The transposon *Bari-Jheh* in *Drosophila melanogaster*: adaptive phenotypes, molecular mechanisms, and genetic inheritance

Lain Guio Leiman

The transposon *Bari-Jheh* in *Drosophila melanogaster*: adaptive phenotypes, molecular mechanisms, and genetic inheritance

Lain Guio Leiman

2016

TESIS DOCTORAL
UNIVERSITAT DE BARCELONA

FACULTAT DE BIOLOGIA
DEPARTAMENT DE GENÈTICA
PROGRAMA DE DOCTORAT EN GENÈTICA

The transposon *Bari-Jheh* in *Drosophila melanogaster*: adaptive phenotypes, molecular mechanisms, and genetic inheritance

"El transposón *Bari-Jheh* en *Drosophila melanogaster*: fenotipos adaptativos, mecanismos moleculares y herencia genética"

Memoria presentada por Lain Guio Leiman para optar al grado de doctor por la Universitat de Barcelona

Trabajo realizado en el Instituto de Biología Evolutiva (CSIC-Universitat Pompeu Fabra)

Lain Guio Leiman
Barcelona 7 de enero de 2016

Tutor
Dr. Francesc Mestres Naval
Universitat de Barcelona

Thesis advisor
Dra. Josefa González Pérez
Instituto de Biología Evolutiva

Acknowledgments

Cada una de las palabras que componen esta tesis las escribí con la ayuda de Pepi González. Además de permitirme formar parte de este proyecto y de enseñarme a pensar y actuar como científico, me dio la oportunidad de volver a esta actividad que creía olvidada. Todas las gracias del mundo no son suficientes para explicar mi deuda con ella por ser la locomotora imparable que ha permitido realizar esta tesis. Gracias por enseñarme a ser crítico con los resultados y a aceptar la evidencia. Gracias Pepi por tu tiempo, tu paciencia, tu pasión, tu sabiduría y por tu inteligencia para hacer de mí un poquito mejor persona y científico. Cada acierto de este trabajo es mérito tuyo y los errores son sólo achacables a quien suscribe. Detrás de toda gran mujer a veces hay un hombre, y este es el caso, gracias también a vos, Roberto, por todo.

Quiero agradecer también a David Martín y su mente preclara con sus sabios consejos en aspectos muy variados entre los cuales también se incluye la Biología. Gracias David por ayudarme a recuperar el placer de la conversación inteligente y demostrarme la importancia de pensar y reflexionar antes de diseñar cualquier experimento, por sencillo que sea. Porque hacer ciencia es principalmente pensar y discutir y proponer ideas antes de usar las pipetas. Las mismas palabras le dedico a Xavier Franch, indispensable compañero de discusiones científicas y de las otras, gracias por las interminables horas de conversaciones sobre historia y política pero especialmente por tus enciclopédicos conocimientos de genética y de Drosophila que han sido más útiles de lo que puedas imaginar.

Gracias también a José Luis Maestro, maestro en el arte de las PCR cuantitativas. Gracias por enseñarme los detalles de esta técnica, sin vos probablemente estos experimentos no habrían llegado nunca a buen puerto y son una parte fundamental de esta tesis. También te agradezco todo el tiempo que me ayudaste con la Hormona Juvenil y tu paciencia ante mi ignorancia. Pero especialmente gracias por ser el señor de la bondad y el optimismo imperturbable a pesar de las condiciones adversas. Gracias también a la hermosa dama del lado oscuro, Elena Casacuberta, por las células y la ayuda con las transfecciones, por todo el tiempo que me dedicó en el arte de inmunoprecipitar cromatina, por los reactivos, los protocolos y las horas durante las cuales lograste vencer mi testarudez para lograr entender cómo hacer bien las cosas. Gracias por estar siempre dispuesta a ayudar y a

animarme incluso en las noches más oscuras. Pero especialmente gracias por recordarme lo hermoso del lado salvaje.

Thanks to Cristina Vieira and the laboratory of transposable elements in Lyon, France, thank you for your suggestions and support. Thank you for teaching me how to dissect flies and the details and secrets of chromatin immunoprecipitation. I am very grateful to Marie Fablet, Nelly Burlet, Clèment Goubert and all the people in the lab for the kind support and beer sessions every Friday afternoon.

I would also like to thank my friend Hung Le for showing me how similar we all are despite cultural differences, for all the presents and all the help taking care of the flies. Thank you for counting flies with me in the lab and for the company and the beers out of the lab.

Eskerrik asko Maite Barrón por los análisis de secuencia de *Bari-Jheh* y la búsqueda de elementos asociados a estrés oxidativo en Drosophila. Gracias especialmente por mostrarme una visión completamente diferente a la mía en tantos aspectos académicos y no académicos y, especialmente, gracias por transmitirme la pasión por la genética de poblaciones y las mutaciones neutras.

Gracias a mis chicas hermosas del laboratorio, Lidia, flamante madre de Laia y brillante científica. Gracias por tu pasión y por tu inteligencia. Gracias también por tu rigor y perfeccionismo y al mismo tiempo por tu despiste constante. Has sido el hermoso hallazgo de un diamante en bruto que poco a poco disfruto viendo como se está convirtiendo en un brillante de mil kilates. A Annita preciosa, por su incansable fuerza. Contra viento y marea rema desde hace tres años a nuestro lado sin importar los obstáculos y eso la hace indispensable. A Miriam, que es como una hermanita pequeña, siempre alegre y optimista. A las tres gracias por compartir estos años y por tantas horas separando mosquitas. También a todos los que han pasado por el laboratorio y que de alguna manera me han ayudado en mayor o menor medida: Eduard, Paula, Marta, Georgina, Quirze, Karine, Jaione, Iván y algunos de los que (¡lo siento!) no recuerdo el nombre. A Ion o Jon, por su ayuda con los mutantes y la Hormona Juvenil y porque creo que al final logramos entendernos. Thank you Vivien, although we did not spend too much time together, I really enjoyed our conversations.

Al otro lado de la poyata de las moscas vivían unas hermosas mariposas que tuvieron que soportar y aún soportan nuestra invasión mosqueril. Pero la soportan con alegría y eso merece un agradecimiento. Raluca, con su fina ironía y sutil inteligencia ha sido un placer de conocer. Margarita con su paciencia infinita con nuestra destrucción y desorden con los geles y las cajas de moscas también ha sido un ejemplo de dulzura. A las dos muchas gracias. También a Roger, Vlad y Gerard que, aunque estuvieron poco en el labo, también nos sufrieron.

Fuera del laboratorio, pero dentro del instituto de biología evolutiva, quiero agradecer a Cristina de Miguel por su ayuda con las clonaciones, por su dulzura y por su hermosa amistad. A Neus Bota, porque es una fuerza imparable de la naturaleza que tuve la suerte de conocer y de disfrutar sus exquisitas tartas (hechas con ayuda de Héctor). A Enric, que me mostró que ser un grande no está reñido con ser humilde y, finalmente a nuestro ángel del renacimiento, Silvia, reciente pero brillante incorporación que estoy seguro que dará que hablar en un futuro cercano. A toda la gente del IBE en general que, de una u otra manera, me han ayudado con esta tesis, Elisenda, Pepi Cruz, Eulalia, Songul, Jesús, Cristina Olivella, Cristina Manjon, Josep, Cecilia, Blanca, Rita, Emiliano, Vicente, Anna y tantos más que son demasiados para nombrar. Todos me habéis enseñado algo y todos me habéis ayudado para llegar a este punto. Gracias por todas esas cosas que un estudiante necesita, incluso sin saber que las necesita.

En este mundo todo se mueve con dinero, así que agradezco también a la AGAUR por la beca doctoral de estos tres años, al campus de excelencia de la Universidad de Barcelona por la beca de tres meses en Lyon y al Plan Nacional del ministerio de economía y competitividad y a la beca de la comisión Europea que le dieron a Pepi porque de ahí salió la pasta para los experimentos y para los últimos meses de la tesis.

Quiero agradecer a la gente del IRTA, especialmente a Miquel Peris por permitirnos recoger moscas en sus campos de manzanas y peras y por su colaboración en la aplicación de insecticidas. Dentro de la UB quiero agradecer muy especialmente a mi tutor Francesc Mestres por toda la ayuda y optimismo desde el primer día que lo conocí hasta el día de entrega de este documento; muchas gracias por estar siempre pendiente de mi actividad. Aprovecho también para agradecer a María Rosa del departamento de Genética y a toda la gente de la administración de la UB por su buena predisposición siempre.

Fuera del ámbito académico quiero agradecer a mis padres, Alicia y Ángel, a mi hermano Esteban y a mi hermana Lucrecia por su incondicional y constante apoyo y por su amor infinito del cual no sé si soy merecedor. También le agradezco a Glòria, mi compañera de la vida, porque con su infinita paciencia lo ha soportado todo y aún permanece a mi lado, obligándome a no darme por vencido jamás. Nunca habría empezado esta aventura si no hubiese sido por tu apoyo y tu amor incondicional. Gracias Glòria, mi amor, por tanto Gracias también por ayudarme con la edición. Y gracias por Julia, sin acento. Empecé estos agradecimientos con la mujer que me llevó de la mano hasta aquí los cuatro últimos años así que quiero terminarlos con la mujer con la que empecé este camino hace 15 años. Sin saber lo que quería, ni hacia donde iba, Naymé me acompaño sin dudarlo un instante, confiando en que este día llegaría. A pesar de las vicisitudes siempre estuvo allí y nuestra amistad perdurará siempre. Gracias Nay. Gracias a todos, esta tesis es el logro de todos ustedes, yo solo escribo.

Content

1 Introduction

1 Introduction

1.1 The concept of adaptation

Adaptation is the evolutionary process whereby a population goes towards the phenotype that best fits the environment it inhabits (Orr 2005). Adaptation could be also described, in terms of population genetics, as the process that changes the frequency of those alleles that will produce phenotypes with higher fitness in a population under a specific environment (Orr 2002, 2009). The concept of adaptation is strongly associated with the concept of fitness. Fitness is defined as the ability to survive and/or reproduce of each member of one specific population in a specific environment (Orr 2009). Since the environment has a limited capacity, the likelihood of success for one specific organism will be determined by its *relative* fitness, among other factors, defined as the differential ability to survive and/or reproduce of each member of a specific population in comparison with other members of the same population (Graur and Li 2000).

Although the concept of adaptation has been clear for some years, the mechanisms behind this process are poorly understood. Mapping adaptive mutations to their relevant phenotypic effects is far from trivial due to the existence of epistatic interactions between alleles, the pleiotropic effects of genes, and the polygenic nature of some adaptive traits (Mackay 2010; Barrett and Hoekstra 2011; Lehner 2013; Orgogozo, et al. 2015). For these reasons, understanding how organisms adapt to their environment remains an open question in Biology. It is also important to stress out that understanding adaptation is important not only for evolutionary biology but also for agriculture and medicine (Hunter 2007; Gluckman, et al. 2011). Besides underlying the ability of species to survive in changing environments, adaptation is also the process behind resistance to pesticides, insecticides, and antibiotics.

1.2 The evolutionary process of adaptation

The process of adaptation is based on three principles: i) variability among the individuals of a population, ii) heredity of this variability, and iii) selective pressure of the environment on the individuals that will allow survival and/or reproduction of only a few members of a population. These three principles summarize of the theory of evolution by natural selection proposed by Charles Darwin, and they perfectly match the process of

adaptation. However, it is important to mention that adaptation is not natural selection. Although natural selection is necessary for adaptation, it is relevant that not all naturally selected traits are adaptive; neither have all adaptive traits been naturally selected (Gould and Lewontin 1979). Moreover, natural selection acts even after optimum fitness is reached, so there is a chance of optimum loss despite the organism's good fit (Orr 2005; Thompson 2013).

Heritable variability could be produced by recombination, migration, and mutations, but because without mutations, recombination and migration will not produce variability, we could assert that mutations are the primary source of genetic variation. Mutations are random in their fitness effect. Other evolutionary forces, such as genetic drift could change its frequency no matter the adaptive or deleterious effect of the new mutation. When a mutation is adaptive, the fate of this adaptive allele will be determined by four main factors: the effective population size (N_e), the size effect of the mutation, the frequency of the mutation in the population at the moment that positive selection affects the mutation, and the dominance effect of the mutation (Olson-Manning, et al. 2012).

The effective population size is a concept that was strictly defined by Wright in 1931 as "the size of an idealized population that would have the same effect of random sampling on allele frequencies as that of the actual population" (Wright 1931). There are several factors that can influence the N_e, such as migrations, bottlenecks, isolation, sex ratio, offspring number, inbreeding, and mode of inheritance, among others (Graur and Li 2000; Charlesworth 2009). Since mutations are random, bigger populations have higher chances of occurrence of an adaptive mutation. Moreover evolutionary forces, such us genetic drift are N_e dependent: the bigger the N_e the lower the effect of genetic drift, avoiding the chance of loss of an adaptive mutation because of this reason (Olson-Manning, et al. 2012).

The size of the effect of an adaptive mutation explains the effect of the molecular change on the phenotype of an organism and consequently on its fitness (Graur and Li 2000). When the mutations affect genes with a small role in a specific phenotype, they are considered small size effect mutations. When the mutations affect major genes with a critical role in a specific phenotype they are considered large size effect mutations. The geometrical model by R.A. Fisher predicts a high frequency of small size effect mutations and a very low frequency of large effect size mutations. However, small size mutations

are more prone to disappear before selection. Taking into account both factors, Motoo Kimura proposed that mutations of medium effect are the most likely to contribute to adaptation. Thus, the size effect affects the fate of an adaptive mutation (Reviewed in Orr 2005 and Olson-Manning, et al. 2012).

The adaptation of a population to a new environment can be caused by new mutations but also by standing genetic variation. The frequency of a new adaptive mutation in a diploid organism population is always 1/2N being N the number of individuals in a population. This new adaptive allele, present at a very low frequency, could disappear in the next generations due to genetic drift. On the other hand, a neutral or mildly deleterious mutation could be in the population at low or intermediate frequencies. When a population faces an environmental change, this allele could turn to be beneficial rising up very rapidly in frequency, and even going up to fixation. Thus, the frequency of an adaptive mutation at the moment of selection will determine its fate (Revised in Orr 2005 and Olson-Manning, et al. 2012).

The last main factor that affects the fate of an adaptive mutation is the dominance effect. John B.S. Haldane proposed in 1927 that the fixation probability of a new mutation it is higher when is dominant rather than when it is recessive (Haldane 1927). This effect is known as "Haldane's sieve" (Turner 1981; Charlesworth 1992). A new adaptive mutation, if dominant, will produce an adaptive phenotype in the first member of a population where the mutation appears and about 50% of its offspring will carry this mutation increasing the frequency of the new allele. The spread of this mutation will be faster than that of a recessive mutation. A recessive mutation must wait until at least two members of the population of opposite sex carry one allele of the mutation to obtain about 25% of the offspring with the adaptive phenotype. Until this situation occurs, the chance of loss of the mutation is high because the frequency of the mutation will remain very low during several generations (see BOX1 for a detailed explanation of the dominance effect).

The idea that adaptive mutations are dominant is widely accepted (Bourguet, et al. 1997; Charlesworth 1998; Orr 2010; Zhang, et al. 2011; Joseph, et al. 2014). However, new studies propose an alternative to this idea. Orr and Betancourt (2001) demonstrated that when an adaptive mutation has originated from mutation-selection equilibrium, the dominance effect does not affect the likelihood of fixation, proposing an alternative to the bias against recessive mutation of the Haldane's sieve.

BOX1: THE DOMINANCE EFFECT

The fitness of a new allele will depend on the mode of interaction between this new mutation and the old allele. If we assume that the fitness of the old allele is 1 then the new allele will change the fitness in "***s***", so if ***s*** is positive the new allele will be advantageous and if ***s*** is negative it will be detrimental. Thus, for a diploid population, with two different alleles A_1 and A_2, where A_1 is the old allele with fitness equal to 1, five models of interaction can be considered:

1) Complete dominance:

A_1A_1	A_1A_2	A_2A_2
1	$1 + s$	$1 + s$

In this situation, one allele or two show the same effect

2) Complete recessive:

A_1A_1	A_1A_2	A_2A_2
1	1	$1 + s$

In this situation, two alleles are necessary to see the change in the fitness

3) Codominance:

A_1A_1	A_1A_2	A_2A_2
1	$1 + s$	$1 + 2s$

In this situation, the fitness effect is increased with the number of alleles.

4) Overdominance:

A_1A_1	A_1A_2	A_2A_2	
1	$1 + s$	$1 + t$	$s > 0$ and $s > t$

This situation is also known as balancing selection or stabilizing selection, heterozygous individuals will present a higher fitness than any homozygous individual. If t=0 then the frequency equilibrium will be probably at 50%.

5) Underdominance:

A_1A_1	A_1A_2	A_2A_2	
1	$1 + s$	$1 + t$	$s < 0$ and $s < t$

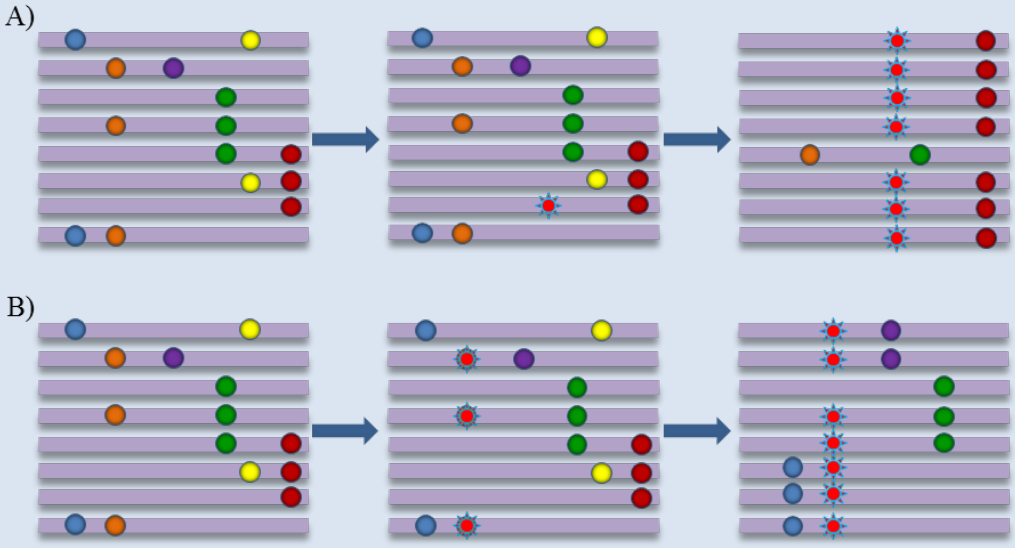
In this situation, the population reaches an unstable equilibrium where any variation in frequency will cause one of the alleles to be eliminated from the population. (Adapted from Graur and Li 2000).

More recently Sellis et al. (2011) showed the importance of heterozygous advantage in adaptation. They demonstrated that in a situation of overdominance (See BOX1), for large effect size mutations, heterozygous advantage will be highly frequent in adaptation (Sellis, et al. 2011). In the light of these new results, the importance of the study of the dominance effect in adaptation takes on a new significance. The evolution of mating systems, the rate of adaptation in diploids, and the magnitude of inbreeding depression are some of the biological processes affected by the dominance effect. Most of the dominance effect studies have been done in recessive mutations, while there are not many

studies of dominance effect of natural adaptive mutations. If an adaptive mutation is dominant, positive selection should drive it to fixation. This rapid fixation leaves a molecular signature known as hard selective sweeps (see BOX2).

BOX2: SELECTIVE SWEEPS

When a new mutation is positively selected, the flanking regions of the locus are also selected reducing the variability of this region in the population. This effect is known as "selective sweep" (Barton 2000). When the selection is strong and affects a new mutation that is fixed in a few numbers of generations the nucleotide variability near the new mutation is drastically reduced and the signature is named "hard sweep" (A). By contrast, if the mutation is old, standing genetic variation has had enough time to recombine during several generations and in the moment of the selective pressure the variability of the flanking regions will not be so low as mentioned above, in this case the signature is called "soft sweep" (B). Small populations or low mutation rates stimulate hard sweeps, whereas large populations or high mutation rates are associated with soft sweeps (Pennings and Hermisson 2006a, b).



Light violet lines represent genomic DNA. Coloured circles represent different polymorphisms. Red lighting circle represent an adaptive mutation. In A) is a "*de novo*" mutation and in B) is a present allele that became adaptive.
Adapted from (Scheinfeldt and Tishkoff 2013)

Despite recent efforts to detect hard selective sweeps in nature, most of the findings showed partial selective sweeps (Garud, et al. 2015), a molecular signature associated with overdominance. However partial selective sweeps are not only associated with

overdominant mutations, other types of adaptive processes, such as polygenic adaptation, could leave the same molecular signature. Thus, to explicitly test whether an adaptive mutation is dominant or overdominant, we need to measure the fitness of heterozygous individuals and compare it with the fitness of homozygous individuals.

1.3 Molecular mechanisms of adaptation

Finding the link between the genotype and the phenotype of adaptive mutations is a key challenge in evolutionary biology. The reverse genetic strategy identifies adaptive mutations in the genome and then searches for the phenotype associated with that mutation. However, the link between the adaptive allele and its phenotype is not simple. Both epistasis and pleiotropy complicate the genotype-phenotype map (Mackay 2010; Barrett and Hoekstra 2011; Lehner 2013; Orgogozo, et al. 2015). Although these effects were traditionally considered as exceptions, they should be treated as fundamental components of genetic analysis (Tyler, et al. 2009). Thus, to understand the genotype-phenotype relationship it is essential to understand the molecular mechanisms that explain this connection (Pavlicev, et al. 2011; Wagner and Zhang 2011; Ostman, et al. 2012).

Genes affected by adaptive mutations could be part of molecular pathways leading to adaptive traits. Mutations affecting upstream genes of a molecular pathway that block downstream mutations were named *epistatic mutations* (reviewed in Phillips 2008). Although epistasis was first described as the discrepancy between the prediction of segregation ratios of genes and the observed phenotype, today the definition of epistasis involves several research fields and it is used to describe complex interactions among genetic loci (Phillips 2008). Epistasis must be considered in adaptation studies since it is now known to be pervasive and to play an important role in quantitative traits affecting heritability and reducing replicability (Mackay 2014). Because the genetic background of the individual where the adaptive mutation appears determines the fitness effect of the mutation, the effect of an adaptive mutation must be tested in different genetic backgrounds (Weinreich, et al. 2005; Olson-Manning, et al. 2012).

Besides epistasis, another very important challenge to link the genotype to the phenotype is pleiotropy. Briefly, pleiotropy is the situation where one gene is involved in more than one trait. There are seven types of pleiotropy, which vary depending on the type of mutation and the role of the product of the gene in the different pathways. For example,

antagonistic pleiotropy produces opposite effects in the fitness of the individual. The same mutation could be beneficial under a specific environment but could be detrimental if the environment changes (Hodgkin 1998). Despite the fact that pleiotropy was discovered several years ago (Grüneberg 1938), it is difficult to detect it. If we ignore pleiotropy, it can lead to misleading interpretations of adaptive mutations. However, there are now many tools to detect pleiotropic effects in the analysis of the phenotype of adaptive mutations (reviewed in Paaby and Rockman 2013).

Despite difficulties to link genotype to phenotype, in recent years, several genome wide studies associated different mutations with adaptive phenotypes (Hancock, et al. 2008; Hancock, et al. 2010; Fournier-Level, et al. 2011; Bergland, et al. 2014). However, only a few adaptive mutations have been identified and functionally characterized. Some of these adaptive mutations affect the coding region of genes. For example, a single amino acid substitution in the *PGI* gene increases metabolic performance in *Colias* butterflies increasing flight abilities (Wheat, et al. 2006). Another single amino acid substitution in *Mc1r* gene affects melanogenesis in *Peromyscus polionotus* mice protecting them against predators (Hoekstra, et al. 2006; Vignieri, et al. 2010). A third example of an amino acid substitution in *Ace-1* is associated with loss of sensitivity to organophosphates and carbamates in *Culex pipiens* mosquitoes increasing resistance to insecticides (Labbe, et al. 2007). Besides changes in the coding regions, mutations in regulatory regions could also produce adaptive phenotypes: Stepwise substitutions in one enhancer of *ebony* increases pigmentation in *D. melanogaster* optimizing thermal regulation (Pool and Aquadro 2007; Rebeiz, et al. 2009). Mutations in regulatory regions of *LCT* gene change its expression level in humans extending the ability to digest lactose to the adult stage (Tishkoff, et al. 2007). Moreover, mutations other than nucleotide substitutions can also induce adaptive phenotypes: a deletion affecting the Open Reading Frame of *FRI* delay flowering time in *Arabidopsis thaliana* improving winter survival (Le Corre, et al. 2002; Korves, et al. 2007). Deletions in the enhancer of *Pitx1* reduce pelvic structure in *Gasterosteus aculeatus,* the three spine stickleback fish, allowing it to adapt to fresh water environments (Chan, et al. 2010). Transposable element insertions are mutations that could also be adaptive. The insertion of a transposable element in the coding region of *CHSD* affects flower color in *Ipomea purpurea*, morning glory, improving self-fertilization (Habu, et al. 1998). The insertion of a transposable element in the coding region of *CHKov1* disrupts a choline kinase in *D. melanogaster* increasing resistance to

insecticides (Aminetzach, et al. 2005). Adaptive insertions of transposable elements in non-coding regions have also been described. An insertion in the 3'-UTR of *CG11699* interferes with the polyadenylation site of the gene, leading to a shorter transcript that is overexpressed (Mateo, et al. 2014). Overexpression of *CG11699* leads to increase *ALDH-III* enzymatic activity that is associated with increased resistance to xenobiotics in *D. melanogaster* (Mateo, et al. 2014). Finally, an intronic insertion in *Sarah* gene is associated with overexpression and with shorter developmental time in *D. melanogaster* (Ullastres, et al. 2015). Thus, all these different types of mutations inducing different molecular mechanisms and producing different adaptive phenotypes highlight the great diversity and complexity of the adaptive process.

However, not all molecular mechanisms involved in adaptation are directly associated with mutations in the DNA sequence. Chromatin state also has an important role in gene regulation. Traditionally chromatin was divided into heterochromatin, associated with gene silencing, and euchromatin, associated with active genes (Filion, et al. 2010; Boros 2012). To regulate chromatin state, cells use epigenetic mechanisms like DNA methylation or acetylation, non-coding RNAs, and histone post translational modifications. Nowadays, several studies have performed a finer classification and up to 30 different chromatin states according to the enrichment of each genome region for specific histone marks have been defined (Filion, et al. 2010; modENCODE, et al. 2010; Kharchenko, et al. 2011; Brown and Bachtrog 2014). For example, in Drosophila, methylation of Histone3 (H3) could be associated to different chromatin states (Greer and Shi 2012). H3K4me3 is usually associated with open chromatin and it is frequently present in active transcription start sites, promoters and enhancers. H3K9me3 is associated with constitutive heterochromatin that is silent most of the life time of the organism (Boros 2012; Greer and Shi 2012; Rebollo et al. 2012a). Finally H3K27me3 is associated with facultative heterochromatin. Basically, facultative heterochromatin has the ability to activate gene expression depending on developmental time and environmental effects. For example in *D. melanogaster*, *Hox* genes expression is controlled by this type of chromatin controlled by the Polycomb Group proteins and the Trithorax group proteins (Trojer and Reinberg 2007). Besides the three histone marks described above, there are more than 16 different histone post translational modifications in *D. melanogaster* (Greer and Shi 2012).

Besides gene regulation and genomic stability, the epigenetic mechanisms mentioned above have been widely described in many eukaryotic organisms as a defence system against transposable element (TE) activity. Host organisms have developed a battery of epigenetic pathways to protect its genome from the deleterious effect of TEs (Slotkin and Martienssen 2007; Fedoroff 2012; Rebollo et al. 2012b). It was proposed that epigenetic mechanisms first appeared as TEs and virus protection system and were then co-opted for gene regulation (Lisch and Bennetzen 2011). TEs are inserted mostly in the heterochromatic region of the genome in a silent state; however, some of them are inserted in the euchromatic region of the genome (Rebollo et al. 2012b; Sentmanat and Elgin 2012; Sentmanat, et al. 2013; Lee 2015). When a TE is inserted in the euchromatic region, epigenetic mechanisms identify it and silence it (Sentmanat and Elgin 2012; Lee 2015). This silencing of TEs in euchromatin could be extended to genes located nearby downregulating their expression (Rebollo et al. 2012b; Lee 2015). As a result, epigenetic mechanisms in euchromatic regions imply an evolutionary trade-off between the protection of the organism against TE transposition and the detrimental effect of gene silencing (Hollister and Gaut 2009).

1.4 Transposable elements and adaptation

Transposable elements were discovered by Barbara McClintock almost 70 years ago (Mc Clintock 1950). TEs -also known as mobile genetic elements- are fragments of DNA with the ability to move around the genome with the possibility of making copies of a in some cases. Some TEs contain the machinery to move themselves but there are many TEs, called non-autonomous TEs, which require the enzymes coded by others TEs to move. The size of TEs is between a few nucleotides and up to more than 20 kb (Wicker et al. 2007) (See Box 3). They are present in almost all the prokaryote and eukaryote organisms currently analyzed (Finnegan 1985; Feschotte and Pritham 2007; Chenais, et al. 2012). The amount of TEs in the genome varies among species, from 1-3% in some fungi and bacteria to up to 80% in some plant species (Bennetzen 2005; Biemont and Vieira 2006; Siguier, et al. 2006; Feschotte and Pritham 2007). The Human genome has more than 66% of TEs (de Koning, et al. 2011) while *Drosophila melanogaster* has approximately 20% of TEs in its genome (Barron, et al. 2014).

TEs are a main source of mutation and enhancers of variability in natural populations (Kidwell and Lisch 1997). Although most of the TEs are placed in heterochromatic

regions of the genome, there is a small fraction that is inserted nearby genes in the regulatory regions or even in the coding region of genes (Belyayev 2014). These insertions in the euchromatin are usually deleterious but there can also be neutral or adaptive. Recently, several examples of the adaptive role of TEs in different organisms have been reported (Habu, et al. 1998; Morgan, et al. 1999; Daborn, et al. 2002; Aminetzach, et al. 2005; McCue, et al. 2012; Mateo, et al. 2014; Schrader, et al. 2014; Ullastres, et al. 2015). TEs can produce many different types of mutations. The position of the insertion in the genome and the specific sequence of the TE inserted determine the type of mutation (Figure 1.1) (reviewed in Belyayev 2014 and Goodier and Kazazian 2008).

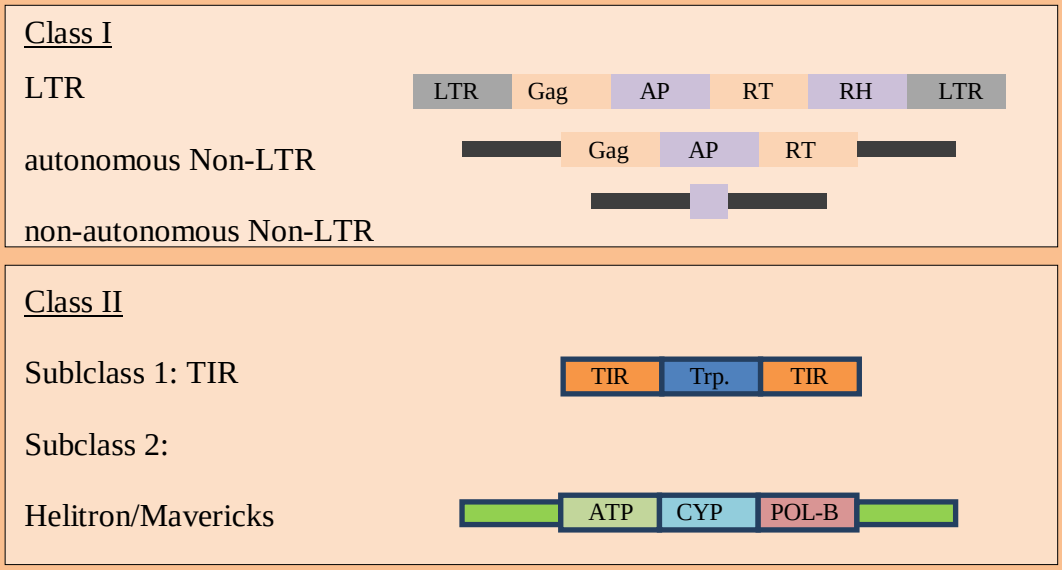
Transposable element activity can be affected by the environment (reviewed in Piacentini, et al. 2014), There is evidence that strongly suggests that stress induces transposition of TEs in bacteria (discussed in Fablet and Vieira 2011), mammals (Liu, et al. 1995; Li, et al. 1999) and plants (Zeller, et al. 2009; Tittel-Elmer, et al. 2010). This property of TEs suggests that population exposed to new mild-stress environments will show an increased number of TEs in the genome (Vieira, et al. 1999). It has been argued, that this new TEs will increase the variability of the population allowing higher evolvability and facilitating adaptation to new environments (Vieira, et al. 1999; Fablet and Vieira 2011; Stapley, et al. 2015).

Although the activation of TEs due to changes in the environment is well described, the mechanisms of its activation are not so clear yet, in particular in eukaryotes (Fablet and Vieira 2011). TE activation by stress is a complex process in which epigenetic repression of TEs plays a role: specific types of stress activate only specific families of TEs due to de-repression of particular epigenetic pathways (Rebollo, et al. 2010; Fablet and Vieira 2011; Elgart, et al. 2015; Hunter, et al. 2015).

BOX3: TRANSPOSABLE ELEMENTS CLASSIFICATION

Transposable elements can transpose using RNA as an intermediary, or they can transpose without intermediary. A first attempt to classify the TEs used this characteristic to divide them in two classes (Finnegan 1989). Class I includes the transposable elements that use RNA as an intermediary, this class is also named "retrotransposons" or RNA mediated transposons. Class II includes the transposable elements that do not use RNA, this class is also known as DNA transposons. This classification was updated to include Subclasses, Orders, Superfamilies, and Families (Wicker, et al. 2007). The authors also include new DNA transposons that can copy themselves during the transposition in a new subclass that includes Helitrons and Mavericks. A recent update included in the classification mobile elements that until now were not properly classified such as inteins and Group I introns (Piegu, et al. 2015).

Class I, retrotransposons are classified in two orders Long Terminal repeats (LTR) and non-LTR. Class II transposons present Terminal inverted Repeats (TIRs) flanking one gene that codify for a transposase. This transposase recognize TIRs and cleaves the element. Helitrons and Mavericks do not encode a transposase gene but they encode other genes that allow them to make copies of themselves (Wicker et al. 2007).



Gag (Capside protein); AP (Aspartic proteinase); RT (Reverse transcriptase); RH (RNAase H); Trp. (Transposase); ATP (Packaging ATPase); CYP (Cysteine protease) and POL-B (DNA polymerase B).(Adapted from Wicker et al. 2007).

Biotic and abiotic environmental stresses activate different protection systems of the cell to face specific stresses. This response is translated in the activation of sets of genes that will protect the organism against the hostile environment. As mentioned above, stress activates also TEs transposition. There is a connection between these two processes. When a gene is being transcribed the chromatin region is open. Since the stress related genes are being transcribed at the same moment that TEs are active, some TEs will preferably insert close to the 5' regions of stress response genes (Shilova, et al. 2006; Naito, et al. 2009; Fablet and Vieira 2011). Consequently, there are several TEs associated with the regulatory regions of stress related genes (Feschotte 2008). Nevertheless, not all insertions occur in the regulatory regions (see Figure 1.1) nor are associated with stress response (Casacuberta and Gonzalez 2013).

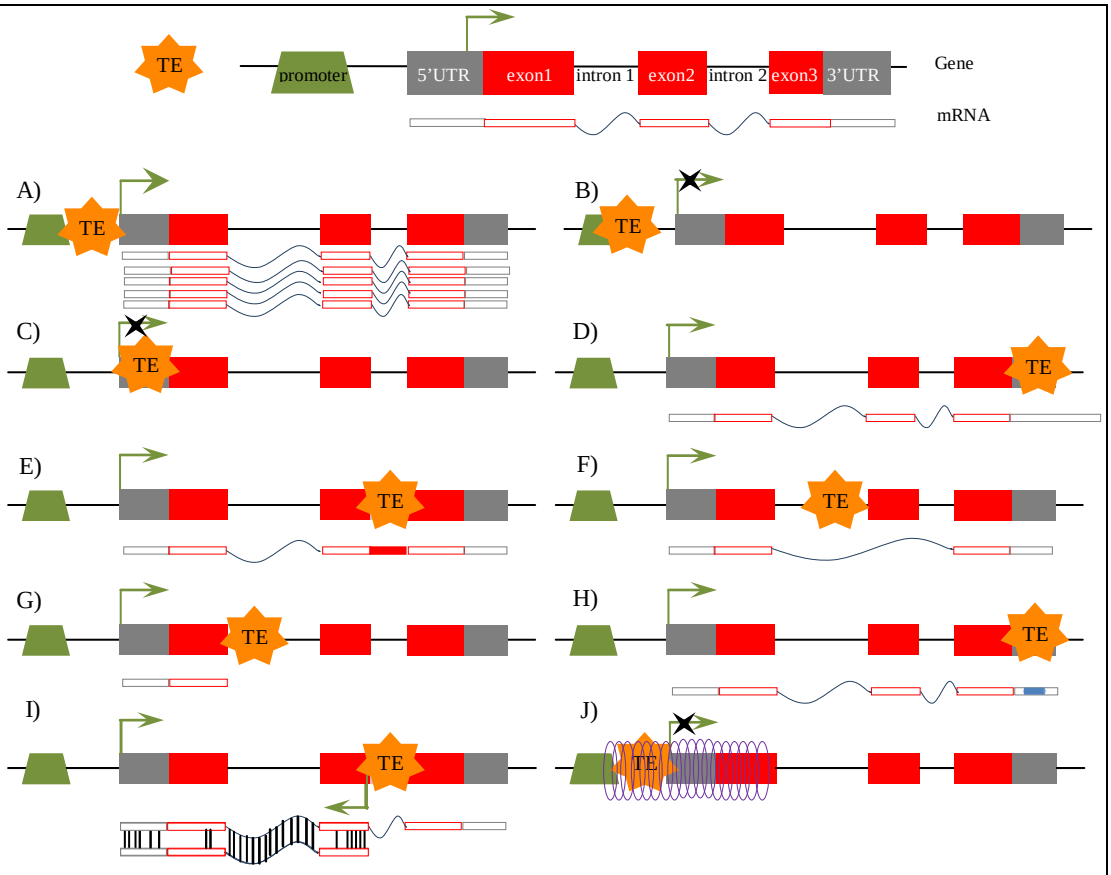


Figure 1.1. Effects of TE insertions on gene regulation and gene products. A) Cis increased expression B) Cis expression disruption C) 5'UTR disruption D) New polyadenylation site E) Exonization F) Alternative splicing G)Premature end H) miRNA target I)Antisense transcription J) Gene silencing. (Adapted from Casacuberta and González 2013).

To summarize, TEs are present in almost all living organisms analyzed so far. They are a major source of variability, generating mutations that could affect the regulation and structure of genes and their products. TEs are sensitive to the environment enhancing the diversity of a population. This increased variability could allow organisms to survive under new selective pressures (Fablet and Vieira 2011; Hua-Van, et al. 2011; Casacuberta and Gonzalez 2013). For these reasons, we argue that transposable elements are an excellent tool to understand the molecular mechanisms underlying the evolutionary process of adaptation. *D. melanogaster* is one of the organisms with the best annotated genomes including transposable elements. These characteristics make *D. melanogaster* an excellent tool to study the role of TEs in adaptation.

1.5 *Drosophila melanogaster* as a model organism in adaptation

Drosophila melanogaster was selected as a model organism for genetic studies one century ago (Morgan 1915a; Morgan 1915b; Morgan, et al. 1925). Today, *D. melanogaster* is one of the best known eukaryotic organisms and it is used as a model organism in different fields of research: from population genetics and development, to human diseases modelling, among others (Rubin 1988; Celotto and Palladino 2005; Stocker and Gallant 2008; Griffith 2012). Currently there are many research communities with *D. melanogaster* as a model organism providing countless tools to work with this species such as: The BDGP gene disruption project (Bellen, et al. 2004); The *D. melanogaster* Genetic Reference Panel (DGRP) (Mackay, et al. 2012); Flybase (dos Santos, et al. 2015); Vienna *Drosophila* Resource Center (VDRC) (Dietzl, et al. 2007); and many others. But *D. melanogaster* was not selected arbitrarily. *D. melanogaster* is small (3mm), so large amounts of individuals can be maintained in a small space. Its developmental time is short (about 11 days), so many generations can be studied in a few months. Its genome size is small (180 Mb) and a lot of functional information is available (Matthews, et al. 2015).

D. melanogaster is a good model organism to study adaptation also because of its population history. The origin of *D. melanogaster*, as a species, was dated around 17-20 million years ago in the tropical area of the African continent and today this species is cosmopolitan (David and Capy 1988; Lachaise, et al. 1988; Begun and Aquadro 1993; Andolfatto 2001; Stephan and Li 2007). The invasion of Europe and Asia took place during the last 10.000 – 16.000 years (Li and Stephan 2006; Thornton and Andolfatto

2006; Duchen, et al. 2013) and America and Australia were invaded during the last 100-200 years (Sturtevant 1920; Keller 2007). Thus, in a very short period of time *D. melanogaster* spread all around the world and it is likely that the expansion out of Africa was accompanied by frequent adaptive mutations. This rapid adaptation to new environments should leave signatures of selection in the genome such as selective sweeps. These sweeps are easy to detect with current sequencing techniques. Moreover, natural populations are very easy to find and collect worldwide making very easy the sequencing of several individuals from different populations.

1.6 The transposable element *Bari-Jheh*

In order to clarify whether adaptive transposable element insertions are exceptions or whether they are frequent in nature, Gonzalez, et al. (2008) performed the first genome-wide screening for adaptive TEs in *D. melanogaster* (González, et al. 2008). They analyzed North American and African populations to compare the frequency of all the TEs annotated in the genome at that moment and identified 18 putatively adaptive TEs (Gonzalez, et al. 2008; Gonzalez, et al. 2010). To identify these putatively adaptive TEs, the authors considered TEs with the following features: 1) TEs present at low frequency in African populations and at high frequency in non-African populations, 2) TEs inserted in high recombination regions of the genome, and 3) TEs that belong to non-neutral families. Furthermore for a subset of the TEs, the authors analyzed the flanking regions looking for signatures of selective sweeps. In all cases, they found that the regions around the putatively adaptive TEs analyzed showed signatures of a selective sweep. The element that was present at a higher frequency in North American populations, and that showed a signature of a selective sweep, was *FBti0018880* (González, et al. 2008).

FBti0018880 (henceforth *Bari-Jheh*) is a full-length transposon of 1,728 base pairs that belongs to the *Bari1* family. The *Bari1* family belongs to the Class II TIR DNA transposons. It was identified first in *D. melanogaster*, in the heterochromatic region near the centromere of chromosome 2 and in a few copies in the euchromatin (Caizzi, et al. 1993). The putative encoded protein of this element shows high similarity to the transposase encoded by *Tc1*, a family of TEs from *Caenorhabditis elegans* that belongs to the *Tc1/Mariner* superfamily (Plasterk, et al. 1999). Years later, *Bari1* was found in other species of *Drosophila* genus with highly homology copies in the *Sophophora* subgenus (Moschetti, et al. 1998). The transposase of this TE is transcribed, but a post

transcriptional modification produces a non-functional protein unable to perform the transposition (Palazzo, et al. 2013).

Further analysis of the flanking region of *Bari-Jheh* showed that this TE was the only mutation responsible for the selective sweep further suggesting its adaptive role (Gonzalez, et al. 2009). No signatures of clinal differentiation were found for this TE (González, et al. 2010), indicating that the adaptive trait related with this insertion is not associated with temperate environment adaptation. *Bari-Jheh* was present at higher frequencies in all the out-of-Africa populations analyzed (González, et al. 2008; González, et al. 2009; González et al. 2010), and later studies showed that *Bari-Jheh* is present at high frequencies in all the populations analyzed until today including African populations (Anna Ullastres, personal communication).

Bari-Jheh is placed in the 0,7-kb intergenic region between the *Juvenile hormone epoxide hydrolase 2 (Jheh2)* and *Jheh3* genes and close to *Jheh1* (Figure 1.2). Each *Jheh* gene encodes one enzyme with an epoxide hydrolase domain. These genes belong to the group of epoxide hydrolases enzymes present in mammals, insects, plants and bacteria (Barth, et al. 2004). More than 100 epoxide hydrolases enzymes have been identified and classified in 12 families. Although all of them catalyse the same reaction, they show very low sequence similarity. This diversity allows the enzymes to perform many different functions and bind to different types of substrates (Barth, et al. 2004). In *D. melanogaster* it seems that *Jheh2* was the original gene and later duplications produced *Jheh1* and *Jheh3* genes (Taniai, et al. 2003). These three genes are present in all the Drosophila genomes sequenced until today (González, et al. 2009) indicating that the duplication event was happened before the divergence of the genus Drosophila. *Jheh* genes are known, in *D. melanogaster,* as Juvenile Hormone hydrolysing enzymes (Campbell, et al. 1992); however, there are no enzymatic studies that confirm this role (Taniai, et al. 2003).

Juvenile Hormone is associated with a number of very important processes in the life of insects: development, metamorphosis, egg production, reproduction, diapause, stress resistance, behavior and aging (Flatt, et al. 2005). Based on these data, González, et al. (2009) performed phenotypic assays that showed that *Bari-Jheh* is associated with slower developmental time and reduced viability. Moreover, they analyzed the expression of *Jheh2* and *Jheh3* genes and observed that *Bari-Jheh* is associated with the down-regulation of the two genes, which is consistent with an increased JH titre that would lead

to the phenotypic effects identified (González, et al. 2009). Thus, the authors identified the cost of selection of this mutation while the adaptive effect of *Bari-Jheh* remained unknown.

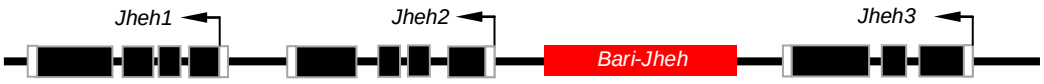


Figure 1.2: The transposon *Bari-Jheh* and its nearby genes in the *Drosophila melanogaster* genome. This region is placed in chromosome 2R between the positions 18.853.700 and 18.862.200 according to Flybase, Release 6.07. The intergenic region between *Bari-Jheh* and *Jheh2* is 45 base pairs. In previous genome annotations, the intergenic region was 350 base pairs. The intergenic region between *Bari-Jheh* and *Jheh3* is 355 base pairs. The transcription sense of *Bari-Jheh* is the same as the nearby genes.

The association of *Bari-Jheh* with detrimental effects in developmental time and viability, and its association with downregulation of *Jheh2* and *Jheh3* suggest that these genes are indeed involved in Juvenile Hormone metabolism (González et al. 2009). Besides this role, other studies identified *Jheh1* and *Jheh2* as candidate genes to be involved in stress response (Girardot, et al. 2004; Sun, et al. 2006; Willoughby, et al. 2007; Misra, et al. 2011). Moreover, Taniai, et al. (2003) showed that JHEH2 does not degrade Juvenile Hormone III but it degrades a xenobiotic compound named cis-stilbene-oxide. Because of that, the authors proposed to change the name of *Jheh2* to *DmEH* for *D. melanogaster Epoxide Hydrolase*. However, there are evidences that suggest that Juvenile Hormone activity in *D. melanogaster* could be due Juvenile Hormone III bisepoxide instead of Juvenile Hormone III (Richard, et al. 1989; Richard, et al. 1990). Thus, *Jheh2* could be involved in xenobiotic response and Juvenile Hormone metabolism. JHEH2 is a microsomal enzyme suggesting a possible role of *Jheh2* and maybe *Jheh1* and *Jheh3* in the detoxification response (Taniai, et al. 2003). Previous studies showed the role of epoxide hydrolase enzymes in the xenobiotic and detoxification response in mammals, specifically the microsomal epoxy hydrolases seems to be the most important in this role (Oesch 1973; Oesch, et al. 2000).

Thus, since *Jheh* genes seem to play other functional roles besides Juvenile Hormone metabolism, we hypothesize that *Bari-Jheh* could play an adaptive role related to the other functions of *Jheh* genes.

2 Objectives

2. Objectives

The main objectives of the present thesis are:

- To find the adaptive phenotype of *Bari-Jheh* insertion. To do this, we will analyse the *Bari-Jheh* sequence and the function of the nearby genes to construct plausible hypotheses about the adaptive effect of this insertion. We will test the candidate adaptive traits in flies, with and without the insertion, from different genetic backgrounds. This objective will allow us to identify the adaptive traits in which *Bari-Jheh* is involved.

- To unravel the molecular mechanisms associated with *Bari-Jheh*. To do this, we will analyse the expression of the genes nearby *Bari-Jheh*. We will search for cis-regulatory elements in the *Bari-Jheh* sequence. Moreover, we will analyse whether *Bari-Jheh* is introducing epigenetic marks. This objective will allow us to extend our current knowledge of the molecular mechanisms underlying adaptation.

- To characterize the dominance effect of *Bari-Jheh*. Because recent findings provide evidence supporting the idea that adaptive mutations are overdominant challenging the traditional idea of adaptive mutations being dominant, we will test the dominance effect of *Bari-Jheh* by comparing the phenotype of heterozygous flies with the phenotype of homozygous flies. This objective will allow us to identify the dominance effect of an adaptive mutation.

3 Thesis structure

3. Thesis structure

The results of the thesis were structured in three chapters, each corresponding to a published or submitted article.

Chapter 4: "The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila"

In this chapter, we identified oxidative stress resistance as the adaptive phenotype of the transposable element *Bari-Jheh*. We also associated *Bari-Jheh* with up regulation of the expression of genes *Jheh1* and *Jheh2* and down regulation of *Jheh3*. Moreover we found strong evidence that indicates that *Bari-Jheh* adds Antioxidant Response Elements (AREs). This could be the molecular mechanism by which *Jheh1* and *Jheh2* increase their expression. Finally, we found additional transposable elements with AREs, suggesting that other transposable elements could be involved in the regulation of the oxidative stress response in Drosophila.

Chapter 5: "The dominance effect of the adaptive transposable element insertion *Bari-Jheh* depends on the genetic background"

In this chapter, we showed that the dominance effect of *Bari-Jheh* in oxidative stress environments is dominant in one of the backgrounds analyzed. However, in two additional backgrounds females showed a dominant behavior while males showed an overdominant behavior. We analyzed the expression levels of the genes near the insertion. We found that oxidative stress resistance is associated with the upregulation of the expression of at least one of the two genes: *Jheh1* or *Jheh2*. Moreover we showed that those heterozygous males with overdominant behavior produce an intermediate level of transcript *Jheh2* compared with homozygous males for the presence and the absence of *Bari-Jheh*.

Chapter 6: "Stress affects the epigenetic marks added by the adaptive transposable element *Bari-Jheh* in Drosophila"

In this chapter, we showed how *Bari-Jheh* adds different histone marks in the intergenic region of *Jheh* genes. Moreover we showed how this marks change when we stressed the flies with oxidative stress agents. Finally, we observed that *Bari-Jheh* is associated with

increased tolerance to *Pseudomonas entomophila* ingestion. This is the first study that analyzes the chromatin state of one specific copy of a transposon in gut under two different conditions.

The main findings of all three articles are then discussed (Chapter 7) and a Conclusions section (Chapter 8) are also included. References quoted in Chapters 1 and 7 are included at the end of the document, while Chapters 4, 5 and 6 each include their own References section.

4 The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila

The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila

Resumen

Elucidar los efectos sobre la eficacia biológica de las variantes genéticas naturales es uno de los grandes retos actuales de la biología evolutiva. Entender la interacción entre el genotipo, el fenotipo, y el ambiente es necesario para hacer predicciones precisas de procesos biológicos importantes, tales como la resistencia al estrés, el rendimiento en las plantas y animales de importancia económica, y las enfermedades en seres humanos. El elemento móvil *Bari-Jheh*, que está insertado en la región intergénica de los genes que codifican para las epoxi hidrolasas de la Hormona Juvenil (*Jheh*), se identificó como putativamente adaptativo en base a su frecuencia poblacional y a las huellas de la selección a nivel de ADN. Sin embargo, el efecto adaptativo de esta mutación no se había encontrado. En este trabajo, integramos información sobre los sitios de unión de factores de transcripción, datos disponibles de ChIP-Seq, análisis de expresión génica y ensayos fenotípicos para identificar el efecto adaptativo y el mecanismo molecular por el que actúa *Bari-Jheh*. Mostramos que *Bari-Jheh* añade elementos adicionales de respuesta al estrés oxidativo aguas arriba de los genes *Jheh1* y *Jheh2*. *Bari-Jheh* está asociado con un incremento de la expresión de *Jheh1* y *Jheh2* y con resistencia a estrés oxidativo inducido por dos compuestos diferentes que son ambos relevantes para poblaciones naturales de *Drosophila melanogaster*. Además, mostramos que aparte de *Bari-Jheh*, otros elementos móviles podrían estar jugando un papel en la respuesta de *D. melanogaster* al estrés oxidativo. En general, nuestros resultados contribuyen a la comprensión de la resistencia al estrés oxidativo en poblaciones naturales y ponen de relieve el papel de los elementos móviles en la adaptación ambiental. La reproducibilidad de los efectos fenotípicos en diferentes fondos genéticos también sugiere que las interacciones epistáticas no parecen dominar la arquitectura genética de la resistencia al estrés oxidativo.

MOLECULAR ECOLOGY

Molecular Ecology (2014) doi: 10.1111/mec.12711

The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila

LAIN GUIO, MAITE G. BARRÓN and JOSEFA GONZÁLEZ

Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Maritim de la Barcelona, 37-49, 08003 Barcelona, Spain

Abstract

Elucidating the fitness effects of natural genetic variants is one of the current major challenges in evolutionary biology. Understanding the interplay between genotype, phenotype and environment is necessary to make accurate predictions of important biological outcomes such as stress resistance or yield in economically important plants and animals, and disease in humans. Based on population frequency patterns and footprints of selection at the DNA level, the transposable element *Bari-Jheh*, inserted in the intergenic region of *Juvenile Hormone Epoxy Hydrolase* (*Jheh*) genes, was previously identified as putatively adaptive. However, the adaptive effect of this mutation remained elusive. In this work, we integrate information on transcription factor binding sites, available ChIP-Seq data, gene expression analyses and phenotypic assays to identify the functional and the mechanistic underpinnings of *Bari-Jheh*. We show that *Bari-Jheh* adds extra antioxidant response elements upstream of *Jheh1* and *Jheh2* genes. Accordingly, we find that *Bari-Jheh* is associated with upregulation of *Jheh1* and *Jheh2* and with resistance to oxidative stress induced by two different compounds relevant for natural *D. melanogaster* populations. We further show that TEs other than *Bari-Jheh* might be playing a role in the *D. melanogaster* response to oxidative stress. Overall our results contribute to the understanding of resistance to oxidative stress in natural populations and highlight the role of transposable elements in environmental adaptation. The replicability of fitness effects on different genetic backgrounds also suggests that epistatic interactions do not seem to dominate the genetic architecture of oxidative stress resistance.

Keywords: antioxidant response elements, cap'n'collar, environmental adaptation, insecticide resistance, Juvenile Hormone Epoxy Hydrolase, regulatory networks

Received 30 December 2013; revision received 27 February 2014; accepted 28 February 2014

Introduction

Oxidative stress affects all aerobic organisms. It is caused by the excessive production of reactive oxygen species (ROS) that arise from internal metabolism or from exogenous sources such as radiation, UV light or electrophilic xenobiotics. ROS cause damage to multiple cellular components, such as lipids, proteins, and DNA, which can be causative or aggravative factors in numerous diseases including respiratory and neurodegenerative diseases and cancer (Sykiotis & Bohmann 2010; Taguchi *et al*. 2011; Kansanen *et al*. 2013).

Organisms from fruit flies to humans share a common oxidative stress response signalling pathway (Van Straalen & Roelofs 2012). Response to oxidative stress is mediated primarily at the transcriptional level by the coordinated upregulation of antioxidant and detoxifying genes through the *cis* antioxidant response elements (AREs; Fig. 1). AREs across organisms are bound by the same family of transcription factors, the *Cap'n'collar* (*Cnc*) family, and share a highly conserved sequence. Indeed, AREs' 9bp core sequence has been repeatedly shown to be necessary and sufficient to induce the

Correspondence: Josefa González, Fax: 34 93 2211011; E-mail: josefa.gonzalez@ibe.upf-csic.es

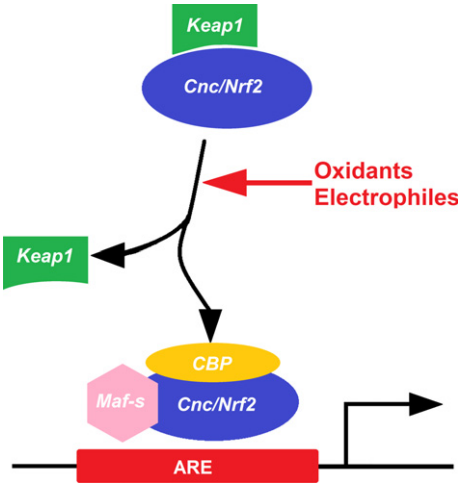


Fig. 1 ARE-mediated response to oxidative stress pathway is conserved from flies to humans. In unstressed conditions, *Nrf2* (*NF-E2-Related Factor 2*) in mammals, and *CncC* (*Cap'n'collar C*) in Drosophila are repressed by *Keap1* (*Kelch-like EΩCH-Associated Protein 1*), which also functions as a sensor of oxidants and electrophilic compounds (Nioi *et al.* 2003; Sykiotis & Bohmann 2008). Under oxidative stress conditions, the inhibition of *Nrf2*/*CncC* by *keap1* is abolished allowing these transcription factors to bind, together with other proteins, to ARE sequences upregulating downstream genes.

upregulation of downstream genes in Drosophila and in mammals (Jaiswal 2004; Janssens *et al.* 2007; Sykiotis & Bohmann 2008; Chatterjee & Bohmann 2012). Although we have a good understanding of the cellular response to oxidative stress, the mechanisms underlying oxidative stress resistance remain elusive (Zhao *et al.* 2011).

In mammals, the role of epoxy hydrolases in oxidative stress response has been clearly demonstrated (Oesch 1973; Oesch *et al.* 2000). In Drosophila, evidence is also starting to accumulate. *Juvenile Hormone Epoxy Hydrolase* (*Jheh*) genes, more specifically *Jheh1* and *Jheh2*, have been repeatedly identified as candidate genes in xenobiotic and oxidative stress response studies (Girardot *et al.* 2004; Sun *et al.* 2006; Willoughby *et al.* 2007; Misra *et al.* 2011). Misra *et al.* (2013) showed that constitutive activation of the *CncC*/*keap1* pathway (Fig. 1) leads to upregulation of *Jheh1*, further suggesting that this gene plays a role in oxidative stress response. Additionally, *Jheh2* is able to metabolize cis-stilbene oxide, a xenobiotic compound that induces oxidative stress in rats (Oguro *et al.* 1997; Taniai *et al.* 2003). Therefore, other than being involved in the catabolism of juvenile hormone (Casas *et al.* 1991; Khlebodarova *et al.* 1996; Flatt *et al.* 2005), *Jheh1* and *Jheh2* genes could also be involved in oxidative stress response in *Drosophila melanogaster*.

Interestingly, a putatively adaptive transposable element (TE) insertion, named *Bari-Jheh*, is inserted 351 bp upstream of *Jheh2* (González *et al.* 2008, 2009). *Bari-Jheh* is a 1.7-kb full-length *Tc1*-like DNA transposon that belongs to the *Bari1* family and is transcriptionally silent (Caizzi *et al.* 1993; Deloger *et al.* 2009). In unstressed conditions, this TE affects the expression level of its two nearby *Jheh* genes and it is associated with reduced viability and extended developmental time (González *et al.* 2009). These phenotypic effects likely represent the cost of selection of this mutation, and the adaptive effect of *Bari-Jheh* remains unknown. Because TEs are known to be responsive and susceptible to environmental changes and to confer this ability to nearby genes (Schmidt & Anderson 2006; Casacuberta & González 2013), it is possible that *Bari-Jheh* enhances the response to oxidative stress by affecting *Jheh* genes.

In this work, we further analysed *Bari-Jheh* insertion, and we found that it adds extra AREs upstream of *Jheh1* and *Jheh2*. Accordingly, under oxidative stress conditions *Bari-Jheh* is associated with increased expression of these two genes and with increased resistance to oxidative stress. Our results provide one of the few examples in which a natural TE mutation has been linked to its ecologically relevant phenotypic effect (Daborn *et al.* 2002; Aminetzach *et al.* 2005; Schmidt *et al.* 2010; Magwire *et al.* 2011).

Material and methods

Prediction of Nrf2 binding sites

Prediction of mammalian *Nrf2* binding sites was used as a proxy to predict *D. melanogaster CncC* binding sites. Both mammalian *Nrf2* and fruit fly *CncC* transcription factors recognized the same binding sequence, the antioxidant response element (ARE). As such, prediction of an *Nrf2* binding sites in *D. melanogaster* indicates the presence of an ARE in that particular region and thus a *CncC* binding site.

To predict *Nrf2* binding sites inside *Bari-Jheh* and in the *Jheh* upstream regions, we downloaded these regions from Flybase (http:/flybase.org, version r5). We used two independent software programs that use different databases and algorithms to identify transcription factor (TF) binding sites. ConSite (Sandelin *et al.* 2004) uses high-quality TF models from the JASPAR database to predict TF binding site. ConSite was run with default parameters. TFBIND (Tsunoda & Takagi 1999) uses positional weight matrices from the TRANSFAC R 3.4 database to accurately predict TF binding sites. Only TF binding sites identified by the two software programs were considered.

Searching for evidence of CncC transcription factor binding inside Bari-Jheh

We downloaded from modENCODE (http://data.modencode.org) the ChIP-seq *fastq* files for *cnc* binding sites, available for adult females and 16–24 h embryos. We first transformed *fastq* sequence to *fasta* using *seqtk* program (https://github.com/lh3/seqtk/) and then created a database for blast with these *fasta* reads (Altschul *et al.* 1990, version 2.2.28). Using Flybase reference genome version r5.53 and Perl *ad hoc* scripts, we generated six 100-bp query sequences for blast: a sequence including the region around the *Nrf2* binding site in the upstream region of (i) *Jheh1* and (ii) *Jheh2*, a sequence including (iii) the left junction of *Bari-Jheh*, and (iv) the right junction of *Bari-Jheh*. Finally, we also identified two diagnostic SNPs that allowed us to differentiate *Bari-Jheh* from other *Bari1* elements. SNPs are located nearby (v) the left *Nrf2* binding site inside the TE, and (vi) the right *Nrf2* binding site inside the TE. Regions (i) and (ii) were used as positive controls and regions (iii) to (vi) were used to find *Bari-Jheh*-specific reads (Table S1, Supporting information).

We used the default parameters of *blastn* using *megablast* task to map the *cnc* reads to the six query sequences. Reads mapping to these six regions were filtered using the following conservative criteria: e-value = 4×10^{-15}, 100% identity, and no mismatches with the query sequence (Table S1, Supporting information). A sample of the selected reads was manually inspected.

Fly stocks

Outbred populations. We used North Carolina (USA) strains from the Drosophila Genetic Reference Panel project (Mackay *et al.* 2012). We created a fly population homozygous for the presence of *Bari-Jheh* insertion by mixing flies from five inbred lines containing this particular TE insertion: RAL-21, RAL-405, RAL-911, RAL-502 and RAL 138. Ten virgin females and ten males of each strain were placed in a fly cage to create an outbred population with the TE insertion. After the first generation, we randomly mated brothers and sisters during at least 10 generations before performing experiments. We also selected five stocks homozygous for the absence of the *Bari-Jheh* insertion (RAL-40, RAL-461, RAL-822, RAL-439 and RAL-908) and repeated the procedure to create an outbred population without the TE insertion. The census size of each population was $n \approx 800$ per generation.

Introgressed strains. We used the introgressed flies described in González *et al.* (2009). The parental strains *Wi3* and *Wi1* were collected in California (USA). The isogenicity of these two stocks was further tested by cloning and sequencing the intergenic region where *Bari-Jheh* is inserted. Primers spanning the whole intergenic region were designed, and PCR products were cloned using the TOPO-TA cloning kit from Invitrogen and Sanger-sequenced (GenBank Accession nos KJ439574 and KJ439575).

qRT-PCR Expression analysis

We quantified the expression of *Jheh1*, *Jheh2* and *Jheh3* in unstressed and stressed conditions induced by paraquat. Five-day-old flies were separated by sex and transferred to normal food or to food containing 10 mM paraquat for 8 h before freezing them with liquid N_2. We did three biological replicas for each sex and condition. We purified total RNA using Trizol reagent and synthesized cDNA using 1 μg of RNA after treatment with DNase. Then, we used the cDNA for quantitative PCR analysis using *Act5C* as a housekeeping gene. Results were analysed using ddCT method. Primers used and qRT-PCR raw data are given in File S1 (Supporting information).

Oxidative stress resistance experiments

We used two different oxidative stress agents: paraquat (Sigma-Aldrich) and malathion (Sigma-Aldrich). Paraquat is one of the most widely used herbicides in agricultural settings including orchards, a natural habitat for *D. melanogaster* (http://www.epa.gov). Malathion is an organophoshate insecticide typically used for control of adult mosquitos and fruit flies in agricultural settings (http://www.epa.gov).

To induce oxidative stress with paraquat, we used two different protocols. (i) Flies were starved for 2 h before starting the experiment. For unstressed conditions, we put the flies in vials with a filter paper soaked with 150 μL of a 5% sucrose solution. To induce stress, we added different concentrations of paraquat, 20 mM and 5 mM, to the 5% sucrose solution. (ii) For unstressed conditions, flies were put in regular food, and to induce stress, paraquat was added as a solution to the food at a final concentration of 10 mM. Four- to five-day-old flies were used in all experiments. In each experiment, 20 replicas of 20 flies each, per sex and per strain, for unstressed and stressed conditions were performed (400 flies × 8 = 3200 flies in total). To estimate the concentration at which paraquat is used in the field, we used the information provided by one of the paraquat manufacturers (Syngenta): final concentration of 8 mM.

Malathion resistance assays were performed using 2% agar and 5% sucrose medium. For stressed conditions,

we added malathion dissolved in isopropanol up to a final concentration of 10 μM. To estimate the concentration at which malathion is used in the field, we used the information provided by one of the malathion distributors (Basf): final concentration of 3–8 mM. Twenty replicas of 20 individuals each per sex, per strain and per condition were performed (400 flies × 8 = 3200 flies in total).

Log-rank tests were performed to test for differences in survival between flies with and without *Bari-Jheh* taking into account the global survival of the 400 flies (20 replicas of 20 flies each).

Searching for putative AREs in annotated TEs

We developed a Perl pipeline to detect putative AREs in the 5.409 annotated TEs in the *D. melanogaster* reference genome. TE fasta sequences and annotations were downloaded from Flybase website (http://flybase.org, version r5.53). First, we search for the conserved core of ARE sequences in all TEs in both orientations (TGAYNNNGC). Then, we chose as candidates the TEs located <1 kb upstream of a gene (regardless of TE and ARE orientation). Finally, we discarded the TEs with a estimated population frequency lower than 5% estimated in 141 strains of a North American population (Mackay *et al.* 2012) using the *T-lex* program (Fiston-Lavier *et al.* 2011, version 2: A-S Fiston-Lavier, MG Barrón, DA Petrov and J González, personal communication). We ended up with a data set of 73 TEs (Table S2, Supporting information).

To check whether any of the 83 genes located downstream of the 73 TEs has a GO functional annotation related to oxidative stress response, we used the *Batch downloaded* tool from Flybase to retrieve their GO annotations. TEs with GO biological process categories – 'response to oxidative stress' and 'response to hypoxia' – were selected.

To check for *CncC* binding sites located between the TEs and the downstream genes, we downloaded from Flybase the track named 'TFBS – other' from *Noncoding Features* and selected the information about *cnc* binding sites. With *ad hoc* Perl scripts, we identified the TEs that have *cnc* binding sites located between the TE and the nearest downstream gene. We also checked using *ad hoc* Perl scripts whether any of these genes have previously been described as a candidate gene for oxidative stress response in the three available GWAS studies (Girardot *et al.* 2004; Zhao *et al.* 2011; Weber *et al.* 2012). All the scripts used to identify AREs in annotated TEs are available in File S2 (Supporting information).

Results

Bari-Jheh introduces extra AREs upstream of Jheh1 and Jheh2 genes

To shed light into the plausible adaptive effect of *Bari-Jheh* insertion, we looked for regulatory elements in the TE sequence and in the *Jheh* intergenic regions where this TE is inserted. Using a combination of prediction softwares (see Material and methods), we found that *Bari-Jheh* contains two putative *Nrf2* binding sites (Fig. 2A). *Nrf2* is the mammalian homolog of the fruit fly *CncC* transcription factor that specifically regulates the transcriptional activation of downstream genes involved in the antioxidant defence response by binding to AREs (Fig. 1; Sykiotis & Bohmann 2008). Interestingly, we also identified putative *Nrf2* binding sites in the upstream regions of *Jheh* genes, further suggesting that these genes are involved in the antioxidant defence response as has been previously reported (Fig. 2A; Sun *et al.* 2006; Misra *et al.* 2011).

To further validate these predictions, we looked for experimental evidence for the presence of *CncC* binding sites in the *Jheh* intergenic regions and inside *Bari-Jheh* using modENCODE ChipSeq data (Nègre *et al.* 2011). There is evidence for the binding of *CncC* to the AREs located upstream of *Jheh1* and *Jheh2* genes, but not to the ARE located upstream of *Jheh3* (Fig. 2B). Furthermore,

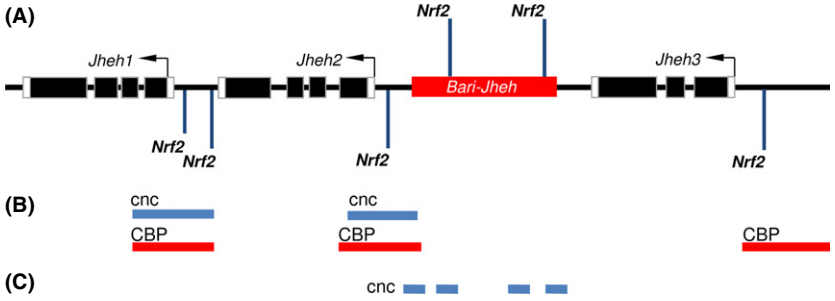


Fig. 2 *Nrf2/CncC* binding sites in *Bari-Jheh* and in the intergenic regions of *Jheh* genes. (A) Black boxes represent exons, white boxes represent 5′ and 3′ UTRs and a red box represents *Bari-Jheh* insertion. ConSite and TFbind *Nrf2* binding sites predictions are depicted. (B) Localization of *cnc* and *CBP* binding sites experimentally determined by modENCODE. (C) Localization of ChIP-seq modENCODE *cnc* reads mapping to the *Bari-Jheh*.

there is also evidence for the binding of *CREB binding protein* (*CBP*) to the upstream sequences of *Jheh1* and *Jheh2*. *CBP* is known to increase the promoter-specific DNA binding of *Nrf2* (Sun *et al.* 2009), further suggesting that *Jheh1* and *Jheh2* are regulated by AREs (Fig. 2B).

modENCODE only takes into account unique reads, and therefore, it only annotates regulatory elements in the nonrepetitive portion of the genome (Nègre *et al.* 2011). Thus, to validate the presence of AREs inside *Bari-Jheh*, we used the modENCODE *cnc* raw reads (see Material and methods). Because reads are short (44 bp), to make sure that the reads identified belong to *Bari-Jheh* insertion and not to any other TE with sequence homology to *Bari-Jheh*, we looked for (i) reads mapping to the *Bari-Jheh* left and right junctions and (ii) reads containing a diagnostic SNP that is only present in *Bari-Jheh* and not in any other *Bari1* elements. We chose the diagnostic SNPs located closest to each one of the two *Nrf2* binding sites inside *Bari-Jheh* (Fig. 2C, Table S1, Supporting information). We found several reads mapping to each one of these four regions, suggesting that *CncC* binds to *Bari-Jheh* (Fig. 2C).

Taken together, our analysis showed that *Jheh1* and *Jheh2*, but not *Jheh3* contain functional AREs in their promoter regions, suggesting that *Jheh1* and *Jheh2* are involved in the oxidative stress response. Our results also show that *Bari-Jheh* is adding extra ARE sequences upstream of *Jheh1* and *Jheh2* genes.

Bari-Jheh is associated with upregulation of Jheh1 and Jheh2 under oxidative stress conditions

The presence of additional ARE sequences in *Bari-Jheh* prompted us to investigate whether *Bari-Jheh* is associated with increased expression of *Jheh1* and *Jheh2*, because there is a positive correlation between the number of AREs in the promoter region and the increase in the level of expression of the downstream genes (Wang *et al.* 2006). We performed qRT-PCR experiments for these two genes in introgressed flies that have been previously shown to differ mostly by the presence/absence of *Bari-Jheh* insertion (González *et al.* 2009). We further confirmed by sequencing that no polymorphisms other than the TE insertion were present in the *Jheh* intergenic region in these introgressed strains (File S3, Supporting information, see Material and methods). We found that in unstressed conditions, there are no differences in the level of expression of *Jheh1* between flies with and without the insertion (Fig 3A). However, under oxidative stress conditions induced by paraquat, both males and females with the insertion show a higher level of expression of *Jheh1* compared with flies without the insertion (*t*-test, males *P*-value = 0.0014 and females *P*-value = 0.046) (Fig. 3A). For *Jheh2*, we found that in unstressed conditions, the level of expression was lower in males with the insertion compared with males without the insertion as has been previously reported (*t*-test *P*-value = 0.014) (Fig. 3B; González *et al.* 2009). On the other hand, under oxidative stress conditions, the presence of *Bari-Jheh* is associated with increased *Jheh2* expression in males (*t*-test *P*-value = 0.043; Fig. 3B). In females, *Jheh2* showed no differences in the level of expression between flies with and without the insertion, neither in control nor in the presence of oxidative stress (Fig. 3B). Although we cannot discard the hypothesis that the increased expression in response to oxidative stress of *Jheh2* is sex-specific, differences between males and females could also be explained for example, by differences in the time point in which *Jheh2* expression increases in response to stress.

Finally, we also analysed the expression of *Jheh3*. We do not expect *Jheh3* to be upregulated in oxidative stress conditions because we did not find experimental evidence for the presence of AREs in the upstream region of this gene (Fig. 2). As we have previously observed for *Jheh2*, we found that *Jheh3* was downregulated in flies with the insertion compared with flies without the insertion in control conditions (*t*-test, males *P*-value = 0.02 and females *P*-value <0.001; Fig. 3C). Under oxidative stress conditions, the level of expression of *Jheh3* is also lower in flies with *Bari-Jheh* compared with flies without *Bari-Jheh* (*P*-value = 0.035 and <0.001 for males and females respectively) (Fig. 3C). These results corroborate

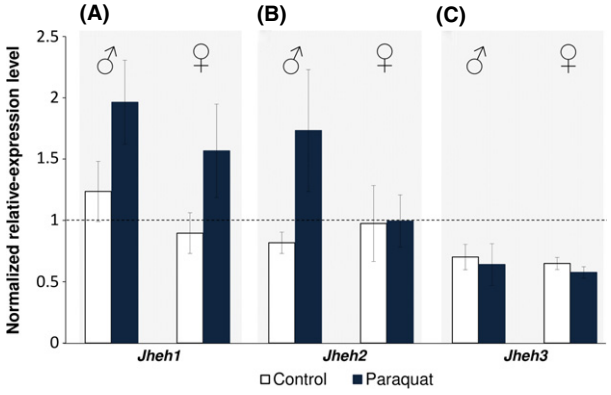


Fig. 3 *Bari-Jheh* is associated with changes in expression levels of *Jheh1, Jheh2 and Jheh3*. Normalized expression ratios (expression level of flies carrying *Bari-Jheh*/expression level of flies lacking *Bari-Jheh)* for *Jheh1* (A), *Jheh2* (B) and *Jheh3* (C) in males and females, under unstressed (white) and stressed with 10 mM Paraquat (blue) experimental conditions. The discontinuous line is the ratio expected if there are no differences in the level of expression of flies carrying and lacking *Bari-Jheh*. Error bars represent the SEM based on the three biological replicas performed.

that the presence of AREs is necessary to induce upregulation of downstream genes.

Overall, we found that *Bari-Jheh* is associated with upregulation of its two downstream genes under oxidative stress conditions, as expected if *Bari-Jheh* is adding functional AREs. The lack of upregulation of *Jheh3* under oxidative stress conditions further confirms that AREs are necessary to induce expression of downstream genes.

Bari-Jheh confers resistance to paraquat, a commonly used oxidative stress-inducing agent

Because *Bari-Jheh* is associated with increased expression of *Jheh1* and *Jheh2* genes under oxidative stress conditions, we hypothesized that flies with *Bari-Jheh* insertion could be more resistant to oxidative stress compared with flies without this insertion. To test this hypothesis, we used paraquat to induce oxidative stress and compared the mortality of flies with different genetic backgrounds. We used two outbred populations created in our laboratory: one homozygous for the presence and the other homozygous for the absence of *Bari-Jheh* (see Material and methods). We also used two introgressed strains that differ mainly in the presence/absence of *Bari-Jheh* (González *et al*. 2009; see Material and methods).

We performed three independent oxidative stress resistance experiments using different concentrations of paraquat and different administration protocols. We first analysed the survival of outbred populations exposed to 20 mM paraquat in a 5% sucrose solution. We found that while there are no differences between outbred flies with and without the insertion in control conditions, both males and females with the insertion were more resistant to paraquat than flies without the insertion: males and females log-rank test P-value $\ll 0.001$, males odds ratio: 6.30 (confidence interval: 4.44–8.95) and females 6.69 (4.62–9.40; Fig. 4A). The selective pressure imposed by this concentration of paraquat is high because after only 28 h 80–98% of the sensitive flies died. We thus performed another resistance stress experiment using flies with a different genetic background, introgressed flies, exposed to a lower concentration of paraquat: 5 mM. Note that this concentration is more similar to the one used in nature (see Material and methods). Consistent with the previous experiment, flies with the insertion were more resistant to paraquat than flies without the insertion: males and females log-rank test P-value $\ll 0.001$, males odds ratio: 2.16 (1.62–2.88) and females 11.30 (7.58–16.83; Fig. S1, Supporting information). However, we observed that mortality of control flies increased after 25–30 h, suggesting that other than oxidative stress, mortality could also be explained by the stressful effects of a diet high in sucrose as the one we used in this experiment (Rzezniczak *et al*. 2011). We thus repeated the experiment by feeding the introgressed strains with 10 mM paraquat added to regular fly food. Flies with the insertion showed increased resistance to paraquat compared with flies without the insertion: males and females log-rank test P-value $\ll 0.001$, males odds ratio: 3.21 (2.37–4.34) and females 5.78 (4.12–8.11; Fig. 4B). Control

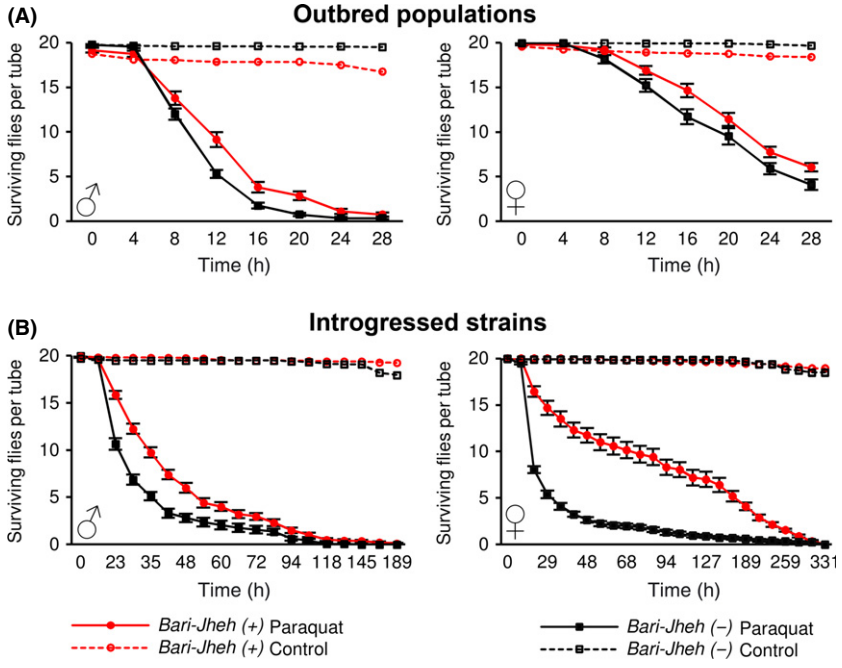


Fig. 4 *Bari-Jheh* is associated with increased resistance to paraquat. Survival curves in unstressed (discontinuous lines) and stressed (continuous lines) conditions for flies with *Bari-Jheh* (red lines) and without *Bari-Jheh* (black lines). Each data point represents the average of surviving flies for 20 replicas of 20 individuals each with error bars indicating the SEM of (A) outbred populations exposed to 20 mM paraquat, without food and (B) introgressed strains exposed to 10 mM paraquat with food.

flies showed no mortality, indicating that mortality of treated flies is due to the presence of paraquat in the food and not to any other factor.

Taken together, our results show that flies with *Bari-Jheh* are more resistant to oxidative stress induced by paraquat than flies without this insertion. Consistent results were obtained for flies with different genetic backgrounds, outbred populations and introgressed flies (Fig. 4), suggesting that the insertion is the causal mutation of this adaptive phenotype.

Bari-Jheh confers resistance to the insecticide malathion

To further confirm that *Bari-Jheh* confers resistance to oxidative stress, we also compared the survival of flies with and without the insertion under oxidative stress conditions using a different oxidative stress-inducing agent: the insecticide malathion. As in the previous experiments, we used two outbred populations and two introgressed strains (see Material and methods). We found that outbred flies with the insertion were more resistant to malathion compared with outbred flies without the insertion [Fig. 5A; males and females log-rank test P-value $\ll$0.001, males odds ratio: 2.15 (1.61–2.87) and females 2.36 (1.77–3.15)]. Results consistent with a role of *Bari-Jheh* in malathion resistance were also found in introgressed flies (Fig. 5B; males and females log-rank test P-value $\ll$0.001, males odds ratio: 4.69 (3.40–6.48) and females 1.56 (1.18–2.07)), further suggesting that this TE insertion confers resistance to oxidative stress.

TEs other than Bari-Jheh add AREs to the promoter regions of D. melanogaster genes

To investigate whether AREs are present in TEs other than *Bari-Jheh*, we search for the presence of core consensus ARE sequences in all the TEs annotated in the *D. melanogaster* genome (see Material and methods). Among the TEs that add AREs in the 1kb region upstream of coding genes, 73 are present at a population frequency higher than 5% and might therefore play a role in oxidative stress response in *D. melanogaster* populations (Table S2, Supporting infomation).

We looked for evidence suggesting a role of the genes nearby these 73 TEs in oxidative stress response by checking (i) whether there is experimental evidence for *CncC* binding sites in their upstream regions, (ii) whether any of these genes have been previously identified in genome-wide association studies (GWAS) for oxidative stress resistance, and (iii) whether their functional annotation is related to oxidative stress response. A total of 10 genes nearby 10 different TEs show additional evidence for their role in stress response (Table 1). Besides the *Cnc*C binding sites introduced by the identified TEs, *CG3262* and *Tsp39D* genes have other *CncC* binding sites in their upstream regions. Seven other genes have been previously identified as candidates in response to oxidative stress analyses, being *CG3987* identified by two different GWAS (Table 1). Two of these seven genes have also been functionally classified as involved in response to hypoxia (*Hsp70Bbb*) and response to DNA damage

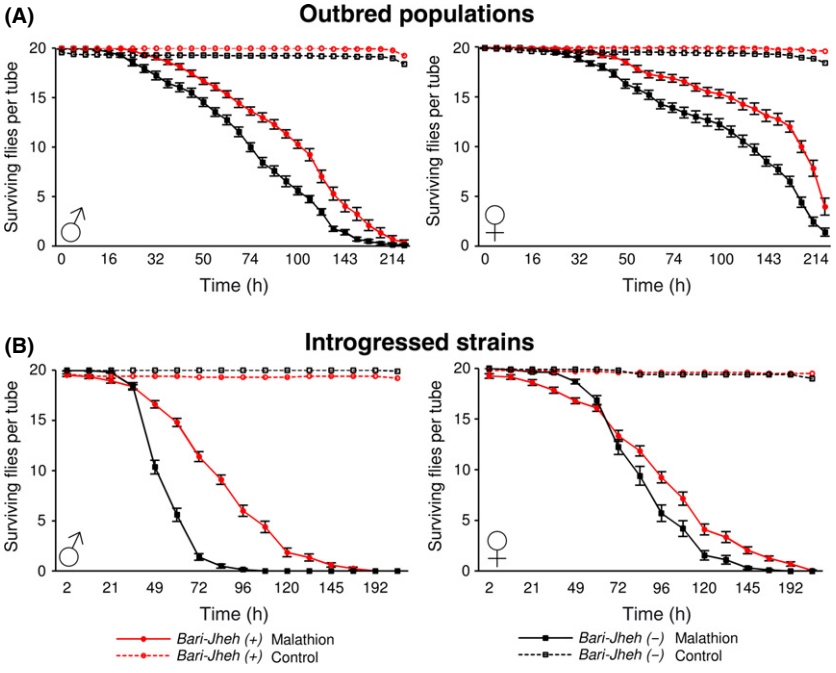


Fig. 5 *Bari-Jheh* is associated with increased resistance to malathion. Survival curves in unstressed conditions (discontinuous lines) and stressed with 10 μM malathion (continuous lines) for flies with *Bari-Jheh* (red lines) and without *Bari-Jheh* (black lines) for outbred populations (A) and introgressed flies (B). Each data point represents the average of surviving flies for 20 replicas of 20 individuals each with error bars indicating the SEM.

Table 1 TEs containing AREs, present at >5% frequency in DGRP populations, and located in the 1 kb upstream region of genes with additional evidence for a role in oxidative stress (O.S.) response

TE flybase identifier	Chr arm	TE family/ order	AREs inside TE	Downstream gene	Distance to gene (bp)	Frequency in DGRP strains (%)	*CncC* binding sites	Candidate O.S. response gene in GWAS analyses	GO Biological process
FBti0060356	2L	diver2/LTR	5	*CG3262*	930	100	1	No	No information
FBti0062156	2L	INE-1/TIR	1	*Tsp39D*	894	100	1	No	No information
FBti0064279	2R	INE-1/TIR	1	*CG14591*	343	100	0	Girardot *et al.* (2004)	No information
FBti0019362	3R	S-element/ TIR	1	*Hsp70Bbb*	709	100	0	Girardot *et al.* (2004)	Response to hypoxia
FBti0020015	3L	412/LTR	3	*LysS*	510	99.34	0	Girardot *et al.* (2004)	Antimicrobial humoral response
FBti0060127	3L	INE-1/TIR	1	*CG14569*	584	100	0	Weber *et al.* (2012)	No information
FBti0062048	3L	Quasimodo/ LTR	9	*Spn77Bc*	290	100	0	Girardot *et al.* (2004)	Multicellular organism reproduction
FBti0062170	3R	Burdock/LTR	1	*CG3987*	400	100	0	Girardot *et al.* (2004), Zhao *et al.* (2011)	Mesoderm development
FBti0063392	2L	INE-1/TIR	1	*CG9272*	269	100	0	Girardot *et al.* (2004)	Response to DNA damage stimulus
FBti0019335	3R	G2/non-LTR	2	*puc*	487	6.38	0	No	Response to oxidative stress

(*CG9272*), further suggesting that they might be involved in oxidative stress response. Additionally, *LysS* gene is involved in antimicrobial humoral response, which is related to oxidative stress response (Zhao *et al.* 2011). Finally, *puc* gene has been functionally classified as involved in response to oxidative stress (Table 1). These results suggest that TEs other than *Bari-Jheh* might also be enhancing the response to oxidative stress by upregulating the expression of nearby genes.

These 10 TEs belong to different families representing the three main orders of TEs in *D. melanogaster:* LTR, non-LTR and TIR. Nine of them are fixed in the DGRP population and only *FBti0019335* is polymorphic (Table 1). Note that *FBti0019335* is not affecting the reported *Bari-Jheh* phenotypes because it is not present in any of the strains analysed in this study.

Discussion

Our results show that *Bari-Jheh* mediates resistance to oxidative stress in *D. melanogaster* providing one of the few examples in which a natural TE insertion has been connected to its adaptive phenotypic effect (Daborn *et al.* 2002; Aminetzach *et al.* 2005; Schmidt *et al.* 2010; Magwire *et al.* 2011). Previous evidence for the adaptive role of this mutation came from molecular population analyses that showed that *Bari-Jheh* is surrounded by a selective sweep that decays on both sides of the TE and that it does not extend beyond *Jheh2* and *Jheh3* genes. *Bari-Jheh* was the only mutation linked to the sweep and, as such, the likely causal adaptive mutation (González *et al.* 2009). However, previous analyses could only identify the cost of selection of this mutation, whose adaptive effect remained unknown. In this work, by combining information on transcription factor binding sites prediction, available ChIP-Seq data, gene expression analyses and phenotypic assays, we have constructed plausible hypotheses about the adaptive effect of *Bari-Jheh* that we have confirmed experimentally.

We showed that *Bari-Jheh* adds extra AREs to the upstream region of *Jheh1* and *Jheh2* (Fig. 2) and that it is associated with increased expression of these two genes in oxidative stress conditions (Fig. 3A and B). The lack of upregulation of *Jheh3* located upstream of *Bari-Jheh* confirms that the presence of AREs is necessary to induce upregulation of downstream genes in response to oxidative stress (Fig. 3C). Consistent with the increased expression of *Jheh1* and *Jheh2*, we found that outbred flies with *Bari-Jheh* insertion have a higher survival rate than outbred flies without *Bari-Jheh* under oxidative stress conditions induced by two different compounds that are used in *D. melanogaster* natural environments: the herbicide paraquat (Fig. 4A) and the insecticide malathion (Fig. 5A).

We explored the replicability of this genotype–phenotype association by performing the oxidative stress resistance experiments in introgressed flies that differ

mainly by the presence/absence of *Bari-Jheh* (González *et al.* 2009). For a second time, we found a significant association between the presence of *Bari-Jheh* and the resistance to oxidative stress induced by paraquat (Fig. 4B and Fig. S1, Supporting information) and by malathion (Fig. 5B), suggesting that *Bari-Jheh* is the causal mutation. Additional evidence for the role of *Bari-Jheh* in resistance to oxidative stress comes from experiments with two isofemale strains, RAL-391 and RAL-783 that were exposed to oxidative stress induced by H_2O_2 in our laboratory (L. Mateo, A. Ullastres, and J. González, personal communication). RAL-391 flies, homozygous for the presence of *Bari-Jheh*, are more resistant to oxidative stress induced by H_2O_2 than RAL-783 flies, homozygous for the absence of *Bari-Jheh* [male and female log-rank P-value $\ll 0.001$, male odds ratio: 5.14 (3.74–7.07) and females 16.39 (10.30–26.08)]. Note that other than being activated by electrophilic chemical compounds, such as paraquat and malathion, ARE-mediated response to stress pathway is also activated by H_2O_2 (Nguyen *et al.* 2009). Taken together, these results strongly suggest that the natural TE insertion *Bari-Jheh* is the causal mutation conferring resistance to oxidative stress because it is unlikely that the detection of a robust significant effect of *Bari-Jheh* on oxidative stress resistance using three different genetic backgrounds and three different oxidative stress agents would occur spuriously (Gruber *et al.* 2007). The replicability of *Bari-Jheh* effect on different genetic backgrounds also suggests that epistatic interactions do not seem to dominate the genetic architecture of oxidative stress resistance, as has been described for other quantitative traits (Huang *et al.* 2012). Nevertheless, the differences in the magnitude of the effect of *Bari-Jheh* insertion in different genetic backgrounds strongly suggest that loci other than *Bari-Jheh* influence the resistance phenotype (Figs 4 and 5).

Taken together, our results suggest that different *Jheh* genes might be playing different functional roles. The three *Jheh* genes, *Jheh1*, *Jheh2* and *Jheh3*, are thought to be involved in the catabolism of juvenile hormone (Casas *et al.* 1991; Khlebodarova *et al.* 1996). However, Taniai *et al.* (2003) demonstrated that *Jheh2* does not hydrolyse this hormone but is rather involved in the metabolism of xenobiotics such as cis-stilbene oxide, an oxidative stress agent. Our results show that both *Jheh1* and *Jheh2* are oxidative stress response genes, while *Jheh3* is not upregulated in response to stress. Further analyses are needed to elucidate the individual functions of each one of these genes in the different biological processes in which they might be involved (Flatt *et al.* 2005).

Our results provide an example of how environment influences the fitness effects of a natural mutation (Lynch 1999; Martin & Lenormand 2006). *Bari-Jheh* reduces the viability and extends the developmental time of flies in unstressed conditions (González *et al.* 2009), while it increases fly survival under oxidative stress (Figs 4 and 5). The fitness costs in unstressed conditions are outweighed by the fitness advantage in stressed conditions resulting in the increase in frequency of *Bari-Jheh* in natural populations (González *et al.* 2008, 2010). The previously identified fitness costs of this insertion also provide a plausible explanation for the lack of fixation of this adaptive mutation, present at high frequencies in natural populations worldwide (González *et al.* 2008, 2009, 2010). These results highlight the importance of exploring different phenotypes to fully characterize the effects of mutations (Olson-Manning *et al.* 2012).

It is plausible that TEs other than *Bari-Jheh* might also be contributing to the response of *D. melanogaster* to oxidative stress (Table 1). By providing AREs to nearby genes, these TEs might be coordinating the expression of genes dispersed throughout the genome in response to an oxidative stress challenge. There are a few clear examples in which TEs have been shown to rewire regulatory networks in animals and in plants (Naito *et al.* 2009; Macfarlan *et al.* 2012; Schmidt *et al.* 2012). Taking into account the conservation of oxidative stress response pathways in organisms from flies to human and the ubiquity of transposable elements, the role of TEs in oxidative stress resistance may not be restricted to *D. melanogaster*.

Overall, our results contribute to the understanding of resistance to oxidative stress in natural populations and highlight the role of TEs in environmental adaptation. Because genes tend to work in evolutionary conserved pathways, genotype–phenotype insights obtained in *D. melanogaster*, as the ones described in this work, provide valuable information that might be relevant for other organisms as well (Lehner 2013).

Acknowledgements

We thank Elena Casacuberta and members of the González laboratory for critically reading the manuscript. We also thank the editor and reviewers for their constructive comments. L.G. is a FI/DGR fellow (2012FI-B-00676). J.G. is a *Ramón y Cajal* fellow (RYC-2010-07306). This work was supported by grants from the European Commission (Marie Curie CIG PCIG-GA-2011-293860) and from the Spanish Government (Fundamental Research Projects Grant BFU-2011-24397).

References

Altschul SF, Gish W, Miller W, Myers EWLD (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Aminetzach YT, Macpherson JM, Petrov D (2005) Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*, **309**, 764–767.

Caizzi R, Cggese C, Pimpinelli S (1993) *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. *Genetics*, **133**, 335–345.

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

Casas J, Harshman LG, Messeguer A, Kuwano E, Hammock BD (1991) In vitro metabolism of juvenile hormone III and juvenile hormone III bisepoxide by *Drosophila melanogaster* and mammalian cytosolic epoxide hydrolase. *Archives of Biochemistry and Biophysics*, **286**, 153–158.

Chatterjee N, Bohmann D (2012) A versatile ΦC31 based reporter system for measuring AP-1 and Nrf2 signaling in Drosophila and in tissue culture. *PLoS ONE* **7**:e34063.

Daborn PJ, Yen JL, Bogwitz MR *et al.* (2002) A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**:2253–2256.

Deloger M, Cavalli FM, Lerat E, Biémont C, Sagot MF, Vieira C (2009) Identification of expressed transposable element insertion in the sequenced genome of *Drosophila melanogaster*. *Gene*, **439**, 55–62.

Fiston-Lavier A-S, Carrigan M, Petrov DA, González J (2011) T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. *Nucleic Acids Research*, **39**, e36.

Flatt T, Tu MP, Tatar M (2005) Hormonal pleiotropy and the juvenile hormone regulation of Drosophila development and life history. *BioEssays*, **27**, 999–1010.

Girardot F, Monnier V, Tricoire H (2004) Genome wide analysis of common and specific stress responses in adult *Drosophila melanogaster*. *BMC Genomics*, **5**, 74.

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.

González J, Macpherson JM, Petrov DA (2009) A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **26**, 1949–1961.

González J, Karasov TL, Messer PW, Petrov DA (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. *PLoS Genetics*, **6**, e1000905.

Gruber JD, Genissel A, Macdonald SJ, Long AD (2007) How repeatable are associations between polymorphisms in *achaete–scute* and bristle number variation in Drosophila? *Genetics*, **175**, 1987–1997.

Huang W, Richards S, Carbone MA *et al.* (2012) Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proceedings of the National Academy of Sciences, USA*, **109**, 15553–15559.

Jaiswal AK (2004) Regulation of antioxidant response element-dependent induction of detoxifying enzyme synthesis. *Methods in Enzymology*, **378**, 221–238.

Janssens TKS, Mariën J, Cenijn P, Legler J, van Straalen NM, Roelofs D (2007) Recombinational micro-evolution of functionally different metallothionein promoter alleles from *Orchesella cincta*. *BMC Evolutionary Biology*, **7**, 88.

Kansanen E, Kuosmanen SM, Leinonen H, Levonen A-L (2013) The Keap1-Nrf2 pathway: mechanisms of activation and dysregulation in cancer. *Redox Biology*, **1**, 45–49.

Khlebodarova TM, Gruntenko NE, Grenback LG *et al.* (1996) A comparative analysis of juvenile hormone metabolyzing enzymes in two species of Drosophila during development. *Insect Biochemistry and Molecular Biology*, **26**, 829–835.

Lehner B (2013) Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews Genetics*, **14**, 168–178.

Lynch M (1999) Estimating genetic correlations in natural populations. *Genetical Research*, **74**, 255–264.

Macfarlan TS, Gifford WD, Driscoll S *et al.* (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**, 57–63.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* genetic reference panel. *Nature*, **482**, 173–178.

Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM (2011) Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a Duplication. *PLoS Genetics*, **7**, e1002337.

Martin G, Lenormand T (2006) The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution*, **60**, 2413–2427.

Misra JR, Horner MA, Lam G, Thummel CS (2011) Transcriptional regulation of xenobiotic detoxification in Drosophila. *Genes & Development*, **2**, 1796–1806.

Misra JR, Lam G, Thummel CS (2013) Constitutive activation of the Nrf2/Keap1 pathway in insecticide-resistant strains of Drosophila. *Insect Biochemistry and Molecular Biology*, **43**, 1116–1124.

Naito K, Zhang F, Tsukiyama T *et al.* (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, **461**, 1130–1134.

Nègre N, Brown CD, Ma L *et al.* (2011) A cis-regulatory map of the Drosophila genome. *Nature*, **471**, 527–531.

Nguyen T, Nioi P, Picket CB (2009) The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *Journal of Biological Chemistry*, **284**, 13291–13295.

Nioi P, McMahon M, Itoh K, Yamamoto M, Hayes JD (2003) Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochemical Journal*, **374**, 337–348.

Oesch F (1973) Mammalian epoxide hydrases: inducible enzymes catalysing the inactivation of carcinogenic and cytotoxic metabolites derived from aromatic and olefinic compounds. *Xenobiotica*, **3**, 305–340.

Oesch F, Herrero ME, Hengstler JG, Lohmann M, Arand M (2000) Metabolic detoxification: implications for thresholds. *Toxicologic Pathology*, **28**, 382–387.

Oguro T, Kaneko E, Numazawa S, Imaoka S, Funae Y, Yoshida T (1997) Induction of hepatic heme oxygenase and changes in cytochrome P-450s in response to oxidative stress produced by stilbenes and stilbene oxides in rats. *Journal of Pharmacology and Experimental Therapeutics*, **280**, 1455–1462.

Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for the theoretical predictions. *Nature Review Genetics*, **13**, 867–877.

Rzezniczak TZ, Douglas LA, Watterson JH, Merritt TJS (2011) Paraquat administration in Drosophila for use in metabolic studies of oxidative stress. *Analytical Biochemistry*, **419**, 345–347.

Sandelin A, Wasserman WW, Lenhard B (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Research*, **32**, W249–W252.

Schmidt AL, Anderson LM (2006) Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biological Reviews*, **81**, 531–543.

Schmidt JM, Good RT, Appleton B *et al.* (2010) Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genetics*, **6**, e1000998.

Schmidt D, Schwalie PC, Wilson MD *et al.* (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, **148**, 335–348.

Sun W, Margam VM, Sun L *et al.* (2006) Genome-wide analysis of phenobarbital-inducible genes in Drosophila melanogaster. *Insect Molecular Biology*, **15**, 455–464.

Sun Z, Chin YE, Zhang DD (2009) Acetylation of Nrf2 by p300/CBP augments promoter-specific DNA binding of Nrf2 during the antioxidant response. *Molecular and Cellular Biology*, **29**, 2658–2672.

Sykiotis GP, Bohmann D (2008) Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. *Developmental Cell*, **14**, 76–85.

Sykiotis G, Bohmann D (2010) Stress-activated cap'n'collar transcription factors in aging and human disease. *Science Signaling*, **3**, 1–45.

Taguchi K, Motohashi H, Yamamoto M (2011) Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes to Cells*, **16**, 123–140.

Taniai K, Inceoglu AB, Yukuhiro K, Hammock BD (2003) Characterization and cDNA cloning of a clofibrate-inducible microsomal epoxide hydrolase in *Drosophila melanogaster*. *European Journal of Biochemistry*, **270**, 4696–4705.

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

Van Straalen NM, Roelofs D (2012) *An Introduction to Ecological Genomics*, 2nd edn. pp. 195–244. Oxford University Press, Oxford.

Wang XJ, Hayes JD, Wolf CR (2006) Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of Nrf2 by cancer chemotherapeutic agents. *Cancer Research*, **66**, 10983–10994.

Weber AL, Khan GF, Magwire MM, Tabor CL, Mackay TFC, Anholt RRH (2012) Genome-wide association analysis of oxidative stress resistance in Drosophila melanogaster. *PLoS ONE*, **7**, e34745.

Willoughby L, Batterham P, Daborn PJ (2007) Piperonyl butoxide induces the expression of cytochrome P450 and glutathione S-transferase genes in Drosophila melanogaster. *Pest Management Science*, **63**, 803–808.

Zhao HW, Zhou D, Haddad GG (2011) Antimicrobial peptides increase tolerance to oxidant stress in *Drosophila melanogaster*. *Journal of Biological Chemistry*, **286**, 6211–6218.

L.G. designed research, performed research, analysed data and drafted the article. M.G.B. designed research, performed research, contributed new reagents or analytical tools, analysed data and drafted the article. J.G. designed research, analysed data and wrote the article.

Data accessibility

DNA sequence data were submitted to GenBank and is available under the following accession numbers: KJ439574 and KJ439575. An alignment of these two sequences is provided in File S3 (Supporting information). qPCR raw data are available in File S1 (Supporting information). Perl scripts used to identify AREs in *D. melanogaster* TEs are available in File S2 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Introgressed strain with *Bari-Jheh* is associated with increased resistance to 5 mM paraquat.

File S1 qPCR raw Data and primer sequences.

File S2 Perl scripts used to identify AREs in *D. melanogaster* TEs.

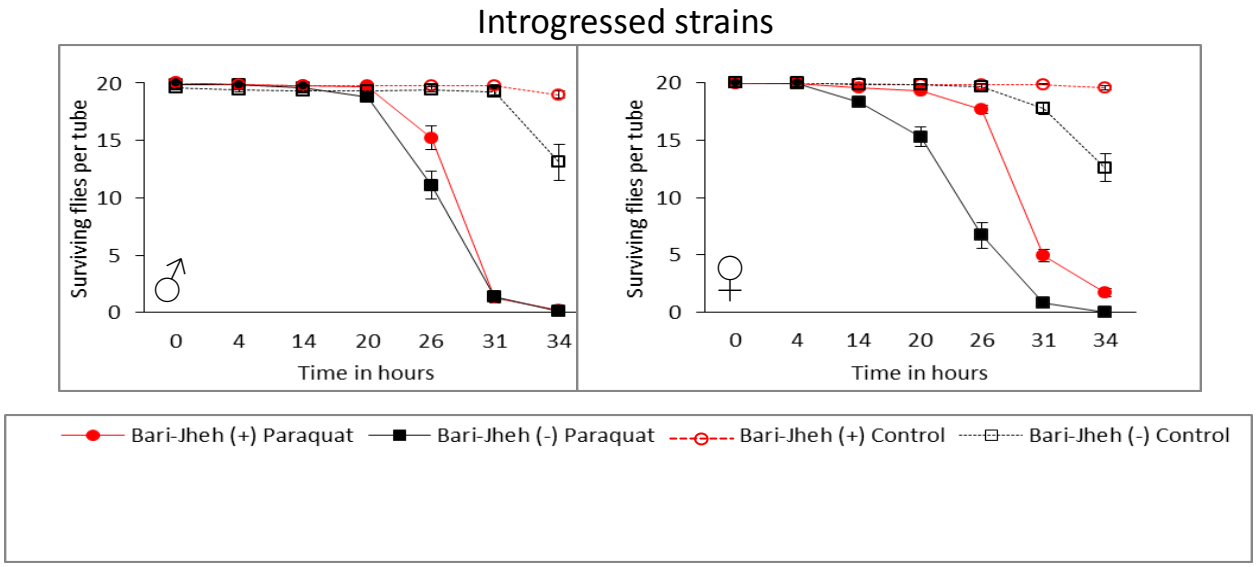
File S3 Alignment of KJ439574 and KJ439575 sequences of introgressed strains.

Table S1 Description of the regions used to look for *cnc* reads and number of *cnc* selected reads per developmental stage and replica (rep).

Table S2 TEs with AREs located in the 1 kb upstream regions of genes.

Supplementary Material

Fig. S1 Introgressed strain with *Bari-Jheh* is associated with increased resistance to 5 mMparaquat



Primers used for expression analysis.

Gene	Left primer	Right primer
Jheh1	CAGTCTCCACCCTGGATAAAGAA	AGTTTACCAGGTTATGGCTGGTC
Jheh2	AGGCCATCCTACCCTTCGACATCA	ATTGGAAACCCACACCCTCCAGTG
Jheh3	CATCACAGTGGCCATTTCAG	CATGCTAACCAAGCACTCAAAC
Act5C	ATGTCACGGACGATTTCACG	GCGCCCTTACTCTTTCACCA

Raw data of expression analysis

Abbreviations:

M	Males
F	Females
B+	Strain with *Bari-Jheh*
B-	Strain without *Bari-Jheh*
Cont.	Control conditions
Pq	Paraquat 10mM conditions
1, 2, 3	number of Biological replicas

Jheh1 males

		DATA INPUT				Total Expression				
	1	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV (variation coefficient)	error copies
Replica 1	M B+ Pq 1	23.57	0.180	23.35	0.066	-0.21184828	0.192319656	**1.158171002**	0.133305828	0.15439094
	M B- Pq 1	24.30	0.085	24.79	0.040	0.492197033	0.093433924	**0.710941603**	0.064763461	0.04604304
	M B+ Cont 1	25.04	0.075	26.97	0.039	1.9302636	0.084477422	**0.262381226**	0.058555287	0.01536381
	M B- Cont 1	24.85	0.042	26.62	0.148	1.772139058	0.153389668	**0.292774324**	0.106321616	0.03112824
		DATA INPUT				Total Expression				
	2	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	M B+ Pq 2	27.78	0.060	28.17	0.056	0.385979781	0.081806416	**0.76525911**	0.056703887	0.04339317
	M B- Pq 2	26.65	0.050	27.44	0.106	0.790570157	0.11691104	**0.578115574**	0.081036558	0.0468485
	M B+ Cont 2	24.20	0.108	25.82	0.022	1.622557058	0.109748842	**0.324759344**	0.0760721	0.02470513
	M B- Cont 2	26.11	0.016	27.90	0.076	1.796122791	0.078047417	**0.287947402**	0.054098347	0.01557748
		DATA INPUT				Total Expression				
	3	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	M B+ Pq 3	23.21	0.139	23.24	0.032	0.032975224	0.142788312	**0.977402552**	0.098973316	0.09673677
	M B- Pq 3	25.97	0.088	26.80	0.054	0.83294337	0.102952552	**0.561382747**	0.071361271	0.04006099
	M B+ Cont 3	26.96	0.153	28.98	0.063	2.016828257	0.16531324	**0.247100827**	0.114586406	0.0283144
	M B- Cont 3	25.71	0.153	27.23	0.058	1.515993297	0.164089384	**0.349655645**	0.113738094	0.03976917

Jheh1 females

		DATA INPUT				Total Expression				
	1	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV (variation coefficient)	error copies
Replica 1	F B+ Pq 1	22.92	0.130	23.45	0.050	0.526069778	0.139620714	**0.694443979**	0.096777704	0.06720669
	F B- Pq 1	22.63	0.148	23.67	0.053	1.044781821	0.156983455	**0.484718207**	0.108812639	0.05274347
	F B+ Cont 1	22.94	0.068	24.84	0.135	1.908037287	0.151243427	**0.266454797**	0.104833955	0.02793351
	F B- Cont 1	22.78	0.090	24.64	0.077	1.859573958	0.11830104	**0.275557642**	0.082000032	0.02259574
		DATA INPUT				Total Expression				
	2	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	F B+ Pq 2	22.19	0.139	22.84	0.019	0.643497044	0.140128285	**0.640159342**	0.097129526	0.06217837
	F B- Pq 2	22.58	0.138	23.50	0.123	0.916028112	0.184514657	**0.529966065**	0.127895815	0.06778044
	F B+ Cont 2	22.22	0.098	23.89	0.064	1.676167996	0.117625309	**0.312912675**	0.081531651	0.02551229
	F B- Cont 2	22.59	0.115	24.50	0.046	1.913904575	0.12357081	**0.265373354**	0.085652758	0.02272996
		DATA INPUT				Total Expression				
	3	Ct Act	Ct Std. Dev	Jheh1	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	F B+ Pq 3	21.78	0.177	22.25	0.155	0.465020037	0.235079319	**0.724461015**	0.162944567	0.11804699
	F B- Pq 3	22.86	0.043	23.79	0.077	0.932699776	0.088422002	**0.523877071**	0.061289461	0.03210814
	F B+ Cont 3	21.54	0.134	23.19	0.061	1.653019557	0.146963301	**0.317973941**	0.101867198	0.03239111
	F B- Cont 3	22.18	0.200	24.00	0.076	1.814863265	0.213986626	**0.284231181**	0.148324227	0.04215837

Jheh2 males

		DATA INPUT				Total Expression				
	1	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 1	M B+ Pq 1	21.76	0.145	21.78	0.110	0.013720389	0.181955975	**0.99053483**	0.126122271	0.1249285
	M B- Pq 1	22.09	0.079	22.76	0.053	0.675173934	0.095040469	**0.626256712**	0.065877033	0.04125593
	M B+ Cont 1	22.25	0.059	23.21	0.080	0.965456904	0.099055446	**0.512116198**	0.068660003	0.0351619
	M B- Cont 1	22.31	0.152	23.13	0.189	0.822047632	0.242639762	**0.565638556**	0.168185067	0.09513196
		DATA INPUT				Total Expression				
	2	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	M B+ Pq 2	22.04	0.050	21.84	0.022	-0.1928837	0.055165449	**1.143046183**	0.038237775	0.04370754
	M B- Pq 2	21.73	0.030	22.35	0.049	0.618521869	0.057188366	**0.651337922**	0.039639955	0.02581901
	M B+ Cont 2	22.55	0.154	23.60	0.054	1.0499413	0.163328608	**0.482987816**	0.113210764	0.05467942
	M B- Cont 2	22.43	0.111	23.28	0.053	0.848622944	0.123404005	**0.555314532**	0.085537138	0.04750002
		DATA INPUT				Total Expression				
	3	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	M B+ Pq 3	22.41	0.122	21.78	0.062	-0.62431667	0.136753562	**1.541480532**	0.094790346	0.14611747
	M B- Pq 3	21.91	0.002	22.15	0.057	0.246545257	0.05681361	**0.842912477**	0.039380194	0.03319406
	M B+ Cont 3	22.39	0.114	23.60	0.014	1.212946179	0.114439131	**0.431386765**	0.079323161	0.03421896
	M B- Cont 3	22.52	0.130	23.20	0.052	0.682821785	0.140161916	**0.622945655**	0.097152837	0.06052094

Jheh2 females

		DATA INPUT				Total Expression				
	1	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 1	M B+ Pq 1	22.78	0.065	23.66	0.061	0.877829042	0.089812608	**0.544185703**	0.062253356	0.03387739
	M B- Pq 1	22.66	0.098	23.49	0.039	0.830245626	0.105829324	**0.562433477**	0.073355298	0.04125748
	M B+ Cont 1	23.04	0.171	23.84	0.077	0.802291534	0.18784506	**0.573437622**	0.130204274	0.07466403
	M B- Cont 1	22.51	0.157	23.49	0.104	0.983008634	0.188158347	**0.505923573**	0.130421428	0.06598327
		DATA INPUT				Total Expression				
	2	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	M B+ Pq 2	22.21	0.178	23.27	0.040	1.057887025	0.1826304	**0.480335045**	0.126589747	0.06080549
	M B- Pq 2	22.18	0.040	23.34	0.101	1.158141613	0.109011129	**0.448089364**	0.075560757	0.03385797
	M B+ Cont 2	22.86	0.067	24.41	0.189	1.55223677	0.199937545	**0.340980993**	0.138586146	0.04725524
	M B- Cont 2	22.45	0.116	23.77	0.054	1.314545622	0.128204911	**0.402052102**	0.088864872	0.03572831
		DATA INPUT				Total Expression				
	3	Ct Act	Ct Std. Dev	Jheh2	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	M B+ Pq 3	22.30	0.065	23.61	0.050	1.310005542	0.082541717	**0.40331933**	0.057213559	0.02307533
	M B- Pq 3	22.88	0.084	24.12	0.058	1.234740713	0.10157294	**0.424918861**	0.070404997	0.02991641
	M B+ Cont 3	22.34	0.094	23.75	0.034	1.410501317	0.099490871	**0.376180946**	0.068961817	0.02594212
	M B- Cont 3	22.84	0.150	24.11	0.030	1.266405261	0.152694609	**0.415694262**	0.105839838	0.04399701

Jheh3 males

		DATA INPUT				Total Expression				
	1	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 1	M B+ Pq 1	24.60	0.178	25.76	0.028	1.15826416	0.180267885	**0.448051303**	0.124952176	0.05598499
	M B- Pq 1	24.59	0.099	25.22	0.082	0.632616067	0.128976515	**0.645005753**	0.089399708	0.05766333
	M B+ Cont 1	25.00	0.098	25.53	0.042	0.529964398	0.106895059	**0.692571825**	0.074094009	0.05131542
	M B- Cont 1	24.93	0.108	24.64	0.077	-0.29231887	0.132601822	**1.224607034**	0.091912579	0.11255679
		DATA INPUT				Total Expression				
	2	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	M B+ Pq 2	24.57	0.024	25.99	0.096	1.413475669	0.098439945	**0.375406187**	0.06823337	0.02561523
	M B- Pq 2	24.63	0.123	25.40	0.028	0.772193917	0.12618648	**0.585526383**	0.087465803	0.05121354
	M B+ Cont 2	24.76	0.060	25.13	0.023	0.372468601	0.064468802	**0.772459607**	0.044686368	0.03451841
	M B- Cont 2	24.69	0.044	24.66	0.032	-0.03125054	0.054431343	**1.021897529**	0.037728932	0.0385551
		DATA INPUT				Total Expression				
	3	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	M B+ Pq 3	24.21	0.000	26.05	0.087	1.837055154	0.086651102	**0.279892522**	0.060061967	0.0168109
	M B- Pq 3	23.84	0.164	24.86	0.059	1.023204075	0.174469392	**0.492022407**	0.120932967	0.05950173
	M B+ Cont 3	24.53	0.172	24.89	0.032	0.353301679	0.17487417	**0.782790591**	0.121213538	0.09488482
	M B- Cont 3	24.20	0.114	24.27	0.151	0.065094818	0.189040168	**0.955882492**	0.131032659	0.12525182

Jheh3 females

		DATA INPUT				Expressió total				
	1	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 1	M B+ Pq 1	24.60	0.178	25.76	0.028	1.15826416	0.180267885	**0.448051303**	0.124952176	0.05598499
	M B- Pq 1	24.59	0.099	25.22	0.082	0.632616067	0.128976515	**0.645005753**	0.089399708	0.05766333
	M B+ Cont 1	25.00	0.098	25.53	0.042	0.529964398	0.106895059	**0.692571825**	0.074094009	0.05131542
	M B- Cont 1	24.93	0.108	24.64	0.077	-0.29231887	0.132601822	**1.224607034**	0.091912579	0.11255679
		DATA INPUT				Expressió total				
	2	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 2	M B+ Pq 2	24.57	0.024	25.99	0.096	1.413475669	0.098439945	**0.375406187**	0.06823337	0.02561523
	M B- Pq 2	24.63	0.123	25.40	0.028	0.772193917	0.12618648	**0.585526383**	0.087465803	0.05121354
	M B+ Cont 2	24.76	0.060	25.13	0.023	0.372468601	0.064468802	**0.772459607**	0.044686368	0.03451841
	M B- Cont 2	24.69	0.044	24.66	0.032	-0.03125054	0.054431343	**1.021897529**	0.037728932	0.0385551
		DATA INPUT				Expressió total				
	3	Ct Act	Ct Std. Dev	Jheh3	Ct Std. Dev	CG FR-Act	Std. dev delta	**nº copies**	CV	error copies
Replica 3	M B+ Pq 3	24.21	0.000	26.05	0.087	1.837055154	0.086651102	**0.279892522**	0.060061967	0.0168109
	M B- Pq 3	23.84	0.164	24.86	0.059	1.023204075	0.174469392	**0.492022407**	0.120932967	0.05950173
	M B+ Cont 3	24.53	0.172	24.89	0.032	0.353301679	0.17487417	**0.782790591**	0.121213538	0.09488482
	M B- Cont 3	24.20	0.114	24.27	0.151	0.065094818	0.189040168	**0.955882492**	0.131032659	0.12525182

Sript to gather 6 sequences of 100pb from reference genome (mentioned in Materials and Methods "Searching for cnc transcriptor binding site in Bari-Jheh")

```
#!/usr/bin/perl
use warnings;
use strict;

$ini_1 = 14964132; #initial annotation
$length = 100; #length of the query sequence
$ini_2 = 14964132 - $length; #end of annotation

open (REF, "3R.raw") or die "Error reference sequence"; #reference genome one chromosome at a time

my $ref = <REF>;
my $seq1 = '';
my $seq2 = '';
$seq1 = (substr $ref, $ini_1, $length);
$seq2 = (substr $ref, $ini_2, $length);

print "$seq1\n";
print "$seq2\n";
```

2) Scripts to search the consensus ARE secuence reference TEs (mentioned in Materials and Methods "Searching for putative AREs in annotated TEs")

```
#!/usr/bin/perl -w
#Eliminar repetidos
use strict;
use warnings;

#Find ARE consensus core in TE according to the fasta file provided in 5'-3' direction.
open (LIST, "/homes/users/mbarron/scratch/AREs/dmel-all-transposon-r5.53_2.txt") or die "Can't open list"; #TE annotations
open (OUTPUT, ">/homes/users/mbarron/scratch/AREs/TE_results_5410_AREsCore_5-3_good.txt") or die "Can't create file"; #Output file

print OUTPUT "TE_name\tTE_length\tstart:stop\t#ofAREs\n";

while (<LIST>){
    $i=0;
    $m=0;
    $start = 0;
    $end=0;
    $ARE_length='';
    if ($_ =~ /(\>.*)\s\s\s(.*)/){
        $name = $1;
        $sec = $2;
        $ARE_length = length ($sec);
        print OUTPUT "$name\t";
        print OUTPUT $ARE_length."\t";
        if ($sec =~ /.*TGA(T|C)[ATGC]{3}GC.*/){#5'-3'
            $sec =~ s/TGA(T|C)[ATGC]{3}GC/X/g;#5'-3'
        #if ($sec =~ /.*CG[ATGC]{3}(T|C)AGT.*/){#3'-5'
            #$sec =~ s/CG[ATGC]{3}(T|C)AGT/X/g;#3'-5'
            @nucl = split ("", $sec);
            foreach $pos (@nucl){
                $m++;
                if ($pos eq "X"){
                    $i++;
                    $start = $m + ($ARE_length*($i-1));
                    $end = $start + $ARE_length;
                    print OUTPUT "$start:$end;";
                    #push (@start, $start);
                    #push (@end, $end);
```

```
                }
            }
        }
        print OUTPUT "\t$i\n";
    }

}


#!/usr/bin/perl -w
use strict;
use warnings;

#Find ARE consensus core in TE according to the fasta file provided in 3'-5'
direction.
open (LIST, "/homes/users/mbarron/scratch/AREs/dmel-all-transposon-r5.53_2.txt")
or die "Can't open lista"; #TE annotations
open (OUTPUT, ">/homes/users/mbarron/scratch/AREs/TE_results_5410_AREsCore_3-
5.txt") or die "Can't create file";#Output file

print OUTPUT "TE_name\tTE_length\tstart:stop\t#ofAREs\n";

while (<LIST>){
    $i=0;
    $m=0;
    $start = 0;
    $end=0;
    $ARE_length='';
    if ($_ =~ /(\>.*)\s\s\s(.*)/){
        $name = $1;
        $sec = $2;
        $ARE_length = length ($sec);
        print OUTPUT "$name\t";
        print OUTPUT $ARE_length."\t";
        #if ($sec =~ /.*TGA(T|C)[ATGC]{3}GC.*/){#5'-3'
            #$sec =~ s/TGA(T|C)[ATGC]{3}GC/X/g;#5'-3'
        if ($sec =~ /.*CG[ATGC]{3}(T|C)AGT.*/){#3'-5'
            $sec =~ s/CG[ATGC]{3}(T|C)AGT/X/g;#3'-5'
            @nucl = split ("", $sec);
            foreach $pos (@nucl){
                $m++;
                if ($pos eq "X"){
                    $i++;
                    $start = $m + ($ARE_length*($i-1));
                    $end = $start + $ARE_length;
                    print OUTPUT "$start:$end;";
                    #push (@start, $start);
                    #push (@end, $end);
                }
            }
        }
        print OUTPUT "\t$i\n";
    }

}

3) Scripts to select the TEs with ARE consensus sequence in a distance lower to
1kb upstream a gene (mentioned in Materials and Methods "Searching for putative
AREs in annotated TEs")

#Finds the genes <1kb upstream of the candidates TEs
#!/usr/bin/perl -w
#use strict;
use warnings;

open (TE,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/TE_with_AREs.txt
") or die "Can't open file";# List of TEs with ARE consensus sequence
```

```
while (<TE>){
        #print $_;
    @col = split ("\t", $_);
    $FBti = $col[6];
    $start_TE = $col[10];
    $stop_TE = $col[11];
    $chr_TE = $col[9];
    $TE_ori = $col[7];
    $TE_name = $col[12];
    $TE_family = $col[13];
    $TE_ARE_number = $col[5];

    $new_stop = $stop_TE + 5000;
    $new_start = $start_TE - 5000;

    #print "$TE_ori\t";
    open (GENE,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/noncoding_gene/d
mel-r5.53.coding.gff") or die "Can't open file"; # Reference annotations
        while (<GENE>){
                @gff = split ("\t", $_);
                $chr_gene = $gff[0];
                $start_gene = $gff[3];
                $stop_gene = $gff[4];
                $ori_gene = $gff[6];
                @info = split (";", $gff[8]);
                        @name = split ("=", $info[0]);
                        $FBgn = $name[1];
                if ($ori_gene eq "+" && $chr_TE eq $chr_gene && $start_gene >
$stop_TE && $start_gene < $new_stop){
                #if ($TE_ori eq "+"){
                        print
"$FBti\t$chr_TE\t$start_TE\t$stop_TE\t$TE_ori\t$TE_name\t$TE_family\t$TE_ARE_num
ber\t$FBgn\t$start_gene\t$stop_gene\t$ori_gene\n";
                }elsif ($ori_gene eq "-" && $chr_TE eq $chr_gene && $stop_gene
< $start_TE && $stop_gene > $new_start){
                        print
"$FBti\t$chr_TE\t$start_TE\t$stop_TE\t$TE_ori\t$TE_name\t$TE_family\t$TE_ARE_num
ber\t$FBgn\t$start_gene\t$stop_gene\t$ori_gene\n";
                }
        }
        }

#Remove the noncoding genes from the list optained by the above script
#!/usr/bin/perl -w
use strict;
use warnings;

open (GENE1,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/area_de_trabajo"
) or die "Can't open file";#List of genes <1kb upstream of cadidate TEs

while (<GENE1>){
        $_ =~s/\n//g;
        @col = split ("\t", $_);
        $FBgn_gwas = $col[0];
        open (GENE2,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/noncoding_gene/d
mel-all-noncoding-r5.53.gff") or die "Can't open file";#reference annotation
        while (<GENE2>){
                $TE_info = $_;
                $TE_info =~ s/\n//g;
                @pos = split ("\t", $TE_info);
                @inf = split (";", $pos[8]);
                @name = split ("=", $inf[2]);
                $FBgn_TE = $name[1];
                #print "$FBgn_gwas\t$FBgn_TE\n";
```

```
                if ($FBgn_gwas eq $FBgn_TE){
                        print "$FBgn_gwas\t$TE_info\n";
                }
        }
}
```

4) Script for selecting the TEs with a determined population frequency (mentioned in Materials and Methods "Searching for putative AREs in annotated TEs")

```
#!/usr/bin/perl -w
use strict;
use warnings;

#Secuencias de los 5425TEs; No en fasta si no que cada linea es un TE. Si hace falta hacer otro perl para transformaciÃ³n.
open (TE,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/gene_1kb_TEwithARE.txt") or die "Can't open file1"; #File of the candidate TEs

while (<TE>){
    $FBti_core = $_;
    $FBti_core =~s/\n//g;
    $FBti_core =~s/>//g;
    @inf = split ("\t", $FBti_core);
    $FBti = $inf[0];
    open (FREQ_1,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/Tfreq_5425knownREannotTEs_141DGRPstrains") or die "Can't open file2";# TE frequency file
    while (<FREQ_1>){
        $_ =~s/\n//g;
        @col = split ("\t", $_);
        $te = $col[0];
        #$total_results = $col[6];
        #$freq = $col[7];
        if ($FBti eq $te){
            print "$FBti_core\t$_\n";
        }
    }
}
```

5) Script for searching cnc bind sites between the TE and the gene (mentioned in Materials and Methods "Searching for putative AREs in annotated TEs")

```
#!/usr/bin/perl -w
use strict;
use warnings;

open (GENE,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/TE_ARE_1kb/TEs_ARE_1kb_all_nov.txt") or die "Can't open file";#candidate gene list

while (<GENE>){
    @first_sep = split ("\t", $_);
    $gene_info = $_;
    $gene_info =~ s/\n//g;
    $chr_gene = $first_sep[15];
    $start = $first_sep[16];
    $end = $first_sep[17];
    #$orient = $first_sep[19];
    #if ($orient eq "+"){
    $start_gene = $start - 10000;
    $end_gene = $end + 10000;
        #}else{
        #$start_gene = $start + 10000;
        #$end_gene = $end - 10000;
        #}
    #print "$gene_info\n";
```

```
    open (TF,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/TFBS/tfbs_cnc_al
l.gff3") or die "Can't open file";#cnc_binding sites annotations
    while (<TF>){
        #print $_."\t";
        @sep = split (" ", $_);
        $TF_info = $_;
        $TF_info =~s/\n//g;
        $chr_tf = $sep[0];
        $start_tf = $sep[3];
        $end_tf = $sep[3];
        #$i++;
        #print "$TF_info\n";
        if ($chr_gene eq $chr_tf && $start_gene <= $start_tf && $end_tf <=
$end_gene){
            print "$gene_info\t$TF_info\n";
        }elsif ($chr_gene eq $chr_tf && $start_tf <= $end_gene && $end_tf >=
$end_gene){
            print "$gene_info\t$TF_info\n";
        }elsif ($chr_gene eq $chr_tf && $start_tf <= $start_gene && $end_tf >=
$start_gene){
            print "$gene_info\t$TF_info\n";
        }
    }
}
```

6) Script to select TEs that are also part of a GWAS study (mentioned in Materials and Methods "Searching for putative AREs in annotated TEs")

```
#!/usr/bin/perl -w
use strict;
use warnings;

open (GENE1,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/GWAs_comp.txt")
or die "Can't open file";#candidate gene list of GWAs studies

while (<GENE1>){
        $_ =~s/\n//g;
        @col = split ("\t", $_);
        #$FBgn_gwas = $col[0];#girardot
        #$FBgn_gwas = $col[4];#weber
        $FBgn_gwas = $col[1];#weber
        open (GENE2,
"/Users/GonzalezLab/Documents/Maite/AREs_finding/AREs_finding_2/GWAs_comp/gene_1
kb_TEwithARE.txt") or die "Can't open file";# gene list <1kb of the TEs with ARE
consensus sequence
        while (<GENE2>){
                $TE_info = $_;
                $TE_info =~ s/\n//g;
                @pos = split ("\t", $TE_info);
                $FBgn_TE = $pos[8];
                #print "$FBgn_gwas\t$FBgn_TE\n";
                if ($FBgn_gwas eq $FBgn_TE){
                        print "$FBgn_gwas\t$TE_info\n";
```

Table S1. Description of the regions used to look for *cnc* reads and number of *cnc* selected reads per developmental stage and replica (rep)

Description	**Query start position**	**Query end position**	**Number of reads selected**			
			embryo		**adult**	
			rep1	**rep2**	**rep1**	**rep2**
(i) 100bp region around the *Nrf2* binding site in the upstream region of *Jheh1*	2R: 14.742.812	2R: 14.742.911	6	25	42	29
(ii) 100bp region around the *Nrf2* binding site in the upstream region of *Jheh2*	2R: 14.745.406	2R: 14.745.505	67	24	166	79
(iii) 100 bp region including the left junction of *Bari-Jheh* (50bp of unique flanking region+first 50bp of *Bari-Jheh)*	2R: 14.745.747	2R: 14.745.846	26	15	34	20
(iv) 100 bp region including the right junction of *Bari-Jheh* (last 50bp of *Bari-Jheh* + 50bp of unique flanking region)	2R: 14.747.475	2R: 14.747.574	3	7	11	8
(v) Diagnostic SNP (*Bari-Jheh* has an A and all other *Bari1* TEs have a G) located 278bp upstream of the left *Bari-Jheh Nrf2* binding site	2R: 14.746.239-	2R: 14.746.338	15	6	14	6
(vi) Diagnostic SNP (*Bari-Jheh* has an G and all other *Bari1* TEs have a A) located 371bp upstream of the right *Bari-Jheh Nrf2* binding site	2R: 14.746.962	2R: 14.747.061	6	7	8	5

Table S2. TEs with AREs located in the 1kb upstream regions of genes

TE flybase identifier	Chromosomal arm	TE family	Number of AREs inside TE	Downstream gene	Frequency in DGRP strains[1]
FBti0019530	X	baggins	1	CG3021	100
FBti0062170	3R	Burdock	1	CG3987	100
FBti0059655	4	Cr1a	2	eIF4G	100
FBti0060356	2L	diver2	5	CG3262	100
FBti0020310	3R	G5	2	glob3	100
FBti0060088	X	G5A	1	CG32500	100
FBti0061033	3L	gypsy2	4	CG17454	100
FBti0020147	3L	gypsy2	2	Met75Cb	100
FBti0019555	X	hopper	1	CG16781	100
FBti0061521	2R	INE-1	1	CG33680	100
FBti0060651	3L	INE-1	1	CG32453	100
FBti0063713	2R	INE-1	2	Or42a	100
	2R	INE-1	2	Tsp42A	100
FBti0060655	3L	INE-1	1	CG14451	100
	3L	INE-1	1	mael	100
FBti0061676	3R	INE-1	1	CR33294	100
FBti0062898	4	INE-1	1	CG2219	100
FBti0061998	4	INE-1	1	CG11155	100
FBti0063392	2L	INE-1	1	CG9272	100
	2L	INE-1	1	CR33319	100
FBti0061194	X	INE-1	1	Cda4	100
FBti0060127	3L	INE-1	1	CG14569	100
FBti0061543	2L	INE-1	1	elfless	100
	2L	INE-1	1	rdo	100
FBti0062273	X	INE-1	1	CG10597	100
	X	INE-1	1	f	100
FBti0060764	4	INE-1	1	fd102C	100

FBti0064279	2R	INE-1	1	CG14591	100
FBti0060447	3R	INE-1	1	CG34303	100
FBti0063387	4	INE-1	1	NfI	100
FBti0062156	2L	INE-1	1	Tsp39D	100
FBti0061379	2L	INE-1	1	squ	100
	2L	INE-1	1	grp	100
FBti0061758	3L	INE-1	2	CG34031	100
FBti0061445	2R	INE-1	1	PebII	100
FBti0060913	3R	INE-1	1	Arfip	100
FBti0061869	3L	INE-1	1	CG33687	100
FBti0020293	3L	invader1	3	Ago3	100
FBti0018999	2R	mariner2	2	Gr43a	100
FBti0020428	4	mariner2	1	CG9935	100
FBti0019322	3R	micropia	10	CG43290	100
FBti0062048	3L	Quasimodo	9	Spn77Bc	100
FBti0019282	3R	Quasimodo	3	CG1092	100
FBti0020401	4	Rt1b	5	JYalpha	100
FBti0019653	X	Rt1c	1	CR33498	100
FBti0063349	2L	Rt1c	3	CR43158	100
FBti0019664	X	Rt1c	1	Ntf-2	100
FBti0019295	3R	S	1	CG9769	100
FBti0019480	4	S	1	CG31999	100
	4	S	1	yellow-h	100
FBti0019362	3R	S	1	Hsp70Bbb	100
FBti0020231	3L	S2	1	RpL10	100
FBti0019124	2L	S2	1	mRpL28	100
FBti0062821	2L	springer	1	CR42545	100
FBti0019651	X	Rt1c	1	CR33496	99.65
FBti0019652	X	Rt1c	1	CR33487	99.65
FBti0063368	4	gypsy8	9	JYalpha	99.64
FBti0020173	3L	Rt1b	1	CG6434	99.64
FBti0019650	X	Rt1c	1	CR33491	99.64
FBti0060075	2R	gypsy9	7	CG44104	99.63

FBti0062858	X	Rt1c	3	CR43377	99.63
FBti0062491	3L	gypsy6	2	CG32350	99.62
FBti0064270	2R	INE-1	1	dream	99.62
FBti0062860	X	Rt1c	3	CR43379	99.62
FBti0062927	X	Rt1c	3	CR43381	99.62
FBti0019647	X	Rt1c	3	CR32821	99.61
FBti0061742	2R	rooA	1	CG18446	99.59
FBti0060834	3R	S	1	CG33494	99.58
FBti0020015	3L	412	3	LysP	99.34
	3L	412	3	LysS	99.34
FBti0019062	X	S	1	Lsd-2	98.98
FBti0060717	3L	297	1	mael	93.01
FBti0019082	X	Rt1b	3	CG18259	91.43
	X	Rt1b	3	CR6900	91.43
FBti0059710	X	roo	2	CG12680	22.98
FBti0019354	3R	17.6	4	CG14692	20.7
FBti0019012	2R	pogo	2	fab1	18.91
	2R	pogo	2	CG33981	18.91
FBti0020152	3L	Doc	1	term	15.62
FBti0019367	3R	Doc	3	CG44142	7.94
FBti0019217	2L	Quasimodo	1	Oseg5	7.42
FBti0019335	3R	G2	2	puc	6.38
FBti0020117	3L	gypsy5	2	Best4	6.37

[1]A-S Fiston-Lavier, M.G. Barron, D.A. Petrov and J. González, personal communication

CLUSTAL 2.1 multiple sequence alignment of the Bari-Jheh intergenic region in introgressed flies with (Wi3/Bari+) and without (Wi3/Bari-) the insertion

```
Wi3/Bari+
TACCAAACGGATTTTGTACTTTCGCTTGGGAATCGGGAGGCACATGAGATAGCTATTAAA 60
Wi3/Bari-
TACCAAACGGATTTTGTACTTTCGCTTGGGAATCGGGAGGCACATGAGATAGCTATTAAA 60

************************************************************

Wi3/Bari+
TTGGCAAGGCGACAAAAGGTCAACGTATTTTAAGTACGGTTTTGCAGTAAGGTGGTTGGG 120
Wi3/Bari-
TTGGCAAGGCGACAAAAGGTCAACGTATTTTAAGTACGGTTTTGCAGTAAGGTGGTTGGG 120

************************************************************

Wi3/Bari+
ACCACCTCAGTACCACCATAACGTGGTTACATAATATTTAACAAATGTGGAAGGGAAATA 180
Wi3/Bari-
ACCACCTCAGTACCACCATAACGTGGTTACATAATATTTAACAAATGTGGAAGGGAAATA 180

************************************************************

Wi3/Bari+
GAAGGGGGTAAGACAAAATCTTTTTTAATTACATTACAATAAAACTAATAACTATAAATA 240
Wi3/Bari-
GAAGGGGGTAAGACAAAATCTTTTTTAATTACATTACAATAAAACTAATAACTATAAATA 240

************************************************************

Wi3/Bari+
GTAGGTATGGTTAAGAAGTAATTGCTTGTAATCCATGTTTGTAATTGACCGCAAACTAAA 300
Wi3/Bari-
GTAGGTATGGTTAAGAAGTAATTGCTTGTAATCCATGTTTGTAATTGACCGCAAACTAAA 300

************************************************************

Wi3/Bari+
AAAATATTATTGATTCTTTATTTATATATATGTATATATGTATATTTATATATATACAGT 360
Wi3/Bari-       AAAATATTATTGATTCTTTATTTATATATATGTATATATGTATATTTATATATAT-
---- 355
                ***************************************************

Wi3/Bari+
CATGGTCAAAATTATTTTCACAAAGTGCATTTTTGTGCATGGGTCACAAACAGTTGCTTG 420
Wi3/Bari-       ----------------------------------------------------------
----


Wi3/Bari+
TGCAGCAAGTGGGGGGAGGTGAAATGCAAAAAAACTTTTGCTTTTGCAAATTCAAACCTA 480
Wi3/Bari-       ----------------------------------------------------------
----


Wi3/Bari+
TGCAGAGTCAGATGAAAGAAGAATTGAAAAAATAACTGTTCCTATGCGCAAGGAAGAGGC 540
```

```
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
AAATGAAGAGATCTTTATCAGTTGTCAGAAGTATTTGCACACGGTTTCGTCGCATCACAA 600
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TTATTTTCACAACGCAATTTCTTCTTCAGTGATTGGTTTAGAGTGACAAGTGCCGGTTTG 660
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TTTGCTTAAATACATTTAAATTATTGAATAAAAATTAGATTTAATCATTTTCCTATTACA 720
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GTTATTAAATAAAATGCCCAAAACAAAAGAGTTAACAGTTGAGGCCCGGGCTGGTATTGT 780
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TGCTAGGTTTAAAGCCGGTACACCTGCGGCCAAAATAGCTGAAATATATCAAATTTCGCG 840
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TAGAACTGTCTACTACTTAATAAAAAAGCTTGATACAGTTGGCACATTAAAAAATAAAAA 900
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
AAGATCAGGCCGAAAACCTGTGCTGGACCAAAGGCAATGCAGGCAAATACTTGGAGTTGT 960
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GGCGAAGAATCCTAGTGCCAGTCCGGTAAAAATTGCCTTAGAATCAAAAAATACAATTGG 1020
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
CAAACAAGTTAGTAGTTCTACAATTCGTCGCAGGCTAAAAGAAGCTGATTTTAAGACATA 1080
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
CGTTGTTCGCAAAACGATTGAGATCACACCAACCAACAAAACAAAACGTCTTCGATTTGC 1140
```

```
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GTTGGAATATGTTAAGAAGCCTCTTGACTTTTGGTTTAATATTTTATGGACTGATGAGTC 1200
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TGCATTTCAGTACCAGGGGTCATACAGCAAGCATTTTATGCATTTGAAAAATAATCAAAA 1260
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GCATTTGGCAGCCCAGCCAACCAATAGATTTGGTGGGGGCACAGTCATGTTTTGGGGATG 1320
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TCTTTCCTATTATGGATTCGGAGACTTGGTACCGATAGAAGGAACTTTAAATCAGAACGG 1380
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
ATACCTTCTTATCTTAAACAACCATGCTTTTACGTCTGGAAATAGACTTTTTCCAACTAC 1440
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TGAATGGATTCTTCAGCAGGACAATGCTCCATGCCATAAGGGTAGGATACCAACAAAATT 1500
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TTTAAACGACCTTAATCTGGCGGTTCTTCCGTGGCCCCCCCAAAGCCCAGACCTTAATAT 1560
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
CATTGAAAACGTTTGGGCTTTTATTAAAAACTAACGAACTATTGATAAAAATAGAAAACG 1620
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
AGAGGGAGCCATCATTGTAATAGCGGAGATTTGGTCCAAATTGACATTAGAATTTGCACA 1680
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
AACTTTGGTAAGGTCAATACCAAAAAGACTTCAAGCAGTTATTGATGCCAAAGGTGGTGT 1740
```

```
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TACAAAATATTAGTATTGTATTTATATAAAATAAAAAAAATTCTTATGTTGAAATTAGAT 1800
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GTTAAGCTGAAATTTACTAAATTAAGTTGAGTGAAAATACTTTTGAAGCGCAATAAACAT 1860
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GTGAAAATACTATTGACAACTTGCATGCATATTTTCTTTTGCTTTAAGCTTTGTACTATG 1920
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
AACCGTTATCTTTCGTATTTCTTTTCGACTACCTTCTGCATAGATCAAGCTAAGCGATAA 1980
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
GAACTATTTCAGGCAAATCGGACAACAACAAGAAGAAATATAACAAAAAGAAGTTGAAGT 2040
Wi3/Bari-       ------------------------------------------------------
----


Wi3/Bari+
TTGCAAATATTGTGCGTTGTGAAAATACTTTTGACCACCTCTGTATATATTAAGGGCTCC 2100
Wi3/Bari-       ------------------------------------------
ATATATTAAGGGCTCC 371

***************

Wi3/Bari+
GCGTGTTGGTAATTCGAGTTCTTAATCATTATTAATTAATTAAATCAGTATTTAGTTAAA 2160
Wi3/Bari-
GCGTGTTGGTAATTCGAGTTCTTAATCATTATTAATTAATTAAATCAGTATTTAGTTAAA 431

************************************************************

Wi3/Bari+
TGTCATATAACAATTTCATTTTAAGCATGAATCGTTTCTTGTTCACTTTACTTTCGTGGA 2220
Wi3/Bari-
TGTCATATAACAATTTCATTTTAAGCATGAATCGTTTCTTGTTCACTTTACTTTCGTGGA 491

************************************************************

Wi3/Bari+
TTGATAAATGGAACTGCTTGATCATCTTCCTAAACTAAATGTAAATTTTAAGTACAAAAA 2280
Wi3/Bari-
TTGATAAATGGAACTGCTTGATCATCTTCCTAAACTAAATGTAAATTTTAAGTACAAAAA 551

************************************************************
```

```
Wi3/Bari+
TTGCTCTCACTCAATTTGTTGCCAAAAGCTCGCCGAAATTCTCAAATTATTTGTCCAATC 2340
Wi3/Bari-
TTGCTCTCACTCAATTTGTTGCCAAAAGCTCGCCGAAATTCTCAAATTATTTGTCCAATC 611

************************************************************

Wi3/Bari+
ATGCTCGCATTGCATTGCCGTGTGGAATACGATCCACTTGAAATCCACAAGCCAACAAAA 2400
Wi3/Bari-
ATGCTCGCATTGCATTGCCGTGTGGAATACGATCCACTTGAAATCCACAAGCCAACAAAA 671

************************************************************

Wi3/Bari+       AGCTTTTGTTTGCCGGTAGCTTGCGCTTTACAAA 2434
Wi3/Bari-       AGCTTTTGTTTGCCGGTAGCTTGCGCTTTACAAA 705
                ***********************************
```

5 The dominance effect of the adaptive transposable element insertion *Bari-Jheh* depends on the genetic background

The dominance effect of the adaptive transposable element insertion *Bari-Jheh* depends on the genetic background

Resumen

Si bien las mutaciones adaptativas son frecuentemente consideradas dominantes, recientemente se ha mostrado que una proporción importante de mutaciones adaptativas deberían mostrar ventajas en heterocigosis. En este trabajo utilizamos un elemento móvil recientemente caracterizado como adaptativo en respuesta a estrés oxidativo en *Drosophila melanogaster*, para estudiar la dominancia de una mutación adaptativa. La comparación de las curvas de supervivencia de moscas heterocigotas con los correspondientes homocigotos indica que la dominancia de *Bari-Jheh* depende del fondo genético. Tanto en moscas homocigotas como heterocigotas, *Bari-Jheh* está asociado con la sobre expresión de los genes *Jheh1* (Epoxi Hidrolasa de la Hormona Juvenil 1) y/o *Jheh2*. Nuestros resultados añaden al limitado número de estudios en los que se ha estimado empíricamente la dominancia de una mutación adaptativa y destacan la complejidad de su herencia.

GBE

Genome Biology and Evolution

SMBE

The Dominance Effect of the Adaptive Transposable Element Insertion *Bari-Jheh* Depends on the Genetic Background

Lain Guio and Josefa González*

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

*Corresponding author: E-mail: josefa.gonzalez@ibe.upf-csic.es.

Accepted: April 22, 2015

Abstract

Although adaptive mutations are often considered to be dominant, it has been recently shown that a substantial proportion of adaptive mutations should display heterozygote advantage. In this work, we take advantage of a recently characterized transposable element insertion mediating oxidative stress response in *Drosophila melanogaster* to test the dominance effect of an adaptive mutation. The comparison of the survival curves of heterozygous and the two corresponding homozygous flies indicated that the dominance effect of *Bari-Jheh* depends on the genetic background. Both in homozygous and in heterozygous flies, *Bari-Jheh* was associated with upregulation of *Jheh1* (*Juvenile Hormone Epoxyde Hydrolase 1*) and/or *Jheh2* genes. Our results add to the limited number of studies in which the dominance effect of adaptive mutations has been empirically estimated and highlights the complexity of their inheritance.

Key words: adaptive mutation, selective sweep, heterozygote advantage, oxidative stress, *Drosophila*, dominance effect.

Dominance Effect of Deleterious and Adaptive Mutations

Understanding the dominance effect of mutations has consequences for several important biological processes, such as the magnitude of inbreeding depression, the evolution of mating systems, and the rate of adaptation in diploids (Charlesworth B and Charlesworth D 1998; Lynch et al. 1999; Manna et al. 2011). To date, most of our knowledge on the dominance effect of mutations comes from the study of deleterious mutations (Charlesworth B and Charlesworth D 1998; Garcia-Dorado et al. 1999; Lynch et al. 1999). These studies, mostly based on mutation-accumulation experiments in flies, showed that the majority of deleterious mutations is recessive to their wild-type allele (Simmons and Crow 1977; Wilkie 1994; Houle et al. 1997; Chavarrias et al. 2001; Fry and Nuzhdin 2003). In contrast, the study of the dominance effect of adaptive mutations has lagged behind, mostly due to the difficulty to identify adaptive mutations. However, and although few studies have empirically determined their dominance effects, adaptive mutations are often considered to be dominant (Bourguet et al. 1997; Charlesworth 1998; Orr 2010; Zhang et al. 2011; Joseph et al. 2014). This notion derives from Haldane (1927) who showed that when a mutation is rare, as it is the case of new mutations, it is more likely to be fixed if it is dominant. This is so because recessive mutations are phenotypically expressed only in homozygotes and, when the mutation is rare, the corresponding homozygotes are even rarer assuming a large outbred population. Therefore, selection has little chance of acting on recessive mutations as most of the mutant alleles are found in heterozygotes. Based on the assumption that adaptive mutations are likely to be dominant, positive selection should drive these mutations to high population frequency removing genetic variation at linked sites and thus leaving characteristic molecular signatures of complete selective sweeps. Until recently most genomic scans for positive selection were focused on identifying signatures of complete selective sweeps (Sabeti et al. 2002; Glinka et al. 2003; Voight et al. 2006). However, it has been recently shown that a substantial proportion of adaptive mutations may display heterozygote advantage (Sellis et al. 2011). Sellis et al. (2011) demonstrated that if selection is stabilizing and mutation effects are large enough to overshoot the fitness optimum, heterozygous advantage should be very common in adaptation. If adaptive mutations are overdominant, besides complete selective sweeps, we would also expect to see many incomplete selective sweeps surrounding adaptive mutations. Indeed, incomplete sweeps are common in several organisms (Clark et al. 2007; Gonzalez et al. 2008; Burke and Rose 2009; Coop et al. 2009). However, evidence of incomplete sweeps is not diagnostic of heterozygote advantage as this molecular

signature is also predicted under other scenarios, such as polygenic adaptation and adaptation to specific subhabitats (Messer and Petrov 2013). Thus, to explicitly test the hypothesis of heterozygote advantage, we need to directly measure the fitness of heterozygous individuals and compare it with the fitness of homozygous individuals for the presence and for the absence of the adaptive mutation (Sellis et al. 2011).

The Dominance Effect of *Bari-Jheh* Depends on the Genetic Background

Bari-Jheh is a full-length transposable element insertion located on chromosomal arm 2R in *Drosophila melanogaster*. *Bari-Jheh* is a good candidate to empirically evaluate the dominance effect of an adaptive mutation: It mediates resistance to oxidative stress and it is polymorphic in natural populations (Gonzalez et al. 2009; Guio et al. 2014). Thus, it is possible to measure the survival of heterozygous flies and compare it with the survival of the two corresponding homozygous (Gonzalez et al. 2008, 2009).

To determine the dominance effect of *Bari-Jheh* on oxidative stress resistance, we compared the survival of homozygous flies for the presence of *Bari-Jheh*, homozygous flies for the absence of *Bari-Jheh,* and heterozygous flies obtained from reciprocal crosses of the two homozygous strains. We first analyzed flies from outbred populations previously created in our lab (Guio et al. 2014). As expected, both male and female flies homozygous for the presence of *Bari-Jheh* were more resistant to oxidative stress compared with flies homozygous for the absence of *Bari-Jheh* (fig. 1*A* and table 1; Guio et al. 2014). Because we did not find differences in the survival curves of heterozygous flies from reciprocal crosses, we did not take into account the direction of the cross in our analyses (table 1). We found that survival curves of heterozygous flies were statistically different from survival curves of homozygous flies without *Bari-Jheh* (table 1 and fig. 1*A*). However, we found that survival curves of heterozygous flies were not statistically different from survival curves of homozygous flies with *Bari-Jheh* suggesting that the effect of this adaptive mutation on oxidative stress resistance is dominant (table 1 and fig. 1*A*).

Because the dominance effect of mutations can be affected by the genetic background (Mukai et al. 1966; Simmons and Crow 1977), we repeated the oxidative stress survival experiment with introgressed flies also previously created in our lab (Gonzalez et al. 2009; Guio et al. 2014). We found that both male and female flies homozygous for the presence of *Bari-Jheh* were more resistant to oxidative stress than homozygous flies for the absence of *Bari-Jheh* (fig. 1*B* and table 1), as we have previously reported (Guio et al. 2014). For females, we did not find differences in the survival curves between the heterozygous flies from reciprocal crosses (table 1). However, we found differences in the survival curves of males and thus we analyzed the two crosses separately for males (table 1). We found that heterozygous female flies and males from one of the reciprocal crosses were more resistant to oxidative stress compared with flies without the insertion and showed no differences compared with flies with the insertion suggesting that *Bari-Jheh* is dominant (fig. 1*B*). On the other hand, males from the other reciprocal cross were more resistant to paraquat compared with flies with and without the insertion suggesting that in this particular background *Bari-Jheh* is overdominant (table 1 and fig. 1*B*). To confirm these results, we repeated the experiments with another pair of introgressed flies (see Materials and Methods). We obtained similar results: Heterozygous female flies and males from one of the reciprocal crosses were more resistant to paraquat compared with flies without the insertion and showed no differences compared with flies with the insertion, whereas males from the other reciprocal cross were more resistant compared with flies with and without the insertion (table 1 and fig. 1*C*).

Overall, we found that *Bari-Jheh* dominance effect depended on the genetic background. In outbred populations, *Bari-Jheh* is a dominant mutation. In introgressed strains, *Bari-Jheh* is a dominant mutation in females whereas in males *Bari-Jheh* is dominant or overdominant depending on the reciprocal cross.

Bari-Jheh Is Associated with Upregulation of *Juvenile Hormone Epoxyde Hydrolase 1* and/or *2* in Homozygous and Heterozygous Flies

Bari-Jheh is located in the intergenic region between *Juvenile Hormone Epoxyde Hydrolase 2* (*Jheh2*) and *Jheh3* and 3.2 kb upstream of *Jheh1*. We have previously reported the expression level of these three genes in flies homozygous for the presence and for the absence of *Bari-Jheh* (Guio et al. 2014). In this work, we have analyzed the expression level of these three genes in heterozygous male flies.

In outbred populations, we found that male flies homozygous for the presence of *Bari-Jheh* are associated with upregulation of *Jheh1* and *Jheh2* and downregulation of *Jheh3* genes, as previously described (*t*-test P value = 0.0004, 0.0080, and 0.0033, respectively; fig. 2*A*; Guio et al. 2014). We compared the expression of the three genes in heterozygous males from the two reciprocal crosses and we did not find significant differences (*t*-test P value > 0.05; supplementary fig. S1*A*, Supplementary Material online). Thus, we combined the expression results for the two crosses (fig. 2*A*). Flies heterozygous for *Bari-Jheh* mutation are associated with *Jheh1* upregulation (*t*-test P value = 0.0325; fig. 2*A*). Because flies heterozygous for *Bari-Jheh* are resistant to oxidative stress, these results suggested that upregulation of one of the two genes, *Jheh1* or *Jheh2* may be enough to confer resistance to oxidative stress. Consistent with this hypothesis, in the

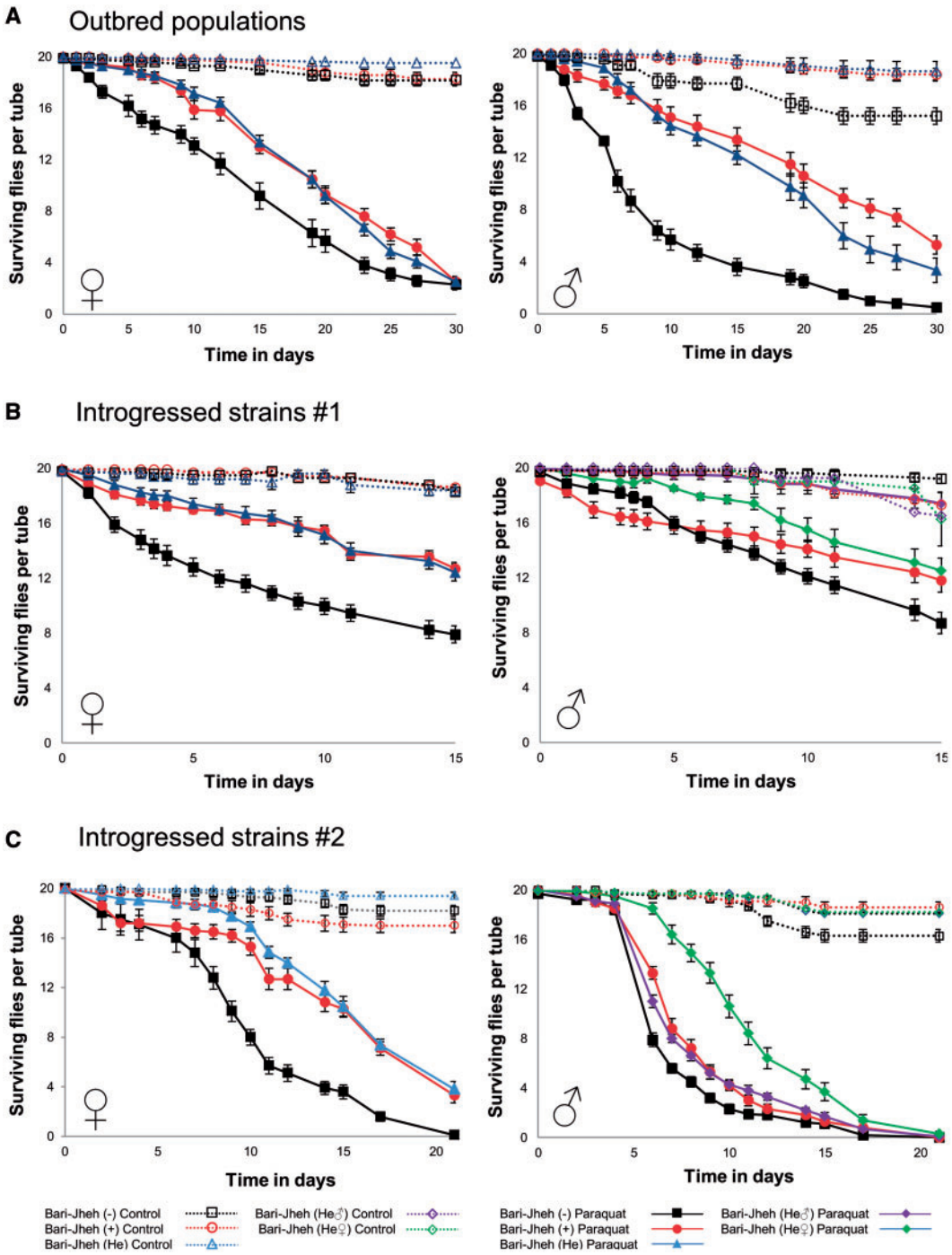


Fig. 1.—Dominance effect of *Bari-Jheh* on oxidate stress resistance in outbred populations (A) and in introgressed strains (B) and (C). Survival curves of homozygous flies with *Bari-Jheh* insertion (*Bari-Jheh* (+)), homozygous flies without *Bari-Jheh* insertion (*Bari-Jheh* (−)), heterozygous flies from crosses in which the father carried the insertion (*Bari-Jheh* (He♂)), heterozygous flies from crosses in which the mother carried the insertion (*Bari-Jheh* (He♀)), and heterozygous flies from the two reciprocal crosses considered together (*Bari-Jheh* (He)).

introgressed strains, flies homozygous for the presence of *Bari-Jheh* are associated with upregulation of *Jheh2* (*t*-test *P* value = 0.0020; fig. 2*B*) and flies heterozygous for *Bari-Jheh* showed upregulation of both *Jheh1* (*t*-test *P* value = 0.0294; fig. 2*B*) and *Jheh2* genes (*t*-test *P* value = 0.0001 and 0.0211 for the two reciprocal crosses; fig. 2*B*).

In introgressed strains, heterozygous flies from the two reciprocal crosses differed in the level of expression of *Jheh2* and

Table 1
Statistical Analyses of the Survival Curves

Genetic Background	Strains Compared[a]	Sex	Logrank Test *P* Value	Odds Ratio (Confidence Interval)
Outbred	*Bari-Jheh* (–) versus *Bari-Jheh* (+)	Females	**1.62×10^{-4}**	2.18 (1.46–3.32)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		0.285	—
	Bari-Jheh (–) versus *Bari-Jheh* (He)		**1.59×10^{-5}**	2.38 (1.59–3.57)
	Bari-Jheh (+) versus *Bari-Jheh* (He)		0.637	—
	Bari-Jheh (–) versus *Bari-Jheh* (+)	Males	**1.06×10^{-27}**	5.66 (3.50–9.17)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		0.433	—
	Bari-Jheh (–) versus *Bari-Jheh* (He)		**1.68×10^{-32}**	8.19 (4.81–13.91)
	Bari-Jheh (+) versus *Bari-Jheh* (He)		**0.0044**	1.44 (0.79–2.63)
Introgressed #1	*Bari-Jheh* (–) versus *Bari-Jheh* (+)	Females	**1.17×10^{-13}**	3.43 (2.53–4.65)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		0.124	—
	Bari-Jheh (–) versus *Bari-Jheh* (He)		**1.79×10^{-13}**	3.15 (2.33–4.26)
	Bari-Jheh (+) versus *Bari-Jheh* (He)		0.771	—
	Bari-Jheh (–) versus *Bari-Jheh* (+)	Males	**2.50×10^{-4}**	1.75 (1.32–2.32)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		**4.15×10^{-4}**	2.50 (1.64–3.79)
	Bari-Jheh (–) versus *Bari-Jheh* (He♂)		**2.84×10^{-15}**	4.29 (2.88–6.39)
	Bari-Jheh (+) versus *Bari-Jheh* (He♂)		**8.24×10^{-7}**	2.24 (1.54–3.27)
	Bari-Jheh (–) versus *Bari-Jheh* (He♀)		**7.66×10^{-6}**	2.03 (1.43–2.89)
	Bari-Jheh (+) versus *Bari-Jheh* (He♀)		0.194	—
Introgressed #2	*Bari-Jheh* (–) versus *Bari-Jheh* (+)	Females	**7.99×10^{-18}**	4.17 (2.66–6.54)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		**0.013**	1.35 (0.91–2.00)
	Bari-Jheh (–) versus *Bari-Jheh* (He)		**5.73×10^{-18}**	8.82 (5.14–15.11)
	Bari-Jheh (+) versus *Bari-Jheh* (He)		0.252	—
	Bari-Jheh (–) versus *Bari-Jheh* (+)	Males	**0.001**	3.04 (2.02–4.57)
	Bari-Jheh (He♂) versus *Bari-Jheh* (He♀)		**7.90×10^{-11}**	10.09 (5.56–18.3)
	Bari-Jheh (–) versus *Bari-Jheh* (He♂)		**0.003**	1.87 (1.25–2.78)
	Bari-Jheh (+) versus *Bari-Jheh* (He♂)		0.942	—
	Bari-Jheh (–) versus *Bari-Jheh* (He♀)		**4.61×10^{-23}**	18.89 (10.4–34.3)
	Bari-Jheh (+) versus *Bari-Jheh* (He♀)		**1.29×10^{-12}**	5.79 (3.67–9.14)

NOTE.—Nomenclature of the strains is the same as in figure 1. Significant *P* values after correcting for multiple testing are given in bold (Benjamini and Hochberg 1995).

were considered separately (*t*-test *P* value = 0.0164; supplementary fig. S1*B*, Supplementary Material online). Heterozygous males from one of the reciprocal crosses showed differences in the level of expression of *Jheh2* compared with the two homozygous strains (*t*-test *P* value = 0.0191 and 0.0211 compared with homozygous for the presence and for the absence, respectively; fig. 2*B*). These heterozygous males also showed differences in survival compared with homozygous flies with and without the insertion (fig. 1*C*). Heterozygous males from the other reciprocal cross only showed differences in expression compared with flies without the insertion (*t*-test *P* value = 0.0001; fig. 2*B*), which is also consistent with these heterozygous flies showing survival differences only with flies without the insertion (fig. 1*C*).

Bari-Jheh: A Case Study on the Dominance Effect of Adaptive Mutations

In this work, we found that the dominance effect of the adaptive transposable element insertion *Bari-Jheh* on oxidative stress resistance depended on the genetic background (fig. 1). The dominance effect of a mutation on a particular trait is influenced by environmental conditions and genetic background (Wool et al. 1982; Bourguet et al. 1996, 1997, 2000). Changes in dominance may arise because of alleles at linked or unlinked loci. This seems to be the case of *Bari-Jheh* mutation that is dominant in one of the backgrounds investigated (outbred populations; fig. 1*A*) and overdominant in males of one of the two reciprocal crosses in the other background (introgressed strains; fig. 1*B* and *C*). Our results highlight the complexity of the inheritance of adaptive mutations.

Our results add to the limited number of studies in which the dominance effect of adaptive mutations has been estimated. Previous empirical evidence focused on mutations conferring resistance to insecticides that most commonly occur through target-inactivation or metabolic detoxification (Ffrench-Constant 2013). *Bari-Jheh* mediates resistance to oxidative stress most likely through increase enzymatic activity of *JHEH2* (Taniai et al. 2003) as well as through changes in juvenile hormone titer (Campbell et al. 1992; Rauschenbach et al. 1996; Taniai et al. 2003; Flatt et al. 2005; Guio et al. 2014). As

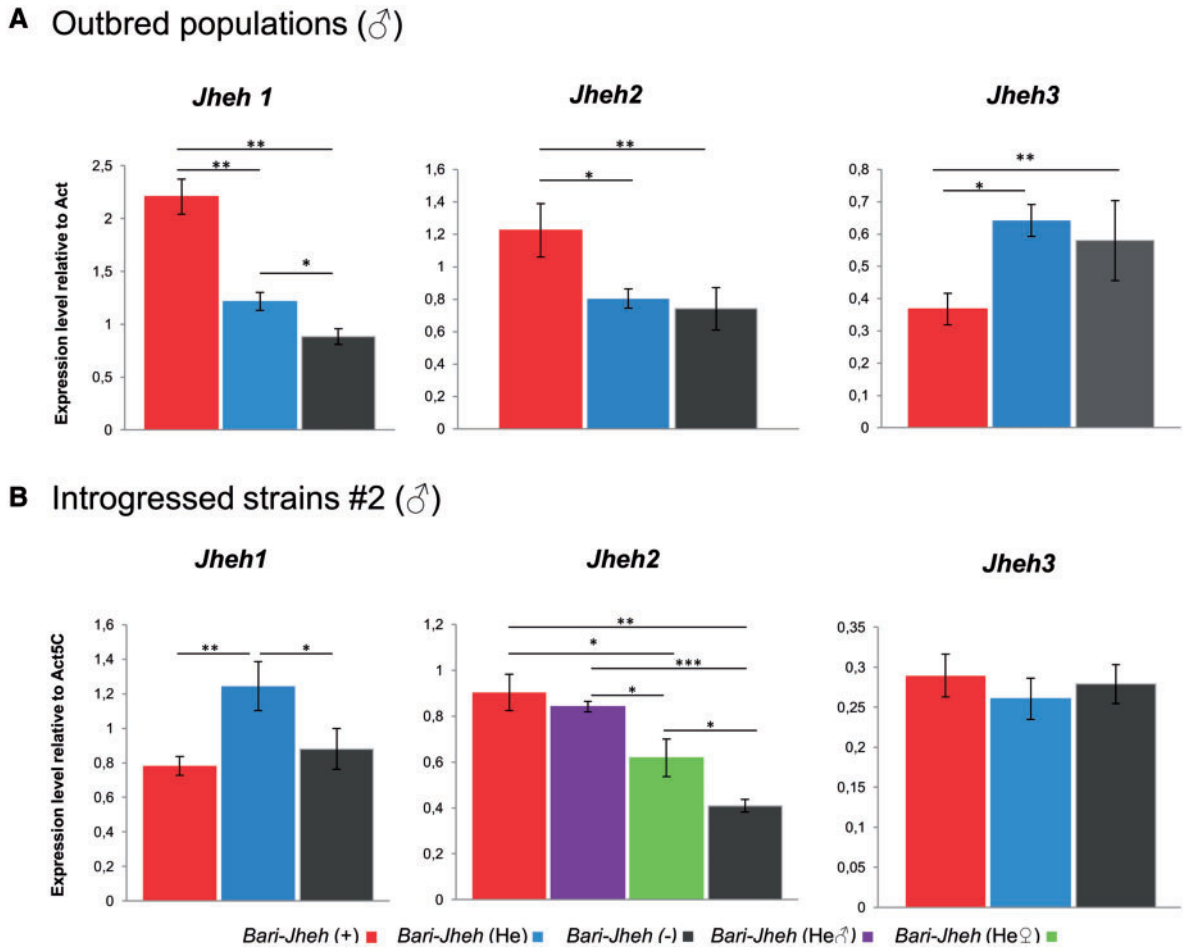


Fig. 2.—Expression level of *Jheh1*, *Jheh2*, and *Jheh3* genes in flies heterozygous for *Bari-Jheh* and in the two corresponding homozygous. Normalized expression level under oxidative stress conditions of male flies from outbred populations (*A*) and from introgressed strains #2 (*B*). In red, expression level of flies homozygous for *Bari-Jheh* insertion. In gray, expression level of flies homozygous for the absence. In blue, expression level of heterozygous flies from the two reciprocal crosses considered together. In purple, expression level of heterozygous flies from crosses in which the father carried *Bari-Jheh* insertion. In green, expression level of heterozygous flies from crosses in which the mother carried the insertion. Error bars indicate standard error of the mean based on the three biological replicas performed.

such, *Bari-Jheh* is an adaptive mutation with a more complex molecular mechanism and phenotypic effect than the other adaptive mutations previously characterized.

To try to shed light on the molecular mechanism behind the dominance effect of *Bari-Jheh*, we compared the expression of *Jheh* genes between heterozygous and homozygous flies. *Bari-Jheh* is associated with upregulation of *Jheh1* and/or *Jheh2* in homozygous flies and heterozygous flies suggesting that upregulation of one of these two genes may be enough to confer resistance to oxidative stress. Interestingly, heterozygous flies that showed overdominance differed in the level of expression of *Jheh2* compared with the two corresponding homozygous. Further experiments are needed in order to get a comprehensive understanding of the molecular mechanism of this adaptive insertion.

The scarcity of empirical studies testing the dominance effect of adaptive mutations is mostly due to the difficulty of identifying adaptive mutations and their fitness effects. However, the availability of technologies such as next generation sequencing has proven useful for the identification of adaptive mutations at an unprecedented scale (Turner et al. 2010; Jones et al. 2012). Future studies of a comprehensive set of adaptive mutations should help provide a more general view of the dominance effect of adaptive mutations.

Materials and Methods

Fly Stocks

Outbred Populations

Outbred populations were previously created in our laboratory (Guio et al. 2014). Briefly, we used flies from the Drosophila Genetics Reference Panel (Mackay et al. 2012) obtained from the Bloomington Stock Centre. We used lines RAL-21,

RAL-405, RAL-911, RAL-502, and RAL-138 to create an outbred population homozygous for the presence of *Bari-Jheh* element. We collected ten virgin females and ten males from each strain and we placed them in one large fly chamber. After the first generation, the siblings were randomly mated during ten generations before performing the experiments. The population size was ≈800 individuals per generation. We repeated the procedure with five strains homozygous for the absence of *Bari-Jheh* element to create an outbred population without this insertion: RAL-40, RAL-461, RAL-822, RAL-439, and RAL-908 (Guio et al. 2014).

Introgressed Strains

Introgressed strains were previously created in Dr Petrov laboratory at Stanford University (Gonzalez et al. 2009). Briefly, female flies with the element *Bari-Jheh* (*Wi3* strain) were crossed with males homozygous for the absence of the element *Bari-Jheh* (*Wi1* strain). Virgin females from F_1 were crossed with males from *Wi1* strain. F_2 virgin females were also crossed with Wi1 males and after egg laying females were analyzed for the presence of *Bari-Jheh* element. Only crosses in which females carried the element were kept to produce the next generation. The procedure was repeated up to eight generations. After eight generations sibling crosses were performed until homozygous strains were established for the presence and the absence of *Bari-Jheh* (Gonzalez et al. 2009). In this work, we used four different strains obtained after this procedure: Two pairs of strains with and without *Bari-Jheh*.

Heterozygous Strains

To create the heterozygous flies, we collected 100 virgin females homozygous for the presence of *Bari-Jheh* and we crossed them with 100 males homozygous for the absence of the element. We performed the crosses in large fly chambers. We kept the flies 72 h to ensure that females were inseminated and we collected eggs during an interval of 24 h. We repeated the same procedure with 100 females homozygous for the absence of the element *Bari-Jheh* and crossed them with 100 males homozygous for the presence of the element in a different chamber. We performed these reciprocal crosses for the outbred populations and for the introgressed strains.

We synchronized the egg laying period of the heterozygous crosses and the homozygous crosses so that the F_1 could be analyzed when all the flies were 5 days old.

Oxidative Stress Resistance Experiments

We used paraquat (methyl viologen dichloride hydrate; Sigma-Aldrich) as an oxidative stress agent. Paraquat is one of the most widely used herbicides in agricultural settings including tree plantation areas, a natural habitat for *D. melanogaster* (http://www.epa.gov).

To induce oxidative stress, we added paraquat to the regular fly food containing 4.5% (w/v) glucose, 6% (w/v) yeast, 0.7% (w/v) agar, and 3% (w/v) wheat flour. The final concentration of paraquat was 3 mM. For control conditions, we used regular fly food without paraquat (for more details, see Guio et al. 2014). For outbred populations, we analyzed 10 tubes containing 20 flies each, per sex, per strain, and per condition. For introgressed strains, we analyzed 20 tubes for homozygous strains and 10 tubes for each heterozygous cross. Survival was monitored every 24 h.

To analyze the data, we used logrank test. When differences between the strains being compared were significant, we estimated the size of the effect and its confidence intervals. When the differences between reciprocal crosses for heterozygous flies were not significant, we considered both crosses together.

Reverse Transcription Polymerase Chain Reaction Expression Analysis

We quantified the expression of *Jheh1*, *Jheh2*, and *Jheh3* in oxidative stress conditions. To induce oxidative stress, we exposed 5-day-old male flies to food containing 10 mM paraquat during 8 h. After the exposure, we freeze flies with liquid N_2. We purified total RNA using Trizol reagent and we synthesized cDNA using 1 μg of RNA after treatment with DNase. Then, we used the cDNA for quantitative polymerase chain reaction analysis using *Act5C* as a housekeeping gene. Expression assays were performed with three biological replicas. Results were analyzed using dCT method. Primers used were described in Guio et al. (2014)

Supplementary Material

Supplementary figure S1 is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Diamantis Sellis and Maite G. Barrón for helpful comments. L.G. is a FI/DGR fellow (2012FI-B-00676). J.G. is a *Ramón y Cajal* fellow (RYC-2010-07306). This work was supported by grants from the European Commission (Marie Curie CIG PCIG-GA-2011-293860) and from the Spanish Government (Fundamental Research Projects Grant BFU-2011-24397).

Literature Cited

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol. 57:289–300.

Bourguet D, et al. 1997. Variation of dominance of newly arisen adaptive genes. Genetics 147:1225–1234.

Bourguet D, Genissel A, Raymond M. 2000. Insecticide resistance and dominance levels. J Econ Entomol. 93:1588–1595.

Bourguet D, Prout M, Raymond M. 1996. Dominance of insecticide resistance presents a plastic response. Genetics 143:407–416.
Burke MK, Rose MR. 2009. Experimental evolution with *Drosophila*. Am J Physiol Regul Integr Comp Physiol. 296:R1847–R1854.
Campbell PM, Healey MJ, Oakeshott JG. 1992. Characterization of juvenile hormone esterasa in *Drosophila melanogaster*. Insect Biochem Mol Biol. 22:665–677.
Charlesworth B. 1998. Adaptive evolution: the struggle for dominance. Curr Biol. 8:R502–R504.
Charlesworth B, Charlesworth D. 1998. Some evolutionary consequences of deleterious mutations. Genetica 102–103:3–19.
Chavarrias D, Lopez-Fanjul C, Garcia-Dorado A. 2001. The rate of mutation and the homozygous and heterozygous mutational effects for competitive viability: a long-term experiment with *Drosophila melanogaster*. Genetics 158:681–693.
Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. Mol Biol Evol. 24:2081–2090.
Coop G, et al. 2009. The role of geography in human adaptation. PLoS Genet. 5:e1000500.
Ffrench-Constant RH. 2013. The molecular genetics of insecticide resistance. Genetics 194:807–815.
Flatt T, Tu MP, Tatar M. 2005. Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. Bioessays 27:999–1010.
Fry JD, Nuzhdin SV. 2003. Dominance of mutations affecting viability in *Drosophila melanogaster*. Genetics 163:1357–1364.
Garcia-Dorado A, Lopez-Fanjul C, Caballero A. 1999. Properties of spontaneous mutations affecting quantitative traits. Genet Res. 74: 341–350.
Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165: 1269–1278.
Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. PLoS Biol. 6:e251.
Gonzalez J, Macpherson JM, Petrov DA. 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. Mol Biol Evol. 26: 1949–1961.
Guio L, Barron MG, Gonzalez J. 2014. The transposable element Bari-Jheh mediates oxidative stress response in *Drosophila*. Mol Ecol. 23: 2020–2030.
Haldane JBS. 1927. A mathematical theory of natural and artificial selection. Pt. V. Selection and mutation. Math Proc Camb Philos Soc. 23: 833–844.
Houle D, Hughes KA, Assimacopoulos S, Charlesworth B. 1997. The effects of spontaneous mutation on quantitative traits. II. Dominance of mutations with effects on life-history traits. Genet Res. 70:27–34.
Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.
Joseph SB, Peck KM, Burch CL. 2014. Dominance effects of deleterious and beneficial mutations in a single gene of the RNA virus varphi6. PLoS One 9:e97717.
Lynch M, et al. 1999. Perspective: spontaneous deleterious mutations. Evolution 53:645–663.
Mackay TF, et al. 2012. The *Drosophila melanogaster* genetic reference panel. Nature 482:173–178.
Manna F, Martin G, Lenormand T. 2011. Fitness landscapes: an alternative theory for the dominance of mutation. Genetics 189:923–937.
Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 28:659–669.
Mukai T, Yoshikawa I, Sano K. 1966. The genetic structure of natural populations of *Drosophila melanogaster*. IV. Heterozygous effects of radiation-induced mutations on viability in various genetic backgrounds. Genetics 53:513–527.
Orr HA. 2010. The population genetics of beneficial mutations. Philos Trans R Soc Lond B Biol Sci. 365:1195–1201.
Rauschenbach IY, et al. 1996. The role of degradation system of the Juvenile hormone in the reproduction of *Drosophila* under stress. J Insect Physiol. 42:735–742.
Sabeti PC, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837.
Sellis D, Callahan BJ, Petrov DA, Messer PW. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. Proc Natl Acad Sci U S A. 108:20666–20671.
Simmons MJ, Crow JF. 1977. Mutations affecting fitness in *Drosophila* populations. Annu Rev Genet. 11:49–78.
Taniai K, Inceoglu AB, Yukuhiro K, Hammock BD. 2003. Characterization and cDNA cloning of a clofibrate-inducible microsomal epoxide hydrolase in *Drosophila melanogaster*. Eur J Biochem. 270: 4696–4705.
Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat Genet. 42:260–263.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.
Wilkie AO. 1994. The molecular basis of genetic dominance. J Med Genet. 31:89–98.
Wool D, Noiman S, Manheim O, Cohen E. 1982. Malathion resistance in *Tribolium* strains and their hybrids: inheritance patterns and possible enzymatic mechanisms (Coleoptera, Tenebrionidae). Biochem Genet. 20:621–636.
Zhang M, Azad P, Woodruff RC. 2011. Adaptation of *Drosophila melanogaster* to increased NaCl concentration due to dominant beneficial mutations. Genetica 139:177–186.

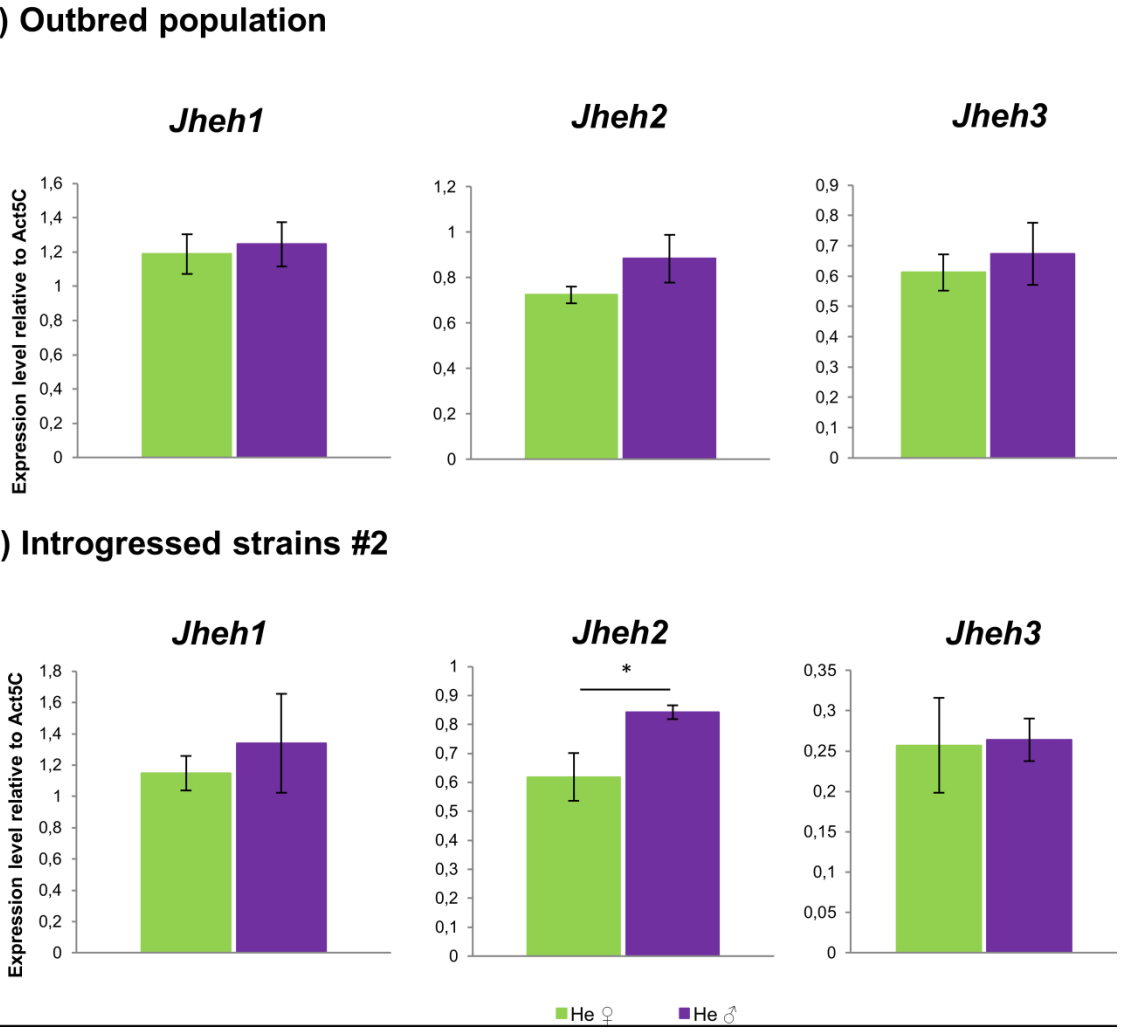
Associate editor: Sarah Schaack

Genome Biology and Evolution SMBE

Supplementary Material

Figure S1. Expression levels of *Jheh1*, *Jheh2* and *Jheh3* in heterozygous males exposed to oxidative stress conditions

Normalized expression level of heterozygous male flies from outbred (A) and introgressed strains (B). In purple, expression level of heterozygous flies from crosses in which the father carried *Bari-Jheh* insertion. In green, expression level of heterozygous flies from crosses in which the mother carried the insertion. Error bars indicate standard error of the mean (SEM) based on three biological replicas performed.



6 Stress affects the epigenetic marks added by *Bari-Jheh:* a natural insertion associated with two adaptive phenotypes in Drosophila

Stress affects the epigenetic marks added by *Bari Jheh*: a natural insertion associated with two adaptive phenotypes in Drosophila

Resumen

Los elementos móviles se están revelando como una importante fuente de secuencias regulatorias y marcas epigenéticas que podrían afectar la expresión génica. Sin embargo, muy pocos estudios han analizado el papel de elementos móviles específicos en la regulación génica. *Bari-Jheh* es un transposón presente en la naturaleza que está involucrado en estrés oxidativo añadiendo secuencias reguladoras. En este trabajo, integramos datos públicos disponibles junto con análisis de inmunoprecipitación de la cromatina y ensayos de respuesta inmune para obtener una comprensión más exhaustiva de los efectos moleculares y funcionales de *Bari-Jheh*. Mostramos que Bari-Jheh está asociado con el enriquecimiento de H3K27me3 que coincide con cambios en la expresión de los genes aledaños. Además, mostramos que el estrés afecta las marcas de histonas introducidas por *Bari-Jheh*, lo cual se correlaciona con cambios adicionales en la expresión. Finalmente, encontramos que las moscas con *Bari-Jheh*, resistentes al estrés oxidativo, son también más tolerantes a la infección bacteriana. Concluimos así que *Bari-Jheh* influye sobre la expresión de los genes aledaños y media en la respuesta a estrés mediante dos diferentes mecanismos: añadiendo secuencias reguladoras y añadiendo marcas de histonas que conducen a cambios en dos fenotipos ecológicamente relevantes.

Stress affects the epigenetic marks added by *Bari-Jheh*: a natural insertion associated with two adaptive phenotypes in Drosophila

Lain Guio[1] Cristina Vieira[2] and Josefa González[1]*

This paper was submmited to Molecular Biology and Evolution

[1]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain

[2]Universite´ de Lyon, Universite´ de Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 bd 11 novembre 1918, F-69622 Villeurbanne cedex, France

*Corresponding author: E-mail: josefa.gonzalez@ibe.upf-csic.es.

Abstract

Transposable elements are emerging as an important source of cis-acting regulatory sequences and epigenetic marks that could influence gene expression. However, few studies have dissected the role of specific transposable element insertions on gene regulation. *Bari-Jheh* is a natural transposon that mediates resistance to oxidative stress by adding cis-regulatory sequences. In this work, we integrated publicly available data with chromatin immunoprecipitation and immune response assays to get a more comprehensive picture of *Bari-Jheh* molecular and functional effects. We showed that *Bari-Jheh* is associated with H3K27me3 enrichment, which is consistent with expression changes in adjacent genes. We further showed that stress affects the histone marks introduced by *Bari-Jheh*, which correlates with further expression changes. Finally, we found that flies with *Bari-Jheh*, which are resistant to oxidative stress, are also more tolerant to bacterial infection. We conclude that *Bari-Jheh* influences gene expression and enables stress response through two different mechanisms, by adding cis-regulatory sequences and by adding histone marks, leading to changes in two ecologically relevant phenotypes.

Keywords: gene expression, histone modifications, ChIP-qPCR, transposable elements, oxidative stress.

Introduction

Gene regulation is an extremely complex process that involves mechanisms at the DNA sequence level and at the epigenetic level. Although genes can acquire novel regulatory mechanisms through different types of mutations, transposable elements (TEs) are emerging as an important source of regulatory variation (Slotkin and Martienssen 2007; Cowley and Oakey 2013). TEs can contain cis-regulatory sequences that affect the expression of nearby genes. Some of the recent examples on the global impact of TEs on gene expression levels include: providing enhancer sequences that contribute to the stress-induced gene activation in maize, adding transcription factor binding sites in the mouse and the human genomes, and providing alternative transcription start sites in Drosophila (Batut, et al. 2013; Sundaram, et al. 2014; Makarevitch, et al. 2015). The epigenetic status of TEs can also affect gene regulation. In *Arabidopsis thaliana*, gene transcription is affected by the methylation status of intragenic TEs (Le, et al. 2015) and correlates with siRNA-targeting of TEs (Wang, et al. 2013). In Drosophila, local spreading of repressive heterochromatin marks from TEs has been associated with gene downregulation (Sentmanat and Elgin 2012; Lee 2015). Although all these studies strongly suggest that TEs may play a role in gene regulation through different molecular mechanisms, detailed analyses that link changes in

expression with fitness effects are needed to conclude that TEs have a functional impact on gene expression.

There are a few examples in which TE-induced changes in gene expression have been shown to be functionally relevant (McCue, et al. 2012; Guio, et al. 2014; Mateo, et al. 2014). One of these cases is *Bari-Jheh*, a *Drosophila melanogaster* full-length transposon providing a cis-regulatory sequence that affects the expression of its nearby genes (Gonzalez, et al. 2008; Gonzalez, et al. 2009). *Bari-Jheh* is associated with downregulation of *Jheh2* and *Jheh3* in nonstress conditions, and with upregulation of *Jheh1* and *Jheh2* and downregulation of *Jheh3* under oxidative stress conditions (Guio, et al. 2014). We have previously shown that *Bari-Jheh* adds Antioxidant Response Elements to the upstream region of *Jheh2* leading to *Jheh2* upregulation under oxidative stress conditions (Guio, et al. 2014; Guio and Gonzalez 2015). How *Bari-Jheh* affects gene expression under nonstress conditions, and how *Bari-Jheh* affects *Jheh3* expression under oxidative stress conditions remains unexplored. In this work, we hypothesized that *Bari-Jheh* could also be affecting the expression of nearby genes by remodeling the local chromatin state. Moreover, because oxidative stress is also caused by Gram-negative bacterial infection, we tested whether flies with *Bari-Jheh* are also more tolerant to bacterial infection (Lemaitre and Hoffmann 2007; Bou Sleiman, et al. 2015).

Results

Bari-Jheh could be affecting the local chromatin

To test whether *Bari-Jheh* could be affecting the local chromatin state, we analyzed the transposon sequence and its flanking regions (Figure 1). We first looked for Trithorax group Response Elements (TREs) that recruit H3K4 methyltransferases, and Polycomb group Response Elements (PREs) that recruit H3K27 methyltransferases (see Material and Methods) (Greer and Shi 2012; Schwartz and Pirrotta 2013). While H3K4me3 is associated with active chromatin, H3K27me3 is associated with facultative heterochromatin.

We found no TREs in the sequence analyzed, but we found one PRE in the *Bari-Jheh* sequence and one PRE in the coding region of *Jheh3*, where modENCODE reports a Polycomb mediated repressive chromatin state (Figure 1A and 1B) (Ringrose, et al. 2003; modENCODE, et al. 2010; Schwartz and Pirrotta 2013).

To further test whether *Bari-Jheh* affects the local heterochromatin state, we also investigated whether *Bari-Jheh* has piRNA binding sites and/or recruits HP1a (see Material and Methods). Sites with homology to piRNAs behave as cis-acting targets for heterochromatin assembly, which is associated with HP1a and H3K9me2/3 (Sentmanat and Elgin 2012). We found that *Bari-Jheh* has sites with homology to piRNAs (Figure 1C). Accordingly, we also found that HP1a specifically binds to the *Bari-Jheh* sequence (Fig. 1D).

Thus, *Bari-Jheh* could be introducing PREs that would be involved in the recruitment of H3K27 methyltransferase enzyme. Additionally, *Bari-Jheh* could also be inducing pi-RNA mediated heterochromatin assembly. These results provide suggestive but not conclusive evidence that *Bari-Jheh* could be introducing heterochromatin histone marks.

Bari-Jheh adds the heterochromatin mark H3K27me3 in nonstress conditions

To experimentally test whether *Bari-Jheh* affects histone marks enrichment, we performed Chromatin Immune Precipitation (ChIP)-qPCR experiments in guts of adult flies with H3K4me3, H3K9me3 and H3K27me3 antibodies (supplementary Figure S1) (see Material and Methods). We compared the histone mark enrichment in both sides of *Bari-Jheh* insertion, *Bari-Jheh2* and *Bari-Jheh3* regions, with the corresponding region in flies without *Bari-Jheh*, *Bari-Absent* region (Figure 1A).

We found no significant differences in H3K4me3 or H3K9me3 enrichment between the strain with and without *Bari-Jheh* (Table 1 and Figure 2A). For H3K27me3, we found significant differences for the *Bari-Jheh3* region but not for the *Bari-Jheh2* region (Table 1 and Figure 2A). Thus, *Bari-Jheh* is associated with the facultative heterochromatin mark H3K27me3 in the *Bari-Jheh3* region.

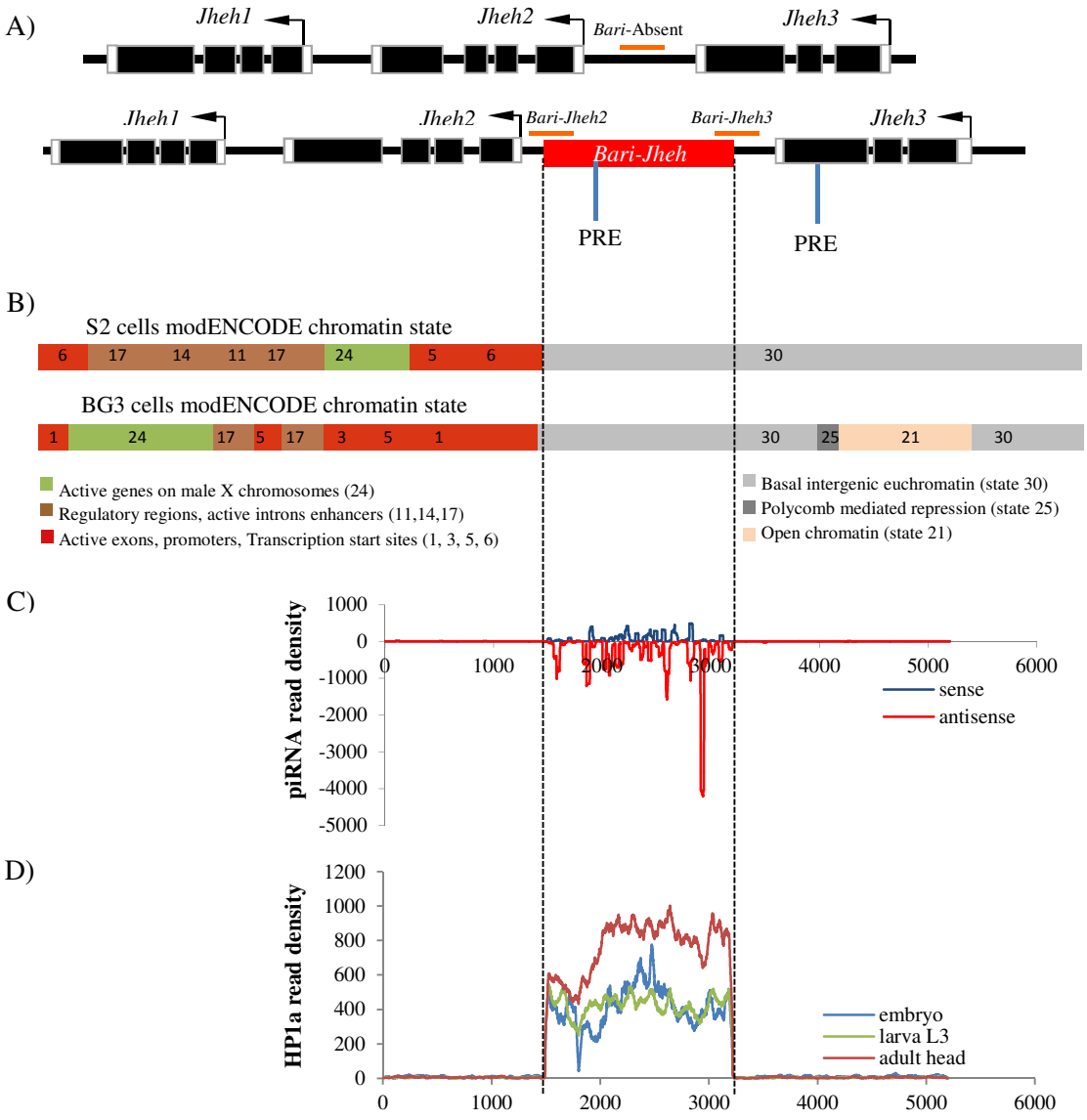


Figure 1. *Bari-Jheh* could be adding heterochromatin marks to the *Jheh* intergenic region.

A) Schematic representation of *Jheh* genes in flies without *Bari-Jheh* and flies with *Bari-Jheh*. Black boxes represent exons, black arrows represent the direction of transcription, white boxes the 5'-UTR and 3'-UTR regions, the black line indicates intergenic or intronic regions and the red box represents Bari-Jheh. Orange lines represent the amplicons of the three regions analyzed using ChIP-qPCR experiments. The blue lines indicate the approximated position of the predicted PREs. B) modENCODE chromatin states in S2 cells and BG3 cells in the region analyzed. S2 cells and BG3 cells are derived from late male embryonic tissues and the central nervous system of male third instar larvae, respectively (modENCODE, et al. 2010). Colours and numbers represent different chromatin states. Note that although *Bari-Jheh* appears to be associated with state 30, modENCODE does not analyzed repetitive regions. The vertical discontinuous lines indicate the location of *Bari-Jheh* insertion. C) Mapping of piRNA reads in the *Bari-Jheh* and 1.5 kb flanking region. Reads mapping in sense orientation are represented in blue, and reads mapping in antisense orientation in red. D) Mapping of HP1a reads in the *Bari-Jheh* and 1.5 kb flanking regions. Reads from embryo stage are represented in blue, reads from larva L3 stage in green and reads from adult head in red.

Table 1. Statistical analyses of histone mark enrichment. Significant values are highlighted in bold.

Condition	Regions compared	Mann-Whitney U-test p-value		
		H3K4me3	H3K9me3	H3K27me3
Nonstress	*Bari-Jheh2 vs Bari-Absent*	0.827	0.827	0.127
	Bari-Jeh3 vs Bari-Absent	0.957	0.927	**0.0495**
Oxidative stress	*Bari-Jheh2 vs Bari-Absent*	0.286	**0.033**	0.088
	Bari-Jheh3 vs Bari-Absent	0.507	**0.031**	**0.033**

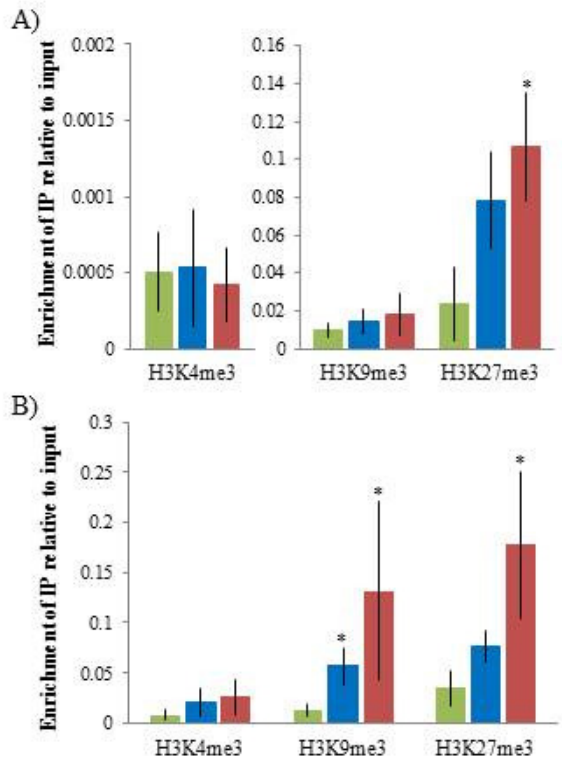


Figure2. Histone mark enrichment in Bari-absent, *Bari-Jheh2* and *Bari-Jheh3* regions.

Enrichment of the histone marks relative to the input of each strain, in A) nonstress conditions and B) oxidative stress conditions. Levels of H3k4me3, H3K9me3 and H3K27me3 in the *Bari*-Absent (green), *Bari-Jheh2* (blue) and *Bari-Jheh3* (red) analyzed regions. Bars give the mean of three to six biological replicas (± SEM). Significant differences between regions are mark with an asterisk (p-value<0.05). C) Fold-enrichment in oxidative stress vs nonstress conditions in the three regions analyzed. The red horizontal line indicates the expected value if there is no enrichment. Significant enrichments are marked with an asterisk.

Bari-Jheh adds the heterochromatin mark H3K9me3 in oxidative stress conditions

To further test whether oxidative stress affects the heterochromatin marks added by *Bari-Jheh*, we performed ChIP-qPCR experiments in flies exposed to paraquat (see Material Methods). We found no significant differences for H3K4me3 between flies with and without *Bari-Jheh* (Table 1 and Figure 2B). H3K9me3 was enriched in *Bari-Jheh2* and *Bari-Jheh3* regions compared with the Bari-Absent region (Table 1 and Figure 2B). Finally, H3K27me3 is only enriched in the Bari-*Jheh3* (Table 1 and Figure 2B).

Overall, these results showed that under oxidative stress conditions *Bari-Jheh* is associated with the constitutive heterochromatin mark H3K9me3 on both sides of the insertion and the facultative heterochromatin mark H3K27me3 only in the *Bari-Jheh3* region.

Flies with *Bari-Jheh* are associated with increased oral infection tolerance

To test whether flies with *Bari-Jheh* were more gut immunocompetent than flies without *Bari-Jheh*, we performed an oral infection experiment using *Pseudomonas entomophila* (Vodovar, et al. 2005). We used flies with three different genetic backgrounds (see Material and Methods). In outbred populations #1 and outbred populations #2, we found that both female and male flies with the insertion were more tolerant to *P. entomophila* infection than flies without the insertion (Figure 3A and 3B, respectively, and Table 2). However, in introgressed strains we found no significant differences in females, while males with *Bari-Jheh* were more sensitive to oral infection than males

without the insertion (Figure 3C and Table 2). Thus, *Bari-Jheh* is associated with tolerance to oral infection with *P. entomophila* in two different outbred populations but not in introgressed strains.

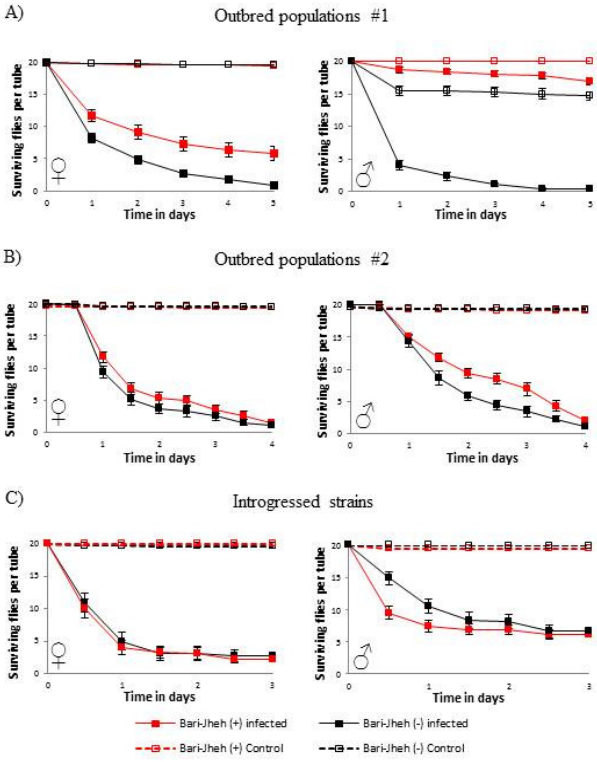


Figure 3. Survival curves after *P. entomophila* infection in female and male flies with and without *Bari-Jheh*.

A) Survival curves for outbred populations #1, B) Survival curves for outbred populations #2, and C) Survival curves for introgressed strains. Survival curves in nonstress conditions are represented as discontinuous lines and survival curves after *P. entomophila* infection are represented as continuous lines. Survival of *Bari-Jheh* flies are represented in red and survival of flies without *Bari-Jheh* are represented in black. Each point represents the mean survival of 10 replicas and the error bars represent the standard error of the mean.

Discussion

In the present work, we combined different sources of information to analyze whether *Bari-Jheh* insertion could be affecting the local chromatin state. We found evidence suggesting that *Bari-Jheh* could be associated both with H3K27 and H3K9 chromatin marks (Figure 1). We tested these predictions by performing ChIP experiments in adult flies, and we found that in nonstress conditions *Bari-Jheh* is associated with H3K27me3 histone mark enrichment (Figure 2A). Previous analyses have shown that different TE families are associated with specific histone marks (Rebollo, et al. 2012). Enrichment for H3K27me3 has previously been reported for the roo family, while other families were enriched both for H3K27me3 and H3K9me2 (Fablet, et al. 2009; Rebollo, et al. 2012; Akkouche, et al. 2013). The enrichment for heterochromatin histone marks is one of the epigenetic silencing mechanisms used to control the activity of TEs (Levin and Moran 2011; Gonzalez and Petrov 2012). However, heterochromatin formation triggered by TEs can also spread into the nearby genes affecting their expression (Sentmanat and Elgin 2012; Lee 2015). Accordingly, we found that *Bari-Jheh* that is associated with H3K27me3 facultative heterochromatin mark is also associated with downregulation of the expression of *Jheh2* and *Jheh3* genes in nonstress conditions (Gonzalez, et al. 2009; Guio, et al. 2014).

Oxidative stress has been associated with increases in several histone methylation marks (Niu, et al. 2015). Consistent with these results, we found that under oxidative stress conditions, besides H3K27me3 enrichment in the *Bari-Jheh3* region, *Bari-Jheh* insertion is associated with H3K9me3 enrichment both in *Bari-Jheh2* and in *Bari-Jheh3* regions (Figure 2B). Although H3K9me3 is often associated with gene silencing, there is evidence suggesting that this histone mark is also associated with gene activation (Vakoc, et al. 2005; Kouzarides 2007). Additionally, we have previously shown that *Bari-Jheh* adds Antioxidant Response Elements (AREs), which are necessary and sufficient to induce the upregulation of downstream genes (Sykiotis and Bohmann 2008; Chatterjee and Bohmann 2012; Guio, et al. 2014). Thus, *Bari-Jheh* could be affecting *Jheh2* expression under oxidative stress conditions both because it adds AREs and because it affects the local chromatin state. Our results also provide a mechanistic explanation for the downregulation of *Jheh3*: *Bari-Jheh* is associated with enrichment of both H3K9me3 and H3K27me3 in the 3' end of this gene. Combined histone marks can have

different roles compared with the same histone marks appearing in isolation (Greer and Shi 2012; Lelli, et al. 2012). Additionally, the effects of histone modifications also depend on the relative position of the histone mark regarding the functional sequence (Vakoc, et al. 2005; Kouzarides 2007; Greer and Shi 2012). Thus it is possible that the combination of these two histone marks in the 3' region of *Jheh3* would lead to downregulation of this gene (Greer and Shi 2012, Lille et al 2012).

Besides elucidating that *Bari-Jheh* could also be affecting gene expression by changing the local chromatin state, we showed that *Bari-Jheh* is associated with increased tolerance to *P. entomophila* infection in two of the three backgrounds analyzed (Figure 3 and Table 1). These results are consistent with previous analysis showing that inbred strains tolerant to *P. entomophila* infection are more resistant to paraquat while inbred strains susceptible to *P. entomophila* could be resistant to paraquat (Bou Sleiman, et al. 2015). Variation in tolerance/susceptibility phenotypes among genetic backgrounds likely reflects the complex nature of immune response, which results from the interplay of many biological processes and it is highly dependent on environmental conditions (Lemaitre and Hoffmann 2007; Lazzaro and Little 2009).

Overall, our results provide further evidence for the complex effects of natural TE insertions on gene regulation and organismal phenotypes. A single mutation, influences gene expression through two different regulatory mechanisms and has fitness consequences on two relevant phenotypes: oxidative stress and immune response.

Table 2. Statistical analyses of the *P .entomophila* infection survival experiments.

Genetic background	Compared Strains	Sex	Logrank test p-value	Odds-ratio (confidence interval)
Outbred #1	*Bari-Jheh* (–) *vs Bari-Jheh* (+)	Females	**<<0.0001**	2.03 (1.36 – 3.02)
	Bari-Jheh (–) *vs Bari-Jheh* (+)	Males	**<<0.0001**	19.46 (10.38 – 36.47)
Outbred #2	*Bari-Jheh* (–) *vs Bari-Jheh* (+)	Females	**0.028**	1.59 (1.07 – 2.36)
	Bari-Jheh (–) *vs Bari-Jheh* (+)	Males	**$3.6x10^{-4}$**	1.91 (1.28 – 2.84)
Introgressed	*Bari-Jheh* (–) *vs Bari-Jheh* (+)	Females	**0.262**	--
	Bari-Jheh (–) *vs Bari-Jheh* (+)	Males	**0.044**	3.23 (2.12 – 4.92)

Materials and Methods

Fly stocks.

We used the outbred populations and introgressed strains described in Guio et al (2014) and a new outbred population created for this work (Supplementary Table S1). All flies were kept in large embryo collection chambers as described in Guio et al (2014).

Oxidative stress exposure.

To induce oxidative stress, we added paraquat to the fly food up to a final concentration of 10 mM. For nonstress conditions, we used regular food. We did three to six replicas, of 50 females each, for each condition and genotype.

Chromatin Immunoprecipitation assays

We performed ChIP assays in guts because the gut is the first barrier against oxidative stress. Guts of 5-day-old females were dissected in 1x PBS with protease inhibitor cocktail. Chromatin immunoprecipitation was performed as described in Silva-Sousa, et al. (2012) with the following changes: guts were resuspended in 2ml buffer A with 1,8% formaldehyde, and sonication was performed with 15 cycles of 30 seconds ON, 30 seconds OFF. All the solutions were made according to the instructions of Magna ChIP G Kit 17-611 from Millipore. Chromatin was immunoprecipitated (IP) with antibodies against H3K4me3 (Catalog # ab8580), H3K9me3 (#ab8898) and H3K27me3 (#ab6002). All the antibodies were ChIP grade and antibody quality was tested before performing the experiments (supplementary Figure S1). We did 3 to 6 biological replicas for each IP.

We quantify the IP enrichment by qRT-PCR normalizing the data using the "input" of each IP as the reference value (ΔCt method, supplementary Table S2). Data was normalized using log transformation before performing ANOVA.

Prediction of PREs and TREs.

We used the database JASPAR (Mathelier, et al. 2015) with 95% threshold to predict the presence of PREs/TREs in the region analyzed: chromosome 2R: 18.856.800-18.861.999 (dos Santos, et al. 2015).

Detection of piRNA reads.

To search for piRNA homology sites in *Bari-Jheh*, we used reads from available piRNA libraries (Li, et al. 2009; Satyaki, et al. 2014) and we followed the methodology described in Ullastres, et al. (2015).

Detection of HP1a binding sites.

To analyze the binding sited for HP1a in the *Bari-Jheh* region we used HP1a modENCODE ChIP-Seq data (Kharchenko, et al. 2011) and we followed the methodology described in Ullastres et al (2015).

Oral infection assays.

We used *Pseudomonas entomophila* to perform oral infection assays (courtesy of Dr. Bruno Lemaitre laboratory) following Neyen, et al. (2014) protocol. We used 10 vials per sex, per strain, and per condition. Before infection, the flies were starved during two hours. The size of the effect (odds-ratio) was measured when the weaker strain arrived at 50% mortality.

Acknowledgments

We thank the Equipe Eléments transposables, Evolution, Populations for their support to L.G. We thank Elena Casacuberta for sharing antibodies and technical advice; Jon Frias for helping to create the new outbred populations, and Anna Ullastres for help with the piRNA and HP1a binding site analyses. We also thank Bruno Lemaitre laboratory for sharing the *P. entomophila* strain. Finally, we thank E. Casacuberta and members of González lab for comments on the manuscript.

This work was supported by the Ministerio de Economia y Competitividad (grant numbers BFU-2011-24397 and RYC-2010-07306 to J.G.); Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (grant number 2014 SGR 201 to J.G. and 2012FI-B-00676 to L.G.); European Commission (grant number PEOPLE-2011-CIG-293860 to J.G.).

References

Akkouche A, Grentzinger T, Fablet M, Armenise C, Burlet N, Braman V, Chambeyron S, Vieira C. 2013. Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. EMBO Rep 14:458-464.

Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. 2013. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res 23:169-180.

Bou Sleiman MS, Osman D, Massouras A, Hoffmann AA, Lemaitre B, Deplancke B. 2015. Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. Nat Commun 6:7829.

Cowley M, Oakey RJ. 2013. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. PLoS Genet 9 (1) e1003234.

Chatterjee N, Bohmann D. 2012. A versatile PhiC31 based reporter system for measuring AP-1 and Nrf2 signaling in Drosophila and in tissue culture. PLoS One 7:e34063.

dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase C. 2015. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43:D690-697.

Fablet M, Lerat E, Rebollo R, Horard B, Burlet N, Martinez S, Brasset E, Gilson E, Vaury C, Vieira C. 2009. Genomic environment influences the dynamics of the tirant LTR retrotransposon in Drosophila. FASEB J 23:1482-1489.

Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. PLoS Biol 6:e251.

Gonzalez J, Macpherson JM, Petrov DA. 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. Mol Biol Evol 26:1949-1961.

Gonzalez J, Petrov DA. 2012. Evolution of genome content: population dynamics of transposable elements in flies and humans. Methods Mol Biol 855:361-383.

Greer EL, Shi Y. 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13:343-357.

Guio L, Barron MG, Gonzalez J. 2014. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Mol Ecol 23:2020-2030.

Guio L, Gonzalez J. 2015. The dominance effect of the adaptive transposable element insertion Bari-Jheh depends on the genetic background. Genome Biol Evol 7:1260-1266.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature 471:480-485.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128:693-705.

Lazzaro BP, Little TJ. 2009. Immunity in a variable world. Philos Trans R Soc Lond B Biol Sci 364:15-26.

Le TN, Miyazaki Y, Takuno S, Saze H. 2015. Epigenetic regulation of intragenic transposable elements impacts gene transcription in Arabidopsis thaliana. Nucleic Acids Res 43:3911-3921.

Lee YC. 2015. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in *Drosophila melanogaster*. PLoS Genet 11:e1005269.

Lelli KM, Slattery M, Mann RS. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet 46:43-68.

Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. Annu Rev Immunol 25:697-743.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615-627.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137:509-521.

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. 2015. Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. PLoS Genet 11.

Mateo L, Ullastres A, Gonzalez J. 2014. A transposable element insertion confers xenobiotic resistance in Drosophila. PLoS Genet 10:e1004560.

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, et al. 2015. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res.

McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. 2012. Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet 8:e1002474.

modENCODE C, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330:1787-1797.

Neyen C, Bretscher AJ, Binggeli O, Lemaitre B. 2014. Methods to study Drosophila immunity. Methods 68:116-128.

Niu Y, DesMarais TL, Tong Z, Yao Y, Costa M. 2015. Oxidative stress alters global histone modification and DNA methylation. Free Radic Biol Med 82:22-28.

Rebollo R, Horard B, Begeot F, Delattre M, Gilson E, Vieira C. 2012. A snapshot of histone modifications within transposable elements in Drosophila wild type strains. PLoS One 7:e44253.

Ringrose L, Rehmsmeier M, Dura JM, Paro R. 2003. Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. Dev Cell 5:759-771.

Satyaki PR, Cuykendall TN, Wei KH, Brideau NJ, Kwak H, Aruna S, Ferree PM, Ji S, Barbash DA. 2014. The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet 10:e1004240.

Schwartz YB, Pirrotta V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. Nat Rev Genet 14:853-864.

Sentmanat M, Elgin SC. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements Proc Natl Acad Sci USA 109:14104-14109.

Silva-Sousa R, Lopez-Panades E, Pineyro D, Casacuberta E. 2012. The chromosomal proteins JIL-1 and Z4/Putzig regulate the telomeric chromatin in *Drosophila melanogaster*. PLoS Genet 8:e1003153.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272-285.

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res 24:1963-1976.

Sykiotis GP, Bohmann D. 2008. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. Dev Cell 14:76-85.

Ullastres A, Petit N, Gonzalez J. 2015. Exploring the Phenotypic Space and the Evolutionary History of a Natural Mutation in *Drosophila melanogaster*. Mol Biol Evol 32:1800-1814.

Vakoc CR, Mandat SA, Olenchock BA, Blobel GA. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell 19:381-391.

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard F, Lemaitre B. 2005. Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. Proc Natl Acad Sci U S A 102:11414-11419.

Wang X, Weigel D, Smith LM. 2013. Transposon variants and their effects on gene expression in Arabidopsis. PLoS Genet 9:e1003255.

Supplementary Material

Supplementary Table 1: List of DGRP lines used to create Outbred populations #2
All the lines used to create the outbred populations were obtained from The Drosophila Genetic Reference Panel project (DGRP) (Mackay et al. 2012).

Lines with *Bari-Jheh*	**Lines without *Bari-Jheh***
RAL-441	RAL-177
RAL-88	RAL-383
RAL-820	RAL-857
RAL-716	RAL-776
RAL-391	RAL-802
RAL-371	RAL-783
RAL-195	RAL-737

Supplementary Table 2: List of primers used in this study.

Note 1: We did not use the positive control values to normalize the results of the qPCR. We only normalized the IPs enrichment with the values of the inputs.

Note 2: All efficiencies were analyzed to validate the quality of the primers.

Gene	Primer	Sequence
Rp49 positive control for H3K4me3 (Rebollo et al. 2012)	*rp49* F	cggatcgatatgctaagctgt
	rp49 R	gcgcttgttcgatccgta
18S positive control for H3K9me3 (Herz et al. 2010)	*18S* F	tttcatgcttgggattgtga
	18S R	gtacaaagggcagggacgta
Ultrabithorax (Ubx) positive control for H3K9me3 (Reddington et al 2013)	*Ubx* F	gaggcctgttcaaagtacgagt
	Ubx R	ggaaaccaattcgtgtgaaatc
Bari-Jheh2	B-J2 Fw	aacacgcggagcccttaata
	B-J2 Rv	atttcaggcaaatcggacaa
Bari-Jheh3	B-J3 Fw	caactgtttgtgacccatgc
	B-J3 Rv	tgtttgtaattgaccgcaaa
Bari-Absent	B-J2 Fw	aacacgcggagcccttaata
	B-J3 Rv	tgtttgtaattgaccgcaaa

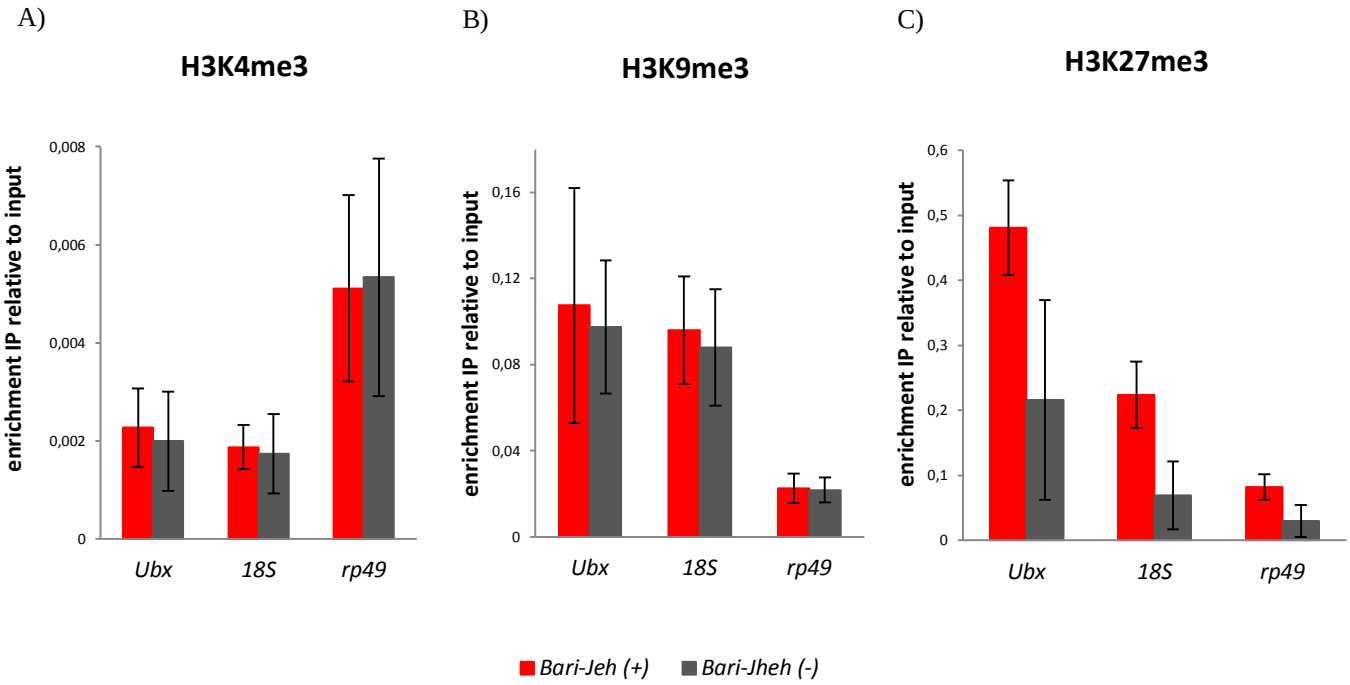


Figure Sup. 1: Positive controls of the ChIP-qPCR experiments for the three histone marks analyzed

To confirm the correct activity of the antibodies we analysed the enrichment in known targets of the three histone marks. A) results for H3K4me3, which is enriched in *rp49* gene as expected, B) results for H3K9me3, which is enriched in *18S* and *Ubx* genes as expected and C) results for H3K27me3, which is enriched in *Ubx* gene as expected. Bar represent the mean of 3 to 6 biological replicates (± SEM). Red bars are for strain with the element and grey bars are for strain without the element. We found no significant differences between strains with and without the element in *rp49* (1A Mann-Whitney p=0.827), *18S* (1B Mann-Whitney p=0.827) or *Ubx* (1C Mann-Whitney p=0.275).

7 Discussion

7. Discussion

7.1 The fitness effect of *Bari-Jheh* insertion depends on the environment

In this thesis, we have shown that the transposon *Bari-Jheh* mediates oxidative stress response and immune response in *Drosophila melanogaster*. *Bari-Jheh* was identified as a putatively adaptive TE in a genome-wide screening designed to identify adaptive insertions (González, et al. 2008). Moreover, sequence analysis looking for signatures of positive selection, in the region flanking the insertion, found a partial selective sweep, and the only mutation associated to this sweep was *Bari-Jheh* (González, et al. 2008). Follow up studies found the cost of selection of this TE: *Bari-Jheh* is associated with increased developmental time and reduced viability (González, et al. 2009). However the adaptive role of *Bari-Jheh* remained unknown. Using different genetic backgrounds, we showed that flies with the insertion have increased tolerance to different xenobiotic compounds that induce oxidative stress (Guio, et al. 2014). Moreover, we showed that in two of three different genetic backgrounds tested, flies with the insertion were more resistant to *Pseudomonas entomophila* (Guio et al. submitted), a bacterial pathogen that is highly virulent and induces a cessation of feeding and gut damage in *D. melanogaster* by oral infection (Vodovar, et al. 2005; Chakrabarti, et al. 2012). Thus, under nonstress conditions *Bari-Jheh* is associated with negative fitness effects while under stress conditions it is associated with positive fitness effects (Figure 7.1A).

7.1.1 *Bari-Jheh* is associated with increased resistance to oxidative stress.

Bari-Jheh downregulates the expression of *Jheh2* and *Jheh3* genes, which together with *Jheh1*, are involved in Juvenile Hormone (JH) catabolism (Campbell, et al. 1992; González, et al. 2009;Guio, et al. 2014). This downregulation could increase the titre of JH, and consistent with an increased JH titre, González, et al. (2009) found that *Bari-Jheh* is associated with reduce viability and longer developmental time (Flatt, et al. 2005; Flatt and Kawecki 2007). However, although *Jheh1*, *Jheh2* and *Jheh3* genes are thought to be involved in JH catabolism, enzymatic studies of JHEH2 showed that this enzyme does not degrade Juvenile Hormone III but cis-stilbene oxide, an oxidant xenobiotic (Taniai, et al. 2003). Finally, several studies associated oxidative stress and xenobiotics stress response with changes in expression of *Jheh1* and *Jheh2* genes (Girardot, et al. 2004; Sun, et al. 2006; Willoughby, et al. 2007; Misra, et al. 2011; Deng and Kerppola 2014).

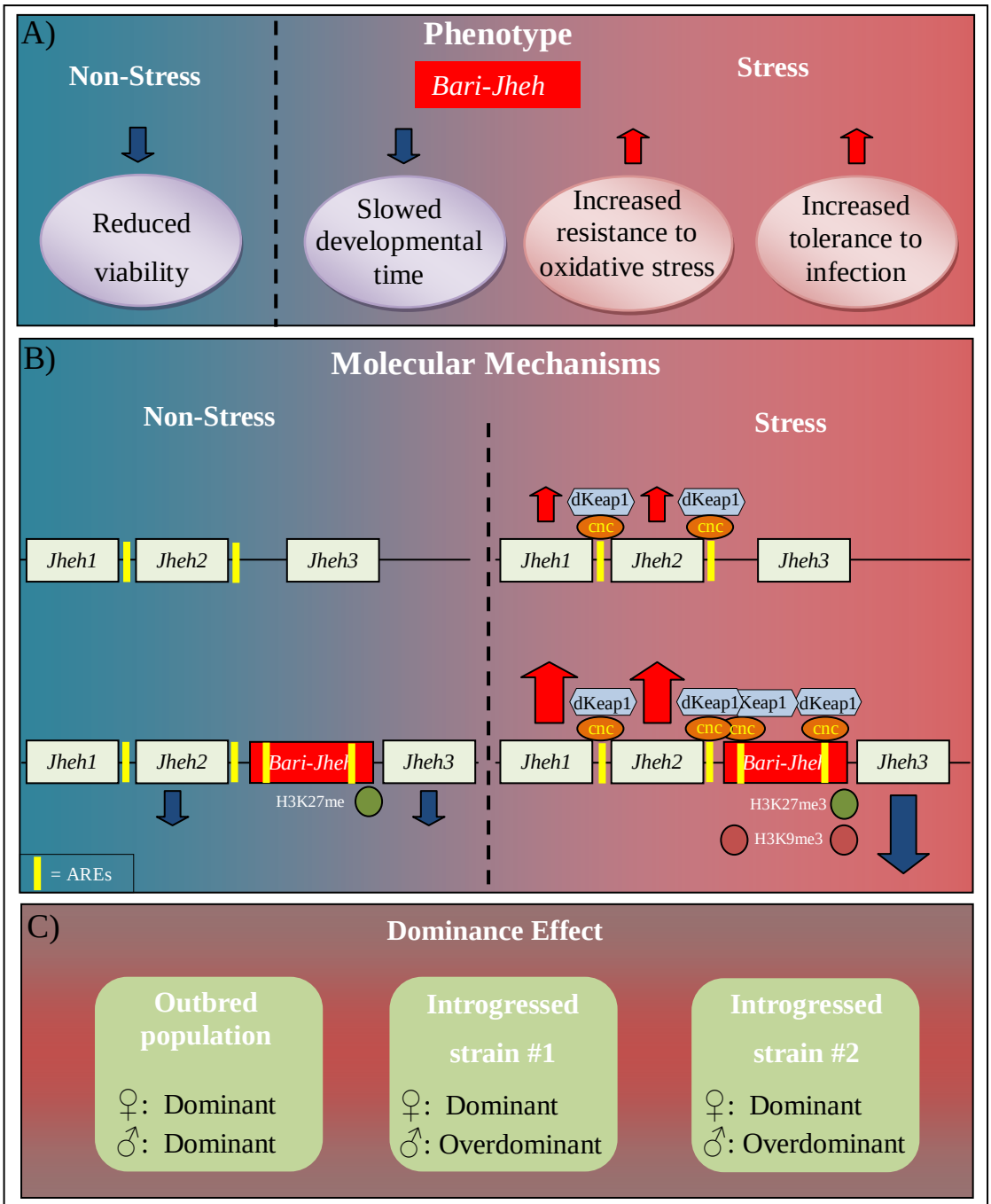


Figure 7.1: Summary of the results described in this thesis. A) *Bari-Jheh* is associated with reduced viability and longer developmental time in non-stressed environments (blue arrows) and with increased resistance to oxidative stress and bacterial infection (red arrows). B) *Bari-Jheh* affects the expression of nearby genes through two molecular mechanisms: *Bari-Jheh* adds AREs that increase the expression of *Jheh1* and *Jheh2* under oxidative stress conditions and *Bari-Jheh* also changes the chromatin state by adding heterochomatin marks. C) The dominance effect of *Bari-Jheh* depends on the genetic background and on the sex of the flies.

Oxidative stress can be defined as an imbalance between the production and the degradation of reactive oxygen species (ROS) in the cell. ROS are produced by natural processes in aerobic organisms, and can also be generated by external sources such as electrophilic compounds or Gram-negative bacterial infection, among others (Lemaitre and Hoffmann 2007; Sykiotis, et al. 2011; Lemaitre and Miguel-Aliaga 2013). Oxidative stress could produce severe damage in DNA, proteins, lipids, and other molecules of the cell and even cell death. Because of that, oxidative stress response is essential for the survival of the organism. Besides the natural processes mentioned above, human activities spread electrophilic substances that increase the amount of ROS in the environment. Therefore, a rapid adaptation to this new environment is critical for a commensal species like *D. melanogaster*.

We have showed that *Bari-Jheh* confers resistance to paraquat, an herbicide that induces oxidative stress in flies with two different genetic backgrounds (Guio, et al. 2014). These results were further confirmed by analyzing additional genetic backgrounds (Guio and Gonzalez 2015). Since oxidative stress is induced by xenobiotics with electrophilic properties, we also tested the oxidative stress resistance of flies with *Bari-Jheh* using Malathion, an organophosphate insecticide that inhibits acetylcholine esterase. We found that flies with the insertion showed increased tolerance to Malathion (Guio, et al. 2014). Finally, Mateo, et al. (2014) tested two different natural fly strains with and without *Bari-Jheh* and they found that flies with the insertion were more resistant to H_2O_2 that strongly induces oxidative stress in the environment (Mateo, et al. 2014).

Overall, *Bari-Jheh* showed a strong association to oxidative stress resistance using three different oxidative stress agents, in flies with four different genetic backgrounds. Thus, we argue that it is very unlikely that the detection of oxidative stress resistance caused by *Bari-Jheh* would occur spuriously (Gruber, et al. 2007). Furthermore, the replicability of the results suggests that epistasis does not have a dominant role in oxidative stress response (Huang, et al. 2012). It is important to mention that we tested paraquat and Malathion at similar levels to those applied in agricultural settings, where *D. melanogaster* lives. We thus conclude that *Bari-Jheh* is the causal mutation of oxidative stress resistance (see figure 7.1A).

7.1.2 *Bari-Jheh* is associated with increased tolerance to Gram-negative bacterial infection.

Because oxidative stress is also caused by Gram-negative bacterial infection, we also tested whether flies carrying *Bari-Jheh* are tolerant to bacterial infection (Guio et al. submitted). Bacterial ingestion triggers several mechanisms of defense in the *D. melanogaster* gut. One of them is the Imd pathway, which produces different antimicrobial peptides, and another defense mechanism is the production of ROS. Bacterial ingestion also triggers a rapid synthesis of ROS in the gut by the specific membrane protein dual oxidase (dDuox) (Lemaitre and Hoffmann 2007; Lemaitre and Girardin 2013; Lemaitre and Miguel-Aliaga 2013). However, as mentioned above, an excess of ROS is dangerous for the host. Because of that, infection also activates the immune response catalase (IRC) that protects the cell against excess of ROS produced by dDuox. This ROS balance is critical for the host survival against bacterial infection (reviewed in Lemaitre and Hoffman 2007) (Figure 7.2). *Pseudomonas entomophila* is a Gram-negative bacterial pathogen that activates the Imd pathway and ROS production in the gut (Vodovar, et al. 2005; Chakrabarti, et al. 2012). In this context, we propose that flies with increased resistance to oxidative stress would show increased tolerance to infection by *Pseudomonas entomophila*.

Using two outbred populations created from DGRP strains, we found that *Bari-Jheh* is associated with increased tolerance to *P. entomophila* (Figure 7A; Guio, et al. submitted). However, a third genetic background created with introgressed isofemale strains showed no difference in infection tolerance or were more susceptible to infection tolerance (Guio, et al. submitted). These results are consistent with a very recent study that showed that DGRP individual strains previously selected for resistance to infection with *P. entomophila* showed increased resistance to paraquat. However, some strains susceptible to *P. entomophila* were resistant to paraquat (Bou Sleiman, et al. 2015) in agreement with our results (Guio, et al. submitted). Bou Sleiman and collaborators (2015) proposed that ROS metabolism is one of the important factors that mediated the variation in gut immunocompetence. Moreover, they analyzed the genetic architecture of this trait and found that it is very complex, probably due to gut immunocompetence being the result of many biological processes, each of them affected by many loci with medium or small effects. The authors proposed that results obtained with standard laboratory strains in

immunological studies may not always be generalizable to all *wildtype* strains, and concluded that some of the observed results would be specific to the analyzed backgrounds (Bou Sleiman, et al. 2015). The complexity of the genetic architecture of immunocompetence is also in agreement with previous studies that associate immune response with distant related processes, such as sleep or food intake, as well as with the environment (Lazzaro and Little 2009). Furthermore, complex eukaryotic organisms coexist with a beneficial or benign microbiome that does not trigger any immune response. As a result, immune response must distinguish pathogenic agents from the non-pathogenic organisms that usually live with the individual (Lazzaro and Rolff 2011).

In the colonization of new environments, pathogen resistance is a major challenge for foreign species. Adaptation of the immune system to this new situation is critical for survival. Thus, *Bari-Jheh*, which is associated with increased tolerance to bacterial infection, could play a role in the adaptation of *D. melanogaster* to new environments.

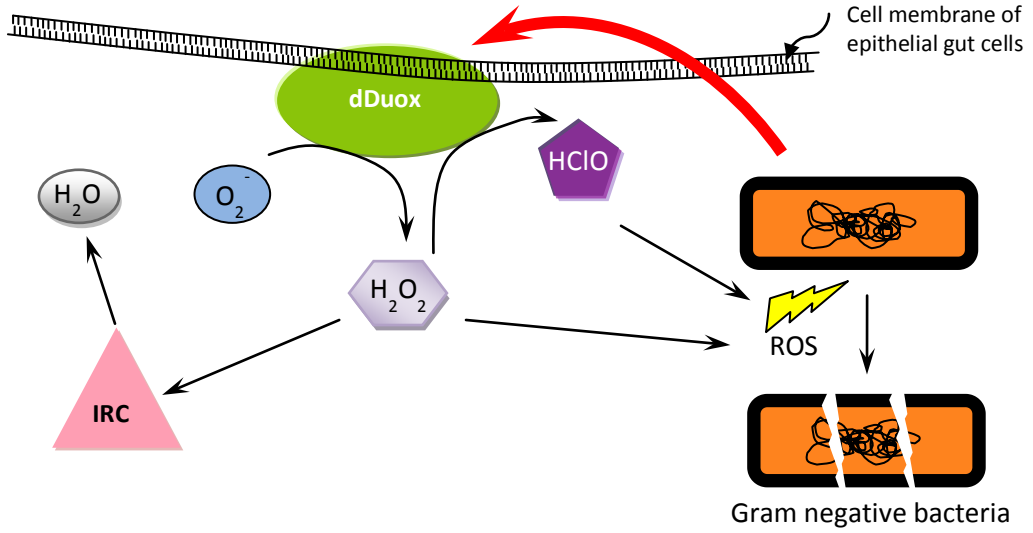


Figure 7.2: ROS Production in the epithelial cells of the *Drosophila* gut. dDuox produce H_2O_2 and the bactericidal substance HClO (hypochlorous acid). IRC reduces the amount of H_2O_2 to maintain the redox balance. H_2O_2 and HClO fight against the bacterial pathogen (adapted from Lemaitre and Hoffman 2007).

7.1.3 *Bari-Jheh* is associated with two adaptive phenotypes.

The results of the survival experiments performed in this thesis allow us to assert that *Bari-Jheh* is an insertion that affects two adaptive phenotypes. This is not the first time that an adaptive transposable element is associated with two different phenotypes.

Doc1420 is a LINE-like element inserted in the coding region of *CHKov1*, a gene that codifies for a choline kinase. *Doc1420* was associated with increased resistance to azinphos-methylphosphate (AZM), an insecticide that inhibits acetylcholine esterase (Aminetzach, et al. 2005). The sequence analysis of the allele where the TE is inserted indicated that the insertion happened approximately 90.000 years ago. However, this allele increased in frequency much recently, 25-250 years ago. Another study associated the same allele with increased resistance to virus infection (Magwire, et al. 2011). Magwire and collaborators proposed that the allele carrying *CHKov1* with the *Doc1420* insertion was selected for their increased virus resistance, and this mutation pre-adapted the flies against organophosphate insecticides in recent times (Magwire, et al. 2011). Moreover, these authors also found an additional step in the evolution of this locus: a duplication of the gene *CHKov1* containing the *Doc1420* and the paralog gene *CHKov2*. Flies with this new allele were more resistant to virus infection than flies carrying the allele with only one copy of *CHKov1* with the insertion of *Doc1420*. Interestingly *CHKov1* was found as an upregulated gene in immune response with RNA-Seq studies in bacterial infection (Bou Sleiman, et al. 2015).

Besides *Doc1420*, another adaptive TE shows an allelic series of mutations. A LTR *Accord* retrotransposon is inserted in the regulatory region of the gene *Cyp6g1*, one of the cytochromes P450 genes, which are involved in detoxification (Daborn, et al. 2002). This *Accord* insertion was associated with increased insecticide resistance and was linked with the upregulation of *Cyp6g1* (Daborn, et al. 2002). Additional analyses of the locus identified copy number variants in different *D. melanogaster* natural populations (Schmidt, et al. 2010). They found six different rearrangements with increased resistance level to the DDT insecticide due to increased expression of *Cyp6g1*. Schmidt and collaborators proposed a four steps series event to arrive to the adaptive optimum and they estimated that this process occurred approximately within 70 years. This rapid process of adaptation is an example of how adaptation occurs in a very fast scale due to the strong pressure of insecticides (Schmidt, et al. 2010). In fact rapid adaptation is not unusual in nature. Insecticide pressure has led to rapid adaptation in insects thanks to copy number variants as well as epigenetic changes that affect transcription regulation of the nearby genes, among other types of mutations (ffrench-Constant 2013). Besides, TEs were proposed to be responsible for rapid adaptation in invasive species (Stapley, et al. 2015).

Thus, we hypothesized that the high frequency of *Bari-Jheh* in natural populations could also be due to a rapid adaptation process. The allele that contains *Bari-Jheh* is between 80 and 800 years old (González, et al. 2008). Human activity during the XIX and XX centuries changed the environment by spreading xenobiotics in the atmosphere. These changes could select those mutations associated to xenobiotic and oxidative stress resistance, such as *Bari-Jheh* mutation. Although we did not looked for the existence of an allelic series in *Bari-Jheh*, the age of the allele where this TE is inserted and its role in oxidative stress resistance suggests that this element could have been subjected to rapid evolution. Further studies would be necessary to test this hypothesis.

To summarize, *Bari-Jheh* is an insertion involved in at least four different phenotypes with different fitness effect, and the different fitness effects depend on the environment. In non-stress conditions, flies carrying *Bari-Jheh* showed longer developmental time and reduced viability. Both phenotypes are associated with reduced fitness. Under oxidative stress conditions, flies carrying *Bari-Jheh* showed increased resistance to oxidative stress agents. Finally, when flies are infected with *P. entomophila*, flies carrying *Bari-Jheh* showed increased tolerance to infection. Both increased resistance to oxidative stress and to bacterial infection are associated with increased fitness. These results highlight the importance of studying different environmental conditions and different phenotypes to understand the effect of natural mutations such as *Bari-Jheh* insertion (Olson-Manning, et al. 2012).

7.2 *Bari-Jheh* affects the expression of *Jheh1*, *Jheh2* and *Jheh3* by adding AREs and by affecting the local chromatin state.

We have shown that *Bari-Jheh* is associated with changes in the expression of its three nearby genes *Jheh1*, *Jheh2* and *Jheh3* under non-stress and stress conditions (Guio, et al. 2014; Guio and González 2015). In non-stress conditions *Bari-Jheh* downregulates *Jheh2* and *Jheh3* and this downregulation could be mediated by the changes in the chromatin state induced by *Bari-Jheh* (Guio, et al. 2014; Guio, et al. submitted). In stress conditions, *Bari-Jheh* upregulates *Jheh1* and/or *Jheh2* and in one of the genetic backgrounds tested it also downregulates *Jheh3* (Guio, et al. 2014, Guio and González 2015). These changes in the expression are associated with the presence of Antioxidant Response Elements (AREs) added by *Bari-Jheh* and changes in the local chromatin state induced by *Bari-Jheh* (Guio, et al. 2014 and Guio, et al. submitted) (Figure 7.1B).

7.2.1 *Bari-Jheh* is associated with expression changes of *Jheh2* and *Jheh3* in non-stress conditions.

TEs inserted close to an active gene could spread heterochromatin marks inducing gene silencing (Sentmanat and Elgin 2012; Lee 2015). We showed that *Bari-Jheh* exhibits matches to piRNA reads (Guio, et al. submitted). piRNAs binding sites could be *cis*-acting targets for heterochromatin assembly which is associated with heterochromatin protein 1a (HP1a) and H3K9me3 (Sentmanat and Elgin 2012). Consistently, we also showed that HP1a binds to *Bari-Jheh* (Guio, et al. submitted). These results suggest that *Bari-Jheh* could be epigenetically silenced by piRNA-induced heterochromatin. However, we found no experimental evidence for the presence of H3K9me3 enrichment in *Bari-Jheh* (Guio, et al. submitted).

Although we found no evidence of constitutive heterochromatin histone mark H3K9me3 enrichment in *Bari-Jheh*, we found enrichment for the facultative heterochromatin histone mark H3K27me3. Specifically, we found enrichment in the region closer to *Jheh3* (Figure 2A Guio, et al. submitted). Experimental evidence for H3K27me3 enrichment is consistent with the prediction of a *zeste* (z) binding site in the *Bari-Jheh* sequence and in the *Jheh3* gene sequence. *zeste* is a member of the Polycomb Group proteins that triggers the trimethylation of H3K27me3, an histone mark associated with facultative heterochromatin (Greer and Shi 2012). This particular chromatin state is transcriptionally silent but retains the potential to change to euchromatin (Trojer and Reinberg 2007). Facultative heterochromatin could adopt different levels of chromatin condensation and could be affected by developmental stage, subnuclear position, non-coding RNA and/or trans-acting protein such as Polycomb group proteins (Trojer and Reinberg 2007).
Overall, we conclude that *Bari-Jheh* downregulates the expression of *Jheh2* and *Jheh3* in non-stress conditions by inducing changes in the local chromatin state.

7.2.2 *Bari-Jheh* is associated with increased expression of *Jheh1* and/or *Jheh2* in stress conditions

Under stress conditions, *Bari-Jheh* is associated with increased expression of *Jheh1* and/or *Jheh2* (Guio, et al. 2014; Guio and González 2015). As was mentioned in section 7.1.1, several previous studies associated these genes with oxidative stress and xenobiotic stress response (Girardot, et al. 2004; Sun, et al. 2006; Willoughby, et al. 2007; Misra, et al. 2011; Deng and Kerppola 2014). We showed that upregulation of *Jheh1* and/or *Jheh2*

could be mediated by the additional AREs added by *Bari-Jheh* to the upstream region of *Jheh2* gene. AREs show a conserved sequence of 20 nucleotides with a core of 9 base pairs: TMAnnRTGAYnnnGCRwwww (Wasserman and Fahl 1997). The 9 base pairs of the core sequence are necessary and sufficient to induce the upregulation of genes placed downstream of the AREs in mammals and insects (Sykiotis and Bohmann 2008). AREs can also activate distal genes in mammals (Mulcahy, et al. 1997), this study support our finding that *Bari-Jheh* up-regulates *Jheh1* located 2,5kb downstream of the AREs added by *Bari-Jheh*. Moreover, an increased number of AREs in the promoter region is associated with higher levels of expression of the downstream genes (Wang, et al. 2006; Chatterjee and Bohmann 2012). We found "*in silico*" several AREs in the sequence of *Bari-Jheh* and in the promoter region upstream of *Jheh1*, *Jheh2* and *Jheh3*. ChIP-Seq experiments showed that *cncC*, the transcription factor that binds AREs in *D. melanogaster*, binds to the upstream regions of *Jheh1* and *Jheh2* (Négre, et al. 2011; Guio, et al. 2014; Deng and Kerppola 2014), as well as *Bari-Jheh* but it does not bind to *Jheh3* (Guio, et al. 2014). Taken together, these results suggest that *Jheh1* and *Jheh2* are involved in oxidative stress response and *Bari-Jheh* could enhance the activity of these genes.

Besides being associated with upregulation of *Jheh1* and *Jheh2* under oxidative stress conditions, in some genetic backgrounds *Bari-Jheh* is also associated with *Jheh3* downregulation (Guio, et al. 2014). Because we have already shown that *Bari-Jheh* is associated with local chromatin changes in nonstress conditions, we tested whether stress further affects *Bari-Jheh* chromatin marks. We found that *Bari-Jheh* is enriched for H3K9me3 in the promoter region of *Jheh2* while it is enriched for H3K9me3 and H3K27me3 in the 3' end of *Jheh3* (Guio, et al. submitted). Although these histone marks are often associated with gene silencing, there is evidence suggesting that H3K9me3 is also associated with gene activation (Vakoc, et al. 2005; Kouzarides 2007). Thus, *Bari-Jheh* could be affecting *Jheh2* expression under oxidative stress conditions by adding AREs and by affecting the local chromatin state. Moreover, the downregulation of *Jheh3* could be explained because *Bari-Jheh* is associated with enrichment of both H3K9me3 and H3K27me3 in the 3' end of this gene. The enrichment for more than one histone mark can have different roles compared with the enrichment of only one histone mark (Greer and Shi 2012; Lelli, et al. 2012). Thus, we propose that the combination of these two histone marks could lead to *Jheh3* downregulation (Greer and Shi 2012; Lelli, et al.

2012). Finally it is important to mention that the gut is the first tissue damaged by oxidative stress in flies exposed to food with paraquat. Usually the gut is the first barrier against biotic and abiotic stress caused by ingestion. Thus, the gut is the first tissue that shows changes in the gene expression related to the stress response and immune response (Lemaitre and Girardin 2013; Lemaitre and Miguel-Aliaga 2013). Thus, major changes in the expression of *Jheh1*, and *Jheh2* may occur in gut and other detoxification tissues such as Malpighian tubules. Although we performed expression analysis with whole fly body, ChIP experiments were performed with fly's guts. Since there are very few ChIP experiments performed in gut we highlight the importance of this tissue for further analysis in oxidative stress or immune response. Furthermore, future expression analysis for these genes under oxidative stress or immune response should be done in gut to increase accuracy in expression changes detection.

In summary, we propose that *Bari-Jheh* is modulating the expression of *Jheh1*, *Jheh2* and *Jheh3* by two different mechanisms (Figure 7.1B). In non-stress conditions, the enrichment of *Bari-Jheh* for H3K27me3 could be associated with downregulation of *Jheh2* and *Jheh3*. In stress conditions, *Bari-Jheh* is enriched for H3K9me3 and H3K27me3 and adds AREs. These two molecular mechanisms could be increasing the expression of *Jheh2* and *Jheh1* and downregulating *Jheh3*.

These results open new questions about the role of *Jheh* genes in oxidative stress response. Because different expression patterns are associated with increased resistance, it would be interesting to analyse whether all three genes are necessary for oxidative stress response, or whether they play different roles. Previous studies provide hints about these questions. The study of the catalytic site showed differences between *Jheh2* (Asp-Glu-His) and *Jheh1* and *Jheh3* (Asp-His-His), suggesting that these enzymes might have different substrates (Taniai, et al. 2003). As mention in the Introduction section, there is experimental evidence for JHEH2 being able to catalyse a xenobiotic compound, providing direct evidence for a role of this gene in oxidative stress. Further experiments with flies in which each one of these genes is mutated could help elucidate the individual function of *Jheh* genes.

7.3 *Bari-Jheh* could be a member of a stress-response regulatory network

In unicellular eukaryotes organisms like *Saccharomyces cerevisiae*, genes involved in the same pathway are usually physically placed together and can be regulated in a coordinated way (Cowley and Oakey 2013). However, in high eukaryotes, genes of the same pathway are not always clustered. To coordinate the simultaneous expression of these genes, they must be activated by the same transcription factor and therefore they should be under the regulation of the same response element. This group of genes could be included in a "transcriptional network" (Cowley and Oakey 2013). The first to propose this idea were Britten and Davidson (1969). They argued that families of TEs with similar regulatory sequences could be recruited by the host for the regulation of the expression of genes included in the same network (Britten and Davidson 1969, 1971). Recent studies provided evidence that support this idea in mammals and plants (Peaston, et al. 2004; Johnson, et al. 2006; Naito, et al. 2009; Macfarlan, et al. 2012; Schmidt, et al. 2012).

In this thesis, we provide evidence that confirm that *Bari-Jheh* is adding AREs that may be affecting the expression of *Jheh1* and *Jheh2*. Moreover, we showed that other TEs add AREs to the *D. melanogaster* genome (Guio, et al. 2014): at least 10 TEs add one or more AREs to the promoter region of downstream genes and two of them showed experimental evidence of *cncC* binding by ChIP-Seq (Guio, et al. 2014). These findings suggest that TEs could coordinate the oxidative stress response by adding AREs. Recent studies in *D. melanogaster* also showed that TEs have an important role in the creation of regulatory networks (Ellison and Bachtrog 2013, 2015). Besides Drosophila, studies in humans and primates also provide strong evidence for TEs being important suppliers of regulatory sequences (Jacques, et al. 2013; Sundaram, et al. 2014). Besides the cis-regulatory elements provided by TEs, it is important to mention that TEs are also domesticated by the host for the transposases as *trans* elements of regulatory networks (reviewed in Feschotte 2008). Thus, TEs could play an important role in creating or/and rewiring regulatory networks both in cis and in trans.

7.4 The dominance effect of *Bari-Jheh* depends of the genetic background

According to Haldane's sieve, dominant adaptive mutations are the most frequent in an ideal natural population (Haldane 1927). Dominant mutations show the adaptive phenotype in the first individual with the adaptive allele thus reducing the chance of

losing the adaptive allele because of genetic drift. However recent studies suggest that overdominance is more relevant in adaptive mutations than previously suggest by J.S.B. Haldane (Orr and Betancourt 2001; Sellis, et al. 2011). In this thesis, we showed that the dominance effect of the adaptive insertion *Bari-Jheh* depends on the genetic background and the sex of the individuals (Guio and González 2015). Flies from an outbred population showed that *Bari-Jheh* is a dominant mutation: homozygous flies for the presence of *Bari-Jheh* and heterozygous flies with only one copy of *Bari-Jheh* showed the same survival level in oxidative stress conditions. Flies of introgressed populations showed that *Bari-Jheh* is dominant in females and in one of the reciprocal crosses in males. However, males from the other reciprocal cross showed that *Bari-Jheh* is overdominant. Interestingly, the *Bari-Jheh* copy came from the mothers in one of the populations and from the fathers in the other population (Guio and González 2015) suggesting that the observed overdominance effects are not due to a maternal effect.

Overall, this study is one of the very few examples in which the dominance effect of an adaptive mutation has been experimentally tested. We showed that the dominance effect of *Bari-Jheh* under oxidative stress environment depends on the genetic background of the flies as has been previously suggested (Wool, et al. 1982; Bourguet, et al. 1996; Bourguet, et al. 1997; Bourguet, et al. 2000).

8 Conclusions

8. Conclusions

From the results obtained in this thesis, we conclude that:

- *Bari-Jheh* shows two adaptive phenotypes: *Bari-Jheh* is associated with increased resistance to oxidative stress conditions, and it is also associated with increased tolerance to *Pseudomonas entomophila* infection.

- Because *Bari-Jheh* has previously been associated with negative fitness effects under nonstress conditions, we can conclude that the effect of *Bari-Jheh* depends on the environmental conditions.

- *Bari-Jheh* is associated with changes in the expression of *Jheh1*, and/or *Jheh2* and/or *Jheh3*, depending on the genetic background analyzed, under oxidative stress conditions. The overexpression of *Jheh1* or *Jheh2* is enough to confer resistance to oxidative stress.

- *Bari-Jheh* affects the expression of *Jheh* genes through two different molecular mechanisms under oxidative stress conditions: by adding stress-related transcription factor binding sites, and by changing the local chromatin state.

- *Bari-Jheh* affects the expression of *Jheh* genes under nonstress conditions by inducing changes in the local chromatin state.

- Besides *Bari-Jheh*, other Transposable Elements adds Antioxidant Response Elements upstream of genes related with oxidative stress response. Thus, *Bari-Jheh* together with these other insertions could be rewiring the regulatory networks for oxidative stress response in *Drosophila melanogaster*.

- *Bari-Jheh* dominance effect depends on the genetic background. *Bari-Jheh* is dominant in one genetic background while in two other genetic backgrounds *Bari-Jheh* is dominant in females but it is overdominant in males.

9 References

9. References

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science 309:764-767.

Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol 18:279-290.

Barrett RD, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12:767-780.

Barron MG, Fiston-Lavier AS, Petrov DA, Gonzalez J. 2014. Population genomics of transposable elements in Drosophila. Annu Rev Genet 48:561-581.

Barth S, Fischer M, Schmid RD, Pleiss J. 2004. Sequence and structure of epoxide hydrolases: a systematic analysis. Proteins 55:846-855.

Barton, NH. 2000. Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci 355: 1553-62

Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365:548-550.

Belyayev A. 2014. Bursts of transposable elements as an evolutionary driving force. J Evol Biol 27:2573-2584.

Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, Evans-Holm M, Hiesinger PR, Schulze KL, Rubin GM, et al. 2004. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics 167:761-781.

Bennetzen JL. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621-627.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genet 10:e1004775.

Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. Nature 443:521-524.

Boros IM. 2012. Histone modification in Drosophila. Brief Funct Genomics 11:319-331.

Bou Sleiman MS, Osman D, Massouras A, Hoffmann AA, Lemaitre B, Deplancke B. 2015. Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. Nat Commun 6:7829.

Bourguet D, Prout M, Raymond M. 1996. Dominance of insecticide resistance presents a plastic response. Genetics 143:407-416.

Bourguet D, Lenormand T, Guillemaud T, Marcel V, Fournier D, Raymond M. 1997. Variation of dominance of newly arisen adaptive genes. Genetics 147:1225-1234.

Bourguet D, Genissel A, Raymond M. 2000. Insecticide resistance and dominance levels. J Econ Entomol 93:1588-1595.

Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. Science 165:349-357.

Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol 46:111-138.

Brown EJ, Bachtrog D. 2014. The chromatin landscape of Drosophila: comparisons between species, sexes, and chromosomes. Genome Res 24:1125-1137.

Caizzi R, Caggese C, Pimpinelli S. 1993. *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics 133:335-345.

Campbell PM, Healy MJ, Oakeshott JG. 1992. Characterisation of juvenile hormone esterase in *Drosophila melanogaster*. Insect Biochemistry and Molecular Biology 22:665-677.

Casacuberta E, Gonzalez J. 2013. The impact of transposable elements in environmental adaptation. Mol Ecol 22:1503-1517.

Celotto AM, Palladino MJ. 2005. Drosophila: a "model" model system to study neurodegeneration. Mol Interv 5:292-303.

Chakrabarti S, Liehl P, Buchon N, Lemaitre B. 2012. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. Cell Host Microbe 12:60-70.

Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302-305.

Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. Am Nat 140:126-148.

Charlesworth B. 1998. Adaptive evolution: the struggle for dominance. Curr Biol 8:R502-504.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195-205.

Chatterjee N, Bohmann D. 2012. A versatile PhiC31 based reporter system for measuring AP-1 and Nrf2 signaling in Drosophila and in tissue culture. PLoS One 7:e34063.

Chenais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 509:7-15.

Cowley M, Oakey RJ. 2013. Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet 9 (1) e1003234.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, et al. 2002. A single p450 allele associated with insecticide resistance in Drosophila. Science 297:2253-2256.

David JR, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet 4:106-111.

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7:e1002384.

Deng H, Kerppola TK. 2014. Visualization of the Drosophila dKeap1-CncC interaction on chromatin illumines cooperative, xenobiotic-specific gene activation. Development 141:3277-3288.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-156.

dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase C. 2015. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43:D690-697.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals African and European admixture in the North American Drosophila melanogaster population. Genetics 193:291-301.

Elgart M, Snir O, Soen Y. 2015. Stress-mediated tuning of developmental robustness and plasticity in flies. Biochim Biophys Acta 1849:462-466.

Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. Science 342:846-850.

Ellison CE, Bachtrog D. 2015. Non-allelic gene conversion enables rapid evolutionary change at multiple regulatory sites encoded by transposable elements. Elife 4:e05899 1-12

Fablet M, Vieira C. 2011. Evolvability, epigenetics and transposable elements. Biomol Concepts 2:333-341.

Fedoroff NV. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. Science 338:758-767.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics 9:397-495.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331-368.

ffrench-Constant RH. 2013. The molecular genetics of insecticide resistance. Genetics 194:807-815.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143:212-224.

Finnegan DJ. 1985. Transposable elements in eukaryotes. Int Rev Cytol 93:281-326.

Flatt T, Kawecki TJ. 2007. Juvenile hormone as a regulator of the trade-off between reproduction and life span in *Drosophila melanogaster*. Evolution 61:1980-1991.

Flatt T, Tu MP, Tatar M. 2005. Hormonal pleiotropy and the juvenile hormone regulation of Drosophila development and life history. Bioessays 27:999-1010.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. Science 334:86-89.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLoS Genet 11:e1005004.

Girardot F, Monnier V, Tricoire H. 2004. Genome wide analysis of common and specific stress responses in adult *Drosophila melanogaster*. BMC Genomics 5:74.

Gluckman PD, Low FM, Buklijas T, Hanson MA, Beedle AS. 2011. How evolutionary principles improve the understanding of human health and disease. Evol Appl 4:249-263.

Gonzalez J, Karasov TL, Messer PW, Petrov DA. 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. PLoS Genet 6:e1000905.

Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. PLoS Biol 6:e251.

Gonzalez J, Macpherson JM, Petrov DA. 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. Mol Biol Evol 26:1949-1961.

Goodier JL, Kazazian HH, Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135:23-35.

Gould SJ, Lewontin RC. 1979. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. Proceedings of the Royal Society B: Biological Sciences 205:581-598.

Graur D, Li W-H. 2000. Fundamentals of molecular evolution. Sunderland, Mass.: Sinauer Associates.

Greer EL, Shi Y. 2012. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13:343-357.

Griffith LC. 2012. Identifying behavioral circuits in *Drosophila melanogaster*: moving targets in a flying insect. Curr Opin Neurobiol 22:609-614.

Gruber JD, Genissel A, Macdonald SJ, Long AD. 2007. How repeatable are associations between polymorphisms in achaete-scute and bristle number variation in Drosophila? Genetics 175:1987-1997.

Grüneberg H. 1938. An analysis of the "pleiotropic" effect of a new lethal mutation in the rat (*Mus norvegicus*). Proc of the Royal Soc of Lon. Series B, Biol Sci 125:123-144.

Guio L, Barron MG, Gonzalez J. 2014. The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila. Mol Ecol 23:2020-2030.

Guio L, Gonzalez J. 2015. The dominance effect of the adaptive transposable element insertion *Bari-Jheh* depends on the genetic background. Genome Biol Evol 7:1260-1266.

Habu Y, Hisatomi Y, Iida S. 1998. Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. Plant J 16:371-376.

Haldane JBS. 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23:838-844.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet 4:e32.

Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, et al. 2010. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A 107 Suppl 2:8924-8930.

Hodgkin J. 1998. Seven types of pleiotropy. Int. J. Dev. Biol. 42:501-505.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101-104.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419-1428.

Hua-Van A, Le Rouzic A, Boutin TS, Filee J, Capy P. 2011. The struggle for life of the genome's selfish architects. Biol Direct 6:19.

Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, et al. 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc Natl Acad Sci USA 109:15553-15559.

Hunter P. 2007. The human impact on biological diversity. How species adapt to urban challenges sheds light on evolution and provides clues about conservation. EMBO Rep 8:316-318.

Hunter RG, Gagnidze K, McEwen BS, Pfaff DW. 2015. Stress and the dynamic genome: Steroids, epigenetics, and the transposome. Proc Natl Acad Sci USA 112:6828-6833.

Jacques PE, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet 9:e1003504.

Johnson R, Gamblin RJ, Ooi L, Bruce AW, Donaldson IJ, Westhead DR, Wood IC, Jackson RM, Buckley NJ. 2006. Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. Nucleic Acids Res 34:3862-3877.

Joseph SB, Peck KM, Burch CL. 2014. Dominance effects of deleterious and beneficial mutations in a single gene of the RNA virus varphi6. PLoS One 9:e97717.

Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. Curr Biol 17:R77-81.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature 471:480-485.

Kidwell MG, Lisch D. 1997. Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704-7711.

Korves TM, Schmid KJ, Caicedo AL, Mays C, Stinchcombe JR, Purugganan MD, Schmitt J. 2007. Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. Am Nat 169:E141-157.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128:693-705.

Labbe P, Berticat C, Berthomieu A, Unal S, Bernard C, Weill M, Lenormand T. 2007. Forty years of erratic insecticide resistance evolution in the mosquito *Culex pipiens*. PLoS Genet 3:e205.

Lachaise D, Cariou M-L, David J, Lemeunier F, Tsacas L, Ashburner M. 1988. Historical Biogeography of the *Drosophila melanogaster* Species Subgroup. In: Hecht M, Wallace B, Prance G, editors. Evolutionary Biology: Springer US. p. 159-225.

Lazzaro BP, Little TJ. 2009. Immunity in a variable world. Philos Trans R Soc Lond B Biol Sci 364:15-26.

Lazzaro BP, Rolff J. 2011. Immunology. Danger, microbes, and homeostasis. Science 332:43-44.

Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol Biol Evol 19:1261-1271.

Lee YC. 2015. The Role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. PLoS Genet 11:e1005269.

Lehner B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. Nat Rev Genet 14:168-178.

Lelli KM, Slattery M, Mann RS. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet 46:43-68.

Lemaitre B, Girardin SE. 2013. Translation inhibition and metabolic stress pathways in the host response to bacterial pathogens. Nat Rev Microbiol 11:365-369.

Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. Annu Rev Immunol 25:697-743.

Lemaitre B, Miguel-Aliaga I. 2013. The digestive tract of *Drosophila melanogaster*. Annu Rev Genet 47:377-404.

Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet 2:e166.

Li T, Spearow J, Rubin CM, Schmid CW. 1999. Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. Gene 239:367-372.

Lisch D, Bennetzen JL. 2011. Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14:156-161.

Liu WM, Chu WM, Choudary PV, Schmid CW. 1995. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res 23:1758-1765.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487:57-63.

Mackay TF. 2010. Mutations and quantitative genetic variation: lessons from Drosophila. Philos Trans R Soc Lond B Biol Sci 365:1229-1239.

Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. Nature 482:173-178.

Mackay TF. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet 15:22-33.

Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. 2011. Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication. PLoS Genet 7:e1002337.

Mateo L, Ullastres A, Gonzalez J. 2014. A transposable element insertion confers xenobiotic resistance in Drosophila. PLoS Genet 10:e1004560.

Matthews BB, dos Santos G, Crosby MA, Emmert DB, St Pierre SE, Gramates LS, Zhou P, Schroeder AJ, Falls K, Strelets V, et al. 2015. Gene model annotations for *Drosophila melanogaster*: Impact of high-throughput Data. G3 (Bethesda) 5:1721-1736.

Mc Clintock B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344-355.

McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. 2012. Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet 8:e1002474.

Misra JR, Horner MA, Lam G, Thummel CS. 2011. Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes Dev 25:1796-1806.

modENCODE C, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330:1787-1797.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 23:314-318.

Morgan TH. 1915a. Localization of the Hereditary Material in the Germ Cells. Proc Natl Acad Sci USA 1:420-429.

Morgan TH. 1915b. The mechanism of Mendelian heredity. New York, Holt publisher 262 p.

Morgan TH, Bridges CB, Sturtevant AH. 1925. The genetics of Drosophila. s'Gravenhage,: M. Nijhoff publisher. 262p

Moschetti R, Caggese C, Barsanti P, Caizzi R. 1998. Intra- and interspecies variation among Bari-1 elements of the *melanogaster* species group. Genetics 150:239-250.

Mulcahy RT, Wartman MA, Bailey HH, Gipp JJ. 1997. Constitutive and -naphthoflavone-induced expression of the human -glutamylcysteine synthetase heavy

subunit gene Is regulated by a distal antioxidant response element/TRE sequence. Journal of Biological Chemistry 272:7445-7454.

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130-1134.

Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton, ML, Loriaux P, Sealfon R. 2011. A cis-regulatory map of the Drosophila genome. Nature 471: 527-531

Oesch F. 1973. Mammalian epoxide hydrases: inducible enzymes catalysing the inactivation of carcinogenic and cytotoxic metabolites derived from aromatic and olefinic compounds. Xenobiotica 3:305-340.

Oesch F, Herrero ME, Hengstler JG, Lohmann M, Arand M. 2000. Metabolic detoxification: implications for thresholds. Toxicol Pathol 28:382-387.

Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet 13:867-877.

Orgogozo V, Morizot B, Martin A. 2015. The differential view of genotype-phenotype relationships. Front Genet 6:179.

Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics 157:875-884.

Orr HA. 2002. The population genetics of adaptation: the adaptation of DNA sequences. Evolution 56:1317-1330.

Orr HA. 2005. The genetic theory of adaptation: a brief history. Nat Rev Genet 6:119-127.

Orr HA. 2009. Fitness and its role in evolutionary genetics. Nat Rev Genet 10:531-539.

Orr HA. 2010. The population genetics of beneficial mutations. Philos Trans R Soc Lond B Biol Sci 365:1195-1201.

Ostman B, Hintze A, Adami C. 2012. Impact of epistasis and pleiotropy on evolutionary adaptation. Proc Biol Sci 279:247-256.

Paaby AB, Rockman MV. 2013. The many faces of pleiotropy. Trends Genet 29:66-73.

Palazzo A, Marconi S, Specchia V, Bozzetti MP, Ivics Z, Caizzi R, Marsano RM. 2013. Functional characterization of the *Bari1* transposition system. PLoS One 8:e79385.

Pavlicev M, Norgard EA, Fawcett GL, Cheverud JM. 2011. Evolution of pleiotropy: epistatic interaction pattern supports a mechanistic model underlying variation in genotype-phenotype map. J Exp Zool B Mol Dev Evol 316:371-385.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7:597-606.

Pennings PS, Hermisson J. 2006a. Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol 23:1076-1084.

Pennings PS, Hermisson J. 2006b. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet 2:e186.

Phillips PC. 2008. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9:855-867.

Piacentini L, Fanti L, Specchia V, Bozzetti MP, Berloco M, Palumbo G, Pimpinelli S. 2014. Transposons, environmental changes, and heritable induced phenotypic variability. Chromosoma 123:345-354.

Piegu B, Bire S, Arensburger P, Bigot Y. 2015. A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. Mol Phylogenet Evol 86:90-109.

Plasterk RH, Izsvak Z, Ivics Z. 1999. Resident aliens: the *Tc1/mariner* superfamily of transposable elements. Trends Genet 15:326-332.

Pool JE, Aquadro CF. 2007. The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. Mol Ecol 16:2844-2851.

Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB. 2009. Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. Science 326:1663-1667.

Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: Towards new species. Gene 454:1-7.

Rebollo R, Horard B, Begeot F, Delattre M, Gilson E, Vieira C. 2012a. A snapshot of histone modifications within transposable elements in Drosophila wild type strains. PLoS One 7:e44253.

Rebollo R, Romanish MT, Mager DL. 2012b. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 46:21-42.

Richard DS, Applebaum SW, Gilbert LI. 1990. Allatostatic regulation of juvenile hormone production in vitro by the ring gland of *Drosophila melanogaster*. Mol Cell Endocrinol 68:153-161.

Richard DS, Applebaum SW, Sliter TJ, Baker FC, Schooley DA, Reuter CC, Henrich VC, Gilbert LI. 1989. Juvenile hormone bisepoxide biosynthesis in vitro by the ring gland of *Drosophila melanogaster*: a putative juvenile hormone in the higher Diptera. Proc Natl Acad Sci U S A 86:1421-1425.

Rubin GM. 1988. *Drosophila melanogaster* as an experimental organism. Science 240:1453-1459.

Scheinfeldt LB, Tishkoff SA. 2013. Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet 14:692-702.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148:335-348.

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, et al. 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. PLoS Genet 6:e1000998.

Schrader L, Kim JW, Ence D, Zimin A, Klein A, Wyschetzki K, Weichselgartner T, Kemena C, Stokl J, Schultner E, et al. 2014. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun 5:5495.

Sellis D, Callahan BJ, Petrov D, Messer PW. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. Proc Natl Acad Sci USA 108:20666-20671.

Sentmanat M, Wang SH, Elgin SC. 2013. Targeting heterochromatin formation to transposable elements in Drosophila: potential roles of the piRNA system. Biochemistry (Mosc) 78:562-571.

Sentmanat MF, Elgin SC. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. Proc Natl Acad Sci USA 109:14104-14109.

Shilova VY, Garbuz DG, Myasyankina EN, Chen B, Evgen'ev MB, Feder ME, Zatsepina OG. 2006. Remarkable site specificity of local transposition into the Hsp70 promoter of *Drosophila melanogaster*. Genetics 173:809-820.

Siguier P, Filee J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9:526-531.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272-285.

Stapley J, Santure AW, Dennis SR. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24:2241-2252.

Stephan W, Li H. 2007. The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity (Edinb) 98:65-68.

Stocker H, Gallant P. 2008. Getting started : an overview on raising and handling Drosophila. Methods Mol Biol 420:27-44.

Sturtevant AH. 1920. Genetic Studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*. Genetics 5:488-500.

Sun W, Margam VM, Sun L, Buczkowski G, Bennett GW, Schemerhorn B, Muir WM, Pittendrigh BR. 2006. Genome-wide analysis of phenobarbital-inducible genes in *Drosophila melanogaster*. Insect Mol Biol 15:455-464.

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res 24:1963-1976.

Sykiotis GP, Bohmann D. 2008. *Keap1/Nrf2* signaling regulates oxidative stress tolerance and lifespan in Drosophila. Dev Cell 14:76-85.

Sykiotis GP, Habeos IG, Samuelson AV, Bohmann D. 2011. The role of the antioxidant and longevity-promoting *Nrf2* pathway in metabolic regulation. Curr Opin Clin Nutr Metab Care 14:41-48.

Taniai K, Inceoglu AB, Yukuhiro K, Hammock BD. 2003. Characterization and cDNA cloning of a clofibrate-inducible microsomal epoxide hydrolase in *Drosophila melanogaster*. European Journal of Biochemistry 270:4696-4705.

Thompson JN. 2013. Relentless evolution. Chicago: The University of Chicago Press.

Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics 172:1607-1619.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31-40.

Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. 2010. Stress-induced activation of heterochromatic transcription. PLoS Genet 6:e1001175.

Trojer P, Reinberg D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell 28:1-13.

Turner JRG. 1981. Adaptation and Evolution in Heliconius: A defense of NeoDarwinism. Annual Review of Ecology and Systematics 12:99-121.

Tyler AL, Asselbergs FW, Williams SM, Moore JH. 2009. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays 31:220-227.

Ullastres A, Petit N, Gonzalez J. 2015. Exploring the phenotypic space and the evolutionary history of a natural mutation in *Drosophila melanogaster*. Mol Biol Evol 32:1800-1814.

Vakoc CR, Mandat SA, Olenchock BA, Blobel GA. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell 19:381-391.

Vieira C, Lepetit D, Dumont S, Biemont C. 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. Mol Biol Evol 16:1251-1255.

Vignieri SN, Larson JG, Hoekstra HE. 2010. The selective advantage of crypsis in mice. Evolution 64:2153-2158.

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard F, Lemaitre B. 2005. Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. Proc Natl Acad Sci USA 102:11414-11419.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204-213.

Wang XJ, Hayes JD, Wolf CR. 2006. Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of *nrf2* by cancer chemotherapeutic agents. Cancer Res 66:10983-10994.

Wasserman WW, Fahl WE. 1997. Functional antioxidant responsive elements. Proc Natl Acad Sci U S A 94:5361-5366.

Weinreich DM, Watson RA, Chao L. 2005. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59:1165-1174.

Wheat CW, Watt WB, Pollock DD, Schulte PM. 2006. From DNA to fitness differences: sequences and structures of adaptive variants of Colias phosphoglucose isomerase (PGI). Mol Biol Evol 23:499-512.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973-982.

Willoughby L, Batterham P, Daborn PJ. 2007. Piperonyl butoxide induces the expression of cytochrome P450 and glutathione S-transferase genes in *Drosophila melanogaster*. Pest Manag Sci 63:803-808.

Wool D, Noiman S, Manheim O, Cohen E. 1982. Malathion resistance in Tribolium strains and their hybrids: inheritance patterns and possible enzymatic mechanisms (Coleoptera, Tenebrionidae). Biochem Genet 20:621-636.

Wright, S. 1931. Evolution in Mendelian populations.Genetics 16: 97-159.

Zeller G, Henz SR, Widmer CK, Sachsenberg T, Ratsch G, Weigel D, Laubinger S. 2009. Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. Plant J 58:1068-1082.

Zhang M, Azad P, Woodruff RC. 2011. Adaptation of *Drosophila melanogaster* to increased NaCl concentration due to dominant beneficial mutations. Genetica 139:177-186.

10 Thesis advisor’s report

10. Thesis advisor's report about authorship and impact factor of the publications of this doctoral thesis presented by Lain Guio Leiman

Publication 1: The transposable element *Bari-Jheh* mediates oxidative stress response in Drosophila
Lain Guio, Maite G. Barrón, and Josefa González
Molecular Ecology 23 (8): 2020-2030 (2014)
Candidate tasks: in this publication the student has designed the research, performed the research, analyzed the data and drafted the article.
Molecular Ecology impact factor is 6.494 and it is ranked the 5th of 46 journals in the "Evolutionary Biology" category (JCR 2014).

Publication 2: The dominance effect of the adaptive transposable element insertion *Bari-Jheh* depends on the genetic background
Lain Guio and Josefa González
Genome Biology and Evolution 7(5):1260-1266 (2015)
Candidate tasks: in this publication the student has designed the research, performed the research, analyzed the data and drafted the article.
Genome Biology and Evolution impact factor is 4.229 and it is ranked 9th of 46 journals in the "Evolutionary Biology" category (JCR 2014).

Publication 3: Stress affects the epigenetic marks added by *Bari-Jheh*: a natural insertion associated with two adaptive phenotypes in Drosophila
Lain Guio, Cristina Vieira and Josefa González
Article submitted
Candidate tasks: In this publication the student has designed the research, performed the research, analyzed the data and drafted the article.
This article was submitted to Molecular Biology and Evolution.
Molecular Biology and Evolution impact factor is 9.105 and it is ranked 4th of 46 journals in the "Evolutionary Biology" category (JCR 2014).

Signed:

Josefa González Pérez Barcelona, January, 5th 2016