos piARNs.

Não se sabe ainda como é que *Wolbachia* confere resistência a vírus e, consequentemente, como poderá estar a afetar a expressão dos transposões. Para testar se o efeito de *Wolbachia* está relacionado com os mecanismos de repressão dos transposões poder-se-iam comparar piARNs de ovários de uma linha infetada com *Wolbachia* e piRNAs da mesma linha mas na qual *Wolbachia* fora removida.

Tendo em conta que as observações feitas neste estudo foram focadas em ovários, seria também interessante testar tecidos da linha germinal masculina (testículos) e em tecidos unicamente somáticos (como os dos tóraxes) para compreendermos se os efeitos ambientais sobre os transposões são específicos de um tecido ou se é um efeito geral, afetando todo o corpo do organismo. Pensamos que é de esperar o favorecimento de um mecanismo que beneficie a geração de variabilidade genética na descendência, em caso de perturbação ambiental, pois alguns desses variantes genéticos poderão ter melhor capacidade de sobrevivência e reprodução nas novas condições ambientais.

English Summary

Transposable elements (TEs) are repetitive DNA sequences capable of changing their position within the genome, potentially causing mutations (Wagner *et al.* 2005). They play an important role in the evolution of novel gene regulation (Cowley *et al.* 2013) and novel traits (Bourque *et al.* 2008). TE mobilization, however, can also be harmful to the host, by disrupting protein-coding genes, altering transcriptional regulatory networks and causing chromosomal breakage and genomic rearrangements (Hedges *et al*. 2007).

Great advances have been made in characterizing TEs and in finding repression mechanisms to control their activity. However, what triggers and changes the transposition that leads to new insertions in natural populations is still largely unexplored. Still, it is recognized that it can be affected by both genetic and environmental factors (Capi *et al.* 2000).

Drosophila melanogaster is a genetic model vastly used for biological research, including in evolution and developmental biology (Arbuthnoot *et al.* 2014, Tiwari *et al.* 2015). There are available genetic tools and information on TEs for this species, including the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012), which contains fully sequenced isogenic lines and *in silico* predictions about TE insertions.

In this study, we used DGRP lines to test the effects of genotype and of environmental perturbations, namely temperature and *Wolbachia*, on TE expression. We first aimed at validating *in silico* predictions for TE insertions and identity in DGRPs, by sequencing insertions described in the panel. We confirmed most of the tested sites. We then searched for associations between TE expression and copy number, finding clear effects of genetic background on TE expression, albeit not the same for all TEs. Lastly, we checked for an effect of temperature and *Wolbachia* on TE expression and found that these factors affect TE expression in a different way for different TEs and genotypes.

Palavras-chave / Key-words

- Variação Genética / Genetic Variation
- Efeitos do Fundo Genético / Genetic Background Effects
- Efeitos Ambientais / Environmental Effects
- Expressão de Transposões / Transposable Elements expression

Index

Acknowledgments

This dissertation would not have been possible without the guidance and support of the amazing people that were always present for me during this journey.

First of all, I want to thank Patrícia Beldade for taking me in her group and trusting in my capacities and will. She was a great supervisor, giving me meaningful advice and autonomy to develop my ideas and being the pragmatic and assertive voice when I needed it.

Secondly, and no less important, I want to thank Marta Marialva. She is the person that most directly tutored me, as my dissertation was a branch of her PhD. I was extremely lucky to have had her by my side. She taught me all the techniques and protocols of this dissertation, being very patient and letting me learn with my mistakes. I felt that my opinions and doubts were valued and I will always be grateful for all the time and effort she dedicated me and this project.

I am also thankful to all the other members of our lab, for all opinions, interest and motivation, and also to the wonderful Evo-Devo community at IGC, especially for their feedback during labmeetings.

Finally, I want to thank my lovely family and friends, always there to motivate and support me in the good and not so good moments.

For them and everyone else that in one way or the other contributed to this project and to my growth as a scientist and as a person, thank you.

Introduction

1.1 – Transposable Elements (TEs)

Discovered by Barbara McClintock in 1948, (McClintock 1951, Pray and Zhaurova 2008) the transposable elements (TEs) are repetitive DNA sequences capable of changing their position and replicating independently in host genomes. They can be found in all eukaryotes studied so far and in about 80% of prokaryotes (Touchon *et al.* 2007), constituting from 1% of genomic sequences in some species, like the fungus *Fusarium graminearum* (Cuomo *et al.* 2007), up to 85% in some plants, like the maize species *Zea mays* and *Zea luxurians* (Tenaillon *et al.* 2011, SanMiguel *et al.* 1996). TEs constitute about 50% of the human genome (Mills *et al.* 2007) and 20% in *Drosophila melanogaster* (Mukamel *et al.* 2013, Barr et al. 2014).

TEs are usually classified according to their transposition mechanism inside the host genome. Wicker *et al*. (2007) proposed a consensual classification system, in which Class I TEs, also called retrotransposons, are similar to retroviruses and transpose via an RNA intermediary. They are first transcribed into RNA, which is then converted to DNA by the Reverse Transcriptase enzyme before re-integration in the genome. This results in a mechanism of *copy-and-paste* for transposition. Retrotransposons can be divided into two major orders - LTRs and non-LTRs, according to whether they possess Long Terminal Repeats or not, respectively (Wicker *et al*. 2007). Class II TEs, or DNA transposons, use a mechanism of *cut-and-paste,* in which TE DNA is cut out of its original location and re-integrated in a new location. Many Class II TEs have terminal inverted repeats (order TIR) (Pray 2008).

TE transposition can cause mutations, adding to the genetic variation in a population when they occur in the germline (Wagner *et al.* 2005). TEs can play an important role in the evolution of gene regulation by providing, for example, novel promoters, splice sites or polyadenylation signals (Cowley *et al.* 2013). They have been implicated in adaptation in laboratory (Sousa *et al.* 2013) and natural populations (González *et al*. 2008) and in the origin of novel traits (e.g. Bourque *et al.* 2008). For example, the primate CYP19 placenta-specific promoter and uterine-specific enhancers of prolactin are derived from transposable elements (Wagner *et al.* 2005). They are also implicated in the evolution of cichlid fish egg-spots (Santos *et al*. 2014) and the pigmentation diversity observed in medaka fish is associated with the TE Tol2 (Pray, 2008).

TEs can be a source of beneficial genetic variation and innovation, and many TE copies end up accumulating mutations at a neutral rate and eventually decay and disappear (Venner et al. 2009). However, TE mobilization and amplification can also be harmful to the host when, for example, they disrupt protein-coding genes, alter transcriptional regulatory networks, or cause chromosomal breakage and large-scale genomic rearrangements (McClintock 1951, Hedges *et al*. 2007). Since, like other mutations, TE insertions are often deleterious, host genomes have evolved mechanisms to control TE activity and protect host genome and health. In *D. melanogaster* somatic cells, TEs are mainly repressed by epigenetic mechanisms, such as methylation and modifications in chromatin condensation (Slotkin *et al.* 2007).

For example, nucleosomes associated with TE insertions can be enriched with methylated histones, repressing transcription in those genomic regions (Slotkin *et al.* 2007).

In the germline and somatic cells of *Drosophila* reproductive organs, one important mechanism of TE silencing is the piRNA pathway (Klattenhoff *et al.* 2007). PiRNAs are a class of small RNAs (Piwi-interacting RNAs), found in clusters throughout the genome (O'Donnell *et al*. 2007) and interacting with Argonaute proteins of the Piwi clade (Piwi, Aubergine, and Argonaute 3) (Aravin *et al.* 2007). There is a positive feedback loop between TE expression and piRNA biogenesis called the *ping-pong cycle* (Brennecke J *et al.* 2007), in which Piwi proteins engage in an amplification loop between piRNA clusters and active TEs. The TE transcript, loaded with antisense piRNA, is cleaved by Aubergine, triggering the production of Argonaute 3-bound piRNAs, which catalyze the production of more competent silencing piRNAs (Malone C D *et al.* 2009). Defects in the piRNA pathway and the consequent increase in TE activity have serious implications for the host genome, including the fragmentation of zygote genome during cleavage stage of embryonic divisions (Khurana *et al.* 2010) and hybrid dysgenesis phenomena, both documented for *D. melanogaster* (Bregliano *et al.* 1980, Brennecke *et al*. 2007).

Great advances have been made in characterizing TEs and in finding host repression mechanisms in somatic and germline tissues. However, what triggers and changes the transposition that leads to new insertions in natural populations is still largely unexplored. Still, it is recognized that, like many other biological processes, the mechanism is affected by genetic and environmental factors, as well as by interactions between them (Capi *et al.* 2000).

1.2 – Study System

- *DROSOPHILA MELANOGASTER* AS MODEL ORGANISM

Drosophila melanogaster, commonly known as fruit fly or vinegar fly, is a model genetics system widely used in biological research, including evolutionary and developmental biology (Arbuthnoot *et al.* 2014, Campos *et al.* 2014, Tiwari *et al.* 2015). The study of TEs in *D. melanogaster* is made easier by the fact that they are well annotated and their main mechanism of repression in the germline, the piRNA pathway, is well described (Olivieri *et al.* 2012, Muerdter *et al.* 2013, Shibata *et al.* 2015). The plethora of genetic tools and resources available for *D. melanogaster* include information about TE composition and position in different genetic backgrounds.

The *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012, Huang *et al*. 2014) consists of around 200 fully sequenced isogenic lines derived from a single natural population collected from Raleigh, North Carolina, USA, in 2003. Mackay *et al.* used whole-genome sequence data of 147 of the DGRP lines to make *in silico* predictions about TE insertions. They identified 149 families of TEs and estimated the number and position of the 147 genomes, for a total of 197,402 insertions. Based on whether they are or not found in the *D. melanogaster* reference genome (version 5.13), these insertions were divided into two groups: "shared" insertions (a total of 148,326) present in the reference genome, and "novel" insertions (49,076) not present in the reference genome. These *in silico* predictions had never been experimentally validated, and here we started to fill in this gap.

We used information about TE insertions in the DGRPs to select in which lines to test different aspects of TE activity. We reasoned that shared insertions, which are also present in most DGRP isogenic lines, are presumably more ancestral and/or of inactive TEs. Conversely, for novel insertions, which are unique to a single isogenic DGRP (around 50%) or present in just a few DGRP lines, we reasoned that they are likely to be more recent and/or of active TEs. We selected target genotypes and TEs to cover a wide range in numbers of novel insertions and quantified TE expression to check for an association with TE number and to test the effects of environmental perturbations.

- ENVIRONMENTAL PERTURBATIONS (ABIOTIC AND BIOTIC FACTORS)

It is well documented that the environment can affect the production of phenotypic variation. This includes effects on mutation rates or types, which result in the generation of novel genetic variants, as well as effects on developmental rates and trajectories, which underlie phenotypic plasticity (Beldade *et al.* 2011). Both abiotic and biotic environmental factors can play a role and, among these, external temperature and interactions with endosymbionts have well described effects in the biology of organisms and which were the focus of this study.

Temperature is an important and well-studied abiotic factor that can affect organisms' fitness traits such as fecundity, viability, and survival (Allen *et al.* 2011, Stoks *et al.* 2011, Ciota *et al.* 2014). Temperature differences have also been implicated in TE dynamics. While in plants it has been described that temperature perturbation can increase TE expression, similar studies in *D. melanogaster* have had mixed results (Grandbastien, 2005). In some studies, higher temperatures were described to increase TE expression (Zabanov *et al.* 1990, Vasilyeva *et al.* 1999, Bubenshchikova *et al.* 2002), while others described no effect whatsoever (Arnault *et al.* 1997, Alonso-González *et al.* 2006, Vázquez *et al.* 2007). The work of Ratner *et al.* (1992) and Arnault *et al*. (1994) are good examples of this. Both studied the effect of increasing temperature on the activity of TE 412, but had different results. While Ratner *et al.* reported an increase in transposition, Arnault *et al.* found no effect. It is unclear to what extent these contradicting results might be explained by genetic background effects or by differences in other environmental factors, such as bacterial infections.

The maternally transmitted endosymbiont *Wolbachia* is a genus of bacteria that infects many species of insects, crustacean and filarial nematodes (Stouthamer *et al.* 1999). *Wolbachia pipientis* is known to protect *Drosophila melanogaster* against viral infections (Teixeira *et al.* 2008, Hedges *et al.* 2008) and, since TEs are in many ways virus-like and some even produce virus-like particles (Schaack *et al.* 2010), we hypothesized that *Wolbachia* might also affect TE activity.

Here, we will use *D. melanogaster* isogenic lines and respective TE information from the DGRP to study the effect of genetic background and of environmental factors (temperature and *Wolbachia pipientis*) on TE expression, as a proxy for TE activity.

2 – Aims & Tasks

- **AIM 1**: TEST THE EFFECT OF GENOTYPE ON TE EXPRESSION

  Task 1: Experimentally validating *in silico* predictions of TE insertions and respective identity in DGRPs, in order to confirm the reliability of the panel's information about TEs.

  Task 2: Checking whether there is an association between TE expression in the ovaries and TE copy number, within different DGRP isogenic lines. We expect to see TE expression depending on number of novel insertions (TEs with higher number of novel insertions, hence higher activity, should be more expressed).

- **AIM 2**: TEST THE EFFECT OF ENVIRONMENTAL PERTURBATION AND GENOTYPE ON TE EXPRESSION

  Task 3: Checking the effect of temperature (an external abiotic factor) and of *Wolbachia* (an internal biotic factor) on TE expression in different DGRP isogenic lines. We hypothesize that temperature has an effect on TE expression, probably not the same for all genotypes, and that the levels of TE expression should be lower in the presence of *Wolbachia*.

Materials & Methods

Fly stocks, transposable elements and genotypes

- FLY STOCKS

We selected *Drosophila melanogaster* lines from the *Drosophila* Genetic Reference Panel (DGRP) (Mackay *et al.*, 2012): RAL-026, RAL-357, RAL-381, RAL-443, RAL-761, RAL-810, RAL-812, RAL-892 and RAL-908 for Task 1, RAL-021, RAL-237, RAL-321, RAL-357, RAL-358, RAL-375, RAL-391, RAL-790, and RAL-908 for Task 2 and RAL-021, RAL-237, RAL-321 and RAL-790 for Task 3.

Flies were reared at 25°C, 60% humidity, 12hr:12hr light:dark cycles and on standard food, unless otherwise mentioned.

In Task 2 and 3, some lines used were naturally infected with *Wolbachia* – RAL-021, RAL-237, RAL-321 and RAL-790, and for them we derived new lines with same genetic background from where we removed *Wolbachia*. Infected lines were treated with 0.05 mg/ml of tetracycline hydrochloride (Sigma) mixed with standard food (Teixeira *et al.*, 2008), administrated by feeding for two generations (Min and Benzer, 1997). After *Wolbachia* removal, we restored the gut flora by sterilizing embryos with 2% sodium hypochlorite and placing them in standard food mixed with gut bacterial *inoculum* from the respective non-treated lines, as described in Chrostek *et al.* (2013). We then confirmed that *Wolbachia* had been completely removed by PCR, using the same primers and amplification conditions as described in Teixeira *et al.* (2008). For Task 2 we only used flies cleaned of *Wolbachia*, for Task 3 we used both stocks, with and without *Wolbachia* infection.

- TASK 1: VALIDATING DGRP *IN SILICO* PREDICTIONS FOR TE INSERTIONS AND IDENTITY

Transposable elements: For each *in silico* prediction of a TE insertion site, there were several TEs possibly inserted, each with an associated probability based on sequence similarity (Mackay *et al.*, 2012). In this study, we considered the TE with highest probability for each insertion position. We then randomly selected 52 predicted novel insertions of ten distinct TEs (TEs in Table 1). For more information about the genomic place of each insertion, TE associated and corresponding probability of identity see supplementary Table S1.

DGRP lines: We used eight DGRP lines to confirm the 52 selected insertions (DGRP lines in Table 1). Some insertions were confirmed in more than one genotype. RAL-026 was used as negative control, as no TE insertion was predicted in the selected positions of that genotype. See also supplementary Table S1 for more information about the insertions confirmed in each DGRP line.

Table 1: TEs and DGRP lines used in Task 1, with respective number of novel insertions studied. The same insertions tested in more than one genotype are represented with the same letter. "-" indicates a line in which we did not study any insertion of that TE.

	Transposable Elements									
Line	**copia**	**opus**	**Transpac**	**I-element**	**Juan**	**F-element**	**Doc**	**hopper**	**pogo**	**hobo**
RAL-357	-	-	1^{B}	2	1^{C}	-	$1^{D}+1^{E}$	$1+1^{F}$	2	2
RAL-381	-	-	-	-	-	-	1^{D}	-	$1^{G}+1^{I}$	-
RAL-443	-	1	1	-	1	-	-	-	-	-
RAL-761	-	-	-	-	1^{C}	-	1^{D}	-	-	1^{J}
RAL-810	1^{A}	-	1	2	1	1	1	1	$1^{G}+1^{H}$	$1+1^{J}$
RAL-812	1^{A}	-	-	-	-	-	-	1^{F}	1^{H}	1^{J}
RAL-892	-	-	1^{B}	-	-	-	1^{E}	-	-	-
RAL-908	2	1	1	1	1	2	2	1	1^{I}	1
Total	4	2	5	5	5	3	8	5	8	7
TE Class	**Class I – LTRs**			**Class I – non-LTRs**				**Class II – TIRs**		

- TASK 2: CHECKING FOR ASSOCIATION BETWEEN TE EXPRESSION AND COPY NUMBER

Transposable elements: We studied the expression of the ten TEs with the highest number of novel insertions in the DGRP lines and Cr1a, an element with many shared insertions but almost no novel insertion in the DGRP (TEs in Table 2).

DGRP lines: We checked TE expression in nine DGRP lines (Table 2), albeit not all lines for each TE. These lines have variable number of novel and shared insertions for the TEs in study (number of insertions of each TE for all lines can be found in supplementary Table S3). We removed *Wolbachia* of the lines that are naturally infected with this endosymbiont.

- TASK 3: CHECKING THE EFFECT OF TEMPERATURE AND *WOLBACHIA* ON TE EXPRESSION IN DIFFERENT DGRP LINES

Transposable elements: We looked at the expression of seven TEs (in Table 3), a subset of the TEs that were also used in Task 2.

DGRP lines: We used DGRP lines naturally infected with *Wolbachia* and the lines derived from them where *Wolbachia* was removed (Table 3).

Experimental setup for temperature perturbation: For each line (with and without *Wolbachia*), five virgin females and two males (both genders were 0-8h old) were kept together in vials for three days at 25°C. Males were then removed and females were placed at different temperatures (21°C, 25°C, and 29°C) for four days (preliminary data using flies from the "wild-type" OregonR line showed changes in TE expression in ovaries at day 4 after temperature perturbation).

Table 2: TEs and DGRP lines used in Task 2, with respective number of novel l total l proportion of novel (ratio novel/total) insertions. "-" indicates a line in which the TE was not studied.

	Transposable Elements					
DGRP Line	**Cr1a**	**1360**	**blood**	**gypsy5**	**hobo**	**I-element**
	non-LTR	**TIR**	**LTR**	**LTR**	**TIR**	**non-LTR**
RAL-21	2|680|0,03	27|128|0,21	-	1|9|0,11	-	-
RAL-237	0|5|0	0|7|0	0|0|x	0|0|x	0|1|0	0|1|0
RAL-321	3|81|0,04	27|168|0,16	-	0|8|0	-	-
RAL-357	2|84|0,02	-	10|11|0,91	0|5|0	37|51|0,73	-
RAL-358	0|38|0	2|32|0,06	1|1|1	0|2|0	1|5|0,20	0|4|0
RAL-375	3|80|0,04	31|160|0,19	12|13|0,92	0|5|0	-	15|24|0,63
RAL-391	0|71|0	-	-	0|4|0	-	-
RAL-790	3|69|0,04	-	-	12|20|0,6	-	-
RAL-908	4|68|0,06	-	-	0|6|0	34|48|0,71	12|21|0,57

	Transposable Elements				
DGRP Line	**INE-1**	**jockey**	**mdg1**	**pogo**	**roo**
	TIR	**non-LTR**	**LTR**	**TIR**	**LTR**
RAL-21	-	-	-	88|91|0,97	69|78|0,88
RAL-237	3|13|0,23	0|0|x	0|1|0	0|0|x	1|4|0,25
RAL-321	218|634|0,34	37|46|0,8	-	-	87|96|0,91
RAL-357	-	-	12|19|0,63	-	70|80|0,88
RAL-358	6|55|0,11	2|3|0,67	0|2|0	1|1|1	3|8|0,38
RAL-375	-	-	-	-	75|85|0,88
RAL-391	249|755|0,33	41|50|0,82	14|17|0,82	-	73|83|0,88
RAL-790	-	-	-	-	58|68|0,85
RAL-908	-	-	12|16|0,75	26|32|0,81	103|113|0,91

Table 3: TEs and DGRP lines used in Task 3, with respective number of novel l total l proportion of novel (ratio novel/total) insertions.

DGRP Line	Transposable Elements			
	1360	blood	gypsy5	I-element
	TIR	LTR	LTR	non-LTR
RAL-21	27|128|0,21	8|8|1	1|9|0,11	10|17|0,59
RAL-237	0|7|0	0|0|x	0|0|x	0|1|0
RAL-321	27|168|0,16	5|5|1	0|8|0	11|21|0,52
RAL-790	15|124|0,12	9|9|1	12|20|0,6	8|17|0,47

DGRP Line	Transposable Elements		
	mdg1	pogo	roo
	LTR	TIR	LTR
RAL-21	7|12|0,58	88|91|0,97	69|78|0,88
RAL-237	0|1|0	0|0|-	1|4|0,25
RAL-321	8|11|0,73	13|19|0,68	87|96|0,91
RAL-790	5|8|0,63	15|21|0,71	58|68|0,85

gDNA extraction, amplification and sequencing for Task 1

A pool of ten males per line was used for genomic DNA (gDNA) extraction, with Qiagen - DNeasy Blood & Tissue Kit, following the manufacturer's instructions. Males were preferred because they have less fat, which ensures a higher efficiency of the DNA extraction and purification procedure, as seen in preliminary studies. We pooled the 10 males into microcentrifuge tubes, killed them on dry ice and then their bodies were disrupted in 180µL of Buffer ATL (Qiagen protocol) using pestles. We used 200µL of Buffer AE to elute gDNA and its concentration was measured in Nanodrop (Nanodrop Technologie, Inc.) and stored at -20°C.

For all 52 selected insertion sites on gDNA from the control line RAL-026, we performed a PCR (total volume of 10µL with 1µL of gDNA, 0.5µL of each primer 10µM) with the following program cycle: 94°C for 4 min; 94°C for 1 min; 35 cycles of 94°C for 30s, primer pair *annealing temperature* for 30s and 72 °C for 30 s; 72 °C for 5 min. We used primers flanking the site of each insertion, designed in PRIMER3 (Untergrasser *et al*. 2012) (primer sequences, PCR *annealing temperature* and expected amplicon sizes without inserted TE in supplementary Table S2).

For the other eight DGRP lines, we performed longPCR (Sigma-Aldrich - Amplification of 5-25 kb DNA with the Expand Long Range dNTPack protocol and cycle program) (total volume of 10µL, 0.2µL of gDNA, 1µL of each primer 5µM) to amplify the 52 TE insertion sequences, and checked amplicon size by gel electrophoresis (primer sequences were the same as used for the control line PCR and can be found in supplementary Table S2).

The identity of inserted TEs was tested by sequencing the products resulting from longPCR, using TermoFisher - BigDye Terminator protocol and following manufacturer's instructions. Sequencing was performed using the forward primers (same primers as used for longPCR).

The 52 sequences obtained were then compared with the *Drosophila melanogaster* transposon sequences on FlyBase.org database (Altschul *et al.* 1997) – canonical set (NT) database of the BLAST program (blastn 2.2.18, 2008). We defined TE identity for each position based on the best hit, regardless of e-score. For TEs identified which did not match DGRP *in silico* predictions or with a BLAST score lower than 200, we sequenced the other end of the corresponding amplicons using the reverse primers used in longPCR. The sequences obtained were processed in the same way as those obtained with the forward primers.

Ovary dissection, RNA extraction, and cDNA synthesis for Tasks 2 and 3

We dissected ovaries from seven day-old females (eight pairs of ovaries per replicate, eight replicates per line per treatment) in cold PBS 1x. Tissues were disrupted in 400µL of Trizol with pestles and samples were stored at -80°C until RNA extraction.

We used Zymo Research - Direct-zol RNA Miniprep kit for RNA extractions, following the manufacturer's instructions. We eluted total RNA in 25µL of RNase-free water (Sigma). RNA purity and concentration of each sample was measured with Nanodrop and typically ranged from 200 to 800 ng/µL.

All RNA samples of the same line (with and without *Wolbachia*) were processed on the same day, first for removal of gDNA contamination and then for cDNA synthesis. We took 1µg of each RNA sample (dilution in RNase-free water) and removed contaminating genomic DNA using DNAse (Promega) treatment, following manufacturer's protocol (total volume of 10µL), and confirmed that all gDNA had been removed by electrophoresis.

For cDNA synthesis, we followed the Reverse Transcription System (Promega) protocol (10µL of not denatured RNA, for a total volume of 25µL reaction), using Oligo dT primers (0.5µM, 1µL) and incubating the reaction at 42°C for 60min. Samples were then heated at 95°C for 5min and then incubated on ice for 5min to inactivate the Reverse Transcriptase. cDNA was diluted in RNAse-free water (1:10) and stored at -20°C until quantitative real-time PCR (qPCR).

Quantitative real-time PCR (qPCR) for Tasks 2 and 3

We measured TE expression by qPCR (BioRad CFX384 thermal cycler) using 5µL SyBR green mix (BioRad), 0.4µM primers, 4µL of diluted cDNA (1:10) and the following program: 50°C for 2min; 95°C for 10min; 40 cycles of 95°C for 30s, 60°C for 1min and 72°C for 30s. Primers for target TEs plus RpL32 (control gene) were designed not spanning an intron (1360, Cr1a, gypsy5, hobo, INE-1, jockey and pogo were designed in PRIMER3, blood was obtained from Handler *et al*. 2011, I-element and roo from Specchia *et al*. 2010, mdg1 from Navarro *et al.* 2009 and RpL32 from Ponton *et al*. 2010) (see primer sequences and amplicon sizes in supplementary Table S4). qPCR melting curves were analyzed to confirm specificity of amplified products.

We used standard curves to calculate the concentration (in nanograms) of amplicon DNA from qPCR Ct values for each gene and qPCR plate. For that, we first obtained a known concentration of each amplicon (for each TE and control gene) by amplifying it from gDNA of OregonR (commonly used "wild-type" *D. melanogaster* line) and the same primers used for the qPCR. PCR amplicons were cleaned using a PCR clean-up kit (Macherey-Nagel - NucleoSpin Gel and PCR Clean-up), following the manufacturer's instructions. DNA was eluted in 15µL of Buffer NE and post-cleanup DNA concentration was measured in Nanodrop. We prepared eight 1:5 serial dilutions of each amplicon, diluting samples in RNAse free water, and used 4µl for qPCR reactions. We included a range of dilutions that included the range of Ct values for the actual qPCR on cDNA template, down until we reached a *plateau* in Ct values corresponding to decreasing template concentrations. The point at which the *plateau* is reached gives us information about the minimum concentration and maximum Ct value that can be detected. For all target TEs and control gene, we obtained Ct values lower than the detection threshold. Standard curves were used to obtain absolute values of expression and those values were normalized to the "housekeeping" gene RpL32, which is commonly used as reference (Parnell *et al.* 2006, Kemp C *et al.* 2007, Becker T *et al.* 2010, Haghaveghi A *et al.* 2010, Wu M *et al.* 2010), to ensure that we were using comparable levels of expression between lines. Standard curves allowed us to control for plate effects.

For each TE, we ran eight biological replicates of all experimental conditions that we wanted to compare (temperature and *Wolbachia* state) in the same plate. For each cDNA sample, we also included two technical replicates in two separate reactions ran on the same plate. In every plate, we also ran the control gene RpL32, four negative controls for each TE and control gene (RNAse free water replacing cDNA) and the standard serial dilutions for the TEs tested and control gene.

qPCR Data analysis

For each qPCR plate, we obtained Ct values using default threshold settings (BioRad CFX Manager). Biological replicates were only considered for the analysis if the standard deviation obtained for Cts of the two technical replicates was inferior to 0.5 (as advised in *Real-Time PCR* – Advanced Methods Series, edited by M. Tevfik Dorak, Oxford: Taylor & Francis, 2006).

We calculated the average Ct of the technical replicates for each biological and transformed those values into absolute expression (ng DNA / µl) using a linear equation obtained for each standard curve, as described in the qPCR BioRad protocol. We then normalized TE expression to RpL32 by dividing TE expression for RpL32 expression.

The statistical analysis of the data was performed in R Studio, version 3.2.2 (rstudio.com). We checked for normal distribution of our data using a Shapiro-Wilk test, and, as it was not normally distributed (alpha=0.05), we chose a general linear model (glm) with quasi-poisson distribution. For Task 2 we used the model glm(TE expression/RPL32 expression ~ genotype, family=quasipoisson). For Task 3 we started with the most complex model glm(TE expression/RPL32 expression ~ temp*wolb, family=quasipoisson), that considers the effect of temperature, *Wolbachia* and the interaction of the two, and compared with the simplified model glm(TEexp/RPL32exp ~ temp+wolb, family=quasipoisson), that does not contemplate the interaction of the factors, using anova. If these two models were not statistically different (p=0.05), we proceeded with the most simplified model, otherwise we used the most complex, to make pairwise comparisons using a least-squares means (lsm) with a Tukey's range test (alpha=0.05).

Results

Most *in silico* predictions for TE position and identity in tested DGRP lines were confirmed

In silico predictions for TE insertions in DGRP lines (Mackay *et al.* 2012) have never been confirmed experimentally. In Task 1, we aimed at validating those predictions so that we could use the data in our experiments.

We sequenced 40 positions in the genome of eight DGRP lines (Table 1). Some of the predicted insertion positions were present in more than one of the selected DGRP genotypes so, in total, we aimed at validating predictions for 52 novel insertions of a total of ten different TEs (Table 1).

We used as negative control the line RAL-026, predicted to not contain any of our target insertions. The PCR amplicons for the control line of the genomic locations of our 40 target insertion positions were all smaller than 400bp. This is shorter that the predicted inserted TEs (all larger than 400bp) and confirmed that there were no TE insertions in those positions in RAL-026 (see supplementary Table S1).

For the eight DGRP lines in study, we amplified the 52 insertions by longPCR and observed that all amplicon sizes were consistent with there being an insertion at the predicted locations. This corresponds to 100% validation of predicted insertion locations. Related to the size of the 52 amplicons, we found 16 amplicons (31%) that had the size corresponding to the length of the predicted full TE (*cf. transposon sequence set* on flybase.org), 13 amplicons (25%) that were longer than the predicted TE (suggesting possible insertion of more than one TE) and 23 insertions (44%) that were longer than the corresponding amplicon from the control line, but smaller than the sequence of the predicted TE (suggesting insertion of incomplete TEs) (see Table 4 and supplementary Table S1).

From sequencing the longPCR amplicons from one or both ends, we could determine if the TE amplified had the identity predicted by the *in silico* analysis of the DGRP genomic data. By blasting the insertion sequences to TE databases, we established that out of 52 amplicons, 48 (92%) had a TE of the correct predicted identity while the remaining four had other TEs inserted. In this situation were a predicted F-element, in one position of a line, which our analysis established that was, in fact, a pogo element, and a Doc element, predicted for one other position in three of the lines, that was a Stalker element (see Table 4 and supplementary Table S1). Therefore, out of the 40 positions in study, we found a total two mismatched TE identities.

Table 4: DGRP *in silico* predictions confirmation for a total of 52 insertions of ten TEs.

TE	Number of insertions tested	Amplicon Size			ID confirmation
		Inferior	Correct	Superior	
copia	4	-	1	3	4
opus	2	2	-	-	2
Transpac	5	-	1	4	5
I-element	5	4	1	-	5
Juan	5	2	1	2	5
F-element	3	-	3	-	2
Doc	8	2	4	2	5
hopper	5	-	4	1	5
pogo	8	8	-	-	8
hobo	7	5	1	1	7
TOTAL	52	23	16	13	48

Looking at the size, the amplicons corresponding to predicted insertions of opus and pogo were always smaller than predicted if those elements were complete. Only for F-element insertions did amplicons always have the correct size for that TE, even though one predicted insertion of this element did not correspond to the predicted TE identity (Table 4). Although *in silico* predictions of TE insertions in the DGRPs might be over-estimating insertions of complete TEs (only 31% with correct size for predicted TEs), we did confirm 100% of predicted insertion sites and 92% of corresponding predicted TE identities. Based on these validations, we were confident about using the *in silico* predictions of TE insertions to select which DGRP lines to study genotype effect on TE expression, testing for an association between TE copy number and levels of TE transcript in adult ovaries.

Genotype affects TE expression, albeit not equally for all TEs

TE expression should not be used as a proxy for TE activity without taking into account TE copy number in the genome. We set to test the correlation between TE copy number and TE expression levels using some selected DGRP lines. We selected lines with different numbers of copies of particular target TEs, paying special attention to the number of "novel insertions" (not shared between DGRP lines and the *Drosophila* reference genome, and typically also not shared between many DGRPs; Mackay *et al.* 2012), because of the expectation that "novel" insertions are more likely to correspond to active TEs than those for which insertions are "shared" (with the reference genome and typically also between DGRPs; see Introduction). We expected that genotypes with more novel insertions would also have higher expression levels of the corresponding TEs.

We chose nine DGRP lines (RAL-021, RAL-237, RAL-321, RAL-357, RAL-358, RAL-375, RAL-391, RAL-790 and RAL-908) and 11 TEs (seven *copy-and-paste*—blood, Cr1a, gypsy5, I-element, jockey, mdg1, and roo, and four *cut-and-paste* elements—1360, hobo, INE-1, and pogo). We expected Cr1a, an element with many shared but almost no novel insertions, to be an inactive TE and to have low expression. Conversely, we expected TEs with more novel than shared insertions (e.g. blood, gypsy5, hobo, I-element, INE-1, jockey, mdg1, pogo and roo) to be active and expressed at higher levels (Table 2).

We did not test all TEs in all DGRP lines, only Cr1a, gypsy5, and roo were tested in all selected DGRPs. For the remaining TEs, we generally quantified expression in four lines, two with high and two with low numbers of novel insertions for those TEs (Table 2). We focused on TE expression in ovaries because this is a tissue where increased TE activity is expected to impact heritable variation in *copy-paste* TE number and lead to increase in *copy-paste* TE copy number between generations.

In general, we observed no association between the number of novel or total insertions and the expression levels for the corresponding TEs in adult ovaries. The only TEs for which we saw higher levels of expression in genotypes with more novel insertions were pogo and gypsy5 (Figure 1). For the other TEs, we either saw decreased expression with increasing copy number (mdg1, roo, 1360) (Figure 2) or no directional trend in the association between copy number and expression (Cr1a, blood, I-element, jockey, INE-1, and hobo) (Figure 3). As complement, see supplementary Figure S1. As expected, Cr1a, with few novel insertions was expressed at low levels relative to most elements tested. However, so did INE-1, for which we had many novel insertions, and Jockey, with a number of novel insertions comparable to 1360, hobo, and pogo (Figures 1-3). We observed high gypsy5 expression in line RAL-790, comparing with other genotypes and TEs (Figure 1). RAL-790 had an unusually high number of gypsy5 insertions (20 total, 60% novel) while most other DGRPs had either close to no insertions of this element or few insertions, that tended to be shared with the reference genome and between DGRP lines.

Relative to the different DGRP lines, we did not find any genotype particularly permissive to general TE activity (judged as high levels of expression). The same genotype could have high expression of some TEs but not for others. RAL-237 and RAL-358 were the lines with the lowest number of novel insertions for the TEs in study and, therefore, we expected TE expression to be lower in them, but that was not the case – TE expression was not especially low in these lines when comparing with the others. In fact, RAL-237 was the line that showed the highest expression levels for 1360, hobo, INE-1, jockey, mdg1, and roo (Figures 2 and 3).

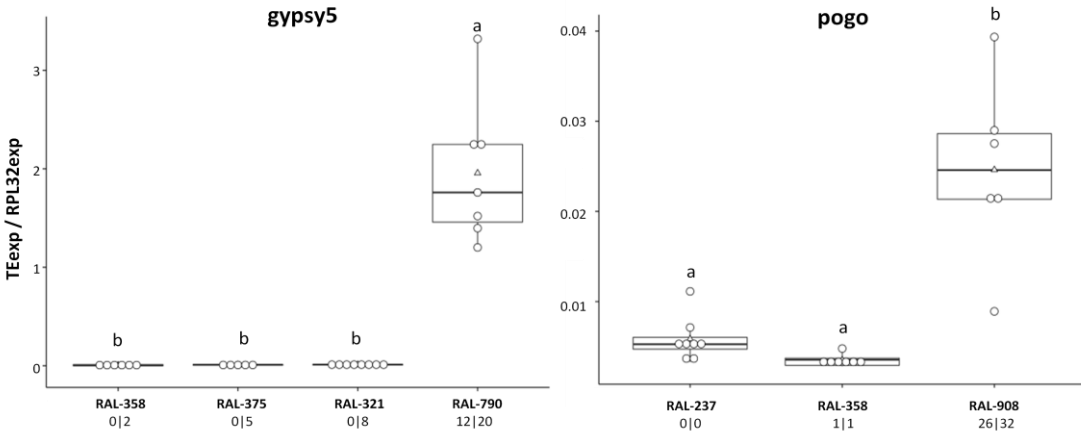

Figure 1: Positive correlation between gypsy5 and pogo expression (normalized to RpL32), in ovaries of DGRP lines, and respective different predicted copy numbers. The DGRPs (RAL lines) in the X-axes are ordered by number of novel insertions. The numbers underneath each DGRP name correspond to the predicted number of novel and total insertions for the corresponding TE. Each white circle corresponds to one biological replicate, white triangles represent the mean and the black line the median. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. Model: glm(TEexp/RPL32exp ~ genotype, family=quasipoisson). Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference.

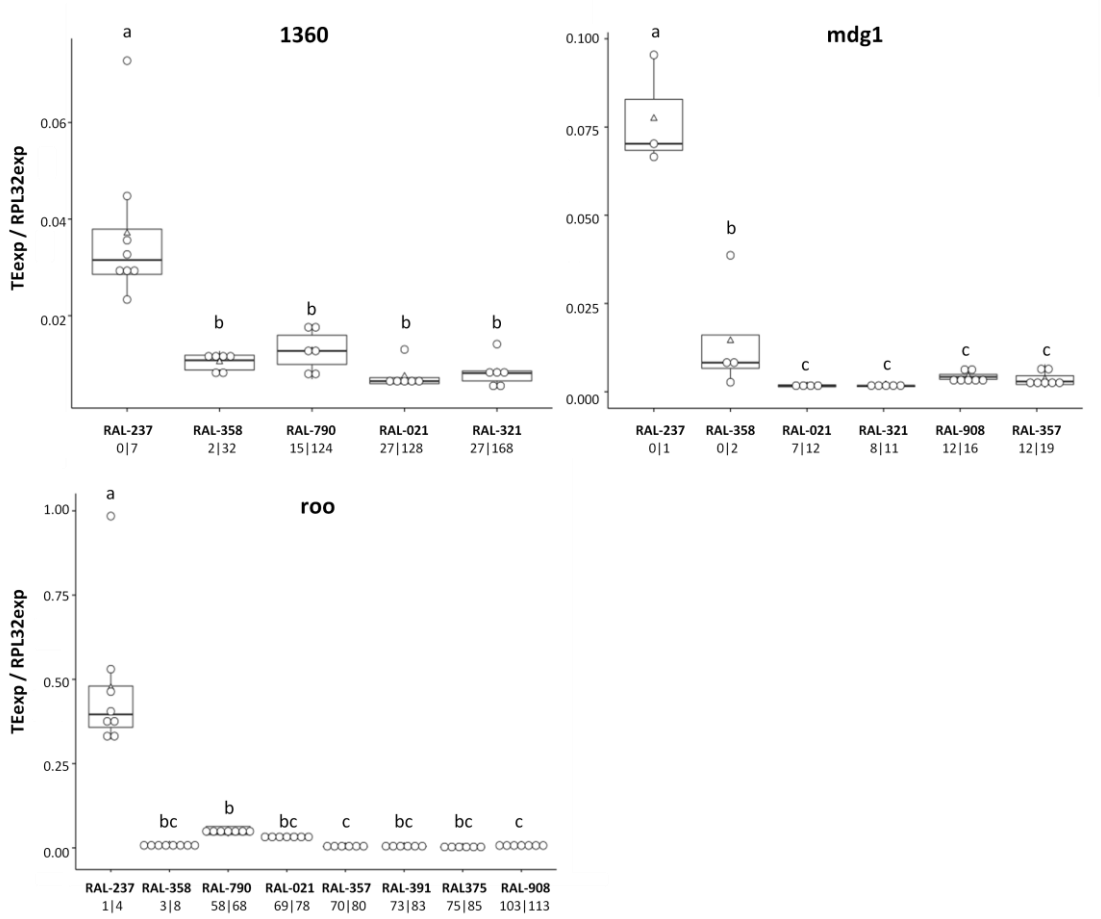

Figure 2: Negative correlation between 1360, mdg1 and roo expression (normalized to RpL32), in ovaries of DGRP lines, and respective different predicted copy numbers. The DGRPs (RAL lines) in the X-axes are ordered by number of novel insertions. The numbers underneath each DGRP name correspond to the predicted number of novel and total insertions for the corresponding TE. Each white circle corresponds to one biological replicate, white triangles represent the mean and the black line the median. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. Model: glm(TEexp/RPL32exp ~ genotype, family=quasipoisson). Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference.

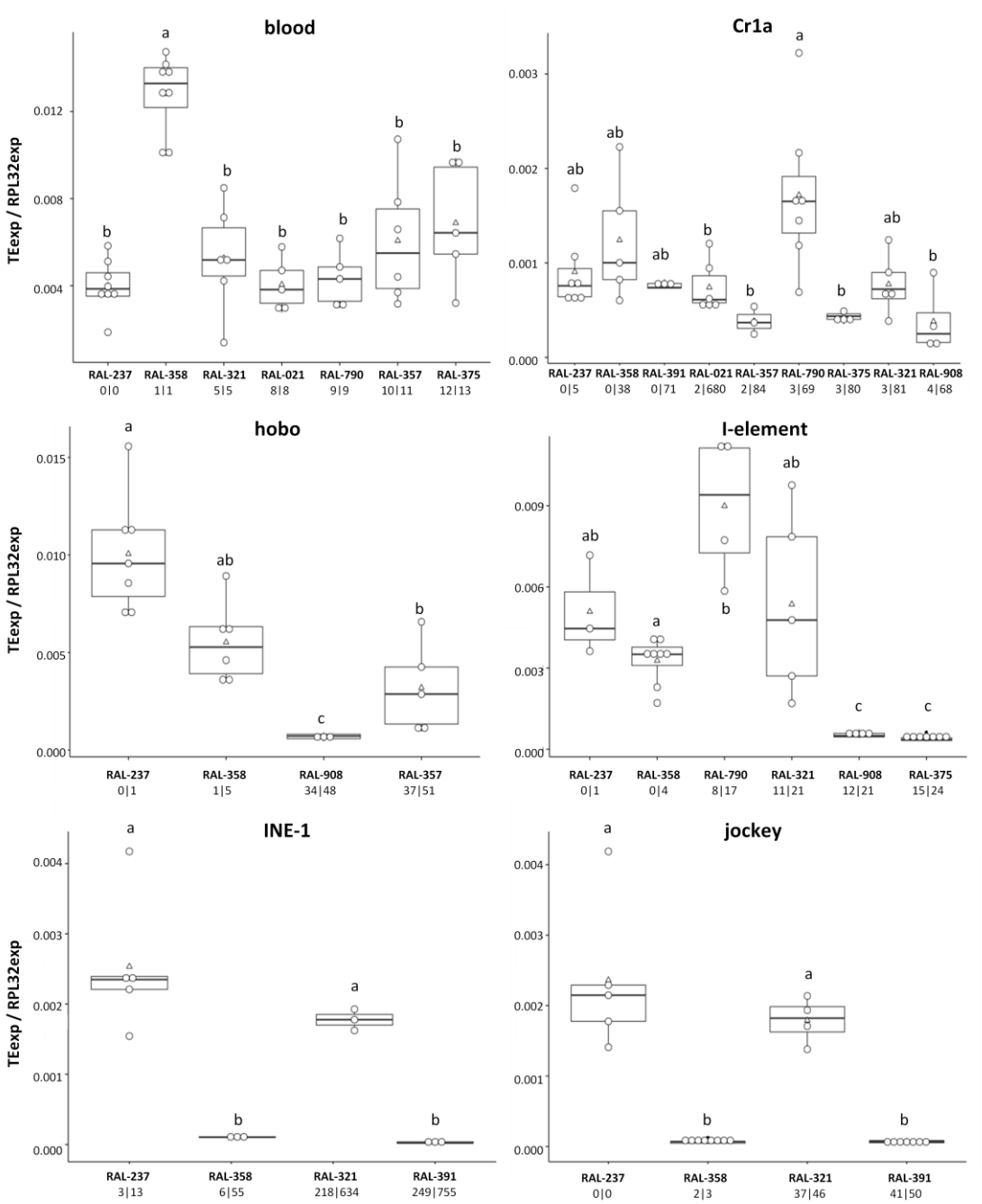

Figure 3: No directional trend observed between blood, Cr1a, hobo, I-element, INE-1 and jockey expression (normalized to RpL32), in ovaries of DGRP lines, and respective different predicted copy numbers. The DGRPs (RAL lines) in the X-axes are ordered by number of novel insertions. The numbers underneath each DGRP name correspond to the predicted number of novel and total insertions for the corresponding TE. Each white circle corresponds to one biological replicate, white triangles represent the mean and the black line the median. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. Model: glm(TEexp/RPL32exp ~ genotype, family=quasipoisson). Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference.

Temperature, *Wolbachia* and the interaction of the two factors have effects on TE expression, depending on the TE and genotype

TE activity can cause mutations that add to the genetic variation of the population. To test whether environmental perturbation of adults affected TE activity in a way that could impact the extent of genetic variation in their progeny we focused on TE expression in ovaries. We chose to study temperature, an external abiotic external factor, and *Wolbachia* infection, an internal biotic internal factor.

We tested the expression of seven TEs – 1360, blood, gypsy5, I-element, mdg1, pogo and roo in females of four different genetic backgrounds – DGRP lines RAL-21, RAL-237, RAL-321, and RAL-790. We compared TE expression under three different temperature treatments – 21°C, 25°C and 29°C – in lines with and without *Wolbachia* (Table 3).

Although our experimental design aimed at characterizing expression of seven TEs in four different genetic backgrounds (28 combinations TE x genotype), several TEs had ovary expression levels undetectable by qPCR. We limited our comparisons to the TEs and lines that showed detectable expression in at least three biological replicates (17 out of 28). Only 1360 and roo had detectable expression levels in all genetic backgrounds. Blood and mdg1 had detectable expression in three genotypes, pogo in two and gypsy5 showed expression in only one DGRP line (RAL-790) (see Figures 4 – 6). I-element did not have enough biological replicates with detectable levels of expression for all conditions in any line, so we could not study this element.

We observed that temperature, *Wolbachia* and the interaction between the two factors had an effect in TE expression, that depended on the TE and genetic background. We discerned temperature effect by looking at differences in TE expression under the three temperatures, 21°C, 25°C and 29°C, within each line with and without *Wolbachia*. In addition, we determined *Wolbachia* effect by looking at TE expression in lines with and without *Wolbachia*, under one particular temperature at a time. Out of the 17 TE x genotype combinations we saw effect of temperature in 10 (59%), *Wolbachia* in 8 (47%) and interaction of the two factor in 3 combinations (18%). Figures 4 – 6 show the temperature effects on the expression relative to control gene for the various TEs, in genetic backgrounds with and without *Wolbachia*.

Overall, looking at temperature effects, we saw more cases in which TE expression levels under 21°C were different from the other temperatures, usually in lines were *Wolbachia* was absent (see Figures 4, 5 and 6). Studying *Wolbachia* effects, we saw again that TE expression levels usually varied at 21°C and, whenever we saw differences, the expression levels were usually higher in the absence of *Wolbachia*, with two exceptions (blood and mdg1, both in RAL-321, with higher expression levels with *Wolbachia*). We only saw statistically significant effect of interaction of the factors for blood in RAL-321 (*p-value*=0.043), mdg1 in RAL-321 (*p-value*=0.021) and pogo in RAL-237 (*p-value*=0.005). The DGRP line in which we saw most variation in TE expression under both environmental factors was RAL-21.

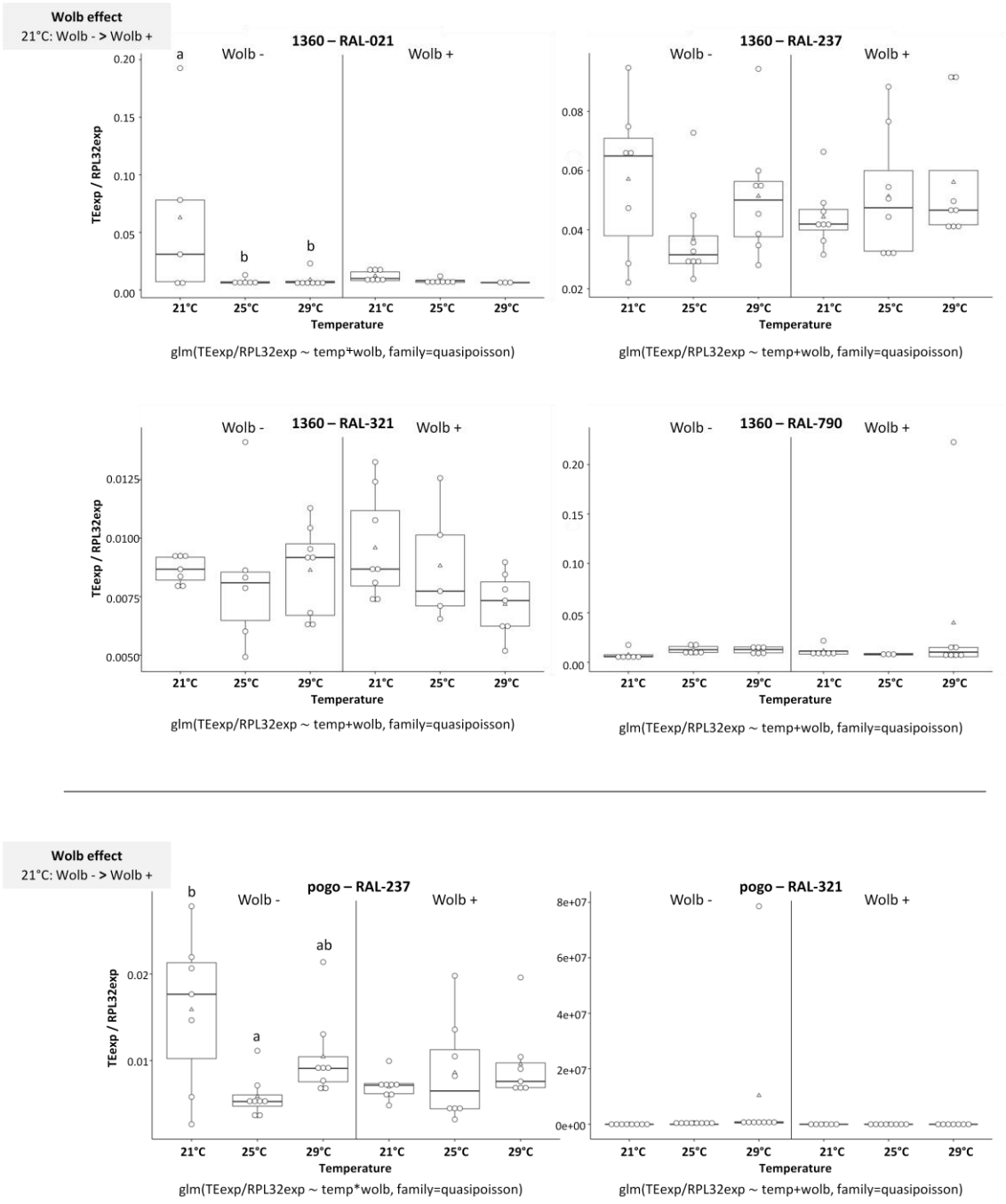

Figure 4: Temperature effects in the expression (relative to control gene RpL32) of 1360 and pogo (*cut-and-paste* TEs) in ovaries of adults of different DGRP lines, with or without *Wolbachia*. The X-axes correspond to the treatments in study – three temperatures (21°C, 25°C and 29°C), with and without *Wolbachia*(Wolb). In the Y-axes, the expression levels are shown separately for the different genotypes (one genotype per panel). Each white dot is a biological replicate, the white triangle represents the mean and the black line the medium of the values. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. The glm model used is under the X-axes, for each panel. Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference. "Wolb effect" grey box shows *Wolbachia* effects in TE expression (least-squares means analysis, alpha=0.05). "<" and ">" indicate in which *Wolbachia* status ("Wolb -":absent, "Wolb +":present) the levels of TE expression were respectively lower or higher, within temperatures in which we found statistically significant difference.

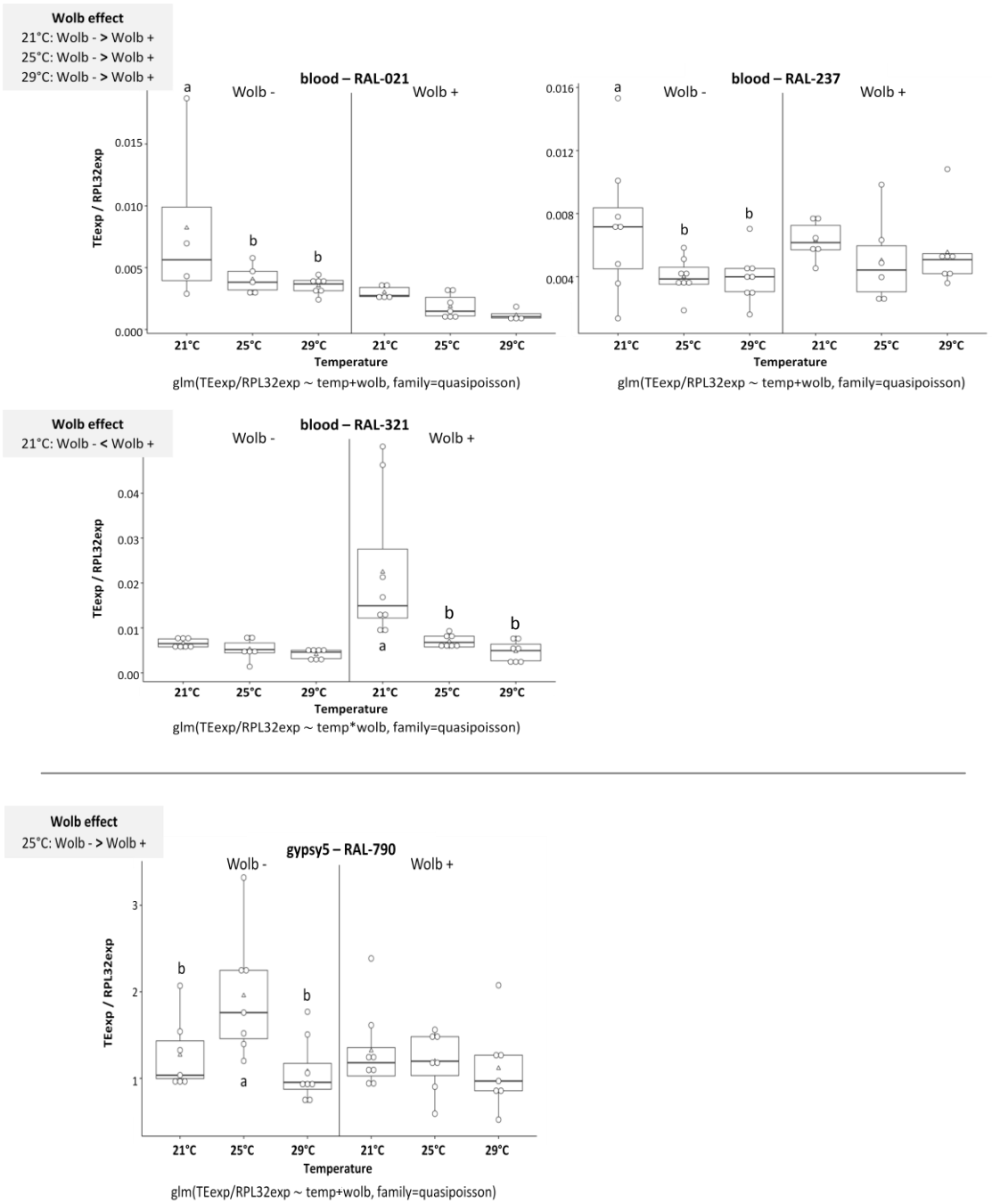

Figure 5: Temperature effects in the expression (relative to control gene RpL32) of blood and gypsy5 (*copy-and-paste* TEs) in ovaries of adults of different DGRP lines, with or without *Wolbachia*. The X-axes correspond to the treatments in study – three temperatures (21°C, 25°C and 29°C), with and without *Wolbachia*(Wolb). In the Y-axes, the expression levels are shown separately for the different genotypes (one genotype per panel). Each white dot is a biological replicate, the white triangle represents the mean and the black line the medium of the values. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. The glm model used is under the X-axes, for each panel. Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference. "Wolb effect" grey box shows *Wolbachia* effects in TE expression (least-squares means analysis, alpha=0.05). "<" and ">" indicate in which *Wolbachia* status ("Wolb -":absent, "Wolb +":present) the levels of TE expression were respectively lower or higher, within temperatures in which we found statistically significant difference.

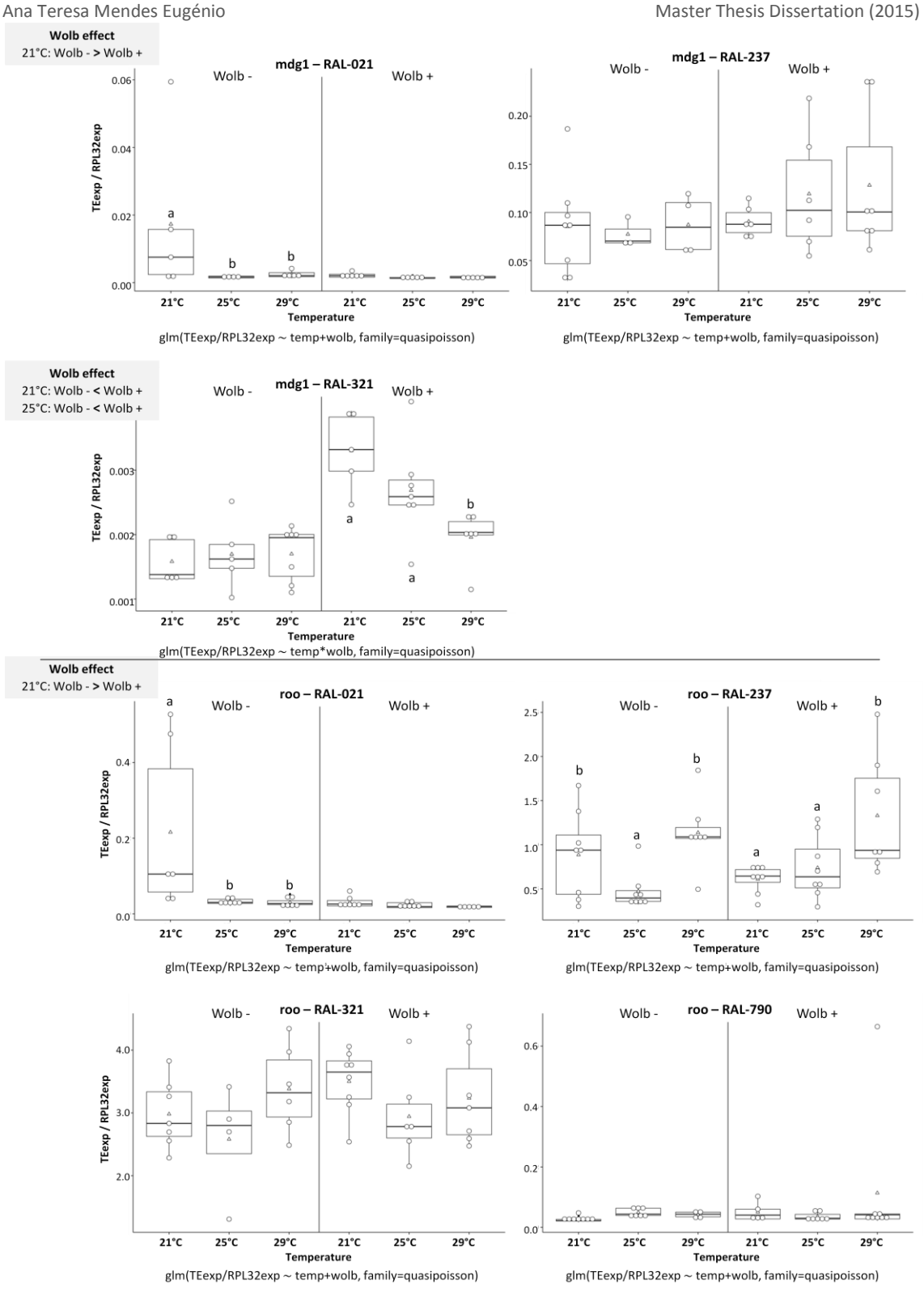

Figure 6: Temperature effects in the expression (relative to control gene RpL32) of mdg1 and roo (*copy-and-paste* TEs) in ovaries of adults of different DGRP lines, with or without *Wolbachia*. The X-axes correspond to the treatments in study – three temperatures (21°C, 25°C and 29°C), with and without *Wolbachia* (Wolb). In the Y-axes, the expression levels are shown separately for the different genotypes (one genotype per panel). Each white dot is a biological replicate, the white triangle represents the mean and the black line the medium of the values. Boxplots include 25-75% percentile and whiskers represent 95% confidence interval. The glm model used is under the X-axes, for each panel. Results of pairwise comparisons between genotypes for each TE (least-squares means analysis, alpha=0.05) are represented with letters: same letter indicates no statistically significant difference and different letters indicates statistically significant difference. "Wolb effect" grey box shows *Wolbachia* effects in TE expression (least-squares means analysis, alpha=0.05). "<" and ">" indicate in which *Wolbachia* status ("Wolb -":absent, "Wolb +":present) the levels of TE expression were respectively lower or higher, within temperatures in which we found statistically significant difference.

Discussion

TEs are phylogenetically widespread, being represented in essentially all genomes sequenced to date. They are an important source of genetic novelty and their contribution to evolutionarily relevant genetic variation has recently started to accumulate experimental evidence. TEs have been implicated in adaptation in natural (González *et al*. 2008) and laboratory (Sousa *et al.* 2013) populations and in the origin of novel traits (Wagner *et al.* 2005, Santos *et al*. 2014, Pray 2008). The molecular mechanisms responsible for TE silencing are topics of very active research and much progress has been made to characterize them on selected model systems like *Drosophila melanogaster*. Nonetheless, our understanding of the factors and mechanisms that can make TEs jump in natural populations is still poorly explored.

In this work, we set out to test the effects of genotype and environment on TE expression. We focused specifically on female ovaries because it is in the germline that TE activity can impact genetic variation in the progeny, which can be the raw material for evolution by natural selection.

We used different genetic backgrounds of *Drosophila melanogaster* genotypes from a panel of wild-derived, isogenic and fully sequenced lines, the DGRPs (Mackay *et al.* 2012). Sequence information available for these lines includes *in silico* predictions of TE insertion sites in the host genome and respective TE identities. The lines differ in how many insertions of different TEs they carry. These insertions have been characterized as “shared” or “novel”, depending on whether they were or not also found in the reference genomic sequence of *Drosophila melanogaster*. Novel insertions tend to also not be common between different DGRPs and we argued that TEs producing them are likely more recent and/or of active. We used this information to guide our choice of DGRP lines, TEs and TE insertion sites to study. We defined two main aims for this work and planned three tasks to meet those aims.

Aim 1: Test the effect of genotype in TE expression

Our Aim 1 was to test the effect of genotype, specifically, the number of copies of a particular TE, on TE expression. To pursue this, we first set out to validate some of the *in silico* predictions of TE insertion sites and identity (Task 1). These had never been experimentally validated before. We validated 100% of the 52 insertion sites we tested and for 92% thereof we also confirmed the identity of the predicted TE, albeit sometimes not in a single copy and often not of the full-size TE (Table 4 and supplementary Table S1). We note that we only tested a very small proportion of all insertions (52 of approximately 50 thousand novel insertions, in about 200 thousand total insertions). It is unclear to what extent our estimated high rates of validation of both position and identity would hold for a more comprehensive sample of insertions, including other TEs and host genotypes.

With more confidence on the *in silico* predictions, we used the DGRP information to guide our selection of lines with more and fewer novel insertions of specific target TEs, with which we set to test whether the number of copies of a TE in a genome could predict its levels of expression (Task 2).

We found differences between genotypes in levels of TE expression in ovaries, albeit not the same for all TEs (Figures 1-3). However, the differences between genetic backgrounds could not be explained by number of insertions (novel or total).

Consequently, we could not confirm our hypothesis that TE expression levels are dependent on number of novel insertions of that TE. It can be that TE activity is not directly proportional to the number of novel insertions, that TE expression is not a good proxy for TE activity, or a combination of both. Importantly, our design suffered from an important limitation, in that the distinct genotypes we studied differ not only in copy number of the target TEs but also in other aspects that might be confounding any potential signal copy number might have on TE expression. We also note that for some TEs (specifically, blood, jockey and pogo) we observed detectable levels of expression in a line (RAL-237) not predicted to have any insertion for these elements (see Table 2 and Figures 1 and 3). This means that the *in silico* annotation probably failed to detect some of their insertions. Less likely seems to be the possibility of an invasion of those elements since the time the lines were sequenced, as the different lines are supposedly kept isolated from other genotypes.

We noted that the gypsy5 TE had unusually high copy number and also unusually high expression in one particular genotype (RAL-790), suggestive of a recent burst and/or very active TE. It is unclear whether this apparent increased activity of gypsy5 in RAL-790 is due to an effect of the genetic background (being especially permissive to jumping of gypsy elements) and/or of the gypsy5 in RAL-790. The effect of the genetic background could be verified by checking the expression of other elements of the gypsy family (assuming that all elements of the gypsy family behave in a similar manner, which is an assumption that would need to be validated), as well as by checking the flamenco piRNA cluster, responsible for silencing gypsy5 (Sarot *et al*. 2010), in RAL-790. An effect of specific properties of the gypsy5 found in RAL-790 could be checked by sequencing all gypsy5 insertions from this line and searching for mutations that might be associated with change in its activity.

Aim 2: Test the effect of environmental perturbation and genotype on TE expression levels

Our Aim 2 was to test the effect of different types of environmental perturbation on TE expression levels. We chose to study the effect on an external abiotic environmental factor (temperature) and one internal biotic factor (*Wolbachia* infection) on TE expression, in different genetic backgrounds. Our choice of environmental factors to test was based on their prevalence in natural populations, as well as on previous studies suggesting their effect on TE activity. Temperature variation had been previously implicated in TE activity in some, but not all previous studies (Capi *et al.* 2000). *Wolbachia*, a very common endosymbiont in natural populations of insects, had been previously shown to confer *D. melanogaster* hosts resistance against infection by virus (Teixeira *et al*. 2008), which have many properties in common with TEs. We found that temperature, *Wolbachia* and their interaction affect TE expression in a different way, depending on the TEs and genotypes (Figures 4, 5 and 6).

The mechanisms underlying the effect of external temperature and of *Wolbachia* infection on TE expression in *D. melanogaster* ovaries have not been explored. Temperature may affect TEs that are under the control of temperature-sensitive enhancers or temperature-responsive transcription factors.To check for this, we could search for temperature-responsive elements in the TE sequences for which we found temperature effects on expression. It is also conceivable that temperature effects on TE activity are mediated by the TE-repressing piRNA pathway. Its key protein Piwi is only phosphorylated and active if associated to Hsp90, a heat-shock protein (Sato *et al*. 2010, Specchia *et al*. 2010, Gangaraju *et al*. 2011). If Hsp90 were ever limiting, its recruitment to chaperone functions involved in the response to temperature stress could limit phosphorylation, and thus activity, of Piwi. To test this hypothesis, we could, for genotypes where TE expression was affected by temperature, investigate temperature effects on Piwi phosphorylation (with a Western-like gel detecting phosphorylated and unphosphorylated Piwi), and/or temperature effects on piRNA levels (by sequencing piRNAs).

For the mechanism whereby *Wolbachia* might affect TE expression, we know even less. In *D. melanogaster, Wolbachia* provides protection against virus (Teixeira *et al*. 2008). We were expecting our results to be consistent with this effect direction (lower TE expression with *Wolbachia*) and that is indeed what we saw. Still, for two TEs in RAL-321 we observed lower expression in the absence of *Wolbachia* and most effects were only seen at a temperature of 21°C. The fact is that the mechanism by which this symbiont confers protection to *D. melanogaster* against virus is still unknown, and, therefore, we do not know how, whatever that mechanism is, can be also affecting TEs. To check whether *Wolbachia* is interfering with the piRNA pathway, a follow-up study could sequence piRNAs in ovaries from a line naturally infected with *Wolbachia* and the same line after *Wolbachia* removal.

Hindsight and perspectives

Aside for the experimental design limitations we identified and discussed above, we can see a number of ways by which our dataset could be improved and better able to solve the biological questions we set forth and/or understand the mechanisms of the phenomena we describe. These are highlighted below.

As the control gene in our qPCR analysis, to assess TE expression levels in ovaries, we used RpL32. This gene is very commonly used as reference gene in measurements of TE expression in *Drosophila melanogaster* (Parnell *et al*. 2006, Kemp *et al*. 2013, Haghayeghi *et al*. 2010). Because finding a control genes suitable for all treatments (three temperatures x two *Wolbachia* status) and genotypes is difficult, we suggest that further studies should include other type of controls. These could be other house-keeping genes that presumably have rather invariable expression levels (e.g. TBP, Lam *et al*. 2012), Hsp70 to account for response environmental perturbations like temperature (Hoekstra *et al*., 2013, Štětina *et al*. 2015) and ovary-specific genes, such as Vasa (Pek *et al.* 2011), to account for potential environmentally-induced changes in ovary development. It would also be interesting to add a *Wolbachia*-specific gene (e.g. surface protein Wsp) to estimate potential temperature, genotype and individual differences in *Wolbachia* load (Poisont *et al.* 1998, Osborne *et al.* 2012).

In this study, we tested the effects of temperature, *Wolbachia* and genotype on TE expression in ovaries. This target organ was chosen because it is in the germline that TE activity has the potential to impact the generation of new genetic variants in the progeny. In order to understand whether the responses we saw are specific to ovaries, future work should investigate TE expression also in testes (male germline) and in somatic-only tissues (like in thoraxes).

To test whether TE expression is, indeed, a good measure of TE activity, we could assess a more direct read-out of TE activity. Focusing on *copy-and-paste* TEs, we could quantify TE copy-number in the next generation (F1) after female exposure to environmental perturbation. If TE expression is a good proxy for activity, we expect to see increased copy-number in F1 genomes when mothers experience environmental conditions that increase TE expression in ovaries. Moreover, this would allow us to determine to what extent environmental perturbation of different genotypes would result in the production of more genetically variable progeny. This could be a mechanism for organisms to ensure production of more variable progeny and maximize chances of some of these being able to deal with the perturbed environment. It has recently been shown that, indeed, *D. melanogaster* facing parasitoids increases recombination rates and produces more genetically variable progeny (Singh *et al*. 2015).

Bibliography

Allen DF, Little TJ (2011). Dissecting the effect of a heterogeneous environment on the interaction between host and parasite fitness traits. *Evolutionary Ecology* **25**: 499-508.

Alonso-González L, Domínguez A, Albornoz J (2006). Direct determination of the influence of extreme temperature on transposition and structural mutation rates in *Drosophila melanogaster* mobile elements. *Genetica* **128**: 11-19.

Altschul SF, Madden TL, Schaffer AA, Miller W, Lipman DJ (1997). Gapped BLAST and PSI - BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Aravin A, Hannon GJ, Brennecke J. (2007). The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**: 761-764.

Arbuthnott D, Dutton EM, Agrawal AF, Rundle HD (2014). The ecology of sexual conflict: ecologically dependent parallel evolution of male harm and female resistance in *Drosophila melanogaster. Ecology Letters* **17**: 221-228.

Arnault C, Dufournel I (1994). Genome and stresses: reactions against aggressions, behavior of transposable elements. *Genetica* **93**: 149-160.

Arnault C, Loevenbruck C, Biémont C (1997). Transposable element mobilization is not induced by heat shocks in *Drosophila melanogaster*. *Naturwissenschatten* **84**: 410-414.

Barr MG, Petrov DA, Gonz J (2014). Population genomics of transposable elements in *Drosophila*. *Annual Review of Genetics* **48**: 561-581.

Becker T, Loch G, Beyer M, Zinke I, Aschenbrenner AC, Carrera P, Inhester T, Schultze JL, Hoch M (2010). FOXO-dependent regulation of innate immune homeostasis. *Nature* **463**: 369-373.

Beldade P, Mateus ARA, Keller RA (2011). Evolution and molecular mechanisms of adaptive developmental plasticity. *Molecular Ecology* **20**: 1347-1363.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* **18**: 1752–1762.

Bregliano JC, Picard G, Bucheton A, Pelisson A, Lavige JM, L'Heritier P (1980). Hybrid dysgenesis in *Drosophila melanogaster*. *Science* **207**: 606-611.

Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089–103 (2007).

Bubenshchikova EV, Antonenko OV, Vasilyeva LA, Ratner VA (2002). Induction of MGE 412 transpositions in spermatogenesis of *Drosophila* males separately by heat and cold shock. *Genetika* **38**: 46-55.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014). The Relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster. Molecular Biology and Evolution* **31**: 1010-1028.

Capy P, Gasperi G, Biémont C, Bazin C (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**: 1365-2540.

Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, Teixeira L (2013). *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genetics* **9**: e1003896.

Ciota AT, Matacchiero AC, Kilpatrick AM, Kramer, LD (2014). The effect of temperature on life history traits of *Culex* mosquitoes. *Journal of Medical Entomology* **51**: 55-62.

Cowley M, Oakey RJ (2013). Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genetics* **9**: e1003234.

Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Kosack KH, Kistler HC *et al.* (2007). The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.

Gangaraju VK, Yin H, Weiner MM, Wang J, Huang XA, Lin H (2011). Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nature Genetics* **43**: 153-158.

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008). High Rate of Recent Transposable Element–Induced Adaptation in Drosophila melanogaster. *PLoS Biology* **6**:e251

Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Costa APP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, van Sluys MA, Mhiri C (2005). Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet Genome Research* **110**: 229-241.

Haghayeghi A, Sarac A, Czerniecki S, Grosshans J, Schock F (2010). Pellino enhances innate immunity in Drosophila. *Mechanisms of Development* **127**: 301-307.

Handler D, Olivieri D, Novatchkova M, Gruber FS, Brennecke J (2011). A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO Journal* **30**: 3977-3993.

Hedges DJ, Deininger PL (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation Research* **616**: 46-59.

Hedges LM, Brownlie JC, O'Neill SL, Johnson KN (2008). *Wolbachia* and virus protection in insects. *Science* **322**: 702.

Hoekstra LA, Montooth K (2013). Inducing extra copies of the *Hsp70* gene in *Drosophila melanogaster* increases energetic demand. *BMC Evolutionary Biology* **13***: 68.*

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, Magwire MM, Blankenburg K, Carbone MA, Chang K, Ellis LL, Mackay TFC *et al.* (2014). Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research* **24**: 1193-1208.

Kemp C, Mueller S, Goto A, Barbier V, Paro S, Bonnay F, Dostert C, Troxler L, Hetru C, Meignin C, Hoffmann JA, Pfeffer S, Imler, J-L (2013). Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in *Drosophila*. *Journal of immunology* **190**: 650–658.

Khurana JS, Xu J, Weng Z, Theurkauf WE (2010). Distinct functions for the *Drosophila* piRNA pathway in genome maintenance and telomere protection. *PLoS Genetics* **6**: e1001246.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Developmental* Cell **12**: 45-55.

Koressaar T, Remm M (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289-91.

Lam KC, Mühlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, Manke T, Akhtar A (2012). The NSL Complex Regulates Housekeeping Genes in *Drosophila*. *PLoS Genetics* **8**: e1002736

Mackay TFC, Richards, S, Stone EA, Barbadilla A, Avroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RRH, Barrón M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Thornton KR, Mittelman D, Gibbs RA *et al.* (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**: 173-178.

Malone, C. D. *et al.* Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522–35 (2009).

McClintock B (1951). Chromosome organization and genic expression. *Cold Spring Harb Symp Quant Biol* **16**: 13-47.

Mills RE, Bennett EA, Iskow RC, Devine SE (2007). Which transposable elements are active in the human genome? *Trends in Genetics* **23**: 183-91.

Min K-T, Benzer S (1997). *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proceedings of the National Academy of Sciences USA* **94**: 10792-10796.

Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Molecular cell* **50**: 736-48.

Mukamel Z, Tanay A (2013). Hypomethylation marks enhancers within transposable elements. *Nature Genetics*. **45**: 717-718.

Navarro C, Bullock S, Lehmann R (2009). Altered dynein-dependent transport in piRNA pathway mutants. *National Academy of Sciences* **106**: 9691-9696.

O'Donnell, K.A. and J.D. Boeke, Mighty Piwis Defend the Germline against Genome Intruders. Cell, 2007. 129(1): p. 37-44.

Olivieri D, Senti K-A, Subramanian S, Sachidanandam R, Brennecke J (2012). The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in *Drosophila*. *Molecular cell* **47**: 954-69.

Osborne SE, Iturbe-Ormaetxe I, Brownlie JC, O'Neill SL, Johnsona KN (2012). Antiviral Protection and the Importance of *Wolbachia* Density and Tissue Tropism in *Drosophila simulans*. Applied and Environmental Microbiology **78**: 6922–6929.

Parnell T J, Kuhn EJ, Gilmore BL, Helou C, Wold MS, Geyer PK (2006) Identification of genomic sites that bind the *Drosophila* suppressor of hairy-wing insulator protein. *Molecular and Cellular Biology* **26**: 5983-5993.

Pek JW, Kai T (2011). A role for Vasa in regulating mitotic chromosome condensation in *Drosophila*. *Current Biology* **21**: 39-44.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO Journal.* **13**: 4401-4411.

Poinsot D, Bourtzis K, Markakis G, Savakis C, Merçot H (1998). *Wolbachia* transfer from *Drosophila melanogaster* into *D. simulans*: Host effect and cytoplasmic incompatibility relationships. *Genetics* **1**: 227-237.

Ponton F, Chapuis M-P, Pernice M, Sword GA, Simpson SJ (2010). Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *Journal of Insect Physiology* **57**: 840–850.

Pray L (2008). Transposons: The jumping genes. *Nature Education* **1**: 204.

Pray L, Zhaurova K (2008). Barbara McClintock and the discovery of jumping genes (transposons). *Nature Education* **1**: 169.

Ratner VA, Zabanov SA, Kolesnikova OV, Vailyeva LA (1992). Induction of the mobile element *Dm-412* transpositions in the *Drosophila* genome by heat shock treatment. *Proceedings of the National Academy of Sciences USA* **89**: 5650-5654.

Reiter L T, Potocki L, Chien S, Gribskov M, Bier E (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research* **11**: 1114–1125.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.

Santos ME, Braash I, Boileau N, Meyer BS, Sauteur L, Böhne A, Belting H-G, Affolter M, Salzburger W (2014). The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change. *Nature Communications* **5**: e5149.

Sarot E, Bucheton A, Pe A (2004). Retrovirus by the *Drosophila melanogaster* flamenco Gene. *Genetics* **1321**: 1313–1321.

Sato K, Siomi H (2010). Is canalization more than just a beautiful idea? *Genome Biology* **11**: 109.

Schaack S, Gilbert C, Feschotte C (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology & Evolution* **25**: 537–46.

Shibata S, Murota Y, Nishimoto Y, Yoshimura M, Nagai T, Okano H, Siomi MC (2015). Immuno-electron microscopy and electron microscopic *in situ* hybridization for visualizing piRNA biogenesis bodies in *Drosophila* ovaries. *Methods in Molecular Biology* **1328**: 163-178.

Singh N, Schlenke T, Skolfield S, Criscoe D, Kohl K, Keebaugh E (2015). Fruit flies diversify their offspring in response to parasite infection. *Science* **349**: 747-750.

Slotkin RK, Martienssen R (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Review Genetics* **8**: 272–85.

Sousa A, Bourgard C, Wahl LM, Gordo I (2013). Rates of transposition in *Escherichia coli*. *Biology letters* **9**: 20130838.

Specchia V, Piacentini L, Tritto P, Fanti L, D'Alessandro R, Palumbo G, Pimpinelli S, Bozzetti MP (2010). Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* **463**: 662-665.

Štětina T, Koštál V, Korbelová J (2015). The role of inducible Hsp70, and other heat shock proteins, in adaptive complex of cold tolerance of the fruit fly (*Drosophila melanogaster*). *PLoS ONE* **10**: e0128976.

Stoks R, Karl I, De Block M, Janowitz SA, Fischer K (2011). Temperature extremes and butterfly fitness: conflicting evidence from life history and immune function. *Global Change Biology* **17**: 676-687.

Stouthamer R, Breeuwer JAJ, Hurst GDD (1999). *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology* **53**: 71-102.

Teixeira L, Ferreira A, Ashburner M (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology* **6**: e2.

Tenaillon MI, Hufford MB, Gaut BS,Ross-Ibarra J (2011). Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biology and Evolution* **3**: 219-229.

Tiwari MD, Wodarz A (2015). Asymmetric cell division and development of the central nervous system in *Drosophila. Cell Polarity **2**: 95-117.*

Touchon M, Rocha EP (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Molecular Biology and Evolution* **24**: 969–81.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.

Vasilyeva LA, Bubenshchikova EV, Ratner VA (1999). Heavy heat shock induced transposon transposition in *Drosophila*. *Genetics Research* **74**: 111-119

Vásquez JF, Albornoz J, Domínguez A (2007). Direct determination of the effects of genotype and extreme temperature on the transposition of *roo* in long-term mutation accumulation lines of *Drosophila melanogaster*. *Molecular Genetics and Genomics* **278**: 653-664.

Venner S, Feschotte C, Biémont C (2009). Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics* **25:** 317–23.

Wagner GP, Lynch VJ (2005). Evolutionary novelties. *Current Biology* **20**: 48–52.

Wu M, Pastor-Pareja JC, Xu T (2010). Interaction between RasV12 and scribbled clones induces tumour growth and invasion. *Nature* **463**: 545–548.

Xiong Y, Eickbush TH (1988). Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Molecular Biology and Evolution* **5**: 675-690.

Zabanov SA, Vasilyeva LA, Ratner VA (1990). Expression of quantitative character radius incompletus in *Drosophila* and localization of mobile elements MDG-1 and copia. *Genetika* **26**: 1144-1153.

Supplementary data

Table S1: Confirmation of DGRP *in silico* predictions (Task 1)

TE	Chrom	Start	End	Prob	Line	TE size	LongPCR size	Sequencing	Neg Ctrl
LTR elements (*copy-paste*)									
copia	3R	4991260	4991264	L, 0.86 R, 0.8	810	5100 bp	✓6000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
copia	3R	4991260	4991264	L, 0.86 R, 0.8	812	5100 bp	✓6000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
copia	3R	13161944	13161944	L,0.85 R,0.86	908	5100 bp	✓6000 bp (Ta=58°C)	✓✓ Fw	✗ 300 bp
copia	2L	4426305	4426310	L,0.7 R,0.95	908	5100 bp	✓5000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
opus	2R	17950637	17950706	L,0.78 R,1	443	7500 bp	+/- 3000 bp (Ta=62°C)	✓✓ Fw	✗ 75 bp
opus	2L	7336324	7336327	L,0.09 R,0.76	908	7500 bp	+/- 700 bp (Ta=60°C)	✓✓ Fw	✗ 75 bp
Transpac	2R	12108723	12108729	L, 1	810	5200 bp	✓6000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
Transpac	X	345626	345630	L, 0.6 R, 0.1	357	5200 bp	✓6000 bp (Ta=58°C)	✓✓ Fw	✗ 200 bp
Transpac	X	345626	345630	L, 0.6 R, 0.1	892	5200 bp	✓5000 bp (Ta=58°C)	✓✓ Fw	✗ 200 bp
Transpac	3L	12856836	12856842	R,1	908	5200 bp	✓6000 bp (Ta=60°C)	✓✓ Fw	✗ 200 bp
Transpac	X	15333499	15333504	R,0.92	443	5200 bp	✓6000 bp (Ta=60°C)	✓✓ Fw	✗ 200 bp
non-LTR elements (*copy-paste*)									
I-element	3R	14271683	14271693	L,0.57 R,1	810	5300 bp	+/- 2000 bp (Ta=56°C)	✓✓ Fw	✗ 200 bp
I-element	2R	13861586	13861600	L,1 R,0.7	810	5300 bp	+/-1000-1500 bp (Ta=60°C)	✓✓ Fw/Rv	✗ 300 bp
I-element	3L	5068011	5068084	L, R,0.66	908	5300 bp	+/- 700 bp (Ta=56°C)	✓✓ Fw	✗ 300 bp
I-element	2R	10573276	10573286	L, 1 R, 0.5	357	5300 bp	+/- 600 - 700 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
I-element	3R	11405981	11405987	L, 0.77 R, 1	357	5300 bp	✓5000 bp (Ta=60°C)	✓✓ Fw	✗ 400 bp
Juan	3L	16130441	16130456	L,0.16 R,0.97	443	4200 bp	+/- 400 bp (Ta=60°C)	✓✓ Fw	✗ 200 bp
Juan	2R	5572279	5572291	L,0.75 R,0.14	810	4200 bp	+/- 600 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
Juan	3L	17698435	17698491	L, 0.11 R, 0.998	357	4200 bp	✓4000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
Juan	3L	17698435	17698491	L, 0.11 R, 0.998	761	4200 bp	+/- 5000 bp (Ta=60°C)	✓✓ Fw	✗ 300 bp
Juan	X	18924423	18924436	L, 0.13 R,0.97	908	4200 bp	✓5000 bp (Ta=58°C)	✓✓Rv	✗ 300 bp
F-element	3R	6947766	6947992	L, 0.1 R, 0.9	810	4700 bp	✓5000 bp (Ta=60°C)	✗ ✗ Fw/Rv (Pogo)	✗ 300 bp
F-element	3R	9994118	9994126	R, 0.97	908	4700 bp	✓5000 bp (Ta=60°C)	✓✓ Fw	✗ 200 bp
F-element	3L	7996067	7996071	L, 0.6 R, 0.25	908	4700 bp	+/-3500 bp (Ta=60°C)	✓✓ Fw/Rv	✗ 300 bp
Doc	3R	7873179	7873180	L, 1	357	4700 bp	+/- 7000 bp (Ta=60°C)	✗ ✗ Fw/Rv (Stalker)	✗ 300 bp
Doc	3R	7873179	7873180	L, 1	381	4700 bp	+/- 7000 bp (Ta=60°C)	✗ ✗ Fw/Rv (Stalker)	✗ 300 bp
Doc	3R	7873179	7873180	L, 1	761	4700 bp	✓5000 bp (Ta=60°C)	✗ ✗ Fw/Rv (Stalker)	✗ 300 bp

TE	Chrom	Start	End	Prob	Line	TE size	LongPCR	Sequencing	Neg Ctrl
non-LTR elements (*copy-paste*)									
Doc	2L	11138677	11138734	L, 0.22 R, 0.87	357	4700 bp	+/- 3000 bp (Ta=60°C)	✓✓ Fw	✗ 400 bp
Doc	2L	11138677	11138734	L, 0.22 R, 0.87	892	4700 bp	✓5000 bp (Ta=60°C)	✓✓ Fw	✗ 400 bp
Doc	2R	7936918	7936946	L, 0.2 R,0.8	810	4700 bp	✓5000 bp (Ta=60°C)	✓✓ Fw	✗ 400 bp
Doc	3R	10301789	10301822	L, 0.1 R, 0.8	908	4700 bp	+/- 3000 bp (Ta=60°C)	✓✓ Fw	✗ 200 bp
Doc	2L	1399467	1399478	R, 0.9	908	4700 bp	✓5000 bp (Ta=60°C)	✓✓ Fw/Rv	✗ 200 bp
TIR elements (*cut-paste*)									
hopper	X	4110151	4110442	L1 R1	810	1400 bp	✓2000 bp (Ta=60°C)	✓✓Fw	✗ 400 bp
hopper	3R	15010776	15010780	L,0.67 R,0.93	908	1400 bp	✓1750 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hopper	3R	13938897	13938901	L, 1 R, 0,44	357	1400 bp	✓1750 bp (Ta=60°C)	✓✓Fw	✗ 200 bp
hopper	3R	13938897	13938901	L, 1 R, 0,44	812	1400 bp	✓1750 bp (Ta=60°C)	✓✓Fw	✗ 200 bp
hopper	2L	2390528	2390533	L, 0.1 R,1	357	1400 bp	✓(Ta=60°C) 1750 bp	✓✓Fw	✗ 300 bp
pogo	2R	11767856	11767857	L,1 R,1	810	2100 bp	+/- 400 bp (Ta=60°C)	✓✓Fw/Rv	✗ 200 bp
pogo	2R	11767856	11767857	L,1 R,1	381	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 200 bp
pogo	3L	21388576	21388578	L,1 R,1	810	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 300 bp
pogo	3L	21388576	21388578	L,1 R,1	812	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 300 bp
pogo	X	6015256	6015259	L,1 R,1	908	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 300 bp
pogo	X	6015256	6015259	L,1 R,1	381	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 300 bp
pogo	3R	11292096	11292097	L, 0.76 R, 0.95	357	2100 bp	+/- 1250 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
pogo	3R	2926760	2926761	L, 1 R, 1	357	2100 bp	+/- 500 bp (Ta=60°C)	✓✓Fw/Rv	✗ 300 bp
hobo	2R	12871101	12871113	L, 0.96 R, 0.8	810	2900 bp	+/- 2000 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3L	19555349	19555354	L, 0.89 R, 1	810	2900 bp	+/- 2000 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3L	19555349	19555354	L, 0.89 R, 1	812	2900 bp	+/- 2000 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3L	19555349	19555354	L, 0.89 R, 1	761	2900 bp	✓3000 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3R	22134885	22134892	L, 1 R, 0.9	908	2900 bp	+/- 1500 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3R	10703373	10703379	L, 0.98 R, 0.85	357	2900 bp	+/- 1500 bp (Ta=60°C)	✓✓Fw	✗ 300 bp
hobo	3R	22133514	22133521	L, 0.94 R, 0.67	357	2900 bp	✓4000 bp (Ta=60°C)	✓✓Fw	✗ 300 bp

Legend: Chrom – chromosome, **Start** – start position of insertion, **End** – end position of insertion, **Prob** – probability that correspondent element is inserted in the position (**L** – probability for left read; **R** – probability for right read), **Line** – DGRP genotype, **Size** – size of the insertion, **Neg Ctrl** – negative control (insertion in line RAL26).

✓ - insertion with right size; ✓✓ - confirmed identity, **+/-** insertion with wrong size; ***X*** – no TE insertion; ***XX*** - wrong TE inserted, **Fw** – sequenced with forward primer, **Rv** – sequenced with reverse primer, **Ta** – annealing temperature (°C); **bp** – base pairs.

Table S2: Primers to confirm DGRP *in silico* predictions (Task 1)

TE	Chrom	Line	Fw/Rv	Sequence	Ta	Frag size
LTR elements						
copia	3R	810	Fw	TCCTCTCCCCCTCTCTGTCT	60,34	238
			Rv	TTAAGCCCAACCACATAGCC	59,96	
copia	2L	908	Fw	CACGTGTCCATAGCCCATTT	60,78	239
			Rv	CTGCTTAACCATTGCGTCCT	60,27	
copia	3R	908	Fw	CTCGAGAGTTCGGAAAGCAT	59,57	219
			Rv	AGGACTCTGGACAGGTGGTG	60,15	
opus	2L	908	Fw	GCATGACGATTACGTGGCTA	59,72	162
			Rv	ACAACCAAACGCTTTTCACC	60,02	
opus	2R	443	Fw	ATATGTCCTCGCCTGACCTG	60,1	153
			Rv	GTTTCCACTGCACAGCCATA	59,72	
Transpac	2R	810	Fw	TTGGTGCCTAACCGAAAAAC	59,97	237
			Rv	TTGTCGCCGTTCTGTAGTTG	59,9	
Transpac	3L	908	Fw	CCCACTTCCTCTTCCACTCA	60,23	212
			Rv	AGTCGACCAGGGACAATGAC	59,97	
Transpac	X	357	Fw	TAACGATGGTGGCTGCTACA	60,28	209
			Rv	AAGGAAAGCGATTCAAGACC	58,39	
Transpac	X	443	Fw	CTGCAACTTTCCATGGCTTT	60,25	150
			Rv	ACAGCTTTCCCCTTCTGGAT	60,07	
Non - LTR elements						
I-element	2R	357	Fw	TCCGTCGGCTCTTATTTGTC	60,21	236
			Rv	CGTCTTACACTCGCAGCAAA	60,19	
I-element	2R	810	Fw	CCCAGATTCGCAATACCAAA	60,83	238
			Rv	AACAAAAGCAACCACCAAGG	60,01	
I-element	3L	908	Fw	TCGAATTGATACAACCCCAAT	59,15	223
			Rv	CTACTACGGCGGTGTTGGTT	60,05	
I-element	3R	357	Fw	GGCAGTGCAAACAAAAACAA	59,75	240
			Rv	CTGAGGCCAAGGACTTATGC	59,84	
I-element	3R	810	Fw	ACCTCATAGGGGGTGCTTTT	59,83	205
			Rv	TTGGAAGTGAAGGCTTTGAA	58,47	
Juan	2R	810	Fw	CTAACACGTTTCCGCCAAGT	60,17	194
			Rv	TTCGAGGGTGTGGGTGTATT	60,23	
Juan	3L	357	Fw	TCAAGTCCCAGATGCACTCA	60,4	246
			Rv	ATGTGGAACTTGGAGGATGC	59,93	
Juan	X	908	Fw	TCGAAGCCATTGCTATTTTTG	60,21	203
			Rv	TGACACCTATTCCTCAGACTCG	59,35	
Juan	3L	443	Fw	CAATCGCCTAGATCGCTTGT	60,37	168
			Rv	AGTAGCAGGTCGCCTTGAAA	60,02	
F-element	3R	810	Fw	TAGGCGCTGTTATTGAAACC	57,93	299
			Rv	CAGTGAAAGTGGGTGCAAA	58,23	
F-element	3L	908	Fw	GGGATTTGCTCTTGCTCTTG	59,96	246
			Rv	GCCATGGTCGAAACAAAACT	59,98	
F-	3R	908	Fw	GCTTGTCAAAGGGTCCAAGA	60,23	156

element			Rv	TGTTATGTGCGCGAACTTGT	60,32	
Doc	2L	357	Fw	AAAATCCATTCGGCAAACTG	59,94	280
			Rv	TCGATCAGCGCCTAGTATCA	59,55	
Doc	3R	357	Fw	ATTGTCTGCGCAACTGTCTG	60,06	213
			Rv	ATGAATTCGTCTGCCTGTCC	60,08	
Doc	2R	810	Fw	CGAAGACATCAGTCCTGCAA	59,98	246
			Rv	CCGCTGACTGTGATTGCTAA	60,01	
Doc	3R	908	Fw	GCACGAGACTCACACAGGAA	60,03	149
			Rv	TTATGGCCATTGTACGCTGA	60,1	
Doc	2L	908	Fw	TGCATCTGTGTGCGTATGTG	60,35	157
			Rv	GCACTTTTTGCCTCTGTTCC	59,86	
TIR elements						
hopper	2L	357	Fw	ACCCATCAGACTTCCACGAC	59,97	238
			Rv	GGAATCGCCTACAGAAGCTG	59,98	
hopper	3R	357	Fw	TCGATTTGGCTGGAAACTCT	59,81	173
			Rv	ATGCTGAACACGATGTGGAA	60,12	
hopper	3R	908	Fw	GGGTACAATCAAATCGAGCTTC	59,97	224
			Rv	GCGAAAACTGCACTCAATCA	60	
hopper	X	810	Fw	CTTCGTTTCATTTGGCCATT	59,94	380
			Rv	TGTGCCAAAAACACAGGCTA	60,29	
pogo	2R	810	Fw	GGCTACGACATTTCCGTTGT	60	165
			Rv	AACCTATTCCTTGCGGACCT	59,96	
pogo	3L	810	Fw	TTCAATACGGATTTGCCACA	59,93	245
			Rv	GCAAAAATAAGGGCCATCCT	60,28	
pogo	3R	357	Fw	GTTGAGCAAACAGACCCACA	59,73	217
			Rv	GGAGCCTCATAATCCGGTCT	60,43	
pogo	3R	357	Fw	AACTCGAATCTGGCTCGAAA	59,96	238
			Rv	AGTGGCCTTATCGATTGGAA	59,53	
pogo	X	908	Fw	GATGTTTCGTGTGGCTGTTG	60,16	247
			Rv	GCAGTCGCTGCAGTTTGATA	60,17	
hobo	3R	357	Fw	CTCCCAAGGATTCTGTCCAA	60,04	168
			Rv	AATGTTTCCCAAAGCTGACG	60,11	
hobo	3R	357	Fw	GGGTCTGAAAGCAGCTATGG	59,84	242
			Rv	CATTGTTCTTGGCTGACGAA	59,84	
hobo	2R	810	Fw	TCAACGCTGAAAAGTATGCAA	59,5	248
			Rv	GCAGATGATGTTGGCTTGAA	59,81	
hobo	3L	810	Fw	AGCTTTAGCCACAGCCACAT	59,9	213
			Rv	GAGAGGCACGCAGGTAAGAC	60,02	
hobo	3R	908	Fw	CAAAGGCAGGGCTAACAAAA	60,24	240
			Rv	CACAAGTGGGAGCATCAACA	60,72	

Legend: Chrom – chromosome, **Line** – DGRP genotype, **Fw** – forward primer, **Rv** – reverse primer, **Ta** – anneling temperature (°C), **Frag size** – fragment size without TE

Table S3: Number of insertions (novel, shared and ratio between both) for Cr1a, 1360, blood, gypsy5, hobo, I-element, INE-1, jockey, mdg1, pogo and roo, for 9 DGRP lines (Tasks 2 and 3)

Line	Cr1a	1360	blood	gypsy5	hobo	I-element
RAL-21	2|680|0,03	27|128|0,21	8|8|1	1|9|0,11	24|38|0,63	10|17|0,59
RAL-237	0|5|0	0|7|0	0|0|-	0|0|-	0|1|0	0|1|0
RAL-321	3|81|0,04	27|168|0,16	5|5|1	0|8|0	23|38|0,61	11|21|0,52
RAL-357	2|84|0,02	20|139|0,14	10|11|0,91	0|5|0	37|51|0,73	7|14|0,5
RAL-358	0|38|0	2|32|0,06	1|1|1	0|2|0	1|5|0,20	0|4|0
RAL-375	3|80|0,04	31|160|0,19	12|13|0,92	0|5|0	20|35|0,57	15|24|0,63
RAL-391	0|71|0	25|152|0,16	7|8|0,88	0|4|0	29|45|0,64	7|15|0,47
RAL-790	3|69|0,04	15|124|0,12	9|9|1	12|20|0,6	22|36|0,61	8|17|0,47
RAL-908	4|68|0,06	16|118|0,14	7|7|1	0|6|0	34|48|0,71	12|21|0,57

Line	INE-1	jockey	mdg1	pogo	roo
RAL-21	155|417|0,37	22|28|0,79	7|12|0,58	88|91|0,97	69|78|0,88
RAL-237	3|13|0,23	0|0|-	0|1|0	0|0|-	1|4|0,25
RAL-321	218|634|0,34	37|46|0,8	8|11|0,73	13|19|0,68	87|96|0,91
RAL-357	209|987|0,21	32|42|0,76	12|19|0,63	22|27|0,81	70|80|0,88
RAL-358	6|55|0,11	2|3|0,67	0|2|0	1|1|1	3|8|0,38
RAL-375	212|906|0,23	34|46|0,74	8|16|0,5	24|29|0,83	75|85|0,88
RAL-391	249|755|0,33	41|50|0,82	14|17|0,82	16|20|0,8	73|83|0,88
RAL-790	155|408|0,38	29|35|0,83	5|8|0,63	15|21|0,71	58|68|0,85
RAL-908	199|513|0,39	32|37|0,86	12|16|0,75	26|32|0,81	103|113|0,91

Legend: novel **|** total **|** proportion of novel insertions (racio novel/total), **Line** – DGRP genotype, light grey background – lines of this study with the highest number of novel insertions for the TE, dark grey background – lines of this study with the lowest number of novel insertions for the TE.

Table S4: Primers for qPCR (Tasks 2 and 3)

TE	Forward (Fw) Reverse (Rv) primers	Sequence	Order	Primers design
1360	Fw	TCTAGCACAACACGCACACT	TIR	Designed in PRIMER3
	Rv	GTGACGGCCAAAATTGCTGT		Designed in PRIMER3
blood	Fw	AACAATAGAAAGAAGCCACCGAAC	LTR	Handler *et al.* (2011)
	Rv	AGTCATGGACTATTGAGGGTGTTG		Handler *et al.* (2011)
Cr1a	Fw	TGGCCGTACAAGTGATGACC	non-LTR	Designed in PRIMER3
	Rv	TCATCTCGTTCGCAACCACA		Designed in PRIMER3
gypsy5	Fw	GCCCAGAGACAACGACAGAA	LTR	Designed in PRIMER3
	Rv	CTGTCTTTGCTGTCCCGGAT		Designed in PRIMER3
hobo	Fw	CATTAAGTCGGAAGGCCAAA	TIR	Designed in PRIMER3
	Rv	CTTGCTCTTCCGCTATCCAC		Designed in PRIMER3
I-element	Fw	CAATCACAACAACAAAATC	non-LTR	Specchia *et al.* (2010)
	Rv	GGTGTTGGTGTGGTTGGTTG		Specchia *et al.* (2010)
INE-1	Fw	GCCGAGTCGATCTTGCCATA	TIR	Designed in PRIMER3
	Rv	TTGTGGACGTTAGAGTGGGC		Designed in PRIMER3
jockey	Fw	GCGGATTAACAAGGGGCTCT	non-LTR	Designed in PRIMER3
	Rv	CCTGGGAGATAGATGCGCTG		Designed in PRIMER3
mdg1	Fw	GTCAGAAGGAGGCCATTCAGGAATTT	LTR	Navarro *et al*. (2009)
	Rv	GTTGCTGGCGGTTTCTGTTATTGTCAA		Navarro *et al*. (2009)
pogo	Fw	CCAGCGATAACGAAGAAAGC	TIR	Designed in PRIMER3
	Rv	GCTGCAAACCCATCCTTAAA		Designed in PRIMER3
roo	Fw	CGTCTGCAATGTACTGGCTCT	LTR	Specchia *et al.* (2010)
	Rv	CGGCACTCCACTAACTTCTCC		Specchia *et al.* (2010)
RPL32 (control gene)	Fw	ATGCTAAGCTGTCGCACAAATG	Ponton *et al*. (2010)	
	Rw	GTTCGATCCGTAACCGATGT	Ponton *et al*. (2010)	

Figure S1: Correlation between number of insertions and TE activity (Task 2)

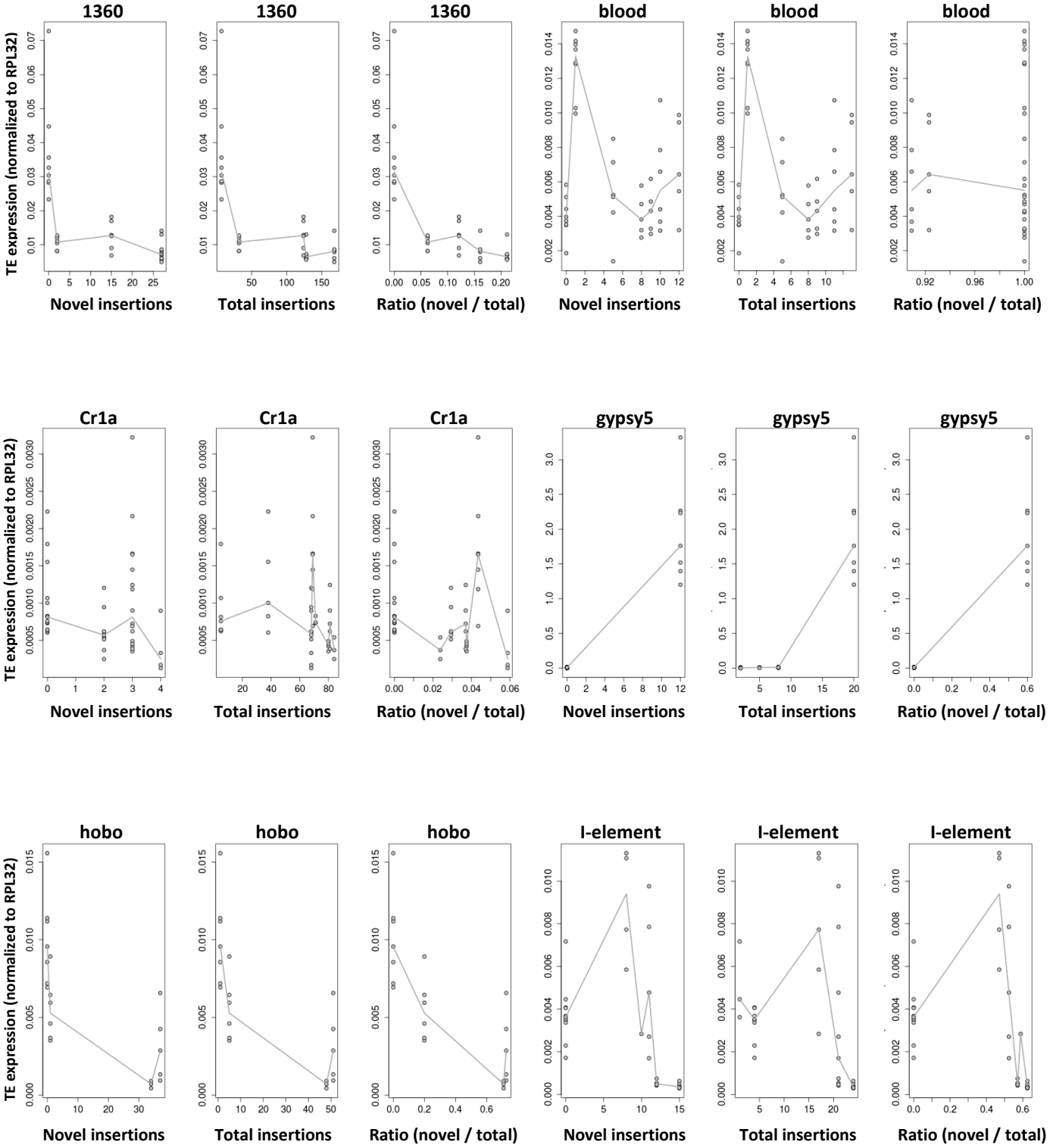

Figure 1: Correlation between number of insertions and TE activity (Task 2)

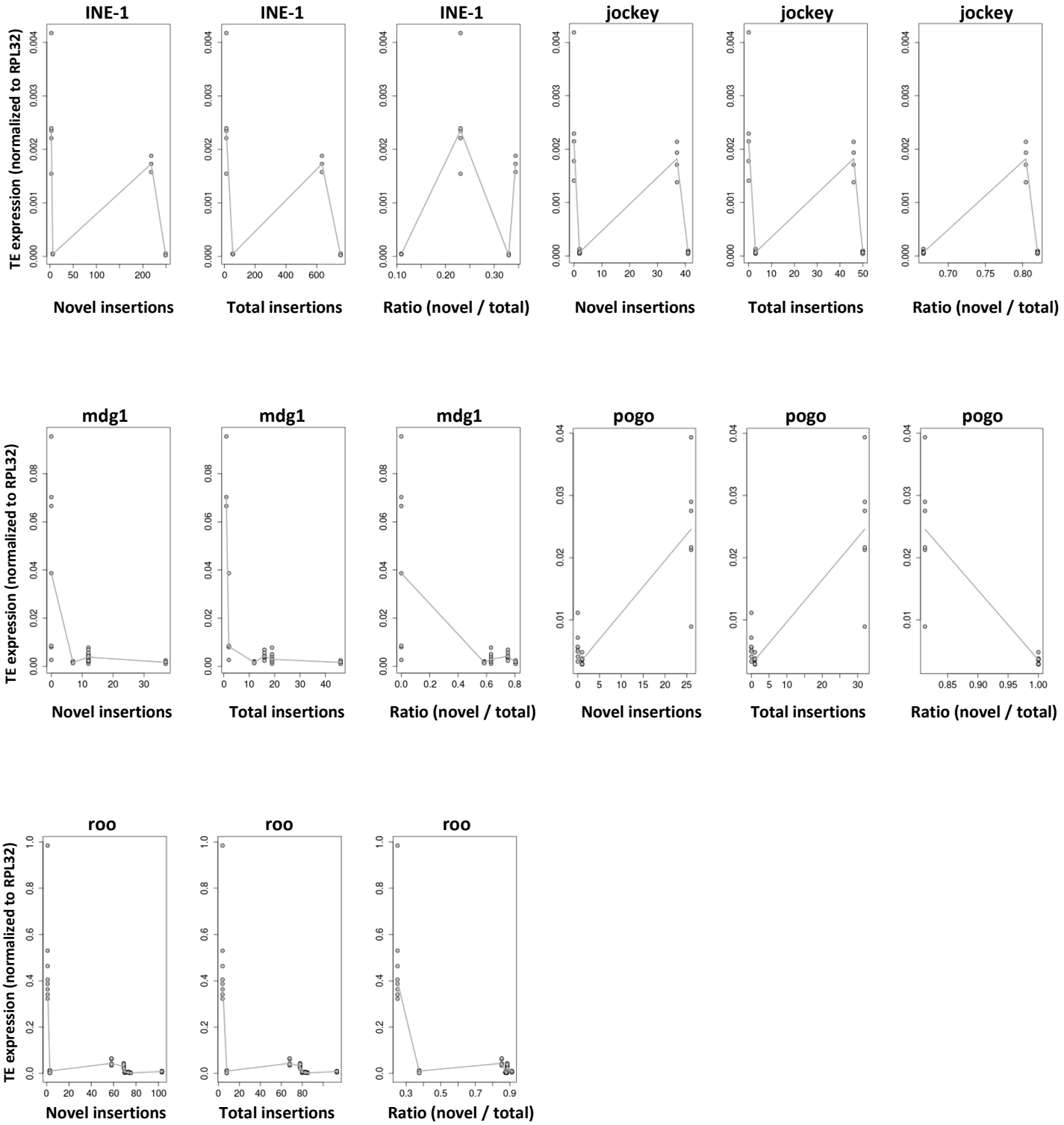

Figure 1: Graphics showing the association between number of insertions (novel, total or ratio novel/total) and TE expression (Task 2). Y-axes represent expression of the TE normalized to RPL32, X-axes represent the number of insertions (novel or total) or the proportion of novel (racio novel/total). Each point represents one biological replicate and the lines were drawn with the median of replicates for each number of insertions