UNIVERSITAT DE BARCELONA

Deciphering the molecular and phenotypic effects of a *Drosophila melanogaster* transposable element located in a unique insertional cluster

Miriam Merenciano González

Deciphering the molecular and phenotypic effects of a *Drosophila melanogaster* transposable element located in a unique insertional cluster

MIRIAM MERENCIANO GONZALEZ

2019

TESI DOCTORAL

UNIVERSITAT DE BARCELONA

Deciphering the molecular and phenotypic effects of a *Drosophila melanogaster* transposable element located in a unique insertional cluster

Memòria presentada per

MIRIAM MERENCIANO GONZÁLEZ

Per optar al grau de Doctora per la Universitat de Barcelona

Programa de Genètica

Treball realitzat a l'Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra)

Doctoranda	**Tutor**	**Directora**
Miriam Merenciano	Dr. Francesc Mestres Naval	Dra. Josefa González Pérez

Barcelona, Desembre 2019

INDEX

AKNOWLEDGEMENTS

Primer de tot voldria agrair l'ajuda en la realització d'aquesta tesi a la Pepi González. Gràcies per donar-me a conèixer el món de la ciència, per ensenyar-me el que és ser un bon científic i en definitiva, per donar-me la oportunitat de poder fer el doctorat. Però sobretot gràcies per demostrar-me que quan treballes amb passió i esforç, no hi ha res que ens pugui separar del nostre objectiu.

****

Aquesta tesi tampoc hauria estat possible sense els companys i companyes que em van veure començar aquest viatge. Lidia, has estat un exemple a seguir en tots els sentits. Gràcies per animar-me a estar aquí. A ti Lain, por ser como un hermano mayor. Por esas charlas con las que, aunque no te lo pienses, me hiciste reflexionar mucho. Gracias por cuidar de mi. Al Quirze, per ser sempre tant optimista i per enrecordar-te sempre de nosaltres tot i estar lluny. Como no, a ti Maite, por tu sinceridad y tu visión critica de la vida. I al Jose. Gràcias por resolverme las dudas bioinformaticas y por las divertides comidas en Japón. Però sobretot voldria donar-li les gràcies a l'Anna. Aquesta tesi no hagués estat possible sense la seva ajuda. Per aquelles converses on discutíem resultats que no sabíem interpretar, per donar-me idees i per haver-me ensenyat a moure'm en un laboratori. Hem compartit viatges, congressos, col·lectes de mosques, frustracions, neguits, rialles… Bàsicament, has estar la millor companya de laboratori, més ben dit amiga, que em podria trobar.

Vull mencionar també a totes aquelles persones que han passat pel laboratori durant aquests quatre anys: Karine, Hung, Jaione, Jon, Lorena, Nico, Marcela. En especial a Camillo, parte de este trabajo es gracias a ti.

****

I no, no me olvido de vosotros, los que me habéis aguantado hasta el final. Al *crack* de Gaby por prestarse siempre a ayudar-me y venga va, también por ese asado tan rico que hay que repetir. Al *beltranero* de Carlos, que siempre nos saca una sonrisa. A Judit, sin hache y con jota jota. Por esa energía contagiosa y por recordarme que un científico debe ir en bata. Thanks Llew, my *phenotyping partner*. A Santi, que aunque haga poquito que nos conocemos estoy segura de que llegaràs lejos. I a tu Laura, per estar en tot i en tot moment. Per l'ajuda científica i la no científica. Thank you all!

I would like to thank Fillip Port for giving me the opportunity to spend three months in the beautiful Heidelberg and to introduce me in the CRISPR world. Part of this thesis is thanks to you. It was also a great pleasure to know you all: the Braintrust. Michi, Michi (Michael), Rene and Leo: you were my family during those three months.

****

Ara bé, tot i haver-me trobat gent científicament brillant en el camí cap al meu doctorat, del que estic més orgullosa es d'haver conegut a les meves *losers*. Les meves amigues. Elles. Per ser allà sempre que necessitava parlar mentre ens fèiem una tapa de braves, o dos... Pels findes, viatges, brava parties, room scapes... Sou genials Anna, Lidia, María, Sílvia i Vivien.

Però muchas gracias especialmente a ti María.

I a la meva confident Sílvia.

****

Al Xavi i el David. Que fan que escrivint aquestes paraules ja se m'escapi el riure. Gràcies per posar el toc d'humor sempre.

També voldria fer especial menció a aquells que entre passadissos han contribuït en que l'IBE sigui com la meva segona casa: Elena Parra, Núria, Alicia, Alfonso, Edu, Alberto, Pepi, Cecília, Josep, Maria Ferrer, Aleksandra, Elena Casacuberta, Gerard, Cristina, Vicente, Anna, Pablo, Blanca, Isabel.

****

A nivell familiar, vull agrair als meus pares, Andreu i Consol, el seu recolzament incondicional. Per haver permès amb el seu esforç que jo hagi arribat fins aquí. Perquè tot i no saber ben bé el què faig, sempre s'han interessat, preguntat i emocionat quan he dit que un experiment ha sortit bé. Jo sí que estic orgullosa de vosaltres.

A la iaia *xatona*, per ser aquella persona a la que li puc explicar tot. Ho sento si t'he fet patir amb els meus viatges, com dius tu, a l'altra punta del món. Una de les grans motivacions d'aquest treball ha estat pensar que tu veuries com ho arribava a estudiar tot, com et vaig dir fa molt. Gràcies.

I al meu *broderito* Andreu. Per ser la persona que m'escoltava assajar les xerrades sense tenir ni idea del que deia ni en quin idioma ho feia. I per ser el millor germà del món.

****

Un dels més sincers agraïments va pels meus *Supercoleguis*. Adrià, Alba, Albert, Carles, Cristina, David, Marc, Soraya i Víctor. Per tindre excuses cada divendres, dissabte, diumenge... i també entre setmana per trobar-nos. Per fer-me passar moments inoblidables. Per ser els meus amics de sempre i per sempre.

Però especialment a tu, Adrià. Saps que sense el teu "*todo está bien*" molts dies haguessin estat més feixucs. Gràcies per escoltar-me sempre, fins i tot les preocupacions i les tonteries que sé que no importen a ningú. Per estar a prop quan estava lluny. Per dir-me que arribaré a fer grans coses. Per confiar en mi. Per la portada, que ha fet que aquesta tesi fos encara una mica més teva. En definitiva, per ser tu, el meu millor amic per sempre.

Hablando de amigos, muchas gracias Victor por estar ahí durante el proceso y darme ánimos en la recta final. Y como no, por tu ayuda en la maquetación de la portada.

I també a les *manoses*. Per acompanyar-me des del bon principi. Per haver-me fet passar la millor època de la meva vida. I per tot el que encara ens queda per fer juntes. Ainoa, Ana, Arnau, Ivan, Maria, Miren, Natalia i Silvia, gràcies de tot cor.

****

També m'agradaria agrair a la Montse i al Cesc. Per preocupar-vos per mi i el meu treball. I per deixar-nos el cotxe per anar a comptar mosques a hores intempestives. Moltes gràcies.

A la Alba i al Pere. Aquests últims anys no haguessin estat el mateix sense vosaltres. Moltes gràcies per deixar-me gaudir de la personeta més maca del món mundial.

****

Voldria acabar per agrair a la persona que veritablement ha permès que tot això fos possible. Aquella persona que ha estat amb mi sempre, agafant-me la mà durant el camí. T'he d'agrair tantes coses... Des de consolar-me el dia que pensava que havia perdut les meves mosques i obligar-me a anar a veure si en quedava alguna viva fins a acompanyar-me a les dotze de la nit al laboratori. I no una vegada, moltes. Gràcies per entendre que vol dir fer un doctorat. Per oferir-te a ajudar-me en qualsevol moment. Per ajudar-me a complir els meus somnis. Per dir-me mil vegades que et faig sentir orgullós. Per estimar-me. Ets el millor que tinc, Jordi.

****

I sobretot, gràcies a tu avi. Perquè un dia et vaig dir que em podrien dir doctora.

ABSTRACT

Understanding how organisms adapt to their environment remains an open question in Biology. So far, most projects focus on the study of single nucleotide polymorphism variants, while other types of mutations, likely to play a role in adaptation, are largely ignored. For instance, the effects of transposable elements (TE), which are potent mutagens that introduce genomic variability in natural populations sometimes driving phenotypic adaptations to environmental conditions, are not so well studied. The aim of this work was to characterize the ecological adaptive effects of the *Drosophila melanogaster roo* insertion *FBti0019985*, which is located in the 5'UTR of the transcription factor *CG18446*. To that end, we studied its possible phenotypic effects and investigated the molecular mechanisms behind those functional changes.

First, we discovered that besides *FBti0019985,* the *CG18446* promoter region harbours 19 independent *roo* insertions. The presence of these recurrent *roo* insertions in the *CG18446* promoter region is likely to be the result of several bursts of transposition. We suggest that chromatin accessibility could be one of the factors explaining the multiple insertions. We also investigated whether the identified insertions were functionally equivalent by performing 5'RACE, gene expression, and cold-stress survival experiments. We found that only *FBti0019985* was associated with *CG18446* up-regulation in embryos and with increased viability in nonstress and under cold-stress conditions.

Second, we further studied the molecular and phenotypic effects of *FBti0019985* in different developmental stages and under different stress conditions. Performing gene expression analysis and *in vivo* enhancer reporter assays we found that *FBti0019985* drives the expression of its nearby gene *CG18446* depending on both the developmental stage and the environmental conditions. We associated the presence of *FBti0019985* with *CG18446* down-regulation. However, in embryos under nonstress conditions and in guts under immune-stress, *FBti0019985* caused a *CG18446* up-regulation. Indeed, we associated this up-regulation with an enhancer activity of the element under these specific contexts. Finally, with the generation of mutant strains with *FBti0019985* deletions in *D. melanogaster* natural populations using the CRISPR/Cas9-mediated homology-directed repair technique, we could also associate the presence of *FBti0019985* with tolerance to *P. entomophila* infection.

Overall, this work gives more evidences of the role of TEs in relevant adaptive traits. Moreover, it also reflects the importance of considering the effect of a candidate adaptive insertion under different contexts to fully characterize its adaptive consequences.

SECTION 1

INTRODUCTION

1. INTRODUCTION

1.1. Transposable elements: brief history

Transposable elements (TEs) are repetitive DNA sequences that have the ability to change their position (transpose) within a genome sometimes increasing their copy number (Bourque et al., 2018). They were first discovered in 1950s by Barbara McClintock in maize (*Zea mays*) under the name of "controlling elements", for what she was awarded with the Nobel Prize in Physiology or Medicine (1983) (McClintock, 1950; McClintock, 1953). She associated the changing colour pattern of maize kernels to some genetic factors that were able to change their position within and between chromosomes. McClintock also described that these "controlling elements" could move along the genome in response to environmental changes, thus modifying the expression of some genes (McClintock, 1950; McClintock, 1953; McClintock, 1956). Later, TE presence was demonstrated in bacteria (Shapiro, 1969) and in Drosophila (Rubin et al., 1982), where the mobilization of specific TEs was associated with the hybrid dysgenesis phenomenon. Moreover, Britten and Davidson (1971) hypothesized that TEs may place regulatory sites across the genome while transposing suggesting an important role in the coordination of gene expression (the "gene battery" hypothesis) (Britten and Davidson, 1971). However, these ideas were rejected and TEs were for many years categorized as "selfish" or "junk" genetic elements that produced detrimental effects (Strobel et al., 1979; Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Hickey, 1982). Although their presence in the genome was commonly considered deleterious, the first evidences of TEs leading to beneficial effects appeared in the following years. For instance, TEs were detected to function as regulatory elements in primates (Samuelson et al., 1990) and protecting broken linear chromosome ends in Drosophila (Biessmann et al., 1992).

Due to the emergence of sequencing techniques we now know that TEs also constitute an important component of genomes from diverse origins (Warren et al., 2015; Deniz et al., 2019). For instance, in humans, half of the genome was found to be constituted by TEs (Lander et al., 2001). Over the last decades, TEs began to acquire even more importance once their influence in recombination rates and in chromosomal rearrangements, their mutagenic abilities, and they role as gene regulators were confirmed (Biémont, 2010; Bourque et al., 2018). Today, there are also some evidences of TEs involved in genome evolution and more specifically, in recent adaptation (Biémont and Vieira, 2006; González et al., 2008; González et al., 2010; Arkhipova, 2018). Considering that, the scientific community now agree that TEs have considerably shaped the structure, function, and evolution of genomes.

1.2. Transposable element classification

The discovery of novel insertions thanks to the increased number of available sequenced genomes, and the introduction of new techniques to detect and classify TE insertions cause a constant evolution and revision of TE classification. Until today, several approaches to categorize TE have been presented based on different criteria.

A first classification divides TEs based on their replication capability in autonomous and non-autonomous insertions. Autonomous TEs encode all the domains needed to move along the genome, i.e. open reading frames (ORFs) and regulatory sequences. However, non-autonomous TEs lack these domains and thus, depend on the enzymes encoded by autonomous TEs to transpose.

Another classification, first proposed long time ago, divides TEs in two major classes based on the nature of their transposition intermediate (Finnegan, 1989). Class I elements, or retrotransposons, mobilize through an RNA intermediate that is reverse-transcribed and integrated elsewhere in the genome. On the other hand, Class II elements, or DNA transposons, mobilize using a DNA intermediate (Finnegan, 1989). While all Class I elements use a retrotranscriptase to transpose, Class II elements encompass several unrelated groups of TEs with different transposition mechanisms, unified only by the absence of an RNA intermediate in their transposition processes. So, inside the Class II group we can find TEs that use a transposase to mobilize (TIRs), TEs that use a tyrosine recombinase (*Crypton*), TEs that use a replication initiator-like protein (*Helitron*), and TEs that use a protein-primed B-type DNA polymerase (*Maverick*).

Later, the "Wicker" and "Repbase" classification systems for eukaryotic TEs used the same basal criteria relying on the presence or absence of RNA transposition intermediates but tried to further unified the current classification (Wicker et al., 2007; Kapitonov and Jurka, 2008). In the "Wicker" classification, the introduction of a more hierarchical structure that includes subclasses, orders, and superfamilies tries to better subclassify elements within classes based on mechanistic and enzymatic criteria (Wicker et al., 2007) (Figure 1.1). Similarly, the "Repbase" classification, besides proposing a universal nomenclature, divides TEs in two main types: Type 1 (DNA transposons) and Type 2 (retrotransposons). These TE types are subclassified in seven

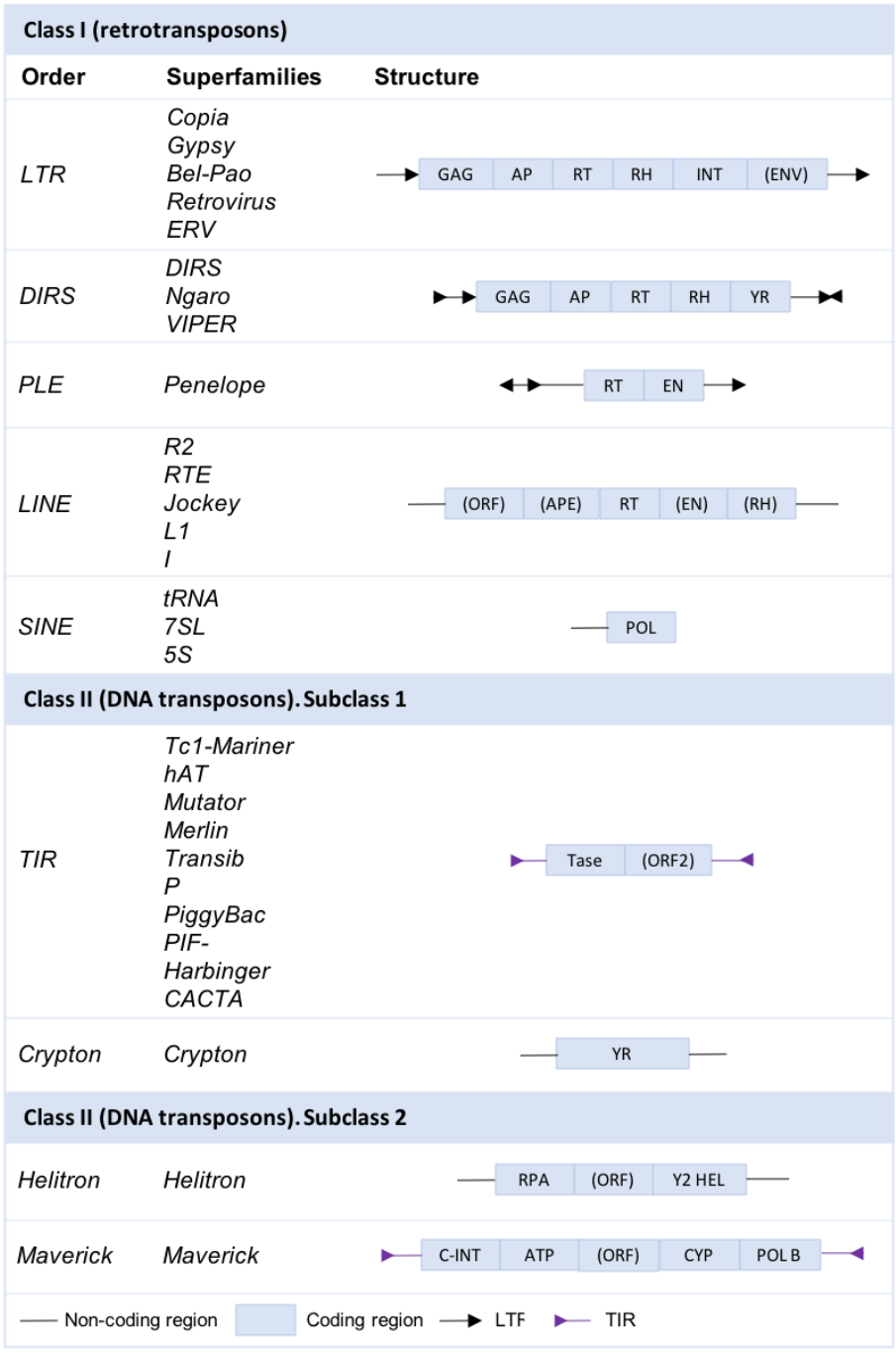


Figure 1.1. Transposable element classification in eukaryotes based on Wicker et al. (2007). Figure adapted from Wicker et al. (2007). Hierarchical classification of eukaryotic TEs proposed in Wicker et al. (2007) which divides TEs in two major classes (Class I and Class II) based on the presence or absence of an RNA intermediate during transposition. These groups are then further divided in subclasses (only Class II elements), orders, superfamilies and families (not shown in the figure). The genetic structure of every order of TEs is represented. Components depicted in brackets are not common for all the TEs from the same order. AP: Aspartic proteinase, APE: Apurinic endonuclease, ATP: Packaging ATPase, C-INT: C-integrase, CYP: Cysteine protease, EN: Endonuclease, ENV: Envelope protein, GAG: Capsid protein, HEL: Helicase, INT: Integrase, ORF: Open reading frame, POL: Polymerase III promoter, POL B: DNA polymerase B, RH: RNase H, RPA: Replication protein A, RT: Reverse transcriptase, Tase: Transposase, YR: Tyrosin recombinase, Y2: YR with YY motif.

classes based on the different enzymes involved in transposition and structural similarities (Kapitonov and Jurka, 2008).

Although being the eukaryotic TE classification proposed by Wicker et al. (2007) well established, it does not reflect the evolutionary history of the different insertions from all kingdoms of life (Piegú et al., 2015). Hence, Piegú et al. (2015) indicated the need of a universal classification system in which all prokaryotic and eukaryotic TEs were present and categorized into taxonomic groups (Piegú et al., 2015). Other authors suggested TE classifications based on the mechanisms they use for moving rooted in an evolutionary context (Curcio and Derbyshire, 2003). However, the latest TE classification suggests a three-component classification scheme in which their replicative, integrative and structural components are considered (Arkhipova, 2017).

1.3. Transposable elements are virtually present in all genomes

TEs are virtually in all eukaryotic and in almost all prokaryotic genomes, being lost only from the reduced genomes of apicomplexan parasites (DeBarry and Kissinger, 2011; Hua-Van et al., 2011). However, they are less ubiquitous in prokaryotes, where complete TEs are not present in more than 20% of the genomes sequenced so far (Touchon and Rocha, 2007; Hua-Van et al., 2011).

TEs usually represent a considerable fraction of the genome, although their abundance is highly variable from one specie to another: from 64-84% in maize to 0.1% in the fungi *Pseudozyma antarctica* (Gao et al., 2016; Sessegolo et al., 2016; Jiao et al., 2017; Guio and González, 2019) (Figure 1.2). It has also been found that TEs are absent in the genome of the parasitic unicellular eukaryote *Plasmodium falciparum* (Gardner et al., 2002). TE abundance can even vary among different species from the same phylogenetic group (Sessegolo et al., 2016). Furthermore, TE families are not equally represented in all species. For instance, *LINE-1* elements are the most abundant type in humans, while *LTR* elements are more abundant in Drosophila (Lander et al., 2001, Hua-Van et al., 2011, Rahman et al., 2015, Sessegolo et al., 2016).

Yet, TE diversity is still biased towards insertions described in reference genomes, missing novel insertions or insertions that differ at population level. This deficiency was recently

pointed out in an extensive genome analysis of 177 *Drosophila melanogaster* strains from the Drosophila Genetic Reference Panel (DGRP) in which, on average, more than 500 novel insertions were found in every strain (Rahman et al., 2015). Despite the increase availability of whole-genome data, TE annotation has been often ignored. This can be due to the fact that a throughout TE annotation requires an important manual curation effort, whereby leaving repetitive regions excluded in genome assemblies (Arkhipova, 2017). Only with the use of third-generation long-read assemblies, a comprehensive TE repertoire will be eventually obtained (Sotero-Caio et al., 2017; Arkhipova, 2017).

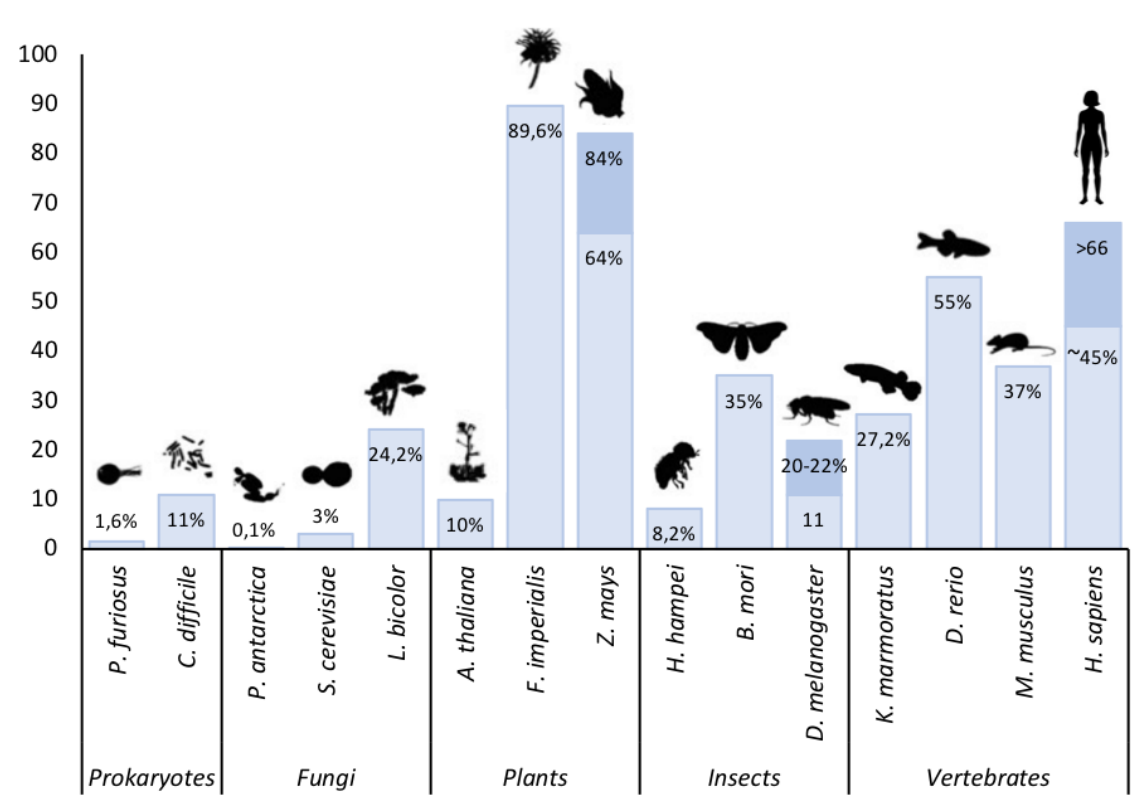


Figure 1.2. Percentage of TE abundance in different organisms. Figure adapted from Guio and González, (2019).

1.4. Transposable elements are not randomly distributed in the genome

TEs are not randomly distributed in the genome, in fact, they are preferentially inserted in heterochromatic, pericentromeric and telomeric regions (Bourque et al., 2018; Arkhipova, 2018). These regions are associated with low recombination rates and low gene content, properties that reduce the potential deleterious impact of TEs (Bourgeois and Boissinot, 2019). In addition, TEs can also be related with low recombination regions because epigenetic modifications promoted by TE silencing mechanism are often negatively associated with recombination (Bourgeois and Boissinot, 2019). Nevertheless, in some cases, TEs present in heterochromatic regions have acquired essential roles in the genomes. For instance, in *D. melanogaster*, telomeres are composed of tandem head-to-tail arrays of transposons, adopting a critical role of telomere structure and maintenance (Mason and Biessmann, 1995).

Despite being preferentially inserted in low recombination regions as shown in *D. melanogaster*, humans and other mammals, the correlation between TEs and recombination not always follows the same pattern (Cridland et al., 2013; Skaletsky et al., 2003; Bourgeois and Boissinot, 2019). Insertion preference for recombination hotspots has been described for some TE families and in different organisms such as *Ficedula* birds and *Caenorhabditis elegans* (Duret et al., 2000; Myers et al., 2008; Kawakami et al., 2017; Laricchia et al., 2017). The shared preference of recombination and transposition machineries for open chromatin and double-stranded DNA breaks are possible explanations for these observations (Bourgeois and Boissinot, 2019).

Insertion preferences for specific euchromatic regions have also been described for some TEs. Retrotransposons in mold and budding and fission yeast have convergently evolved to target upstream regions of genes transcribed by RNA polymerase III, where they can be transcribed by themselves without altering the host gene expression (Spaller et al., 2017; Sultana et al., 2017; Cheung et al., 2018). Thus, these TEs favour their own propagation while ensuring a less detrimental effect for the host. In *Anolis* lizards, a burst of transposition resulted in high numbers of TEs in *Hox* gene clusters, regions that regulate development and morphological adaptation. Considering that, these TEs may have been a source for phenotypic variation, thus contributing to lizard speciation processes (Feiner, 2016). Similarly, some TE families are preferentially found in 5′ gene regulatory regions, such as *P-elements* in Drosophila (Spradling et al., 1995; Liao et al., 2000), *Mutator* elements in maize (Liu et al., 2009), *mPing* MITEs in rice (Naito et al., 2009), and *Tf1* retroelements in fission yeast (Leem et al., 2008). In *D.*

melanogaster, target site motifs for some TE families have been described using population genomic data of 166 strains (Linheiro and Bergman, 2012). However, in some cases, local preference depends more on DNA structure than on primary sequence, being especially true for *P-elements* (Liao et al., 2000). However, the study of *de novo* insertions will be crucial to further investigate the mechanisms of TE integration site preference.

1.4.1. Recurrent insertion of transposable elements in specific genomic regions

Recurrent insertion of TEs in specific genomic regions has been described in *D. melanogaster*. 23 extreme TE-density regions were identified in pericentromeric regions or on chromosome 4 (Bergman et al., 2006). Besides transposition, TE duplication was also one of the factors contributing to the accumulation of insertions in those extreme TE-density regions (Bergman et al., 2006). Multiple insertions in gene regions were also found in the large genes *RNA-binding protein 6* (*Rbp6*) (179.5 kb), *klarsicht* (106.5 kb), and *derailed-2* (24 kb) analysing two different *D. melanogaster* panels of lines: the DGRP and the Drosophila Synthetic Population Resource (DSPR) (Cridland et al., 2013). The vast majority of the insertions found were unique to a single strain or present at very low frequencies. Moreover, all the insertions were located in intronic regions, except two TEs inserted in exons of *klarsicht* and *derailed-2* genes, respectively (Cridland et al., 2013). In the same organism, it has been reported that proximal promoter regions of *hsp* genes are a natural target for *P-element* insertions (Walser et al., 2006). Transposition into *hsp* promoter regions could affect promoter architecture, thus having an impact on *hsp* expression (Lerman and Feder, 2005). Besides *hsp* promoter regions, no other region in *D. melanogaster* genome has shown such number of recurrent insertions of the same TE family.

1.5. Transposable elements are a source of genomic variation

TEs, as mobile DNA sequences, are an incredible source of genomic variation. DNA breaks and insertions associated with transposition events cause an obvious alteration to the host genome. Besides affecting the host genome depending on where they insert, TEs can produce RNAs or proteins altering cell functions. They can also perform chromosomal rearrangements as a result of ectopic recombination between copies dispersed along the genome (Bourque et al., 2018; Bourgeois and Boissinot, 2019) (Figure 1.3 and 1.4).

The consequences of TE mobilization can also be more global while altering gene expression and structure or chromosome dynamics mainly as a result of ectopic recombination. It has been described that ectopic recombination between TE copies account for a number of human disorders (Sasaki et al., 2010). Changes performed by TEs in gene expression or structure can be both genetic (e.g. introducing new cis-regulatory elements) or epigenetic (altering chromatin structure) (Klein et al., 2018; Bourgeois and Boissinot, 2019).

1.5.1. Transposable elements as modulators of gene expression

In line with McClintock predictions, nowadays there are several studies evidencing that TEs have been a source of material for the modulation of gene expression (Bourque et al., 2018). There are myriad mechanisms by which TEs can alter host gene expression both in cis and in trans, transcriptional or post-transcriptionally (Elbarbary et al., 2016; Chuong et al., 2017) (Figure 1.3).

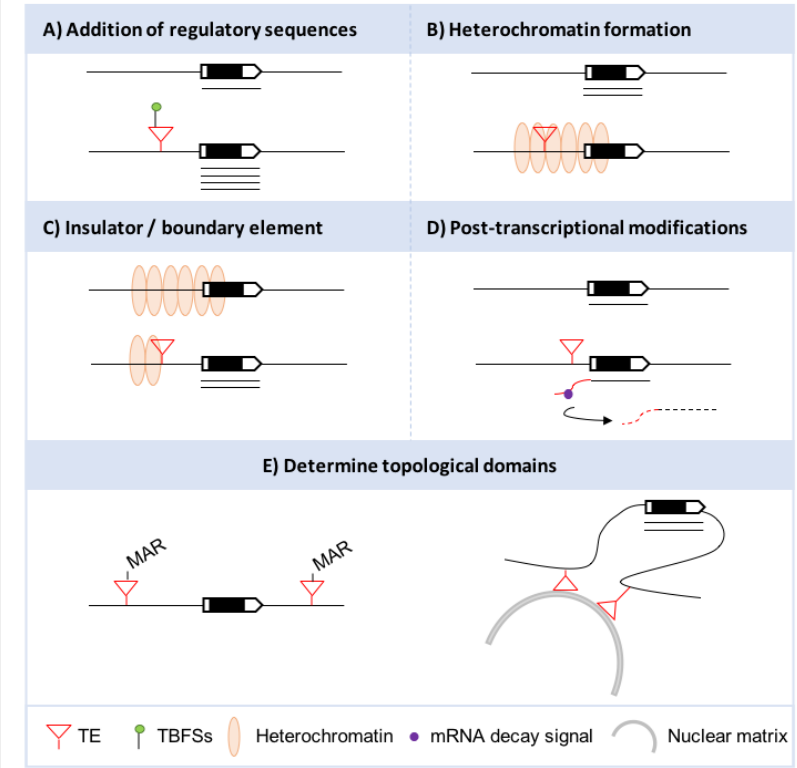


Figure 1.3. TEs can modulate gene expression through different mechanisms. A) TEs can add new regulatory sequences such as promoters, enhancers or repressive elements that fine-tune nearby genes expression. **B)** TEs can contribute to the formation of heterochromatin, hence affecting the expression of their nearby genes. **C)** TEs can act as insulators or boundary elements preventing the spread of heterochromatin. **D)** TEs can modulate gene expression through post-transcriptional modifications. They can provide sequences for mRNA decay, localization or translation efficiency. **E)** TEs can determine topological domains in the nucleus thereby participating in the determination of chromatin loops. Black and white boxes represent exons and UTR regions, respectively. Gene transcripts are represented as short black lines.

TEs can harbour cis-regulatory elements that regulate gene expression. In primates, different reporter assay techniques validated that newly evolved cis-regulatory elements were enriched in *LTR* and *SVA* elements with the capacity to modulate gene expression (Trizzino et al., 2017). This study thus concluded that TEs had an important role in rewiring ape gene regulation (Trizzino et al., 2017). Transcription factor binding sites (TFBSs), important cis-regulatory elements, have been repeatedly identified in TE sequences. Chip-seq studies discovered that TEs contribute a substantial fraction (5-40%) of binding sites across the genome in mammals (Sundaram et al., 2014; Chuong et al., 2017; Bourque et al., 2018). And, in the Drosophila genus, several TFBSs within TE sequences have been identified and often functionally validated demonstrating that TEs can provide material for the emergence of new regulatory networks (Brönner et al., 1995; Carareto et al., 2014; Loreto et al., 2018; Palazzo et al., 2019; Villanueva-Cañas et al., 2019). Among all TE types, *LTR* elements seem to contribute more to disperse TFBSs since they have intrinsic ability to recruit RNA polymerase II and they are more likely to retain the ancestral cis-regulatory activity (Feschotte, 2008; de Souza et al., 2013; Jacques et al., 2013; Thompson et al., 2016; Chuong et al., 2017).

Besides modulating gene expression, TEs can also alter chromatin structure by recruiting heterochromatin proteins that silence nearby genes. Similarly, they have also been described to function as insulators and/or boundary elements, preventing the spread of heterochromatin (de Souza et al., 2013; Chuong et al., 2017). Furthermore, TEs are also contributing regulatory sequences that modulate gene expression post-transcriptionally. For example, they can provide recognition sequences for mRNA decay, localization or translation efficiency (Gong and Maquat, 2011; Shen et al., 2011; Hu et al., 2015; Elbarbary et al., 2016). Finally, TEs may have an important role in the three-dimensional organization of the genome isolating regions and thus controlling the transcriptional regulation of large chromosomal regions (Chuong et al., 2017). In *D. melanogaster*, it has been described that *LTR* elements harbour matrix attachment regions with the ability to determine topological domains (Mamillapalli et al., 2013).

TEs might also play a role in rewiring gene regulatory networks. It has been shown that they participate in important physiological responses such as early development and immune response in mammals (Gerdes et al., 2016; Chuong et al., 2016) or sex dosage compensation in *Drosophila miranda* (Ellison and Bachtrog, 2013).

1.5.2. Transposable elements as modulators of gene structure

There are a great variety of mechanisms by which TEs can impact gene structure depending on whether they are inserted into introns or in gene non-coding regions (Figure 1.4).

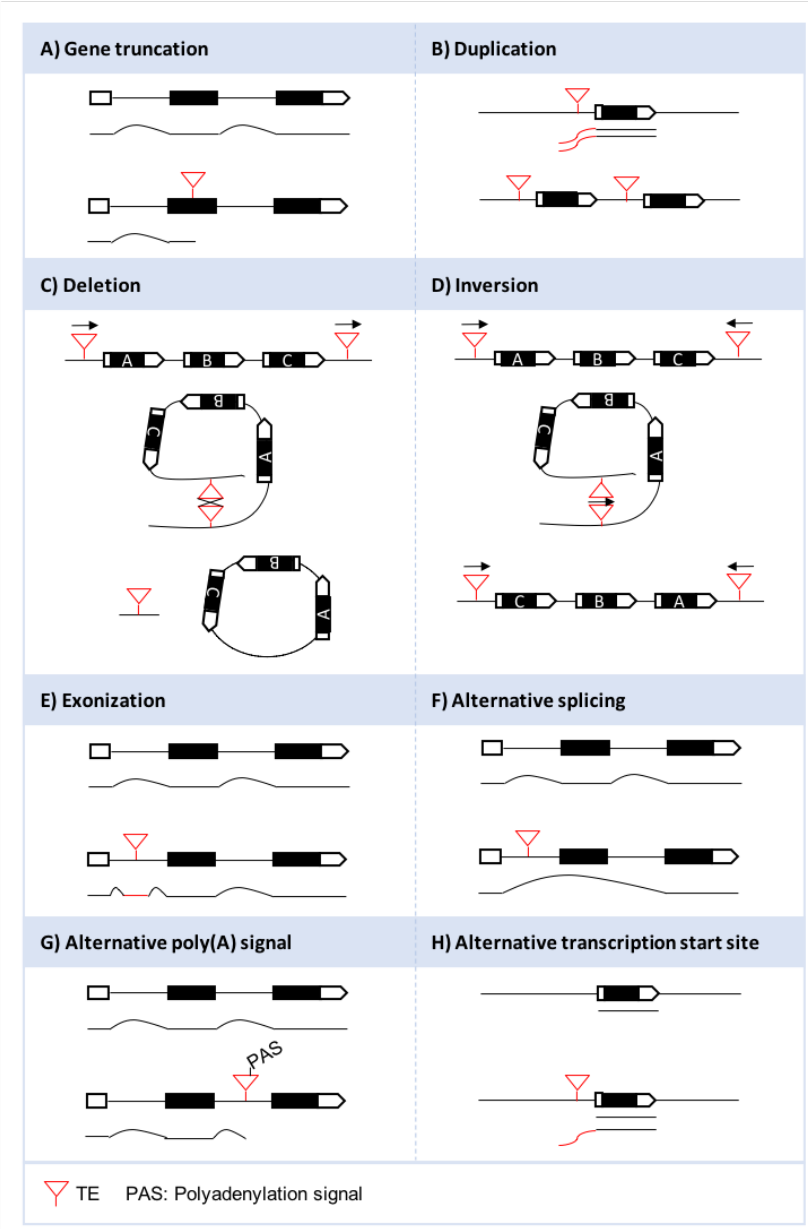


Figure 1.4. TEs can affect gene structure. A) Depending on where they are inserted, TEs can truncate essential sequences such as gene promoters or gene coding regions. TEs can act as substrate for ectopic recombination due to its sequence similarity between copies from the same family. As a result, they can generate **B)** duplications, **C)** deletions and **D)** inversions. When inserted into introns, TEs can generate new transcripts through processes like **E)** exonization, **F)** the addition of alternative splicing sites or **G)** the addition of alternative poly(A) signals. **H)** TEs are also a source of new transcripts by contributing with alternative TSS for a gene. Black boxes represent exons and white boxes represent UTR regions. Gene transcripts are represented as short black lines.

When inserted into introns, TEs can be incorporated as new exons (exonization), provide alternative splicing sites, or introduce alternative poly adenylation (poly(A)) signals (Schmitz and Brosius, 2011; Chenais et al., 2012; Casacuberta and González, 2013; Warren et al., 2015). Exonization defines the process of creation of a new exon due to mutations in intronic regions. In humans, this process was first described when *Alu* elements were found within a large number of mature mRNAs (Makalowski et al., 1994). Exonization was later confirmed in other different studies, showing that these events are more frequent in vertebrates than in invertebrates (Nekrutenko and Li, 2001; Sela et al., 2010). New alternative splice variants are sometimes acquired by the use of alternative splice sites added by TE insertions. It is estimated that 5% of human alternative spliced exons derived from *Alu* sequences (Sorek et al., 2002). Alternative poly(A) can also be introduced by TE insertions leading to a shorter or truncated transcript. It has been shown that nonconserved poly(A) sites are more frequently associated with TEs, suggesting a significant role in poly(A) site evolution and defining the 3′ end of genes (Lee et al., 2008).

Besides being inserted into introns, TEs can also alter gene structure when inserted in other non-coding regions such as 5′UTR regions. There, they can introduce alternative transcription start sites (TSS), hence modulating gene transcript lengths. Studies of transcription initiation mapping using cap analysis of gene expression followed by sequencing (CAGE-seq) showed remarkable amounts of *Pol II* initiation within TEs in humans, mouse as well as in *D. melanogaster* (Faulkner et al., 2009; Batut et al., 2013).

1.5.3. Co-option of transposable element sequences

Some TE-derived sequences are sometimes used by the host genome as elements that modulate gene expression. However, the recruitment of TE coding sequences to generate functional proteins is an evolutionary process that turned out to be surprisingly common. These processes of TE sequence recruitment create innovations at both molecular and phenotypic levels that are beneficial for the host organism. Several terms have been used to described these events such as domestication (Miller et al., 1997), co-option (Sarkar et al., 2003), or exaptation (Brandt et al., 2005), although the term exaptation is better used when the evolved trait has a different usage of nucleotides or a different function (Schrader and Schmitz, 2019).

Domestication of TE proteins has been repeatedly described to function against retroviruses or TEs themselves. In those cases, cells have evolved to reduce the potential deleterious effect of invasive genetic elements like retroviruses and TEs with the recruitment of TE proteins. In *S. pombe*, the control of *Tf2* retrotransposons is achieved by the cooperation of three transposase-derived proteins called Abp1, Cbh1 and Cbh2 with histone deacetylases (Cam et al., 2008; Lorenz et al., 2012). Similarly, in Arabidopsis, the two genes derived from *Ty3/gypsy* retrotransposons *MAIL1* and *MAIN* participate in an epigenetic silencing pathway that supress many TEs (Ikeda et al., 2017). Another example of domestication is the *LINE-1* type transposase domain-containing 1 (L1TD1) in mammals, which has been suggested to protect against *LINE-1* and other TE types (McLaughlin et al., 2014). This domain was co-opted from the open reading frame 1 (ORF1) of *LINE-1* elements by the ancestor of placental mammals (McLaughlin et al., 2014). In addition, it has also been described that the human *L1TD1* element plays an important role in the maintenance of embryonic stem cell pluripotency (McLaughlin et al., 2014).

The process of domestication has also contributed to novel biological processes such as adaptive immune systems (Jangam et al., 2017). A well-known example is the domestication of the RAG1 and RAG2 catalytic proteins and the cis-acting DNA sequences RSSs which are crucial components of V(D)J recombination mechanism (Kapitonov et al., 2005). In jawed vertebrates, V(D)J recombination is a process that creates an illimitable collection of antibodies in B and T cells. In a study performed in the cephalochordate lancelet, it has been postulated that RAG1, RAG2, and RSS descend from an ancestral *Transib* transposon named ProtoRAG (Huang et al., 2016). Thus, it has been suggested that this ancestral element was transmitted vertically through chordate and vertebrate evolution (Huang et al., 2016). In prokaryotes, TE domestication could have been responsible of the emergence of the Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-associated nuclease (CRISPR/Cas) system. Cas proteins have shown sequence similarities to the transposases encoded by the Casposon group (Krupovic et al., 2014). Moreover, CRISPR repeats may derived from TIR sequences in those *Casposon* elements (Béguin et al., 2016; Krupovic et al., 2017). Finally, in humans, the insertion of an *ERV* element (*MER41*) introduces interferon-inducible enhancers to the *AIM2* gene that have been co-opted to regulate inflammation in response to infection (Chuong et al., 2016).

1.6. Transposable elements and stress

Stress conditions have been repeatedly related with TE activation and, consequently, associated with an adaptive role of TEs under challenging environments (McClintock, 1984; Capy et al., 2000; Fablet and Vieira, 2011; Chénais et al., 2012; Casacuberta and González, 2013; Negi et al., 2016; Rey et al., 2016, Horváth et al., 2017). Both biotic and abiotic stresses can promote TE activation increasing both TE transcriptional and mobilization rates (Figure 1.5). This process of TE activation increases population variability potentiating both evolutionary plasticity and evolvability (Horváth et al., 2017). In addition, stress can also activate stress-response regulatory sequences inside TEs, thus spreading stress-responsive sequences throughout the genome and rewiring stress response networks (Cowley et al., 2013) (Figure 1.5).

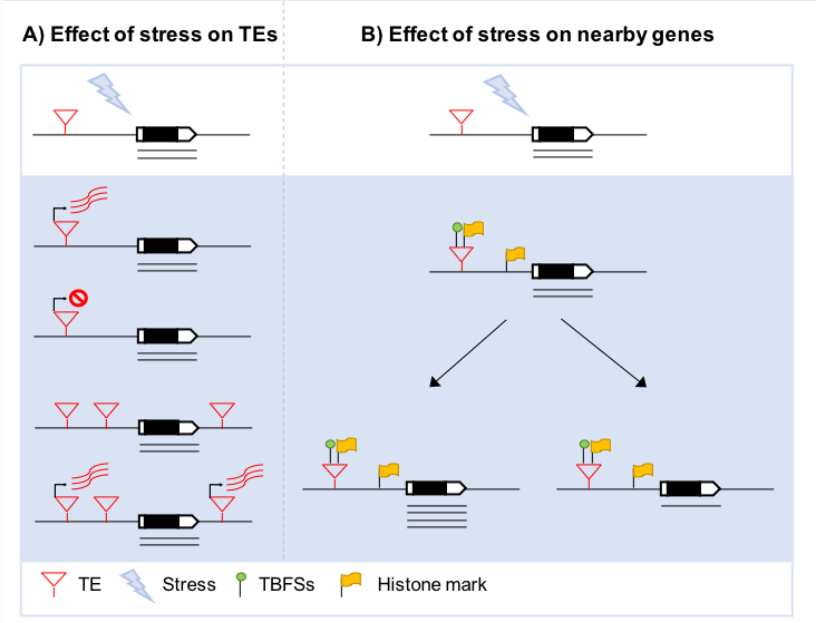


Figure 1.5. Effect of stress on TEs and on nearby genes. A) Stress can affect TEs both promoting and inhibiting their transcriptional activation. Moreover, it can also increase TE transposition. **B)** TEs can harbor stress-response regulatory sequences that may be activated by stress, thus affecting the expression of their nearby genes. Black boxes represent exons and white boxes represent UTR regions. TE transcripts are represented as red lines while gene transcripts are represented as short black lines.

There are several studies evidencing TE activation under stress in different organisms. In *Schizosaccharomyces pombe*, TE mobility was reported under different stressors such as heavy metals, caffeine, and plasticizer phthalate (Esnault et al., 2019) and in the tomato plant, drought stress triggered the accumulation of transcripts from the *Rider* element (Benoit et al., 2019).

Sometimes, the molecular mechanisms behind the activation of TEs under stress have been identified. For example, in humans, stress causes the relocation to DNA damage sites of the longevity regulating protein Sirtuin 6 (SIRT6) that under normal conditions is silencing *LINE-1* elements, therefore leading to TE activation (Van Meter et al., 2014). In the Drosophila germline, the Piwi-interacting RNA (piRNA) pathway operates to control TE activity. Heat-shock stress has been reported to alter piRNA biogenesis through the action of the inducible chaperone *Hsp70* causing an increased TE activation at a post-transcriptional level (Cappucci et al., 2019). However, besides TE activation, stress has sometimes been related with TE repression. The expression of the *Ty3* element in *Saccharomyces cerevisiae* was repressed under heat stress and after an ethanol exposure (Menees and Sandmeyer, 1996). Similarly, a short-term treatment with morphine reduced the expression of *LINE-1* elements in human neuronal cells (Trivedi et al., 2014).

TE activation under stress is often associated with changes in the expression of their nearby genes (Guio et al., 2014; Mateo et al., 2014., Makarevitch et al., 2015; Bouttier et al., 2016; Chuong et al., 2016; Forestan et al., 2016; Persson et al., 2016; Zovoilis et al., 2016; Hummel et al., 2017; Le et al., 2017; Villanueva-Cañas et al., 2019) (Figure 1.5). In a recent study that analyses the role of TEs in six different Drosophila stress regulatory networks, the authors found that TEs have a relevant role in the transcriptional regulation of stress-response genes (Villanueva-Cañas et al., 2019). The authors functionally validate the regulatory activity of 6 different TEs by performing *in vivo* enhancer reporter assays. They found that while some TEs down-regulated the expression of the reporter gene, others were promoting its expression (Villanueva-Cañas et al., 2019). Besides adding cis-regulatory regions, other mechanisms have been described to alter gene expression under stress conditions (Figure 1.5). For instance, the *Bari-Jheh* insertion has been associated with the addition of histone marks in the intergenic region of its nearby genes modulating their expression (Guio et al., 2018).

Finally, a positive fitness effect of TEs under stress conditions is often assumed although only a small number of studies have provided evidence at the phenotypic level (Guio et al., 2014; Mateo et al., 2014; Zhao et al., 2014; Bouttier et al., 2016; Chuong et al., 2016; Horváth et al., 2017). However, the effect of stress on TEs has also been related with negative effects such as cancer and neurodegenerative diseases (Rishishwar et al., 2017).

With all, the relationship between TEs and stress is quite complex and seems context dependent. Nevertheless, there vast majority of studies assume that the activation of TEs under stress generates genome variability upon which natural selection can act.

1.7. Transposable elements as effective drivers of adaptation

Due to its mutagenic nature, the vast majority of TE insertions have deleterious or neutral effects. In humans, the insertion of a *LINE-1* element into one of the exons of the factor VIII gene causes a disruption of the gene that triggers haemophilia (Kazazian et al., 1988). Another *LINE-1* element inserted in the last exon of the adenomatous polyposis coli (APC) gene was associated with the disruption of this tumour suppressor gene producing colon cancer (Miki et al., 1992). Taking advantage of this mutagenic capacity, TEs have been used as mutagenic agents in genetic experiments with Drosophila. Several Drosophila collections exist nowadays with TE mutations in different genes (Peter et al., 2002; Beinert et al., 2004; Bellen et al., 2004; Thibault et al., 2004; Metaxakis et al., 2005; Staudt et al., 2005).

Besides having deleterious or neutral effects, TEs have been sometimes related with adaptive processes. Organisms have to deal with continuous and diverse environmental changes, both biotic and abiotic. One of the mechanisms by which organisms react to these challenges is through adaptive evolution to the new environmental conditions. The process of adaptive evolution takes place by natural selection when organisms that are better suited to the new environment transfer more genes to the next generation. Considering that, the genetic variance conferring the advantage can increase in frequency within the population. Adaptation, as a result of natural selection, relies in the capacity of mutations to generate genetic diversity. TEs are a remarkably source of genetic and epigenetic variation and in the same way, their activity can be susceptible to environmental changes. For this reason, TEs are considered to play a role in adaptation in different organisms. However, they have long been ignored as candidate mutations involved in adaptation.

IS elements in bacteria have been associated to environmental adaptation in several studies. For instance, a cause-effect relationship has been established between *IS* elements and adaptation to high osmolarity (Stoebel et al., 2009; Stoebel and Dorman, 2010) and to conditions with limited amounts of metals and nutrients (Chou et al., 2009; Gaffé et al., 2011).

In plants, adaptation has also been associated to TE-induced mutations. TE insertions are associated with adaptation to high latitudes in the soybean (Liu et al., 2008; Kanazawa et al., 2009) and to changing light environments in Arabidopsis (Lin et al., 2007). In addition, the peppered moth *carb* insertion was found to up-regulate the *cortex* gene resulting in an increased dark coloration of this organism. This phenotype is thought to improve fitness in polluted areas (Van't Hof et al., 2016).

1.7.1. Transposable elements as effective drivers of adaptation in *Drosophila melanogaster*

As mentioned above, the vast majority of TE-induced mutations are deleterious and thus present at low population frequencies. Then, TEs present at high population frequencies are more likely to play a role in adaptation (Barrón et al., 2014). This reasoning was applied in *D. melanogaster* where the efficiency of selection in removing deleterious mutations is high due to its big population size (Kofler et al., 2012; Cridland et al., 2013; Barrón et al 2014., Blumenstiel et al., 2014). *D. melanogaster* was originated from southern Africa and it spread around the word in a short period of time (19,000 years ago) suggesting that recent adaptation to different environments must be common (David and Capy, 1988; Li and Stephan, 2006; Arguello et al., 2019).

In the first genome-wide study analysing TE-induced adaptive mutations, authors showed that TEs have contributed to *D. melanogaster* recent adaptation to out-of-Africa (González et al., 2008). Based on a pooled-PCR screening in North America natural populations, they found 18 putatively adaptive TEs (González et al., 2008; González et al., 2010). Five of them showed signatures of selection in their flanking regions and eight present population differentiation (González et al., 2008). Later, an *in-silico* genome-wide study using pooled DNA sequences from a European population found 13 TE candidates for adaptation (Kofler et al., 2012). These 13 TEs were found at high population frequencies and had evidences of positive selection in their flanking regions (Kofler et al., 2012). Another genome-wide study analysed by PCR two populations from North America and Africa, and identified a total of nine candidate adaptive TEs (Blumenstiel et al., 2014). Finally, a more recent analysis of 1,615 annotated TE insertions in 60 worldwide natural populations identified 300 TEs present at high population frequencies likely due to the action of positive selection (Rech et al., 2019). 84 of these putatively adaptive insertions showed evidences of selection (Rech et al., 2019). Moreover, gene ontology analysis

revealed that their nearby genes are involved in stress response, behaviour, and developmental processes (Rech et al., 2019).

So far, the link between the genotype and the adaptive phenotype has been stablished for some of these putatively adaptive TEs (Daborn et al., 2002; Aminetzach et al., 2005; Schmidt et al., 2010; Magwire et al., 2011; Mateo et al., 2014; Guio et al., 2014; Ullastres et al., 2015; Ullastres et al., 2019) (Figure 1.6). One of these TEs is the *Accord* element inserted in the 5′ end of the *Cyp6g1* gene that has been associated with resistance to pesticides (Daborn et al., 2002). Moreover, duplications of the region containing the *Cyp6g1* allele and additional TE insertions resulted in further resistance to pesticides (Schmidt et al., 2010). Another TE belonging to the *Doc* family inserted into the protein coding region of *CHKov1* has been related to resistance to organophosphate insecticides and to viral infection. The insertion truncates the *CHKov1* gene and generates a functional protein during the process that confers resistance to pesticides (Aminetzach et al., 2005). The same insertion was later associated with resistance to viral infection (Magwire et al., 2011). On top of that, a complex rearrangement that results in a duplication of the allele containing the *Doc* element, further increases resistance to virus (Magwire et al., 2011). The *pogo* transposon *FBti0019627* provides an alternative poly(A) signal to the *CG11699* gene causing a shorter transcript (Mateo et al., 2014). The presence of this insertions was also associated with increased expression levels of *CG11699* leading to resistance to different xenobiotics through an increased ALDH-III activity (Mateo et al., 2014). Altogether, the fact that *FBti0019627* has recently increased its frequency in out-of-Africa populations could be due to positive selection (González et al., 2008; Mateo et al., 2014). Another candidate TE, the *Bari-Jheh* element, was associated with increased expression levels of its nearby genes *Jheh1* and *Jheh2* and with resistance to oxidative stress and tolerance to bacterial infection (Guio et al., 2014; Guio et al., 2016). Finally, the *FBti0019386* insertion belonging to the *invader4* family was associated with increased expression of its nearby gene *sra* and with the addition of an alternative TSS to *Bin1* only under immune-stress conditions. Flies with *FBti0019386* were associated with shorter developmental time and with tolerance to bacterial infection (Ullastres et al., 2015; Ullastres et al., 2019).

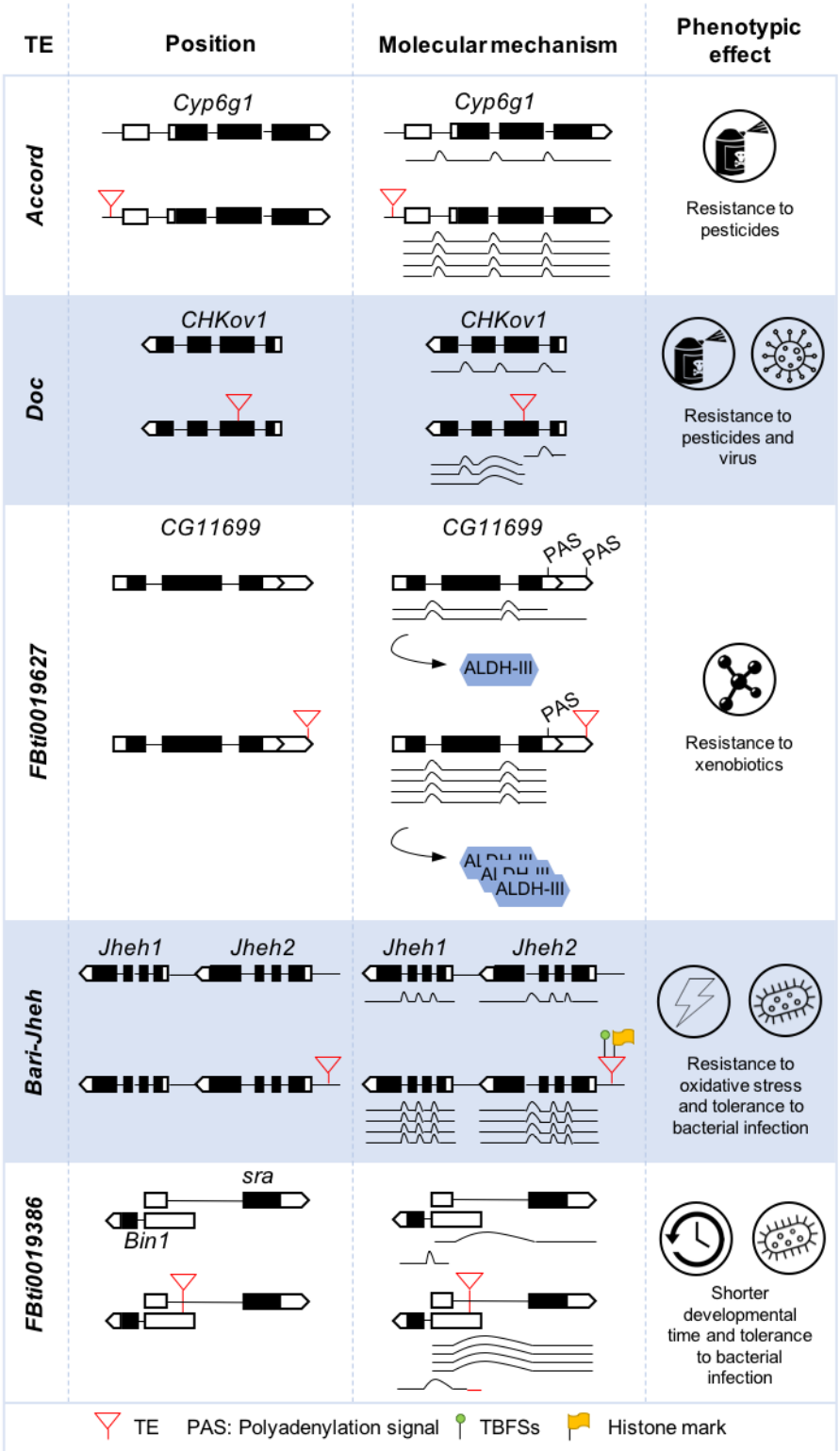


Figure 1.6. Schematic representation of the molecular and phenotypic effect of five adaptive TE insertions described in *D. melanogaster*. The link between the genotype and the adaptive phenotype has been stablished for five putatively adaptive TEs in *D. melanogaster*. Black boxes represent exons and white boxes represent UTR regions. Gene transcripts are represented as short black lines.

All the studies mentioned above were performed based on association studies analysing the organismal phenotypic response under different conditions in populations with and without the candidate TE insertions. However, complementary experiments would be necessary to further confirm the causal links between the TE-induced mutations and their phenotypic effect. The discovery of the novel CRISPR/Cas system has improved the field of genotype to phenotype mapping thanks to its capacity to generate precise cuts on target sequences. The use of CRISPR/Cas system has already been used to precisely excide TEs in different organisms such as *Drosophila simulans* (Ding et al., 2016), humans (Chuong et al., 2016), and mouse (Todd et al., 2019) allowing the characterization of TE-derived effects.

1.7.2. The *Drosophila melanogaster* transposable element insertion *FBti0019985*

FBti0019985 is a 428bp retroelement belonging to the *roo* family found in *D. melanogaster*. It is a non-full-length element comprised of a single *LTR* sequence (*solo-LTR* element) (Thurmond et al., 2019). Thus, *FBti0019985* is a non-autonomous TE with no transposition activity. It is inserted in the 5'UTR region of *CG18446* gene, which overlaps with the first intron of *cbx* gene (Thurmond et al., 2019). It has been found that *FBti0019985* provides an alternative TSS to its nearby gene *CG18446* (Batut et al., 2013).

FBti0019985 was first analysed in a genome-wide screening for adaptive insertions in *D. melanogaster* (González et al., 2008). In this study, authors found that *FBti0019985* was frequent in North American while rare in African population pools (González et al., 2008). However, *FBti0019985* PCR screenings with different individual strains from North America showed variability in the amplicon length, thus do not guaranteeing its high frequency in North America (González et al., 2008). Because the population frequency of *FBti0019985* was not accurately determined, this insertion was not further studied and therefore, not categorized as a putative adaptive TE insertion (González et al., 2008).

In this work, we study in detail the genomic region where *FBti0019985* is inserted in order to figure out why PCR results showed inconsistent band lengths. Furthermore, we also investigate its molecular and phenotypic function trying to link them with positive fitness-effects related to an out-of-Africa adaptation process.

SECTION 2

OBJECTIVES

2. OBJECTIVES

The objectives of the present thesis are:

1. **To study in detail the genomic region where *FBti0019985* is inserted**

   I will study the region where *FBti0019985* is inserted to figure out why previous PCR results with different individual strains from North America showed variability in the amplicon lengths. To that end, I will screen the *CG18446* promoter region in different natural populations from different geographic locations.

2. **To molecularly and phenotypically characterize the effect of *FBti0019985***

   I will explore the possible adaptive phenotypes associated with *FBti0019985* and the molecular mechanisms behind its effects. To do that, I will perform phenotypic experiments, expression analysis, and *in vivo* enhancer assays under different environmental conditions and in different developmental stages. I will also generate mutant fly strains without *FBti0019985* using the CRISPR/Cas9-mediated homology-directed repair technique.

SECTION 3

RESULTS

3. RESULTS

3.1. Structure of the research

The results of the present thesis are divided in three sections. Section 3.3 and section 3.4 correspond to published articles.

Section 3.3. Multiple independent retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in *Drosophila*.

In this section, we analysed by PCR the proximal promoter region of *CG18446* in four different *D. melanogaster* natural populations. We identified nine independent *roo* transposable element insertions in this proximal promoter region. We also investigated whether the identified insertions were functionally equivalent by performing 5'RACE, gene expression, and cold-stress survival experiments. We found that *FBti0019985* and *roo*$_{-7}$ provided an alternative transcription start site to *CG18446*. Finally, we found that only *FBti0019985* was associated with *CG18446* up-regulation in embryos and with increased viability in nonstress and under cold-stress conditions.

Section 3.4. A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila melanogaster*

In this section, we further analysed the proximal promoter region of *CG18446* performing a PCR screening in 218 strains from 15 different natural population. In this exhaustive screening we identified 11 *roo* insertions not described before in the proximal promoter region of *CG18446*. We also suggested that the presence of recurrent *roo* insertions in this region is likely to be the result of several bursts of transposition. Finally, we found that the *roo* insertional cluster in *CG18446* is unique in the *D. melanogaster* genome.

Section 3.5. A versatile transposable element affects the expression of a transcription factor depending on the developmental stage and the environmental conditions in *Drosophila melanogaster*

In this section, we further investigated the molecular and phenotypic effects of *FBti0019985*. Performing *in vivo* enhancer reporter assays and gene expression analysis we found that *FBti0019985* affects the expression of its nearby gene *CG18446* depending on the developmental stage and the environmental conditions. Moreover, we generated *FBti0019985* deletions in *D. melanogaster* natural populations using the CRISPR/Cas9-mediated homology-

directed repair technique. We then associated the presence of *FBti0019985* with tolerance to *P. entomophila* infection.

The main findings of all three sections are then discussed in Section 4 (Discussion) and conclusions are presented in Section 5. References quoted in Section 1 (Introduction) and Section 4 (Discussion) are included in Section 6 (References), while Sections 3.3, 3.4, and 3.5 include their own References section.

3.2. Thesis advisor's report about authorship and impact factor of the publications of this doctoral thesis presented by Miriam Merenciano González

Publication 1: Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*

Miriam Merenciano, Anna Ullastres, M. A. R. de Cara, Maite G. Barrón, Josefa González

PLOS Genetics. 12(8), e1006249. doi:10.1371/journal.pgen.1006249

Candidate tasks: in this publication the student has designed the research, performed the research, analysed the data and drafted the article.

PLOS Genetics impact factor is 6.100 and it is ranked the 16th of 167 journals in the "Genetics and Heredity" category (JCR 2016).

Publication 2: A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila melanogaster*

Miriam Merenciano, Camillo Iacometti, Josefa González

Mobile DNA. 10,10. doi:10.1186/s13100-019-0152-9

Candidate tasks: in this publication the student has designed the research, performed the research, analysed the data and drafted the article.

Mobile DNA impact factor is 3.630 and it is ranked the 50th of 173 journals in the "Genetics and Heredity" category (JCR 2018).

Publication 3 *(in preparation)*: A versatile transposable element affects the expression of a transcription factor depending on the developmental stage and environmental conditions in *Drosophila melanogaster*

Miriam Merenciano and Josefa González

Candidate tasks: in this publication the student has designed the research, performed the research, analysed the data and drafted the article.

This article will be submitted to Molecular Biology and Evolution. Molecular Biology and Evolution impact factor is 14.797 and it is ranked the 2nd of 50 journals in the "Evolutionary Biology" category (JCR 2018).

Josefa González Pérez

Barcelona, December 2019

3.3. Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*

Resum

Els promotors són regions reguladores de gens estructural i funcionalment diversos. La presència o absència d'unitats de seqüència i l'espai que hi ha entre elles defineixen les propietats dels promotors. Anàlisis recents d'usos alternatius de promotors en *Drosophila melanogaster* han revelat que els transposons contribueixen considerablement a promoure'n la seva diversitat. En aquest treball, analitzem en detall un transposó anomenat *FBti0019985* que ha estat incorporat per promoure l'expressió del gen *CG18446*, un gen candidat de resposta a estrès. Es van analitzar soques de diferents poblacions naturals i es va trobar que a part del transposó *FBti0019985*, hi ha unes altres vuit insercions independents a la regió promotora proximal de *CG18446*. Les nou insercions son solo-LTRs que pertanyen a la família *roo*. Es van analitzar les seqüències de les nou insercions *roo* i es va investigar si les diferents insercions eren equivalents funcionalment realitzant 5'-RACE, anàlisis d'expressió gènica i experiments de supervivència al fred. Es va trobar que les diferents insercions tenen diferents conseqüències tant moleculars com funcionals. La posició exacta on els transposons es troben inserits importa, ja que tots ells mostren un alta similitud en la seva seqüència, però només dos de les insercions analitzades proporcionen un inici de transcripció alternatiu i només la inserció *FBti0019985* afecta consistentment l'expressió de *CG18446*. Les conseqüències fenotípiques de les diferents insercions també varien: només *FBti0019985* es va associar amb tolerància al fred. Curiosament, l'única informació anterior sobre transposons inserits repetida i independentment en una regió promotora en *D. melanogaster* també va ser localitzada riu amunt d'un gen de resposta a estrès. Els nostres resultats suggereixen que la validació funcional de variants estructurals individuals és necessària per resoldre la complexitat de les agrupacions d'insercions.

PLOS | GENETICS

RESEARCH ARTICLE

Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*

Miriam Merenciano[1], Anna Ullastres[1], M. A. R. de Cara[2], Maite G. Barrón[1], Josefa González[1]*

1 Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona. Spain, **2** Laboratoire d'Eco-anthropologie et Ethnobiologie, UMR 7206, CNRS/MNHN/Universite Paris 7, Museum National d'Histoire Naturelle, F-75116 Paris, France

* josefa.gonzalez@ibe.upf-csic.es

OPEN ACCESS

Citation: Merenciano M, Ullastres A, de Cara MAR, Barrón MG, González J (2016) Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*. PLoS Genet 12(8): e1006249. doi:10.1371/journal.pgen.1006249

Editor: Cédric Feschotte, University of Utah School of Medicine, UNITED STATES

Received: December 28, 2015

Accepted: July 18, 2016

Published: August 12, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Ministerio de Economía y Competitividad (MINECO: http://www.idi.mineco.gob.es) BFU-2011-24397 and RYC-2010-07306 to JG, and BES-2012-052999 to AU, Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional (MINECO/FEDER:http://www.idi.mineco.gob.es) BFU2014-57779-P, the European Commission (https://ec.europa.eu) FP7-PEOPLE-2011-CIG-293860, the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement

Abstract

Promoters are structurally and functionally diverse gene regulatory regions. The presence or absence of sequence motifs and the spacing between the motifs defines the properties of promoters. Recent alternative promoter usage analyses in *Drosophila melanogaster* revealed that transposable elements significantly contribute to promote diversity. In this work, we analyzed in detail one of the transposable element insertions, named *FBti0019985*, that has been co-opted to drive expression of *CG18446*, a candidate stress response gene. We analyzed strains from different natural populations and we found that besides *FBti0019985*, there are another eight independent transposable elements inserted in the proximal promoter region of *CG18446*. All nine insertions are solo-LTRs that belong to the *roo* family. We analyzed the sequence of the nine *roo* insertions and we investigated whether the different insertions were functionally equivalent by performing 5'-RACE, gene expression, and cold-stress survival experiments. We found that different insertions have different molecular and functional consequences. The exact position where the transposable elements are inserted matters, as they all showed highly conserved sequences but only two of the analyzed insertions provided alternative transcription start sites, and only the *FBti0019985* insertion consistently affects *CG18446* expression. The phenotypic consequences of the different insertions also vary: only *FBti0019985* was associated with cold-stress tolerance. Interestingly, the only previous report of transposable elements inserting repeatedly and independently in a promoter region in *D. melanogaster*, were also located upstream of a stress response gene. Our results suggest that functional validation of individual structural variants is needed to resolve the complexity of insertion clusters.

de la Generalitat de Catalunya (http://agaur.gencat.cat) 2014-SGR-201 to JG, and by the Agence Nationale de la Recherche (http://www.agence-nationale-recherche.fr) ANR-14-CE02-0003 to MARdC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

The presence of several transposable element insertions in the promoter region of a *Drosophila melanogaster* gene has only been described in *heat shock protein* genes. In this work, we have discovered and characterized in detail several naturally occurring independent transposable element insertions in the promoter region of a cold-stress response gene in the fruitfly *Drosophila melanogaster*. The nine transposable element insertions described are clustered in a small 368 bp region and all belong to the same family of transposable elements: the *roo* family. Each individual insertion is present at relatively low population frequencies, ranging from 1% to 17%. However, the majority of strains analyzed contain one of these nine *roo* insertions suggesting that this region might be evolving under positive selection. Although the sequence of these insertions is highly similar, their molecular and functional consequences are different. Only one of them, *FBti0019985*, is associated with increased viability in nonstress and in cold-stress conditions.

Introduction

Promoters are crucial regions for the transcriptional regulation of gene expression. Recent computational and experimental advances in functional genomics techniques have allowed defining the promoter architecture to an unprecedented level. Several core promoter motifs such as the *Initiator* (*Inr*) and the *Downstream core Promoter Element* (*DPE*) have been described, and it is likely that many others remain to be discovered. The presence or absence of the core promoter motifs influences enhancer-promoter communication and thus gene regulation [1]. Promoter regions also harbour transcription factor binding motifs, which are another important component in the regulation of gene expression [2]. Besides cis-regulatory elements that influence the temporal and spatial expression patterns of genes, proximal promoters often contain alternative transcription start sites (TSSs) [1, 3]. Rather than being "biological noise" from imprecise binding of the transcription initiation machinery, genome-wide analyses of TSSs usage showed that alternative TSSs play an important role in the diversification of gene expression patterns [4–8].

Transposable elements (TEs), long proposed to play an important role in gene regulation [9, 10], have recently been found to provide at least 1,300 alternative TSSs in the *Drosophila melanogaster* genome [8]. TEs can also add Transcription Factor Binding Sites (TFBSs) to the promoter of genes as has been recently shown in Drosophila and humans [11–13]. As a result of adding particular sequence elements, many TEs confer their intrinsic regulatory properties to nearby genes demonstrating that they distribute cis-regulatory modules [8]. Finally, TEs inserted in promoter regions can also influence gene expression by disrupting the promoter architecture. This is the case, for example, of naturally occurring *P-element* insertions in the promoter of *heat shock protein* (*hsp*) genes [14].

One of the TEs identified as providing an alternative TSS by Batut *et al* (2013) [8], named *FBti0019985*, was previously reported in a screening designed to identify putatively adaptive TE insertions in *D. melanogaster* [15]. However, this particular TE was not further studied because its population frequency could not be accurately determined [15]. *FBti0019985* is a *roo* solo-LTR inserted in the 5'-UTR of *CG18446* gene, which is nested in the first intron of *crossbronx* (*cbx*) (Fig 1). TEs from the *roo* family have long been proposed to affect the expression of nearby genes by adding and distributing cis-regulatory regions [16–19]. Specifically, *roo* LTRs contain several TFBSs and the *Inr* sequence characteristic of core promoters [8, 20].

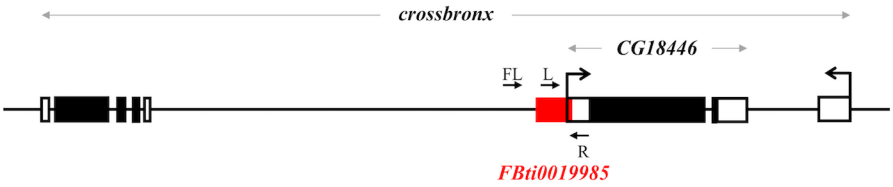


Fig 1. *FBti0019985* is inserted in the first intron of *cbx* gene and it overlaps with *CG18446* 5'-UTR region. Schematic representation of the genomic region where *FBti0019985* is inserted: chromosome 2R: 9,864,510–9,875,072. *FBti0019985* is shown in red. Black boxes represent exons and white boxes represent the 5'-UTRs and the 3'-UTRs. Primers used to check for the presence/absence of *FBti0019985* are depicted as black arrows (FL, R, and L; see Materials and Methods).

doi:10.1371/journal.pgen.1006249.g001

Interestingly, *CG18446* has been identified as a candidate gene for cold resistance: it is upregulated in fly strains that have been selected for increased cold resistance compared with control strains that were not subjected to cold-stress [21]. Cold resistance is an ecologically and evolutionarily relevant trait because it influences the ability of the species to adapt to different climatic conditions and thus, their geographical distribution [22, 23]. There is good evidence suggesting that *D. melanogaster* adapts to cold environments and a growing list of candidate genes involved in this thermotolerance phenotype is being identified [21, 24–28]. However, the molecular variants responsible for the adaptive cold-stress resistance phenotype remain elusive [29].

In this work, we further analyzed the presence/absence of *FBti0019985* in four natural populations of *D. melanogaster*. We found that besides *FBti0019985*, eight other *roo* elements have inserted in a 368 bp region around *CG18446* transcript start site. These *roo* elements differ in the insertion site and in their orientation. On the other hand, all elements have the same size and show high sequence conservation: all cis-regulatory elements previously described in *roo* LTRs are highly conserved [8, 30]. We further investigated whether these different insertions were functionally equivalent by performing 5'-RACE, gene expression, and phenotypic analyses. Our results showed that the functional consequences of the different *roo* insertions depend on the particular position where the element is inserted. Among the nine different *roo* solo-LTR insertions, only *FBti0019985* is consistently associated with increased viability in nonstress and cold-stress conditions across genetic backgrounds.

Results

Besides *FBti0019985*, eight other *roo* solo-LTRs are inserted in the promoter region of *CG18446*

We first aimed at estimating the frequency of *FBti0019985* in non-African natural *D. melanogaster* populations. Thus, we checked using PCR whether this insertion was present, polymorphic, or absent in 28 strains from a natural population collected in North Carolina (North America, DGRP strains [31, 32]) and in 15 strains from a natural population collected in Bari (Italy, Europe [33]) (Table 1). We obtained PCR results for 39 of the 43 strains tested: nine strains produced PCR bands consistent with *FBti0019985* being present, five strains appeared as heterozygous, 13 strains showed unexpected band patterns, and 12 strains appeared as absent (Table 1) (see Material and Methods). To verify these results, we sequenced 32 of the 39 strains including all the strains that showed some evidence of presence (Table 1).

Only four of the nine strains classified as present, according to the PCR results, had the *FBti0019985* insertion. For the rest of this work, we considered the position where *FBti0019985* is inserted as the "reference position". The other five present strains, the five heterozygous strains, and 12 of the 13 strains that gave unexpected PCR bands contained different *roo* solo-

Table 1. The nine *roo* solo-LTR insertions analyzed in this work.

Insertion	Fly strain	PCR results	Sequenced band	Insertion position[a]
FBti0019985	*RAL-639*	Present	FL-R	Reference position
	RAL-802	Present	FL-R / L-R	Reference position
	RAL-810	Present	FL-R	Reference position
	IV68	Present	FL-R / L-R	Reference position
***roo*$_{+7}$**	*RAL-405*	Present	FL-R	+ 7 bp
	RAL-887	Present	FL-R	+ 7 bp
	RAL-911	Present	FL-R	+7 bp
	RAL-441[b]	Larger L-R	FL-R	+ 7 bp
		Larger FL-R		
	RAL-801[b]	Larger L-R	FL-R	+ 7 bp
		Larger FL-R		
***roo*$_{+175}$**	*IV145*	Heterozygous	Larger FL-R	+ 175 bp
***roo*$_{+278}$**	*RAL-502*	Smaller L-R	L-R	+ 278 bp
		No FL-R		
***roo*$_{-19}$**	*IV42*	Present	FL-R	- 19 bp inverted
	IV127	Present	FL-R	- 19 bp inverted
***roo*$_{-28}$**	*IV40*	Heterozygous	Larger FL-R	- 28 bp inverted
***roo*$_{-44}$**	*RAL-195*	Only FL-R	FL-R	- 44 bp inverted
	RAL-383	Only FL-R	FL-R	- 44 bp inverted
***roo*$_{-68}$**	*RAL-75*	Only FL-R	FL-R	- 68 bp inverted
	RAL-716	Only FL-R	FL-R	- 68 bp inverted
	IV69	Heterozygous	Larger FL-R	- 68 bp inverted
***roo*$_{-90}$**	*RAL-21*	Larger L-R	FL-R	- 90 bp
	RAL-88	Larger L-R	FL-R	- 90 bp
	RAL-177	Larger L-R	FL-R	- 90 bp
	RAL-737	Larger L-R	FL-R	- 90 bp
	RAL-820	Larger L-R	FL-R / L-R	- 90 bp
	RAL-857	Heterozygous	FL-R / L-R	- 90 bp
	IV50	Heterozygous	Larger FL-R	- 90 bp
Absent	*RAL-40*	Smaller L-R	FL-R / L-R	Absent
	RAL-371	Absent	FL-R	Absent
	RAL-391	Absent	FL-R	Absent
	RAL-508	Absent	NS	Absent
	RAL-783	Absent	FL-R	Absent
	RAL-822	Absent	NS	Absent
	RAL-855	Absent	NS	Absent
	RAL-908	Absent	FL-R	Absent
	IV22	Absent	FL-R	Absent
	IV49	Absent	NS	Absent
	IV52	Absent	NS	Absent
	IV72	Absent	NS	Absent
	IV75	Absent	NS	Absent
No data	*RAL-776*	No results		
	IV33	No results		
	IV125	No results		
	IV148	No results		

NS, not sequenced

[a] "+" indicates the insertion is downstream of *FBti0019985* and "-" indicates the insertion is upstream of *FBti0019985*

[b]These strains have a 95 bp duplication upstream of the insertion

doi:10.1371/journal.pgen.1006249.t001

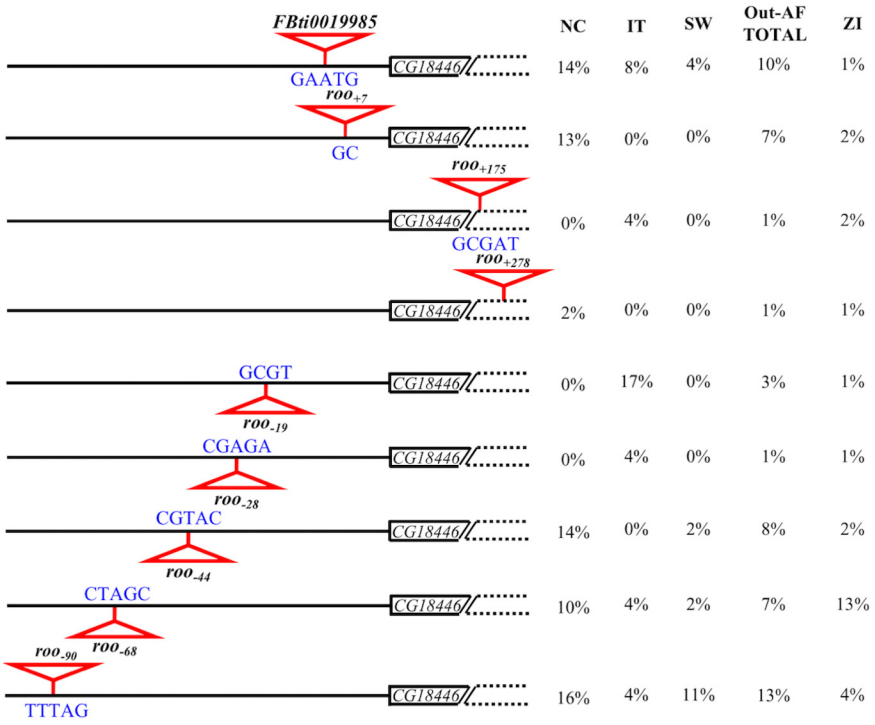


Fig 2. Besides *FBti0019985*, eight other *roo* solo-LTR are inserted in the proximal promoter of *CG18446*. Schematic representation of the genomic region where the nine *roo* solo-LTRs are inserted. *roo* insertions are depicted as red triangles. White boxes represent *CG18446* 5'-UTR. Regions depicted with dotted lines are not drawn to scale. Target Site Duplications (TSDs) are shown in blue. NC, allele frequency (%) in the North American population; IT, allele frequency (%) in the Italian population; SW, allele frequency (%) in the Swedish population; Out-AF total, allele frequency (%) in all the out-of-Africa populations; ZI, allele frequency (%) in the Zambia population.

doi:10.1371/journal.pgen.1006249.g002

LTR insertions (Table 1). Overall, besides *FBti0019985*, we found eight other 428 bp *roo* solo-LTRs inserted in eight different positions (Fig 2). Three *roo* insertions are located downstream of the reference position: roo_{+7}, roo_{+175}, and roo_{+278} (Fig 2). Two of the four strains carrying roo_{+7} have a duplication of the 95 bp region located immediately upstream of the insertion (Table 1). roo_{+175} element is inserted in the 5'-UTR region, and roo_{+278} is inserted in the first exon of *CG18446* gene. Both roo_{+175} and roo_{+278} have a conserved *Inr* motif. If transcription starts in these insertions, flies carrying roo_{+175} would have a 100 bp shorter 5'-UTR, and flies carrying roo_{+278} would have a 35 amino acids shorter *CG18446* protein. The other five *roo* insertions are located upstream of the reference position: roo_{-19}, roo_{-28}, roo_{-44}, roo_{-68}, roo_{-90} (Fig 2). Four of them, roo_{-19}, roo_{-28}, roo_{-44}, and roo_{-68}, are inserted in reverse orientation.

We used *Tlex-2* software to further analyze the frequency of the nine *roo* insertions in 21 additional DGRP strains, in 26 strains from a Swedish natural population, and in 42 strains from a population collected in the ancestral range of the species, Zambia (Fig 2 and S1 Table) (see Material and Methods) [34]. Overall, we found that 67 strains, out of the 128 strains analyzed, contained one of the nine *roo* solo-LTR insertions. The two most common *roo* insertion in out-of-Africa populations are roo_{-90} and *FBti0019985* present in 13% and 10% of the strains tested, respectively (Fig 2). Besides, some insertions are only present in the North Carolina natural population while others are specific to the Italian natural population (Fig 2). Only three of the nine insertions described in North Carolina and Italian populations are present in the Swedish population. However, we did not perform *de novo* discovery of TEs in this population.

Thus, it could be that other private insertions are present in the Swedish population. Finally, all the nine insertions were present in the African population although most of them were present at very low frequencies (Fig 2).

In summary, we have found that besides the *FBti0019985* insertion annotated in the reference genome, eight other 428 bp *roo* solo-LTRs are inserted nearby *CG18446* TSS in natural populations of *D. melanogaster* (Fig 2) [35]. Each one of the strains analyzed contains a single solo-LTR *roo* insertion and most of the analyzed strains contain one of the nine solo-LTR *roo* insertions.

The nine *roo* solo-LTR are independent insertions that occurred at different evolutionary timepoints

We identified the Target Site Duplications (TSD) of the nine different *roo* insertions using data from the 26 present strains sequenced in this work (Table 1). We could identify the TSD for all *roo* insertions except for roo_{+278}. We found that six of the eight TSDs identified are five nucleotides long as has been previously described for this family [36] (Fig 2). However, the TSD sequences did not match the proposed TSD consensus sequence [34, 36, 37]. We thus used all the available *roo* TSD sequences to build a new consensus (S1 Fig). The different *roo* solo-LTR insertions had different TSDs suggesting that they are independent insertions (Fig 2). Furthermore, all the *roo* elements located in a given insertion site have the same exact TSD and are inserted in the same orientation suggesting that each one of them is a unique insertion event (Fig 2).

To test whether these nine insertion events were the result of a burst of transposition, we constructed a phylogenetic tree. We included the nine *roo* insertions sequenced in this work and 115 other *roo* insertions present in the *D. melanogaster* genome (S2 Fig and S1 Text). We found that not all the newly described *roo* insertions clustered together suggesting that they did not insert at the same time (S2 Fig and S1 Text).

All the TEs identified in *CG18446* proximal promoter region belong to the *roo* family. Thus, we also investigated whether *roo* elements annotated in the reference genome are preferentially inserted into gene proximal promoter regions as has been previously described for other TE families [38, 39]. We analyzed the 138 insertions belonging to the *roo* family annotated in the *D. melanogaster* reference genome (v5). We found 21 *roo* insertions located in the 1 kb region upstream of a gene or overlapping the 5'-end of a gene. Thus, only 15.2% of the *roo* elements in the *D. melanogaster* genome are located in gene promoters and/or 5'-UTRs.

In summary, TSD analyses of the nine insertions characterized in this work suggested that they are independent insertions, and confirmed the length but not the sequence previously reported as the TSD consensus for this family. Our results are not consistent with the nine *roo* insertions being the result of a single burst of transposition. Finally, our analyses also suggested that *roo* elements do not preferentially insert in 5' gene regions.

The nine *roo* insertions add the same cis-regulatory sequences

We analyzed multiple sequence alignments of all the *roo* insertions located nearby *CG18446*. We identified TFBSs using the JASPAR database (see Material and Methods). We also specifically looked for conservation of the regulatory regions previously described in the *roo* family [8, 30], and for conserved core promoter motifs [1] (Fig 3A and S2A Table). Overall, there was very little diversity among the nine solo-LTRs (S3A Fig). The five TFBSs and the *Inr* sequence previously identified in the consensus sequence of *roo* LTRs are conserved in all the *roo* copies located in the proximal promoter of *CG18446* [8]. Additionally, we found another four TFBSs that are also highly conserved in all the copies (Fig 3A and S3A Fig). The nine transcription factors are involved in developmental processes. Additionally, *Deaf1* and *Nub* are also involved in immune response [40, 41]. Finally, three previously identified Matrix Associated Regions

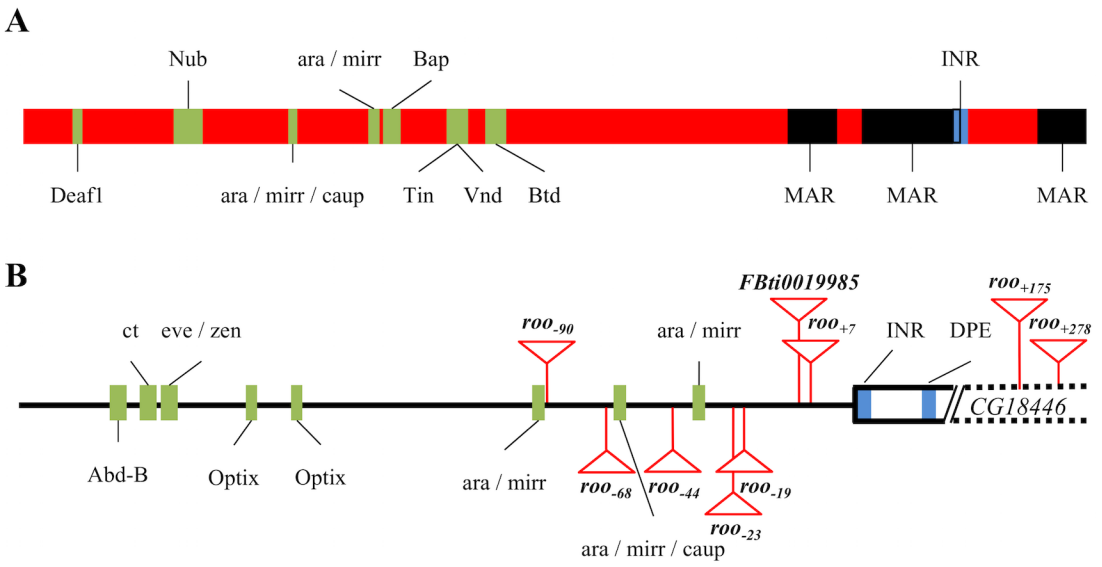


Fig 3. Conserved regulatory regions in the the nine *roo* solo-LTR insertions and in the proximal promoter region of *CG18446*. (A) Location of the nine transcription factor binding sites (green boxes), the *Inr* motif (blue box), and regions with matrix association potential (MARs) (black boxes), in the *roo* solo-LTR consensus sequence. *Deaf1*, *ara*, *mirr* and *caup* TFBS have been identified in this work. (B) Location of the eight transcription factor binding sites (green boxes) and the two core promoter motifs (blue boxes) in the proximal promoter region of *CG18446*. Different *roo* insertions are depicted as red triangles. The positions of roo_{+175} and roo_{+278} are not drawn to scale.

doi:10.1371/journal.pgen.1006249.g003

(MARs) in LTRs from the *roo* family are also highly conserved in the nine insertions (Fig 3A and S3B Fig) [30]. These results suggest that these *roo* solo-LTR insertions are introducing the same cis-regulatory regions in the *CG18446* proximal promoter region. Still, the functional effect of these insertions might be different because they are located in different positions and have different orientations (Fig 2).

roo insertions affect the spacing of Transcription Factor Binding Sites in the proximal promoter region of *CG18446*

We analyzed the proximal promoter region of *CG18446* in the 30 strains sequenced in this work. We could not identify the TATA box suggesting that *CG18446* has a DPE promoter [1]. We identified eight TFBSs in the proximal promoter of *CG18446* (Fig 3B and S2C Table). These eight TFBSs are highly conserved in all the strains analyzed (S3C Fig). The different *roo* insertions characterized in this work do not disrupt any of the identified core promoter motifs or TFBSs (Fig 3B). However, they do affect the spacing between the different regulatory motifs, which might affect the protein-protein interaction at the *CG18446* promoter and thus the expression level of this gene (Fig 3B) [14].

roo insertions could be recruiting the HP1a protein

Besides affecting the spacing of transcription factor binding site, another mechanism by which *roo* insertions could be affecting *CG18446* expression is by recruiting piRNAs that would lead to heterochromatin formation [42, 43]. We mapped piRNA reads from three different available libraries to a 1.4 kb region including *FBti0019985* (Fig 4A) (see Material and Methods) [44–46]. We found that most of the piRNAs mapping to the insertion were sense reads, suggesting that *FBti0019985* is not acting as a target for heterochromatin assembly [42].

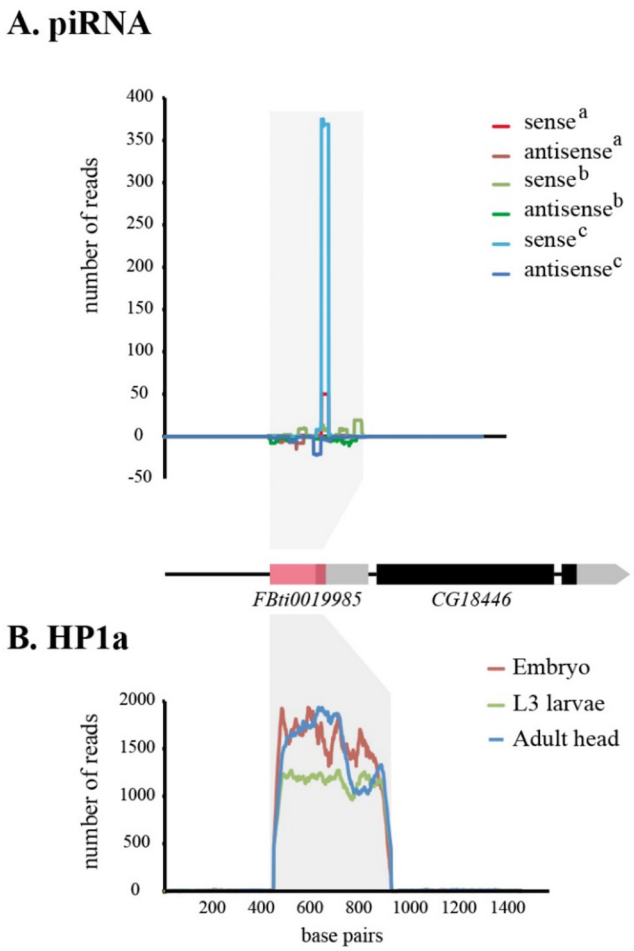


Fig 4. Mapping of piRNA reads and HP1a reads to the *FBti0019985* region. (A) Number of piRNA reads mapped to a 1.4 kb region including *FBti0019985*. [a]Li *et al* (2009) piRNA library, [b]Satyaki *et al* (2014) piRNA library and [c]Shpiz *et al* (2014) piRNA library. (B) Number of HP1a reads mapped to the same 1.4 kb region.

doi:10.1371/journal.pgen.1006249.g004

We also looked for evidence of HP1a binding to *FBti0019985* using modENCODE data (see Material and Methods) [47]. HP1a is a structural chromosomal protein that mediates both gene expression and gene silencing [48]. We did find evidence of HP1a reads binding to *FBti0019985* (Fig 4B). Thus, by recruiting HP1a, *FBti0019985* could be affecting the expression of *CG18446*. The same results were obtained for the other eight *roo* solo-LTR insertions: most of the piRNAs mapping to the insertions were sense reads and we found evidence of HP1a binding to all of them (S3 Table). Overall, our results are suggestive but not conclusive of HP1a binding to the nine *roo* insertions described in this work.

To further investigate the possible functional consequences of the *roo* insertions, we focused on the five insertions present at higher population frequencies in out-of-Africa populations: *FBti0019985*, roo_{+7}, roo_{-44}, roo_{-90}, and roo_{-68} (Fig 2).

Only *FBti0019985* and roo_{+7} affect the transcription start site of *CG18446*

We investigated whether *roo* insertions could be providing an alternative TSS to *CG18446*. Batut *et al* (2013) [8] reported that the TSS of *CG18446* is located inside *FBti0019985*. However,

this finding was obtained using RAMPAGE and was not further validated using 5'-RACE. For this reason, we performed a 5'-RACE with the *RAL-810* strain that carries *FBti0019985* and with the *RAL-783* strain that does not carry any of the nine *roo* solo-LTR insertions. As expected, we found that the TSS of *CG18446* is inside the TE: the first 50 bp of the 276 bp 5'-UTR correspond to *FBti0019985* (Fig 5). Additionally, flies with the insertion have also a shorter transcript, with a 201 bp 5'-UTR, that does not start in *FBti0019985* (Fig 5). Most of the sequenced transcripts start in the *FBti0019985* insertion (14 out of 20 transcripts analyzed). Flies without the *FBti0019985* insertion only have the 201 bp 5'-UTR transcript (Fig 5).

We then checked whether roo_{+7}, located only 7 bp downstream of *FBti0019985*, roo_{-90}, which is the most distal insertion, and roo_{-44}, which is inserted in reversed orientation, also provide an alternative TSS to *CG18446*. We found that roo_{+7} affects the TSS of *CG18446* (Fig 5). Indeed, the TSS in roo_{+7} is in the same nucleotide position as in *FBti0019985*. Thus, *CG18446* transcript in flies with roo_{+7} is 7 bp shorter compared with the transcript in flies with *FBti0019985*. Similarly to *FBti0019985*, most of the sequenced transcripts started in the roo_{+7} insertion (18 out of 22 transcripts analyzed). On the other hand, we did not find evidence of a TSS inside roo_{-90}, which might indicate that the distance of the TE to the nearby gene affects its ability to provide an alternative TSS (Fig 5). Finally, we analyzed two different strains carrying the roo_{-44} insertion in the same position and we could not find evidence for a transcript with the TSS in roo_{-44} (Fig 5).

Overall, we found that only *FBti0019985* and roo_{+7} insertions modify the length of *CG18446* transcript. These two *roo* insertions are located a few nucleotides from the gene and both are inserted in 5' to 3' orientation.

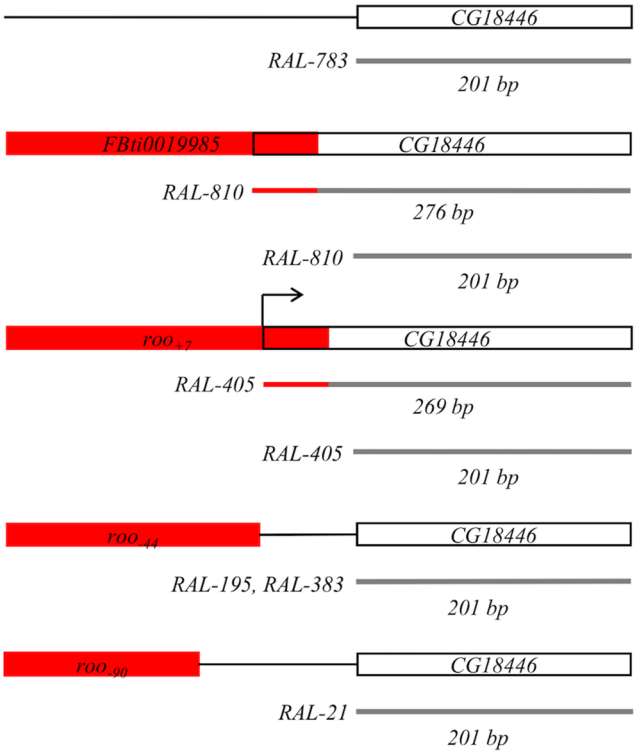


Fig 5. *FBti0019985* and roo_{+7} affect the transcription start site of *CG18446*. Schematic representation of the results obtained using the 5'-RACE technique. Red boxes represent different *roo* insertions and white boxes represent *CG18446* 5'-UTRs. Partial transcripts obtained by 5'-RACE are depicted as grey lines. The region of the transcript that overlaps with a *roo* insertion is shown as a red line. The last 50 bp of *FBti0019985* and roo_{+7} are included in the 5'-UTR of *CG18446*.

doi:10.1371/journal.pgen.1006249.g005

FBti0019985 is associated with changes in embryonic *CG18446* expression

We further analyzed whether different *roo* insertions were associated with changes in *CG18446* expression in embryos, where this gene is highly expressed [49]. For *FBti0019985*, we analyzed the expression of *CG18446* in flies with four different genetic backgrounds. In three of the four backgrounds, *FBti0019985* is associated with upregulation of *CG18446* (Fig 6A). This result is significant in two genetic backgrounds, *RAL-810* and *IV68*, and marginally significant in a third background, *RAL-639* (t-test p-value = 0.045, p-value = 0.005 and p-value = 0.062, respectively) (Fig 6A). On the other hand, only in one of the three genetic backgrounds analyzed for roo_{+7}, the insertion is associated with downregulation of this gene (t-test p-value = 0.015 for *RAL-405*) (Fig 6B).

We also checked the expression of *CG18446* in flies with two *roo* solo-LTR insertions that do not provide an alternative TSS to this gene: roo_{-90} and roo_{-44}. We found that roo_{-90} is only associated with *CG18446* upregulation in one of the three backgrounds analyzed (p-value = 0.001, for *RAL-21*) (Fig 6C). Two different strains with the roo_{-44} solo-LTR insertion

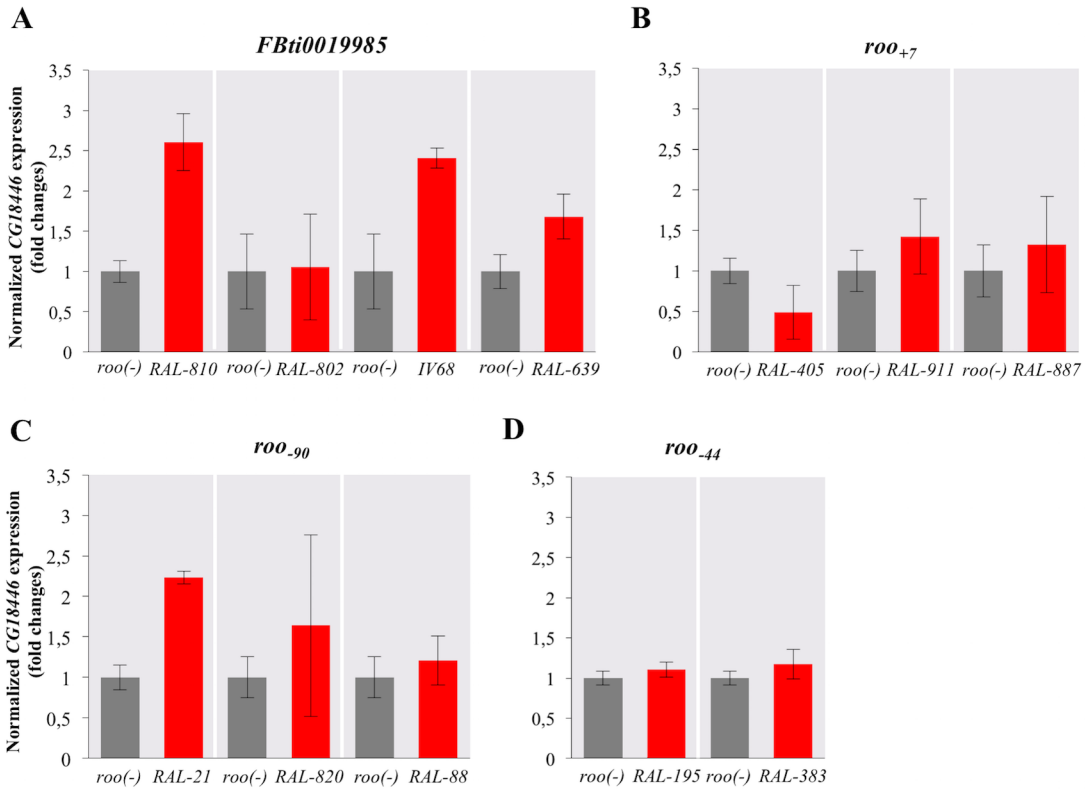


Fig 6. *FBti0019985* is associated with changes in *CG18446* expression. Normalized *CG18446* expression level relative to *Act5C* in embryos without *roo* insertion (grey) and in embryos with different *roo* insertions (red). (A) For *FBti0019985*, we compared the expression of *CG18446* in flies with four different genetic backgrounds. In three backgrounds, the presence of *FBti0019985* was associated with *CG18446* upregulation. These results were significant in two backgrounds, *RAL-810* and *IV68*, and marginally significant in the third background, *RAL-639*. (B) roo_{+7} was only associated with changes of expression in one of the three backgrounds analyzed: *RAL-405*. (C) roo_{-90} was also only associated with changes of expression in one of the three backgrounds analyzed: *RAL-21*. (D) Finally, roo_{-44} was not associated with changes in expression in any of the two backgrounds analyzed. Error bars represent the standard error of the mean (SEM) for the three biological replicates performed for each experiment.

doi:10.1371/journal.pgen.1006249.g006

did not show differences in the level of expression of *CG18446* compared with strains without the insertion (p-values > 0.05 in both cases) (Fig 6D).

Overall, we found that *FBti0019985* is associated with *CG18446* upregulation in three of the four backgrounds analyzed (Fig 6A). In the majority of strains, roo_{+7}, roo_{-90}, and roo_{-44} are not associated with changes in *CG18446* expression level (Fig 6B–6D). However, we can not discard that the presence of these insertions is associated with changes in the expression of *CG18446* in other developmental stages and/or in tissues not analyzed in this work.

FBti0019985 is associated with increased viability in nonstress and in cold-stress conditions

We have shown that *FBti0019985* affects the transcript length and it is associated with upregulation of *CG18446* in most of the genetic backgrounds analyzed (Figs 5 and 6A). Because *CG18446* has been previously identified as a cold-stress candidate gene, we tested whether flies with and without *FBti0019985* differed in their sensitivity to cold-stress [21]. We first compared *RAL-810*, which carries *FBti0019985*, with *RAL-783*, which does not carry any of the nine *roo* insertions (Fig 7A). We performed three biological replicates. ANOVA analyses showed that the experimental condition (nonstress or cold-stress) and the insertion genotype (presence or absence of *FBti0019985*) were significant (Table 2). Flies with *FBti0019985* had a higher viability than flies without this insertion in both nonstress and cold-stress conditions. Furthermore, the interaction between these two factors was also significant suggesting that the effect of the insertion is larger in cold-stress conditions (Fig 7A and Table 2).

We repeated the experiment using flies with different genetic backgrounds: *RAL-802* that carries *FBti0019985* and *RAL-908* that does not carry this insertion (Fig 7B). ANOVA analyses showed that the experimental condition and the insertion genotype are significant while the interaction between these two factors was not significant (Table 2). *RAL-802* flies had a higher egg-to-adult viability in nonstress and in cold-stress conditions compared with flies without *FBti0019985*.

Finally, we tested whether flies from a different population, *IV68* carrying *FBti0019985* and *IV22* without this particular insertion both collected in Italy, also showed significantly increased viability in nonstress and in cold-stress conditions (Fig 7C and Table 2). We found

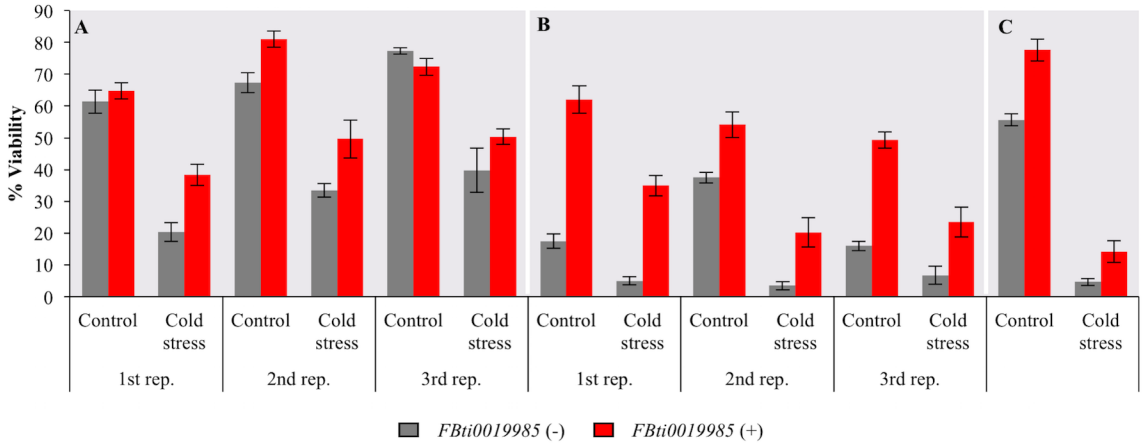


Fig 7. Flies with *FBti0019985* showed increased egg-to-adult viability under nonstress and under cold-stress conditions in three different genetic backgrounds. Egg-to-adult viability of strains without *FBti0019985* (grey) and with the *FBti0019985* insertion (red) in nonstress (control) and in cold-stress conditions. Results of the three replicates performed with (A) *RAL-783* and *RAL-810*, (B) *RAL-908* and *RAL-802*, and (C) *IV22* and *IV68*. Error bars represent the SEM of the different vials analyzed in each experiment.

doi:10.1371/journal.pgen.1006249.g007

Table 2. ANOVA for cold-stress assays in flies with and without different *roo* solo-LTR insertions.

Two-way ANOVA							
		Experimental condition		Insertion genotype		Experimental condition * Insertion genotype	
Insertion	Strains	P-value	Effect size[a]	P-value	Effect size[a]	P-value	Effect size[a]
FBti0019985	*RAL-810 (FBti0019985) / RAL-783 (roo-)*	≪0.001	0.69	≪0.001	0.13	0.034	0.04
	RAL-802 (FBti0019985) / RAL-908 (roo-)	≪0.001	0.56	≪0.001	0.57	0.932	-
	IV68 (FBti0019985) / IV22 (roo-)	≪0.001	0.74	0.003	0.25	0.981	-
roo$_{+7}$	*RAL-405 (roo$_{+7}$) / RAL-783 (roo-)*	≪0.001	0.75	0.001	0.59	0.497	-
	RAL-911 (roo$_{+7}$) / RAL-783 (roo-)	≪0.001	0.76	0.530	-	0.220	-
roo$_{-90}$	*RAL-21 (roo$_{-90}$) / RAL-783 (roo-)*	≪0.001	0.88	0.358	-	0.118	-
	RAL-820 (roo$_{-90}$) / RAL-783 (roo-)	≪0.001	0.71	0.681	-	0.123	-
roo$_{-44}$	*RAL-195 (roo$_{-44}$) / RAL-783 (roo-)*	≪0.001	0.79	0.038	0.31	0.027	0.35
	RAL-383 (roo$_{-44}$) / RAL-783 (roo-)	≪0.001	0.95	≪0.001	0.76	0.991	-
roo$_{-68}$	*RAL-75 (roo$_{-68}$) / RAL-783 (roo-)*	≪0.001	0.66	0.505	-	0.004	0.51
	RAL-716 (roo$_{-68}$) / RAL-783 (roo-)	≪0.001	0.87	0.002	0.56	0.032	0.33

[a]Partial eta-squared values calculated as a measure of effect size.

doi:10.1371/journal.pgen.1006249.t002

that *IV68* flies had a higher viability than flies without the *FBti0019985* insertion in both nonstress and cold-stress conditions (Table 2).

Overall, we found consistent results, across genetic backgrounds from two different natural populations, suggesting that flies with the *FBti0019985* insertion are associated with increased viability compared to flies without this insertion in nonstress and in cold-stress conditions. In all cases, the effect of the presence of the insertion was either medium or large (Table 2). In one of the genetic backgrounds, the effect was larger under cold-stress conditions (Fig 7A) while no interaction between experimental condition and insertion genotype was found in the other two backgrounds (Fig 7B and 7C).

Other *roo* solo-LTR insertions in the proximal promoter of *CG18446* are not consistently associated with cold-stress phenotypes

We further checked whether another four *roo* solo-LTR insertions described in this work are associated with cold-stress phenotypes. For each insertion, we compared the egg-to-adult viability of flies with two different genetic backgrounds with the egg-to-adult viability of *RAL-783* that does not carry any of these insertions (Fig 8). In all cases, we performed ANOVA analyses to check whether the experimental conditions, insertion genotype, and/or the interaction between these two factors were significant (Table 2).

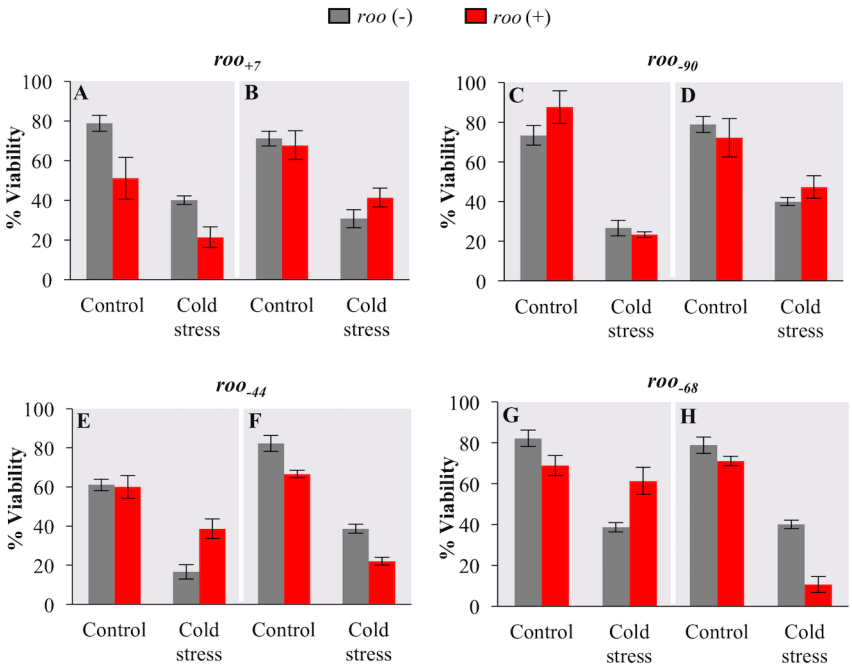


Fig 8. Other *roo* solo-LTR insertions are not consistently associated with cold-stress resistant phenotypes. Egg-to-adult viability in nonstress (control) and in cold-stress conditions of the *RAL-783* strain without any of the nine *roo* insertions (grey) and of different strains with *roo* insertions (red). (A) *RAL-405* (roo_{+7}), (B) *RAL-911* (roo_{+7}), (C) *RAL-21* (roo_{-90}), (D) *RAL-820* (roo_{-90}), (E) *RAL-195* (roo_{-44}), (F) *RAL-383* (roo_{-44}), (G) *RAL- 75* (roo_{-68}), and (H) *RAL-716* (roo_{-68}). Error bars represent the SEM of the different vials analyzed in each experiment.

doi:10.1371/journal.pgen.1006249.g008

We found that the experimental condition had a significant effect on egg-to-adult viability in all the strains tested (Table 2). On the other hand, the effect of the insertion was only significant in some of the genetic backgrounds (Table 2). Among strains that carry the roo_{+7} insertion, the insertion genotype had an effect only in one of the two backgrounds tested (Fig 8A and 8B and Table 2). *RAL-405* flies with roo_{+7} insertion showed decreased viability (Fig 8A and Table 2). The presence/ absence of roo_{-90} did not have a significant effect on egg-to-adult viability (Fig 8C and 8D and Table 2). For roo_{-44}, while the insertion genotype had a significant effect on the two backgrounds tested, results were not consistent. In one background, the presence of the insertion is associated with increased viability under cold-stress conditions and the interaction between the treatment and the insertion genotype is significant (Fig 8E and Table 2), while in the other background the presence of roo_{-44} is associated with decreased viability (Fig 8F and Table 2). Finally, the presence of roo_{-68} significantly affected viability in only one of the two backgrounds tested: *RAL-716* flies carrying roo_{-68} showed decreased viability (Fig 8H and Table 2).

Overall our results suggested that the presence of roo_{+7}, roo_{-90}, roo_{-44}, and roo_{-68} solo-LTR insertions reported in this work was not consistently associated with cold-stress phenotypes (Fig 8). These other insertions could have no phenotypic effect or could be involved in phenotypes not analyzed in this work.

Inference of selection in the region flanking the *FBti0019985* insertion

We looked for evidence of positive selection in the 2 kb region flanking the *FBti0019985* insertion. We analyzed the number of segregating sites (*S*) in this region and estimated Tajima´s D,

iHS, *nSL*, H_{12} and *XP-EHH* (see Material and Methods). We found reduced diversity in the strains with *FBti0019985*: the number of segregating sites in this region is significantly smaller than the number of segregating sites found in 2 kb regions of chromosome 2R, where the *FBti0019985* insertion is located (p-value = 0.015) (S4 Table). We also found that Tajima's D was significantly negative in the 2 kb region where *FBti0019985* is inserted, as expected if this region is under positive selection (p-value = 0.009) (S4 Fig and S4 Table). Finally, we also found significant values of *iHS* and H_{12} in the region flanking the *FBti0019985* insertion (*p-value* = 0.048 and *p-value* = 0.023, respectively) (S5 Fig and S4 Table).

We also looked for evidence of selection taking into account not only the strains in which *FBti0019985* is inserted, but all the strains that contain one of the nine *roo* insertions described in this work. In this case, only *iHS* showed a marginally significant value (*p-value* = 0.049) (S6 Fig).

Overall, our results suggest that the strains carrying *FBti0019985* might be evolving under positive selection while the evidence for positive selection taking into account all the strains with one of the nine *roo* solo-LTRs, was only marginally significant.

Discussion

Besides *FBti0019985*, we have discovered eight other *roo* solo-LTR elements inserted in the 368 bp region nearby the TSS of the cold-stress response gene *CG18446* (Fig 2) [21]. Each strain contained a single *roo* insertion and the population frequency of the different individual insertions varies from 1% to 17% (Fig 2). Full-length elements from the *roo* family are 8.7 kb long. Such long insertions in the proximal promoter of *CG18446* located in the first intron of *cbx*, might be deleterious, which could explain why all the identified insertions were solo-LTR elements. In *D. melanogaster*, repeated insertions of TEs have only been described in the proximal promoters of a particular gene class: *hsp* genes [50]. The susceptibility of *hsp* genes to TE insertions was attributed to their peculiar chromatin architecture: constitutively decondensed chromatin and nucleosome-free regions [51, 52]. However, promoter regions of non-*hsp* genes with similar chromatin architecture are not targets for TE insertions suggesting that chromatin accessibility is not sufficient to explain the susceptibility of *hsp* genes to TE insertions [50]. From a functional point of view, the presence of TEs in the promoter regions of *hsp* genes has been suggested to allow a rapid gene expression response to unpredictable temperature changes [50]. Similarly, the presence of *roo* insertions in the promoter of *CG18446* could also be enhancing the ability of this gene to respond to environmental challenges, although only one of the nine *roo* insertions was associated with cold-stress tolerance (see below). Interestingly, almost 100% of the insertions described in heat-shock genes are *P-element* insertions, and all the insertions described here are *roo* elements. *P-elements* preferentially insert in the 5' end of genes where they recognize a structural motif rather than a sequence motif [38, 39]. While 81% of *P-elements* insert in 5' gene regions, our results showed that only 15.2% of the *roo* elements annotated in the reference genome are inserted in 5' gene regions. Thus, with the data currently available, *roo* insertions do not seem to preferentially insert into 5' gene regions although analyses of *de novo* insertions should shed more light on this issue.

Our results showed that the different *roo* elements inserted in the proximal promoter of *CG18446* differ in their molecular and functional effects (Table 3). We found that the two insertions that are more closely located to *CG18446*, *FBti0019985* and roo_{+7}, provided an alternative TSS to this gene (Fig 5 and Table 3). However, only *FBti0019985* is associated with upregulation of *CG18446* expression (Fig 6 and Table 3). Besides providing an alternative TSS, the effect of the *FBti0019985* insertion on *CG18446* expression could be due to the addition of new regulatory regions (Fig 3A), to the disruption of the spacing of pre-existing ones (Fig 3B), and/

Table 3. Summary of the experimental results obtained in fly strains with five different *roo* solo-LTR insertions.

Insertion	Orientation	Strain	5'-RACE	*CG18446* expression	Effect of the insertion in egg-to-adult viability
FBti0019985	5' to 3'	*RAL-810*	TSS inside TE	Upregulation	Increase
		RAL-802	-	No differences	Increase
		IV68	-	Upregulation	Increase
		RAL-639	-	Upregulation	-
***roo*$_{+7}$**	5' to 3'	*RAL-405*	TSS inside TE	Downregulation	Decrease
		RAL-911	-	No differences	No differences
		RAL-887	-	No differences	-
***roo*$_{-44}$**	3' to 5'	*RAL-195*	TSS outside TE	No differences	Increase
		RAL-383	TSS outside TE	No differences	Decrease
***roo*$_{-68}$**	3' to 5'	*RAL-75*	-	-	No differences
		RAL-716	-	-	Decrease
***roo*$_{-90}$**	5' to 3'	*RAL-21*	TSS outside TE	Upregulation	No differences
		RAL-820	-	No differences	No differences
		RAL-88	-	No differences	-

doi:10.1371/journal.pgen.1006249.t003

or to the recruitment of HP1a protein that could also lead to changes in the expression of *CG18446* (Fig 4B). Finally, we cannot discard that polymorphisms other than the presence/absence of the *FBti0019985* insertion also affect the expression of *CG18446*.

We found that the *FBti0019985* insertion, which is associated with increased *CG18446* expression, is consistently associated with increased viability in nonstress and in cold-stress conditions (Fig 7 and Table 3). Although we cannot exclude that other variants linked to *FBti0019985* contribute to the increased viability phenotypes, we argue that it is unlikely that the association between the *FBti0019985* insertion and increased viability in three different genetic backgrounds from two different natural populations would occur spuriously [53]. These results also suggest that *CG18446* is likely to play a role in cold tolerance as was previously suggested based on cold-stress selection experiments in which this gene was found to be overexpressed [21]. However, *FBti0019985* is present in only 10% of the out-of-Africa natural strains analyzed in this work. Our screening was focused on three out-of-Africa populations, thus we cannot discard that *FBti0019985* is present at higher frequencies in other populations. Alternatively, it is also possible that the relatively low frequency of *FBti0019985* is due to negative fitness effects of this insertion on other phenotypes. Cold-stress resistance has been associated with decreased starvation resistance [54, 55] and reduced fecundity [56, 57]. Therefore, the benefit of flies carrying *FBti0019985* in cold-stress conditions might be a cost, for example, when food resources are scarce.

While *FBti0019985* has a consistent cold-stress tolerance phenotype, four other *roo* insertions also located on the proximal promoter of *CG18446* did not (Fig 8 and Table 3). The insertion that is present at higher frequencies in out-of-Africa populations is *roo*$_{-90}$ (Fig 2). However, this insertion is not associated with changes of expression of *CG18446* in embryos (Fig 6) and was not found to be associated with cold-stress tolerance phenotypes (Fig 8C and 8D and Table 3). It could be that this insertion has no phenotypic effect. Alternatively, *roo*$_{-90}$ could be affecting a phenotype other than cold tolerance. A recent update in FlyBase revealed that *CG18446* is also an ethanol-regulated gene that could contribute to ethanol sensitivity or tolerance [58]. Another possibility is that *roo*$_{-90}$ affects *cbx*. As the other *roo* insertion described in this work and *CG18446* gene, *roo*$_{-90}$ is inserted in the first intron of *cbx* which has been functionally classified as a defense response to bacterium and spermatogenesis gene [59] (Fig 1). Elucidating whether *roo*$_{-90}$ has an adaptive effect is beyond the scope of this paper.

Overall, we did not find evidence of positive selection at the DNA level in the region where the nine *roo* solo-LTR elements are inserted. We did find evidence of reduced diversity in this region when only the strains containing *FBti0019985* were considered (S4–S6 Figs and S4 Table). Further analyses with a bigger dataset of strains is needed in order to determine whether this region shows signals of positive selection at the DNA level.

In summary, our results showed that different TE insertions in the same gene promoter region might have different molecular and functional consequences. Thus, the description of complex regions, as the one reported in this work, should be followed by functional analysis of the structural variants if we want to elucidate which ones are functionally relevant.

Materials and Methods

Fly stocks

We used inbred strains from the Drosophila Genetic Reference Panel (DGRP [31, 32]) and isofemale strains from an Italian population collected in Castellana Grotte (Bari, Italy [33]) to perform the molecular and phenotypic assays.

Analysis of presence/absence by PCR of the nine solo-LTR *roo* insertions

We used a PCR approach to check for presence/ absence of *FBti0019985* in 28 strains from the North Carolina population and in 15 strains from Italy. The primers used were FBti0019985_FL (5'-GGCATCATAAAACCGTTGAACAC-3'), FBti0019985_L (5'-AGTCCCTTAGTGGGA GACCACAG-3') and FBti0019985_R (5'-CGTAGGATCAGTGGGTGAAAATG-3') (Fig 1). Primers FBti0019985_L and FBti0019985_R are expected to give a 616 bp band when the TE is present. Primers FBti0019985_FL and FBti0019985_R are expected to give a 638 bp band when the TE is absent and a 1066 bp band when the TE is present. All PCR bands giving evidence of presence and some of the PCR bands giving evidence of absence were cloned using TOPO TA Cloning Kit for Sequencing (Invitrogen) following the manufacturer's instructions and Sanger-sequenced using M13 forward and/or M13 reverse primers to verify the results. Sequences have been deposited in GenBank under accession numbers KU672690-KU672720.

Analysis of the population frequencies of the nine *roo* solo-LTR insertions using *Tlex2*

We estimated the frequencies of the nine *roo* solo-LTR insertions described in this work using *T-lex2* software [34]. Because *T-lex2* works only for annotated TEs, we constructed eight new reference sequences including each one of the newly described *roo* solo-LTR insertions. The new reference sequences included 500 bp at each side of the TE and the TSD of each insertion.

We run *T-lex2* in strains from three different populations: 50 strains from North Carolina (DGRP [31, 32]), 27 strains from a population collected in Stockholm, Sweden [33], and 67 strains from a population collected in Siavonga, Zambia [60]. As a control, we also run *T-lex2* in the strains for which we have PCR results (S1 Table). We obtained results for 21 out of 50 DGRP strains, 26 out of 27 Swedish strains and 42 out of 67 Zambian strains. In some of the strains, *T-lex2* detects more than one insertion per strain. However, PCR analyses of these strains revealed that only one insertion was present. These results suggest that *T-lex2* cannot accurately estimate the frequency of insertion when they are closely located to each other. We thus discarded *T-lex2* results indicating the presence of more than one insertion per strain. Other factors such as the quality of the reads and the coverage of the different strains could also be affecting *T-lex2* results.

Analysis of target site motifs

Target site motifs were constructed in WebLogo (http://weblogo.berkeley.edu) using six TSDs sequences obtained in this work and 41 TSDs sequences predicted with *T-lex2* software [34].

Phylogenetic analysis

For each *roo* solo-LTR insertion, we constructed a consensus sequence taking into account the 26 strains sequenced in this work using *Sequencher* 5.0 software. We aligned the nine *roo* insertion consensus sequences with 115 of the 137 other *roo* insertions present in the *D. melanogaster* genome using the multiple sequence aligner program MAFFT [61]. The quality sequence of the other 22 *roo* insertions was too low to include them in the alignment. A maximum likelihood tree was inferred using RAxML Version 8 [62] under the general time-reversible nucleotide model and a gamma distribution of evolutionary rates. We use the ETE toolkit Python framework for the analysis and visualization of trees [63].

roo insertions and *CG18446* promoter sequence analysis

We looked for conservation of the Transcription Factor Binding Sites (TFBSs) previously described in the *roo* family [8] in all the *roo* solo-LTRs characterized in this work. First, we downloaded from FlyBase version r6.06 (http://flybase.org) the *fasta* file of *FBti0019985* sequence (genome region 2R: 9,871,090–9,871,523). We also searched for TFBSs in the *roo* insertions and in the *CG18446* promoter regions using all the available JASPAR CORE Insecta matrices (http://jaspar.genereg.net). Only those sites predicted with a relative score higher than 0.995 were considered. We identified four new TFBS in *FBti0019985* sequence: *Deaf1*, *ara*, *mir*, and *caup*. We then look for conservation of the identified motifs in all the *roo* solo-LTR sequences described in this work. For some strains, we used the information available in http://popdrowser.uab.cat [64].

Detection of piRNA reads

We used three piRNA libraries [44–46] to map piRNA reads to a 1.4 kb region including *FBti0019985* and to all the *roo* insertions described in this work following the methodology described in Ullastres *et al* (2015) [33]. Briefly, we used BWA-MEM package version 0.7.5 a-r405 [65] to align the reads and then we used SamTools and BamTools [66] to index and filter by sense/antisense reads. The total read density was obtained using R (Rstudio v0.98.507) [67].

Detection of HP1a binding sites

We used modENCODE ChIP-Seq data [47] to map HP1a reads to a 1.4 kb region including *FBti0019985* and to all the *roo* insertions described in this work following the methodology described in Ullastres *et al* (2015) [33]. We aligned the reads using BWA-MEM package version 0.7.5 a-r405 [65]. The total read density was obtained using R (Rstudio v0.98.507) [67].

5'-RACE experiments

5-to-7 day-old flies were placed in a fly cage with egg-laying medium (2% agar with apple juice and a piece of fresh yeast) during 4 hours. Then, adult flies were separated and embryos were collected following the suspension method described in Schou (2013) [68]. Embryo dechorionation was done by bleach (50%) immersion. Total RNA was extracted using TRIzol Plus RNA Purification Kit (Ambion). RNA was then treated on-column with DNase I (Thermo) during purification, and then treated once more after purification. 5'-RACE was performed with First-Choice RLM-RACE Kit and using Small-scale reaction RNA processing with RNA samples of

RAL-783 (*roo-*), *RAL-810* (*FBti0019985*), *RAL-405* (roo_{+7}), *RAL-21* (roo_{-90}), *RAL-383* (roo_{-44}) and *RAL-195* (roo_{-44}). The gene specific outer primer was 5'-GACACTCTTCGGTTGGTGGA-3' and the gene specific inner primer was 5'-ACAACTGTTCTGTAGGATCGC-3'. The control primer was 5'-TAGTCCGCAGAGAAACGTCG-3'. Inner PCR products were then cloned and Sanger-sequenced as mentioned above. Sequences have been deposited in GenBank under accession numbers KU672721-KU672722.

Quantitative RT-PCR expression analysis

Embryo collection and RNA extraction was performed as described before. Reverse transcription was carried out using 500 ng of total RNA using Transcriptor First Strand cDNA Synthesis Kit (Roche). The cDNA was then used in a 1/50 dilution for qRT-PCR with SYBR green master-mix (Bio-Rad) on an iQ5 Thermal cycler. *CG18446* expression was measured using specific primers (5'-GAGCAGTTGGAATCGGGTTTTAC-3' and 5'-GTATGAATCGCAGTCCAGC CATA-3') spanning 99 bp cDNA in the exon 1/exon 2 junction of *CG18446*. The primer pair efficiency was 99,1% (r^2 larger than 0.99). *CG18446* expression was normalized with *Act5C* expression levels (5'-GCGCCCTTACTCTTTCACCA-3' and 5'-ATGTCACGGACGATTT CACG-3').

Cold-stress resistance assays

Embryo collection was performed as mentioned above. Embryos were put into 50 ml fresh food vials. When embryos were 4–8 hour-old, they were kept at 1 C for 14 hours and then they were kept at room temperature (22–25 C). Simultaneously, control vials were always kept at room temperature (22–25 C) and never exposed to cold-stress. A total of 8–20 vials were analyzed per experiment. The same number of embryos per vial, 30 or 50, were used for all the replicates of a given experiment. Percentage viability was calculated based on the number of emerged flies to the total number of embryos placed in each vial.

Statistical significance was calculated performing two-way ANOVA using SPSS v21. We combined all the data into a full model: experimental condition (stress and nonstress), insertion genotype (presence/absence of the insertion) and interaction between these two factors. For those experiments in which more than one replicate was performed, the replicate effect was also taken into account. Because our dependent variable was a proportion, we used the arcsine transformation of the data before performing statistical analysis. We tested whether the data was normally distributed using Kolmogorov-Smirnov test. When the data was not normally distributed after the arcsine transformation, we applied the rank transformation. When the statistical test was significant, we estimated partial eta-squared values as a measure of the effect size (0.01 small effect, 0.06 medium effect, and 0.14 large effect).

Inferences of selection in the region flanking the *roo* solo-LTR insertions

We estimated the number of segregating sites (*S*), Tajima´s D, *iHS*, *nSL* and *XP-EHH* in the 2 kb region flanking the *FBti0019985* insertion (chromosome 2R: 5758000–5760000) in 10 DGRP strains containing this insertion, in the 23 DGRP strains containing one of the *roo* insertions described in this work, and in the 15 strains that do not contain any insertion in the promoter region of *CG18446*. Note that the coordinates of *FBti0019985* in the r5 of the *D. melanogaster* genome used by the DGRP project to generate the *vcf* files are 2R: 5,758,595–5,759,028. *S* and Tajima´s D are standard mesures of neutrality. *iHS* and *nSL* tests identify hard sweeps although they have some power to detect soft sweeps as well [69, 70]. H_{12} tests for positive selection on new variation and standing genetic variation within a population, that is, it searches both for soft and hard sweeps in a population [71]. Finally, *XP-EHH* is a statistical test

of positive selection in one population that uses between populations comparisons to increase power in regions near fixation in the selected population [72].

We have used *vcftools* to calculate the number of segregating sites, and Tajima´s D using parameters –maf 1/(2*n*), where *n* is the sample size, and –remove-indels. We have obtained *iHS*, *nSL*, and *XP-EHH* using the *selscan* software with default parameters [73]. Finally, we have calculated H_{12} with *ad hoc* scripts. The four latter statistics require phased data. Thus, chromosome 2R of the 205 DGRP strains were phased together using ShapeIt [74].

To calculate the significance for the number of segregating sites, we resampled at random the same number of strains from the 205 DGRP strains available and calculated the distribution of segregating sites in the same 2 kb region. To calculate the significance of Tajima´s D, *iHS*, *nSL* and *XP-EHH*, we have used the empirical distributions of these statistics obtained from chromosome 2R.

Supporting Information

S1 Fig. *roo* consensus Target Site Duplication (TSD). A frequency plot was built with all the TSD identified in this work, except the TSD of roo_{-19} and roo_{+7} that had four and two nucleotides instead of five, respectively, and with the 41 *roo* TSD motifs identified by Fiston-Lavier *et al* (2015) [34] (see Materials and Methods).
(TIFF)

S2 Fig. Phylogenetic tree including the nine *roo* elements analyzed in this work and the 115 *roo* elements annotated in the *D. melanogaster* reference genome. The nine *roo* elements sequenced in this work are depicted in red.
(TIFF)

S3 Fig. Sequence alignments of the regulatory regions identified in *roo* insertions and in the *CG18446* promoter region. Single nucleotide polymorphisms are highlighted in red. (A) Alignment of the different *roo* insertions analyzed in this work. For *RAL-502* and *RAL-857* we could only sequence a partial region of the insertion and thus we only analyzed the *Inr* motif. (B) Alignment of the three regions with matrix association potential. (C) Alignment of the *CG18446* promoter region in the different strains analyzed. Underlined sequences are from popdrowser [64]. For additional details see Fig 3 legend.
(PDF)

S4 Fig. From top to bottom: Tajima's D in the 23 strains with one of the nine solo-LTR insertions, Tajima's D in the 10 strains with the *FBti0019985* insertion, and Tajima's D in the 15 strains without any of the nine insertions.
(PDF)

S5 Fig. From top to bottom, results for *XP-EHH*, H_{12}, *nSL*, and *iHS*. H_{12} was calculated on haplotypes of 40 segregating sites. All results are for the 10 strains with the *FBti0019985* insertion combined with the 15 strains without any of the nine insertions, except for *XP-EHH*, which is calculated between the 10 strains with the *FBti0019985* insertion and the 15 strains without any of the nine insertions. Horizontal dashed lines show significance levels while vertical dashed lines show the region of the insertion.
(PDF)

S6 Fig. Results for *XP-EHH*, H_{12}, *nSL*, and *iHS* from top to bottom calculated with the 23 strains that contain one of the nine *roo* insertions and the 15 strains without any of the *roo* insertions. See legend of S5 Fig for details.
(PDF)

S1 Table. Allele frequency estimates using *T-lex2* for the nine *roo* solo-LTR insertions analyzed. (A) Summary of the *T-lex2* results in all populations. (B) Results for DGRP population. (C) Results for Sweden population. (D) Results for Zambia population. (E) Results for the Italian population for which PCRs were also performed. (F) Results for the DGRP strains for which PCR were also performed.
(XLSX)

S2 Table. Sequence alignments of the cis-regulatory motifs located in *roo* solo-LTR insertions and in the *CG18446* promoter region. (A) Transcription factor binding sites and promoter motifs, and (B) Matrix Associated regions, found in *FBti0019985*. (C) Transcription factor binding sites and promoter motifs found in the *CG18446* promoter region.
(DOCX)

S3 Table. Number of piRNA reads (A) and HP1a reads (B) mapping to each one of the nine *roo* insertions analyzed in this work. The total number of piRNA reads and of HP1a reads per nucleotide position, and the average number of piRNA reads and of HP1a reads per insertion are given.
(XLSX)

S4 Table. Results of the different statistics used to infer positive selection in the region flanking the nine solo-LTR insertions.
(DOCX)

S1 Text. Phylogenetic tree containing the nine *roo* elements sequenced in this work and the 115 *roo* elements annotated in the *D. melanogaster* reference genome.
(TXT)

Acknowledgments

We thank all members of the González lab for comments on the manuscript.

Author Contributions

Conceptualization: JG.

Formal analysis: MM MARdC JG.

Funding acquisition: JG.

Investigation: MM AU MGB MARdC.

Methodology: JG MM.

Project administration: JG.

Software: MGB.

Supervision: JG.

Validation: MM AU MGB MARdC JG.

Visualization: MM JG.

Writing - original draft: JG MM.

Writing - review & editing: MM AU MGB MARdC JG.

References

1. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental biology. 2010; 339(2):225–9. doi: 10.1016/j.ydbio.2009.08.009 PMID: 19682982
2. Haberle V, Lenhard B. Promoter architectures and developmental gene regulation. Semin Cell Dev Biol. 2016.
3. Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, et al. Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res. 2011; 21(2):182–92. doi: 10.1101/gr.112466.110 PMID: 21177961
4. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet. 2006; 38(6):626–35. PMID: 16645617
5. Kawaji H, Frith MC, Katayama S, Sandelin A, Kai C, Kawai J, et al. Dynamic usage of transcription start sites within core promoters. Genome Biol. 2006; 7(12):R118. PMID: 17156492
6. Consortium Fantom, Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet. 2009; 41(5):553–62. doi: 10.1038/ng.375 PMID: 19377474
7. Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, et al. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS genetics. 2011; 7(1): e1001274. doi: 10.1371/journal.pgen.1001274 PMID: 21249180
8. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 2013; 23(1):169–80. doi: 10.1101/gr.139618.112 PMID: 22936248
9. McClintock B. Intranuclear systems controlling gene action and mutation. Brookhaven Symp Biol. 1956;Feb(8:):58–74. PMID: 13293421
10. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969; 165(3891):349–57. PMID: 5789433
11. Guio L, Barron MG, Gonzalez J. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Molecular ecology. 2014; 23(8):2020–30. doi: 10.1111/mec.12711 PMID: 24629106
12. Guio L, Gonzalez J. The dominance effect of the adaptive transposable element insertion Bari-Jheh depends on the genetic background. Genome biology and evolution. 2015; 7(5):1260–6. doi: 10.1093/gbe/evv071 PMID: 25912044
13. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016; 351(6277):1083–7. doi: 10.1126/science.aad5497 PMID: 26941318
14. Lerman DN, Feder ME. Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster. Molecular biology and evolution. 2005; 22(3):776–83. PMID: 15574805
15. Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS biology. 2008; 6(10):e251. doi: 10.1371/journal.pbio.0060251 PMID: 18942889
16. Bronner G, Taubert H, Jackle H. Mesoderm-specific B104 expression in the Drosophila embryo is mediated by internal cis-acting elements of the transposon. Chromosoma. 1995; 103(10):669–75. PMID: 7664613
17. Scherer G, Telford J, Baldari C, Pirrotta V. Isolation of cloned genes differentially expressed at early and late stages of Drosophila embryonic development. Developmental biology. 1981; 86(2):438–47. PMID: 6269930
18. Scherer G, Tschudi C, Perera J, Delius H, Pirrotta V. B104, a new dispersed repeated gene family in Drosophila melanogaster and its analogies with retroviruses. J Mol Biol. 1982; 157(3):435–51. PMID: 6181263
19. Meyerowitz EM, Hogness DS. Molecular organization of a Drosophila puff site that responds to ecdysone. Cell. 1982; 28(1):165–76. PMID: 6279311
20. FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. Comparative genomics of Drosophila and human core promoters. Genome Biol. 2006; 7(7):R53. PMID: 16827941
21. Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. Selection for cold resistance alters gene transcript levels in Drosophila melanogaster. Journal of insect physiology. 2009; 55(6):549–55. doi: 10.1016/j.jinsphys.2009.01.010 PMID: 19232407

22. Hoffmann AA, Scott M, Partridge L, Hallas R. Overwintering in Drosophila melanogaster: outdoor field cage experiments on clinal and laboratory selected populations help to elucidate traits under selection. J Evol Biol. 2003; 16(4):614–23. PMID: 14632225

23. Kellermann V, Loeschcke V, Hoffmann AA, Kristensen TN, Flojgaard C, David JR, et al. Phylogenetic constraints in key functional traits behind species' climate niches: patterns of desiccation and cold resistance across 95 Drosophila species. Evolution. 2012; 66(11):3377–89. doi: 10.1111/j.1558-5646.2012.01685.x PMID: 23106704

24. Goto SG. Expression of Drosophila homologue of senescence marker protein-30 during cold acclimation. Journal of insect physiology. 2000; 46(7):1111–20. PMID: 10817837

25. Goto SG. A novel gene that is up-regulated during recovery from cold shock in Drosophila melanogaster. Gene. 2001; 270(1–2):259–64. PMID: 11404024

26. Greenberg AJ, Moran JR, Coyne JA, Wu CI. Ecological adaptation during incipient speciation revealed by precise gene replacement. Science. 2003; 302(5651):1754–7. PMID: 14657496

27. Qin W, Neal SJ, Robertson RM, Westwood JT, Walker VK. Cold hardening and transcriptional change in Drosophila melanogaster. Insect Mol Biol. 2005; 14(6):607–13. PMID: 16313561

28. Morgan TJ, Mackay TF. Quantitative trait loci for thermotolerance phenotypes in Drosophila melanogaster. Heredity (Edinb). 2006; 96(3):232–42.

29. Hoffmann AA, Blacket MJ, McKechnie SW, Rako L, Schiffer M, Rane RV, et al. A proline repeat polymorphism of the Frost gene of Drosophila melanogaster showing clinal variation but not associated with cold resistance. Insect Mol Biol. 2012; 21(4):437–45. doi: 10.1111/j.1365-2583.2012.01149.x PMID: 22708613

30. Mamillapalli A, Pathak RU, Garapati HS, Mishra RK. Transposable element 'roo' attaches to nuclear matrix of the Drosophila melanogaster. Journal of insect science. 2013; 13:111. doi: 10.1673/031.013.11101 PMID: 24735214

31. Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014; 24(7):1193–208. doi: 10.1101/gr.171546.113 PMID: 24714809

32. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012; 482(7384):173–8. doi: 10.1038/nature10811 PMID: 22318601

33. Ullastres A, Petit N, Gonzalez J. Exploring the Phenotypic Space and the Evolutionary History of a Natural Mutation in Drosophila melanogaster. Molecular biology and evolution. 2015; 32(7):1800–14. doi: 10.1093/molbev/msv061 PMID: 25862139

34. Fiston-Lavier AS, Barron MG, Petrov DA, Gonzalez J. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. Nucleic Acids Res. 2015; 43(4):e22. doi: 10.1093/nar/gku1250 PMID: 25510498

35. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, et al. FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res. 2016; 44(D1):D786–92. doi: 10.1093/nar/gkv1046 PMID: 26467478

36. Bernstein M, Lersh RA, Subrahmanyan L, Cline TW. Transposon Insertions Causing Constitutive Sex-Lethal Activity in Drosophila melanogaster Affect Sxl Sex-Specific Transcript Splicing. Genetics. 1995; 139:631–48. PMID: 7713421

37. Linheiro RS, Bergman CM. Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. PLoS One. 2012; 7(2):e30008. doi: 10.1371/journal.pone.0030008 PMID: 22347367

38. Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM. Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proceedings of the National Academy of Sciences of the United States of America. 1995; 92(24):10824–30. PMID: 7479892

39. Liao GC, Rehm EJ, Rubin GM. Insertion site preferences of the P transposable element in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97(7):3347–51. PMID: 10716700

40. Reed DE, Huang XM, Wohlschlegel JA, Levine MS, Senger K. DEAF-1 regulates immunity gene expression in Drosophila. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(24):8351–6. doi: 10.1073/pnas.0802921105 PMID: 18550807

41. Dantoft W, Davis MM, Lindvall JM, Tang X, Uvell H, Junell A, et al. The Oct1 homolog Nubbin is a repressor of NF-kappaB-dependent immune gene expression that increases the tolerance to gut microbiota. BMC biology. 2013; 11:99. doi: 10.1186/1741-7007-11-99 PMID: 24010524

42. Sentmanat MF, Elgin SC. Ectopic assembly of heterochromatin in Drosophila melanogaster triggered by transposable elements. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(35):14104–9. doi: 10.1073/pnas.1207036109 PMID: 22891327

43. Lee YC. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. PLoS genetics. 2015; 11(6):e1005269. doi: 10.1371/journal.pgen.1005269 PMID: 26042931

44. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009; 137(3):509–21. doi: 10.1016/j.cell.2009.04.027 PMID: 19395009

45. Satyaki PR, Cuykendall TN, Wei KH, Brideau NJ, Kwak H, Aruna S, et al. The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS genetics. 2014; 10(3): e1004240. doi: 10.1371/journal.pgen.1004240 PMID: 24651406

46. Shpiz S, Ryazansky S, Olovnikov I, Abramov Y, Kalmykova A. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. PLoS genetics. 2014; 10(2):e1004138. doi: 10.1371/journal.pgen.1004138 PMID: 24516406

47. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature. 2011; 471(7339):480–5. doi: 10.1038/nature09725 PMID: 21179089

48. Eissenberg JC, Elgin SC. HP1a: a structural chromosomal protein regulating transcription. Trends Genet. 2014; 30(3):103–10. doi: 10.1016/j.tig.2014.01.002 PMID: 24555990

49. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, et al. FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 2015; 43(Database issue):D690–7. doi: 10.1093/nar/gku1099 PMID: 25398896

50. Walser JC, Chen B, Feder ME. Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. PLoS genetics. 2006; 2(10):e165. PMID: 17029562

51. Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. Modification of Heat-Shock Gene Expression in Drosophila melanogaster Populations via Transposable Elements. Molecular biology and evolution. 2003; 20(1):135–44. PMID: 12519916

52. Shilova VY, Garbuz DG, Myasyankina EN, Chen B, Evgen'ev MB, Feder ME, et al. Remarkable site specificity of local transposition into the Hsp70 promoter of Drosophila melanogaster. Genetics. 2006; 173(2):809–20. PMID: 16582443

53. Gruber JD, Genissel A, Macdonald SJ, Long AD. How repeatable are associations between polymorphisms in achaete-scute and bristle number variation in Drosophila? Genetics. 2007; 175(4):1987–97. PMID: 17277365

54. Kenny MC, Wilton A, Ballard JWO. Seasonal trade-off between starvation resistance and cold resistance in temperate wild-caught Drosophila simulans. Australian Journal of Entomology. 2008; 47 (1):20–3.

55. Hoffmann AA, Hallas R, Anderson AR, Telonis-Scott M. Evidence for a robust sex-specific trade-off between cold resistance and starvation resistance in Drosophila melanogaster. J Evol Biol. 2005; 18 (4):804–10. PMID: 16033551

56. Watson MJO, Hoffmann AA. Acclimation, cross-generation effects, and the response to selection for increased cold resistance in Drosophila. Evolution. 1996; 50(3):1182–92.

57. Marshall KE, Sinclair BJ. Repeated stress exposure results in a survival-reproduction trade-off in Drosophila melanogaster. Proc Biol Sci. 2010; 277(1683):963–9. doi: 10.1098/rspb.2009.1807 PMID: 19939842

58. Kong EC, Allouche L, Chapot PA, Vranizan K, Moore MS, Heberlein U, et al. Ethanol-regulated genes that contribute to ethanol sensitivity and rapid tolerance in Drosophila. Alcohol Clin Exp Res. 2010; 34 (2):302–16. doi: 10.1111/j.1530-0277.2009.01093.x PMID: 19951294

59. Ayres JS, Freitag N, Schneider DS. Identification of Drosophila Mutants Altering Defense of and Endurance to Listeria monocytogenes Infection. Genetics. 2008; 178:1807–15. doi: 10.1534/genetics.107.083782 PMID: 18245331

60. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, et al. The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics. 2015; 199(4):1229–41. doi: 10.1534/genetics.115.174664 PMID: 25631317

61. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 2013; 30(4):772–80. doi: 10.1093/molbev/mst010 PMID: 23329690

62. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. doi: 10.1093/bioinformatics/btu033 PMID: 24451623

63. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Molecular biology and evolution. 2016; 33(6):1635–8. doi: 10.1093/molbev/msw046 PMID: 26921390

64. Ramia M, Librado P, Casillas S, Rozas J, Barbadilla A. PopDrowser: the Population Drosophila Browser. Bioinformatics. 2012; 28(4):595–6. doi: 10.1093/bioinformatics/btr691 PMID: 22180410

65. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. bioRxiv 13033997v2. 2013.

66. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011; 27(12):1691–2. doi: 10.1093/bioinformatics/btr174 PMID: 21493652

67. RStudioTeam. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/. 2015.

68. Schou MF. Fast egg collection method greatly improves randomness of egg sampling in Drosophila melanogaster. Fly (Austin). 2013; 7(1):44–6.

69. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS biology. 2006; 4(3):e72. PMID: 16494531

70. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Molecular biology and evolution. 2014; 31(5):1275–91. doi: 10.1093/molbev/msu077 PMID: 24554778

71. Garud NR, Messer PW, Buzbas EO, Petrov DA. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS genetics. 2015; 11(2):e1005004. doi: 10.1371/journal.pgen.1005004 PMID: 25706129

72. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449(7164):913–8. PMID: 17943131

73. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Molecular biology and evolution. 2014; 31(10):2824–7. doi: 10.1093/molbev/msu211 PMID: 25015648

74. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. Am J Hum Genet. 2013; 93(4):687–96. doi: 10.1016/j.ajhg.2013.09.002 PMID: 24094745

3.4. A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila melanogaster*

Resum

Els transposons no estan distribuïts aleatòriament pel genoma. Un anàlisi de tot el genoma de *D. melanogaster* va trobar que les diferències en la densitat de transposons a través de regions genòmiques de 50 kb eren degudes a processos de transposició i duplicació. A escales genòmiques més petites, s'ha vist que les regions promotores dels gens *hsp* i la regió promotora del gen *CG18446* acumulen insercions de transposons. En aquest treball, hem estudiat més detalladament la regió promotora del gen *CG18446.* Es van analitzar 218 soques col·lectades en 15 poblacions naturals i es va trobar que la regió promotora del gen *CG18446* conté 20 insercions independents *roo*. Basat en un anàlisi filogenètic, es suggereix que la presència de múltiples insercions *roo* en aquesta regió pot ser el resultat de diversos esclats de transposició. A més a més, es va trobar que l'agrupació d'insercions *roo* en la regió promotora de *CG18446* és única: no hi ha cap altre regió promotora en el genoma que contingui un nombre semblant d'insercions *roo*. Es va trobar que, semblant a les regions promotores dels gens *hsp*, l'accessibilitat de la cromatina podria ser un dels factors que expliqués les insercions recurrents d'elements *roo* a la regió promotora de *CG18446*.

Merenciano et al. Mobile DNA (2019) 10:10
https://doi.org/10.1186/s13100-019-0152-9

Mobile DNA

SHORT REPORT Open Access

A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila melanogaster*

Miriam Merenciano[1], Camillo Iacometti[1,2] and Josefa González[1*]

Abstract

Transposable elements (TEs) are not randomly distributed in the genome. A genome-wide analysis of the *D. melanogaster* genome found that differences in TE density across 50 kb genomic regions was due both to transposition and duplication. At smaller genomic scales, promoter regions of *hsp* genes and the promoter region of *CG18446* have been shown to accumulate TE insertions. In this work, we have further analyzed the promoter region of *CG18446*. We screened 218 strains collected in 15 natural populations, and we found that the *CG18446* promoter region contains 20 independent *roo* insertions. Based on phylogenetic analysis, we suggest that the presence of multiple *roo* insertions in this region is likely to be the result of several bursts of transposition. Moreover, we found that the *roo* insertional cluster in the *CG18446* promoter region is unique: no other promoter region in the genome contains a similar number of *roo* insertions. We found that, similar to *hsp* gene promoters, chromatin accessibility could be one of the factors explaining the recurrent insertions of *roo* elements in *CG18446* promoter region.

Keywords: Transposable element, Fecundity, Viability, Target site duplication, Recurrent insertion, Natural population

Background

Recurrent insertion of transposable elements in specific genomic regions has been described in the *Drosophila melanogaster* reference genome. The analysis of 50 kb genomic windows identified 23 regions with a high density of TE insertions, most of them located in pericentromeric regions or on chromosome 4 [1]. Transposition and duplication were identified as the two mechanisms generating these high-density TE regions. In recent years, computational pipelines have been developed to analyze the TE content in multiple strains [2–4]. Thus, besides TEs annotated in the reference genome, non-reference TE insertions can now also be analyzed. Based on these population analyses, some genes have also been reported to accumulate many TE insertions, such as the 106.5 kb *klarsicht*, and the 24 kb *derailed-2* that were analyzed in 146 strains of the Drosophila Synthetic Population Resource [5, 6]. At a much finer scale, several insertions in the proximal promoter regions of *hsp* genes have been reported [7, 8]. While the vast majority of these insertions were *P-elements*, insertions from the *Gypsy* and the *Jockey* family were also identified. *P-elements* have a preference to insert in 5′ gene flanking regions [9]. The accumulation of TEs in the promoter of *hsp* genes was explained by the chromatin conformation of this particular region, and by selection favoring the retention of TEs because of their effect on gene expression [8]. More recently, nine *roo* insertions were also described in the promoter region of another stress response gene, *CG18446* that encodes a nucleic acid binding protein [10]. *CG18446* is a cold resistance candidate gene [11] and an ethanol-regulated gene [12] highly expressed in ovaries and in 6–10 h-old embryos [13]. Only one of the nine identified insertions was found to consistently affect the expression of *CG18446*, and it was associated with increased viability in non-stress and cold-stress conditions [10]. However, only 39 strains from two natural populations were screened, and thus it is still an open question whether more *roo* insertions are present in the *CG18446* promoter region. Indeed, *roo* are the most abundant elements in the *D. melanogaster* genome [14, 15]. Thus, it is possible that

* Correspondence: josefa.gonzalez@ibe.upf-csic.es
[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Maritim de la Barceloneta 37,49, 08003 Barcelona, Spain
Full list of author information is available at the end of the article

besides the cluster identified by Merenciano et al. (2016) [10] other similar clusters of *roo* insertions in gene promoter regions are present in the genome. Interestingly, while the majority of strains analyzed so far contain a *roo* insertion (26 out of 39), none of them contains more than one insertion [10].

In this work, we looked for TE insertions in the *CG18446* promoter region in 218 strains from 15 natural populations in Europe, North America, and Africa. In addition, based on the analysis of the reference genome, and on the analysis of 177 DGRP strains, we identified 53 promoter regions that could potentially contain multiple *roo* insertions. Finally, we performed fecundity and viability experiments to investigate why we did not find any fly containing two *roo* insertions in the *CG18446* promoter region.

Results

Twenty *roo* solo LTR insertions are present in the *CG18446* promoter region in natural populations

To check whether there were more *roo* insertions in the *CG18446* promoter region, we performed a PCR screening in 218 strains from 15 natural populations: 13 European, one North American [16], and one African population collected in the ancestral range of the species (Zambia) (Additional file 1) [17]. 143 strains gave a band consistent with the presence of an insertion, in homozygous or heterozygous state, and 75 strains gave a band consistent with the absence of an insertion (Table 1 and Additional file 2A). We sequenced all the obtained PCR bands and we found that besides the nine insertions discovered in Merenciano et al. (2016) [10], there are 11 other 428 bp *roo* solo-LTR insertions in the promoter region of *CG18446* (Fig. 1). All the strains with an insertion contained a single *roo* insertion. Across strains, three of the insertion sites contained *roo* elements inserted in opposite orientations, roo_{-90}, *FBti0019985* and roo_{+7}, suggesting recurrent insertion in the same exact genomic position (Fig. 1). Recurrent insertion in the same exact genomic position has also been described for *P-elements* in *D. melanogaster* [18, 19]. Note that based on the results of *T-lex2* [20], a computational pipeline that estimates presence/absence of insertions based on next generation sequencing data, we previously reported that the first nine *roo* insertions described in the *CG18446* promoter region were present in Zambia [10]. However, PCR analyses of 23 of the 42 strains analyzed with *T-lex2* containing four of these nine insertions indicated that these four insertions are not actually present in any of the strains (Additional file 2B). These four unvalidated insertions were polymorphic according to *T-lex2.* Thus, it could be that these insertions

Table 1 PCR results and *roo* insertions identified in the 218 strains analyzed in this work, and in brackets insertions identified in the 39 strains analyzed in Merenciano et al. (2016) (10)

Population	Strains analyzed	Strains homozygous for the presence of a *roo* insertion	Strains heterozygous for the presence of a *roo* insertion	*roo* insertions identified
Akka, FI	13	3	4	roo_{-90}, roo_{-64}, roo_{-291}
Stockholm, SE	23	9	6	roo_{-44}, roo_{-68}, roo_{-90}, roo_{-393}, roo_{-64}, roo_{-42}, $FBti0019985_{(3'\text{-}5')}$
Lund, SE	6	3	1	roo_{-68}, roo_{-64}
Karensminde, DK	12	5	2	*FBti0019985*, roo_{-19}, roo_{-68}, roo_{-64}, $roo_{-90(3'\text{-}5')}$
Munich, DE	14	6	5	roo_{+175}, roo_{-68}, roo_{-90}, roo_{-378}
Market Harborough, UK	20	5	7	*FBti0019985*, roo_{+37}, roo_{-68}, roo_{-90}, roo_{-291}, roo_{-42}, $roo_{-90(3'\text{-}5')}$
Gotheron, FR	13	3	2	roo_{-68}, roo_{-90}, roo_{-378}, roo_{-64}
Bari, IT	(12)	(3)	(4)	*FBti0019985*, roo_{+175}, roo_{-19}, roo_{-28}, roo_{-68}, roo_{-90}
Gimenells, ES	14	3	9	*FBti0019985*, roo_{+175}, roo_{-44}, roo_{-90}
Tomelloso, ES	15	3	10	roo_{-44}, roo_{-90}, roo_{-291}
Cortes de Baza, ES	13	0	9	roo_{-44}, roo_{-90}, $roo_{-90(3'\text{-}5')}$
Guadix, ES	14	0	11	roo_{-68}, roo_{-90}, roo_{-64}, $FBti0019985_{(3'\text{-}5')}$
San Cristóbal de la Laguna, ES	12	6	2	*FBti0019985*, roo_{-90}, roo_{-291}
Raleigh, US	22 (27)	17 (19)	0	*FBti0019985*, roo_{+7}, roo_{+278}, roo_{-28}, roo_{-44}, roo_{-68}, roo_{-90}, $FBti0019985_{(3'\text{-}5')}$
Siavonga, ZI	27	2	10	roo_{-90}, $roo_{+7(3'\text{-}5')}$, roo_{-56}, roo_{+192}
TOTAL	257	87	82	20 *roo* insertions

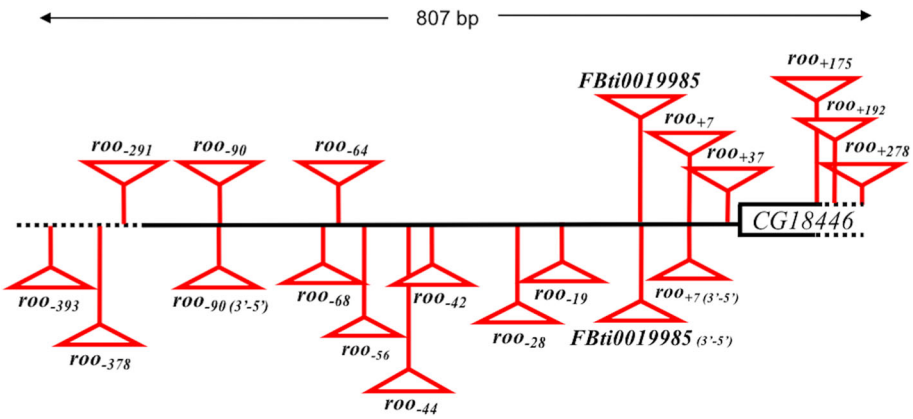


Fig. 1 Twenty *roo* solo-LTR insertions are located in the promoter region of *CG18446* gene in different natural strains. Schematic representation of the *CG18446* promoter region where the 20 *roo* solo-LTRs are inserted. The black line represents the *CG18446* promoter region. Note that although only one insertion was found in any given strain, we have represented them together for simplicity. The white box represents *CG18446* 5′UTR. Regions depicted with dotted lines are not drawn to scale. Insertions present in 5′-3′ orientation are shown above the black line and insertions present in 3′-5′ orientation are shown below the black line

have been lost in the isofemale strains since they were originally sequenced. Errors in genotyping of *T-lex2* could also explain some of these discrepancies, although all the homozygous insertions that *T-lex2* predicted were validated by PCR (Additional file 2B).

The majority of the 20 *roo* insertions inserted in the *CG18446* promoter region were present at very low allelic frequencies, ranging from 0.2% to 16.5% (Fig. 2, Additional file 2C). The two most common insertions were roo_{-90} and *FBti0019985*, with allelic frequencies of 16.5% and 6.3%, respectively (Fig. 2, Additional file 2C). While seven of the insertions were private, roo_{-68} and roo_{-90} were present in nine and 13 out of the 15 populations analyzed, respectively (Additional file 2D). We tested whether European populations at different latitudes differed in the diversity of *roo* insertions or in the total number of strains containing an insertion. Note that we did not considered the strains from Lund (Sweden) as only four strains were analyzed in this population. We found no correlation between latitude and the number of different *roo* insertions (Pearson r^2 = 0.006, *p-value* = 0.793), or between latitude and the number of strains with an insertion (Pearson r^2 = 0.063, *p-value* = 0.388). We also analyzed whether any of the insertions were more frequent in cold, temperate, or arid climates (Additional file 1). We found that roo_{-90} was more frequent in arid climates (*p-value* < 0.001) and roo_{-64} was more frequent in cold climates (*p-value* = 0.003) (Fig. 2).

Overall, we identified 20 *roo* insertions in the *CG18446* promoter region, most of them present at low population frequencies. While the majority of strains (169 out of 257) had one of the 20 *roo* insertions, none of the strains analyzed contained more than one *roo* insertion.

Recurrent insertion is the most likely explanation for the presence of 20 insertions in the promoter region of *CG18446*

We identified the target site duplication (TSD) for 17 of the 20 *roo* insertions located in the *CG18446* promoter region. These 17 *roo* solo-LTR insertions have different TSDs suggesting that they are independent insertion events (Additional file 3). 15 of the 17 identified TSD were five bp-long and the consensus TSD was similar to the one previously described [10, 20, 21] (Additional file 3). Thus, multiple insertions in the *CG18446* promoter region are likely the result of transposition rather than small rearrangements such as duplications or inversions, which would change the location of the insertions but not the TSDs.

We tested whether the multiple insertions could have been the results of a burst of transposition. We constructed a phylogenetic tree for the *roo* insertions present in the reference genome, and the 20 *roo* insertions found in the *CG18446* promoter region (see Material and Methods). Briefly, we estimated the unique number of substitutions shared between the two closest TEs assuming that all the *roo* copies present in the genome derived from a common ancestral sequence [22]. We found four groups of *roo* copies that are identical to each other and thus appeared to be the result of several bursts of transposition (Fig. 3, see Material and Methods). This is consistent with *roo* being one of the most active families in the *D. melanogaster* genome [14, 15, 23, 24].

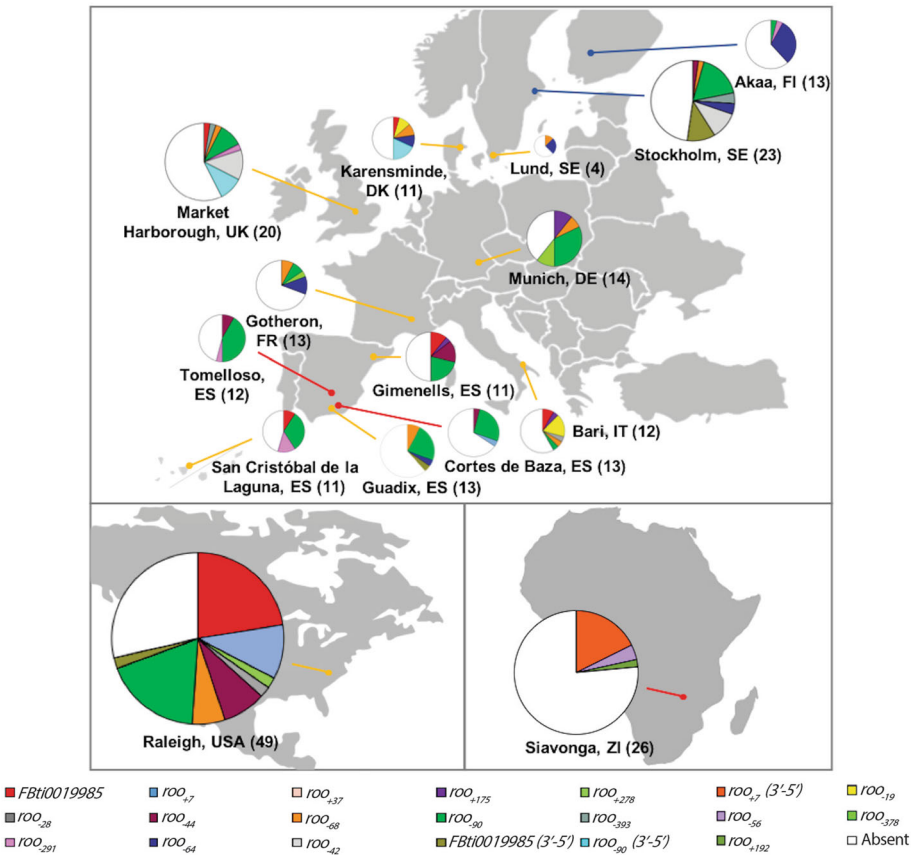


Fig. 2 The majority of the insertions in the *CG18446* promoter region are present at very low allelic frequencies. Allelic frequencies of the 20 *roo* solo-LTRs insertions in populations from Europe, North America, and Africa. The number of strains analyzed in each population is given in parenthesis and the pie chart size is proportional to this number. Blue, yellow, and red lines represent populations with cold, temperate, and arid climates, respectively

We then checked whether *roo* elements have a preference for inserting in 5′ gene regions. We considered as a 5′ gene region the 1 kb upstream of a gene and its 5'UTR region. Considering not only the 138 *roo* insertions annotated in the reference genome but also the 12,745 *roo de novo* insertions found in 177 DGRP strains by TIDAL software [15], we found that only 4.5% (586) of the *roo* elements are inserted in gene promoter regions or/and 5'UTR regions (see Material and Methods). This percentage is smaller than the one found for other TE families with preference for inserting in 5′ gene regions, such as the *P-element* family for which this percentage is > 77% [9, 10]. Thus, we considered that *roo* elements do not have a preference for inserting in 5′ gene regions.

We also checked whether the promoter region of *CG18446* has similarities with the promoter of *hsp* genes that could explain the high number of insertions in this region [8]. We found that, similar to *hsp* genes, *CG18446* is regulated by polymerase pausing [25], and has a high germline transcription activity [13]. Thus, chromatin accessibility could be one of the factors explaining the high TE density in the *CG18446* promoter region.

Finally, we found that transcription factor binding sites, core promoter motifs, and Matrix Associated Regions (MARs) previously described in the *roo* family were highly conserved in all the *roo* sequences described in this work (Additional file 4) [10, 26, 27].

Overall, we found that the presence of the 20 *roo* insertions in the *CG18446* promoter region is likely to be the result of several bursts of transposition (Fig. 3). Thus, recurrent insertions seem the most likely explanation for the presence of *roo* elements in the *CG18446* promoter region. Similar to the cluster of *P-element* insertions in the promoter of *hsp* genes, *roo* elements are also inserted in a promoter region with an open chromatin architecture [8].

The *roo* insertion cluster in *CG18446* is unique

We tested whether other *roo* clusters in gene promoter regions were present in the reference genome. Out of the 137 other *roo* elements present in the reference genome, 26 are inserted in promoters (less than 1 kb

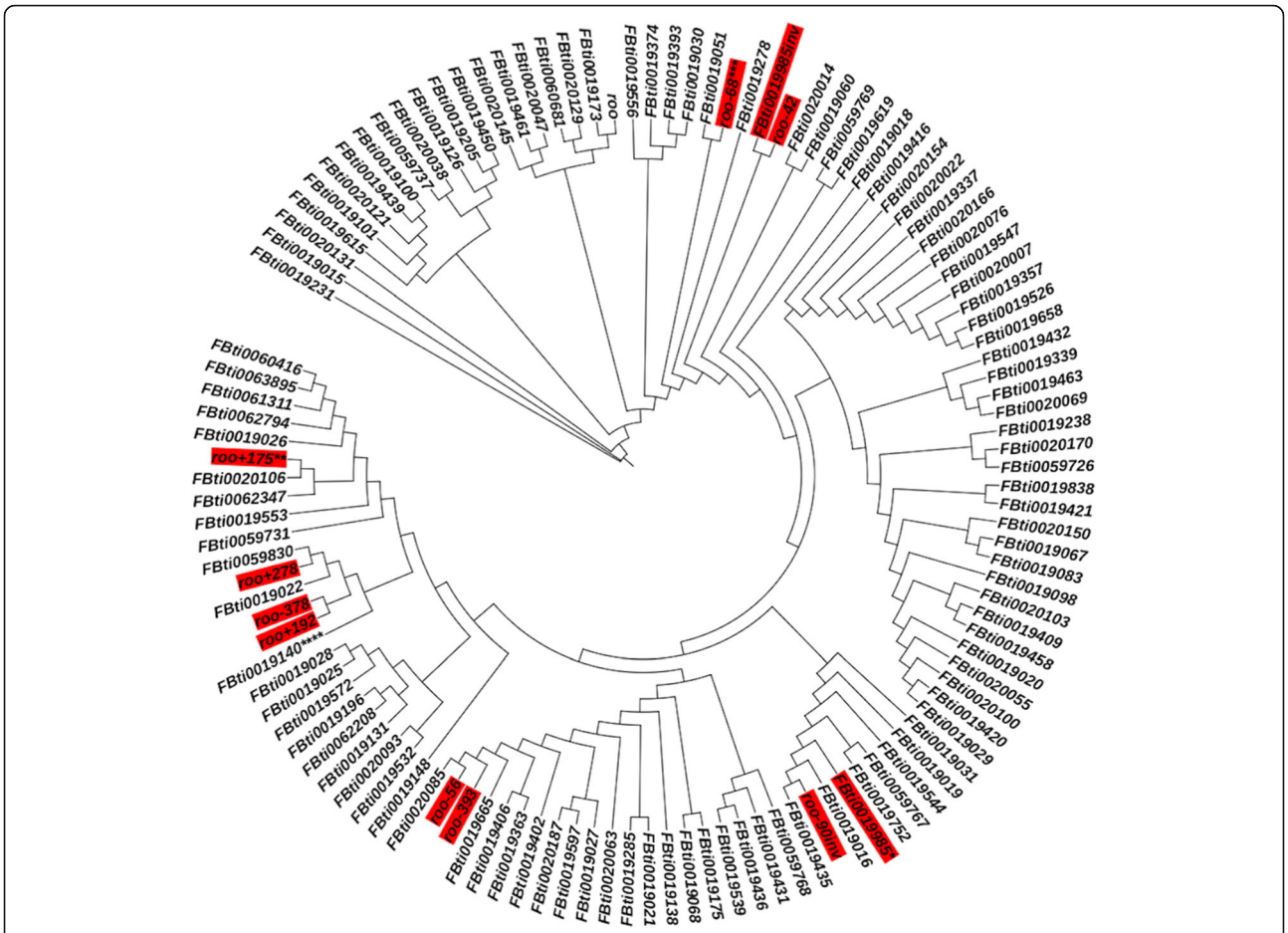


Fig. 3 The 20 *roo* solo-LTR insertions found in *CG18446* promoter region are the result of several bursts of transposition. Phylogenetic tree including the 20 *roo* solo-LTR insertions found in *CG18446* promoter region and 115 other *roo* insertions annotated in the reference genome. *roo* solo-LTRs found in this work are highlighted in red. When several TEs with the same exact sequence were identified, we removed all of them but one. The TEs included in the tree are marked with *. The TEs that were eliminated are: (*) *FBti0019017, FBti0019394, FBti0019438, FBti0020009, FBti0020080,* roo_{+7}, roo_{-90}, roo_{-19}, roo_{-28}, roo_{-64}, roo_{+37}, *and* roo_{+7inv}, (**) roo_{-291}, (***) roo_{-44}, and (****) *FBti0019608*

from a gene) or in 5'UTR regions. These 26 *roo* elements are inserted in 26 different promoter regions, and five of them are solo LTRs (Additional files 5 and 6A). We screened by PCR the presence/absence of insertions in these 26 promoter regions in 10 randomly chosen DGRP strains (see Material and Methods). For 22 of the 26 promoter regions, no other insertion was found in any of the 10 strains. The other four promoter regions contained the same *roo* element present in the reference genome in all the 10 strains analyzed (Additional files 5 and 6A). These results suggest that considering all the *roo* insertions annotated in the reference genome only the *CG18446* promoter region contain a cluster of *roo* insertions.

Besides the *roo* insertions annotated in the reference genome, we also analyzed all the *de novo roo* insertions identified by TIDAL in a set of 177 DGRP strains [15, 16]. There are 559 *roo* elements inserted in promoters or in 5'UTR regions. These 559 *roo* elements are distributed in 421 gene promoter or 5′ UTR regions (Additional file 5). According to TIDAL, the promoter region of *CG18446* has a *roo* insertion in eight different DGRP strains. We focused on the 27 gene promoter regions where TIDAL identifies three or more strains containing a *roo* insertion (Additional files 5, 6B and C). In order to test whether any of the 27 promoter regions harbors different *roo* insertions, we checked by PCR and sequenced the obtained bands of the 27 gene promoters in 95 strains (Additional file 6B). Among the 27 promoter regions analyzed, only four have two different *roo* insertions in different strains (Table 2). For these four genomic regions, we performed further PCR analysis in another 10 randomly chosen strains. We could not detect any other *roo* insertion in these promoter regions, suggesting that they probably harbor only the two *de novo roo* insertions found before.

Finally, it could be that *roo* insertions tend to form clusters, but that these clusters are deleterious when located in promoter regions. We thus also checked whether *roo*

elements cluster in 1 kb regions genome-wide, not necessarily located in gene promoters. We found five 1 kb regions with seven or more *de novo roo* insertions located in chromosomes 2 and 3 (Additional file 7, 8A and B). Because TIDAL does not predict the exact insertion site but rather provides a range of nucleotides where the TE is inserted, it is likely that the total number of *roo* insertions predicted in these windows is an overestimate. Indeed, the two regions with more *roo* insertions, 17 and 13 insertions, overlapped 911 bp and 323 bp respectively with the *roo* cluster in *CG18446* promoter region. Based on the screening reported in this work, we know that there are eight and one insertions respectively in these two regions. We checked by PCR whether all the elements predicted within the five 1 kb regions with more than seven insertions, and two randomly chosen windows with six and four predicted insertions had the same insertion site or not. The two regions overlapping with the *CG18446* promoter region contained five and one insertion (Additional file 8A). The other five regions analyzed contained at most two *roo* insertions (Additional file 8A). Thus, we found that only the 1 kb region that overlaps with the *CG18446* promoter region is actually a *roo* insertional cluster (Additional file 8A).

Flies with two *roo* insertions in the *CG18446* promoter regions are viable and show similar fecundity rates as flies with one *roo* insertion

As mentioned above, none of the 257 strains analyzed contains more than one *roo* insertion in the *CG18446* promoter region. The two *roo* insertions that are present at higher population frequencies are *FBti0019985* and $roo_{\_90}$. Thus, for these two insertions, and depending on the population analyzed, we would expect to find from 0.6% to 8.8% of flies containing these two insertions in different haplotypes (Additional file 2E). Since the number of strains sampled per population is not very high (Additional file 1), it could be that we have not screened enough flies to find one strain containing two insertions.

To discard that flies with two *roo* insertions have reduced egg-to-adult viability or reduced fecundity compared with flies containing only one *roo* insertion, we created flies containing two insertions in the *CG18446* promoter region (see Material and Methods). We found that flies with two *roo* insertions had similar or significantly higher viability compared with flies with only one of the *roo* insertions (ANOVA $p\text{-value} < 0.001$ Fig. 4a). Early fecundity of flies containing two *roo* insertions was not significantly different from that of flies containing only one *roo* insertion (ANOVA $p\text{-value} = 0.068$, Fig. 4b). Similarly, we did not find differences in the average number of eggs laid per day during 18 days between flies with one or two *roo* insertions (ANOVA $p\text{-value} = 0.494$, Fig. 4c). Note that the genetic background of flies containing one or two *roo* insertions is different. Thus, polymorphisms other than the presence/absence of these insertions are likely to be also contributing to the lack of differences observed.

Discussion

Besides the nine *roo* solo-LTRs found in Merenciano et al. (2016) [10], we have discovered 11 new *roo* insertions in the *CG18446* promoter region. It is known that *D. melanogaster* populations differ in their TE content [10, 28–31]. Thus, it could be that analyzing more populations, especially from geographical areas currently underrepresented such as Central and South America or Asia, could lead to the discovery of more *roo* insertions in the *CG18446* promoter region. However, the number of populations analyzed in this work was seven-fold higher than in Merenciano et al. (2016) [10] and the number of new *roo* insertions was only twice that of our previous study, suggesting that it is likely that we have discovered the majority of the *roo* elements in the *CG18446* promoter region.

All 20 *roo* insertions identified in the promoter region of *CG18446* are solo LTR insertions, while the majority (21 out of 26) of the *roo* insertions found in other promoter regions are full-length insertions (Additional file 6 A). Solo LTRs are presumably the result of homologous unequal recombination between the two LTRs of a full-length element [14]. Thus, the recombination region where these TEs are located could influence whether they are full-length elements or solo LTRs. However, only four of the 21 full-length elements are present in regions with a low recombination rate, while the other 17 *roo* insertions are located in regions with a similar recombination rate as the *CG18446* promoter region (Additional file 6A) [32, 33]. Although it is not clear why all the insertions in the *CG18446* promoter region are solo LTRs, the location of this promoter in an open chromatin region could be one of the

Table 2 *de novo roo* insertions found in four gene promoter regions

Promoter region	Number of strains predicted to have an insertion	Genomic coordinates of the insertion sites validated
plum	4	3R: 25,621,076 and 26,521,553
CG11459	3	3R: 6,027,532 and 6,027,608
CG15879	3	3 L: 2,169,152 and 2,169,162
CR44657	3	X: 14,114,700 and 14,115,661

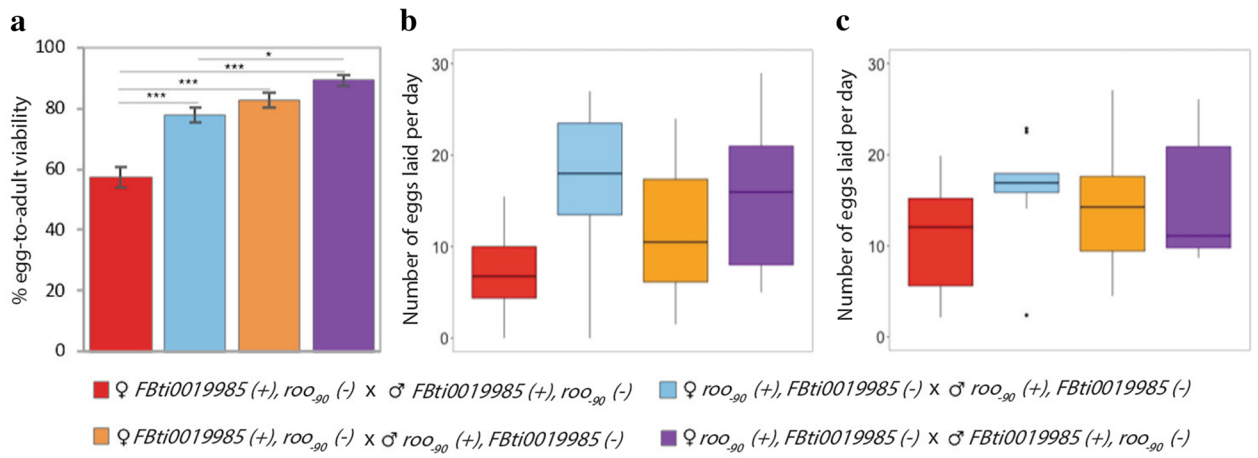


Fig. 4 Egg-to-adult viability and fecundity are not likely to reduce the probability of finding flies with more than one *roo* element in the *CG18446* promoter region. **a**. Percentage of emerged flies from the four crosses tested. **b**. Average number of eggs laid per day during the first 48 h for each of the four crosses analyzed. **c**. Average number of eggs laid per day during 18 days for each of the four crosses analyzed

contributing factors as it has been suggested that chromatin accessibility favors double strand breaks and thus recombination [33].

Phylogenetic analyses revealed that the presence of multiple *roo* insertions in the *CG18446* promoter is likely to be the result of several bursts of transposition rather than small rearrangements or insertion preference for 5′ gene flanking regions. This is consistent with previous data suggesting that *roo* is one of the most active TE families with a high transposition rate [14, 15, 23, 24]. Indeed, it has been suggested that *roo* elements have been able to evade piRNA silencing, because the number of novel *roo* insertions is high despite the presence of a high proportion of piRNAs against this family [15]. Note that the piRNA data analyzed in Rahman et al. (2015) [15] was obtained from ovaries and ovarian cell cultures [34, 35], and it has been suggested that TE activity in female and male germlines might differ due to polymorphisms in the piRNA regulatory genes between sexes [36].

Why *roo* insertions recurrently insert in the promoter region of the *CG18446* gene is not yet completely understood. We showed that there is no other cluster of *roo* insertions in promoter regions or in 1 kb genomic regions genome-wide. Thus, the presence of multiple *roo* insertion in this particular promoter region is probably related to some specific feature of this promoter. We indeed found that chromatin accessibility could be one of the factors explaining the recurrent insertions in this promoter region. In *D. melanogaster*, one other insertional cluster described is also located in the promoter region of stress response genes, which is located in an open chromatin region [8]. Several of the TEs located in the promoter of *hsp* genes have been shown to affect the expression of the nearby genes by altering the promoter architecture [7, 37]. So far, only one of the *roo* insertions in the *CG18446* promoter region, *FBti0019985*, has been shown to affect the expression of this gene by adding a new transcription start site [10]. In this work, we found that $roo_{\_90}$ has an allelic frequency of 16.5% and is significantly more frequent in arid climates. Thus, it would be interesting to test whether this insertion affects expression of the nearby gene and/or is associated with a fitness-related trait that could explain its higher frequencies in arid climate conditions.

Finally, in *Arabidopsis thaliana* recurrent insertion of TEs from the *Copia* family in the first intron of the *FLC* locus have been associated with epigenetic regulation of this locus in response to cold [38]. Thus, not only in *D. melanogaster* but at least also in *A. thaliana*, recurrent insertions of TEs belonging to a single family are associated with stress-related genes, and some of these insertions have fitness-related consequences.

Material and methods

Fly stocks

Fly stocks used for PCR screening are listed in Additional file 2A. One outbred population homozygous for the presence of *FBti0019985*, and one outbred population homozygous for the presence of $roo_{\_90}$ were generated by a round-robin cross of inbred lines from the Drosophila Genetic Reference Panel (DGRP) [39] and isofemale lines from different European populations (Additional file 9). We maintained the population by random mating with a large population size for over five generations before starting the experiments. All flies were reared on fly food medium in a 12:12 h light/dark cycle at 25 °C.

Analysis of TE presence/ absence

We used the same PCR approach as in Merenciano et al. (2016) [10] to check for the presence/absence of TE insertions in the *CG18446* promoter region in 234 natural strains from Europe, North America (DGRP) [16] and Africa (Nexus) [17]. Briefly, genomic DNA was extracted from a pool of 10 female flies of each strain. We performed PCR with two primer pairs. Primer pair Flanking (FL6) (5′-AACAATGCAAGTCCGTGCTC-3′) and Right (R) (5′-CGTAGGATCAGTGGGTGAAAATG-3′) are expected to give an 802 bp band when insertions are absent and a bigger band when there is an insertion. Primer pair Left (L) (5′ -AGTCCCTTAGTGGGAGACCACAG-3′) and R are expected to give a band only when there is a *roo* insertion. When the two PCRs failed, we used the alternative primer R2 (5′-CGGGTACATCTTTGCGGGAT-3′). When the PCR using the FL6 primer failed, we used the alternative primers FL (5′-GGCATCATAAAACCGTTGAACAC-3′), and/or FL7 (5′- TTCGTGCGTGTTCGGTACTT-3′). PCR products were purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) using the manufacturer's instructions and Sanger-sequenced using FL and/or L and R primers to verify the results. PCR failed for 16 strains and thus we could analyze 218 out of the 234.

Consensus motifs

We aligned using Genious 9.1.4 (https://www.geneious.com) the *roo* element sequences from the 114 strains that were fully sequenced in this work. We also included in the alignments the *roo* sequences reported in Merenciano et al. (2016) [10]. We identified in these sequences the nine transcription factor binding sites, the *Inr* promoter motif, and the MARs previously identified by Merenciano et al. (2016) [10]. We constructed the consensus sequence logos using WebLogo [40]. The target site duplication (TSD) consensus was constructed also using Weblogo with 15 out of 17 of the TSDs found in this work and in Merenciano et al. (2016) [10]. The two TSD removed have shorter sequence length.

Phylogenetic analysis

We followed the same approach as in Merenciano et al. (2016) [10]. Briefly, 16 of the 20 *roo* solo-LTR insertions in *CG18446* promoter region were sequenced in several strains (Additional file 2A). For each of these 16 insertions, we aligned the sequences and generated a consensus. We then aligned these 16 consensus sequences, the other four *roo* insertions and the 115 *roo* insertions found in *D. melanogaster* genome using the multiple sequence aligner program MAFFT. We inferred a maximum likelihood tree under the general time-reversible nucleotide model and a gamma distribution of evolutionary rates, using RAxML Version 8 [41] (Additional file 10). We removed from the phylogenetic analysis those TEs with exact identical sequences. The interactive tree of life (iTOL) framework (https://itol.embl.de/) was used for the analysis and visualization of the tree, ignoring branch lengths.

Analysis of other *roo* clusters in promoter regions

We analyzed the region where 27 *roo* elements are inserted less than 1 kb from a gene or in 5'UTR regions in the *D. melanogaster* reference genome (R6.07) in 10 randomly chosen DGRP strains. To determine if 10 strains are enough to detect a cluster, we ran 1000 randomly generated trials using a Python script. This script randomly chose 10 strains among all the DGRP strains screened by PCR in this work and in Merenciano et al. (2016) [10] and counted the number of different *roo* insertions obtained in every iteration. We found that four was the average number of different *roo* insertions that can be found in a screening of 10 randomly chosen DGRP strains. Then, by checking 10 different DGRP strains we expected to find an average of four different *roo* insertions in the case of the presence of an insertion cluster similar to the one found in the *CG18446* promoter region.

For each strain, genomic DNA was extracted from a pool of 10 female flies. Primers (forward and reverse) were design in the flanking region of the insertion amplifying a minimum of 500 bp when the TE is not present (Additional file 11). We also used a combination of primers (*roo*_primer and reverse) that gave a PCR band only when a *roo* element is present (Additional file 11). PCR programs were set according to the length of each TE insertion. In addition, we also considered *de novo* insertions found with TIDAL software in a set of 177 DGRP strains [15]. We first selected all the 559 *roo* elements predicted to be inserted less than 1 kb from the nearest gene or in 5'UTR regions. Then, we grouped the insertions based on the promoter region where they are inserted. Finally, we analyzed by PCR the 27 promoter regions where three or more strains putatively have a *roo* insertion. As before, genomic DNA was extracted from a pool of 10 female flies of each strain. Five combinations of primer pairs were used in order to verify the position of the insertion: one primer pair in the flanking region of the insertion amplifying a minimum of 500 bp when the TE is not present (ClusterF and ClusterR), and other four combinations where one primer was located in the LTR region in both genomic orientations (ClusterF and rooL2, ClusterR and rooL, ClusterF and rooL, and ClusterR and rooL2) (Additional file 11). PCR products were purified and Sanger sequenced as mentioned before. For the four promoter regions for which we found two *roo* insertions, 8, 13, 17, and 23, we

performed additional PCRs following the same approach in ten DGRP strains (RAL-105, RAL-129, RAL-136, RAL-161, RAL-208, RAL-239, RAL-208, RAL-239, RAL-280, RAL-301, RAL-309, and RAL-379).

Analysis of other clusters in the genome

We selected the 12,745 *roo de novo* insertions predicted by TIDAL software in 177 DGRP strains [15]. Since TIDAL software predicts a range of coordinates where the TEs may be inserted, we established as the insertion site the midpoint of the coordinates. For each chromosome arm (except the Y chromosome), we first considered as the same insertion those inserted within 5 bp windows. Thus, we got a total of 9243 *roo de novo* insertions. After that, we counted how many predicted *de novo roo* elements are in windows of 1 kb. We then chose for PCR validation five 1 kb regions with more than seven predicted *roo* insertions, and two additional 1 kb regions with four and six insertions predicted. Every region was validated in 7–10 different DGRP strains. For each strain, genomic DNA was extracted from a pool of 10 female flies. Five combinations of primers were designed following the same approach as before (Additional file 11). PCR products were purified and Sanger sequenced as mentioned before.

Expected genotype frequency calculation

For all the populations analyzed in this work, the expected genotype frequencies of flies containing both *FBti0019985* and *roo_90* insertions were calculated multiplying the observed allelic frequency for *FBti0019985* and the observed allelic frequency for *roo_90*, considering that they are in different haplotypes (Additional file 2E).

Viability assays

We checked the egg-to-adult viability of outbred *FBti0019985* (+) crosses, outbred *roo_90* (+) crosses and their reciprocal crosses. In total, 100 five to seven day-old flies (50 males and 50 virgin females) for each cross were allowed to lay eggs for 24 h on apple juice-agar medium with fresh yeast at 25 °C. Embryos were collected following the protocol described in Schou et al. (2013) [42]. For each cross, we collected a total number of 150 embryos and put them in groups of 30 in empty vials with fresh food. We maintained the vials at 25 °C until adult emergence. The percentage of egg-to-adult viability was calculated as the ratio of the number of emerged flies to the total number of embryos placed in each vial. Statistical significance was calculated performing ANOVA using SPSS v.21 followed by Tukey post-hoc multiple comparison procedure.

Fecundity assays

We checked the fecundity of outbred *FBti0019985* (+) crosses, outbred *roo_90* (+) crosses and their reciprocal crosses. For each cross, 10 virgin females were placed individually with one male in vials with fresh food. Flies were moved to new vials every day during 18 days without CO_2 anesthesia, and dead males were replaced. The number of eggs laid per day was counted every day during this period. The average of the total number of eggs laid per day during the 18 days (total fecundity), and the average of the total number of eggs laid per day during the first 48 h (early fecundity) was compared between crosses. We removed from the analysis those vials where the female died during the experiment. Statistical significance was calculated performing ANOVA using SPSS v.21.

Additional files

Additional file 1: Populations used for the analysis. (XLSX 11 kb)

Additional file 2: **A.** PCR results for the 277 strains analyzed in this work and in Merenciano et al. (2016). Strains used in Merenciano et al. (2016) are highlighted in blue. **B.** *Tlex-2* predictions in Merenciano et al. (2016) compared to PCR results in this work. Correct predictions are highlighted in green. Strains with *roo* insertions not identified in Merenciano et al. (2016) are highlighted in orange. Strains for which no results were obtained either by *T-lex2* or by PCR are highlighted in grey. **C.** Allelic frequencies of all the 20 *roo* insertions in all the populations analyzed. EU: Europe, NA: North America and ZI: Zambia. **D.** Allelic frequencies (%) of the 20 *roo* insertions in the 15 different populations analyzed. Elements only present in one population are highlighted in red. **E.** Expected genotype frequency of heterozygous flies with the two most common insertions, *FBti0019985* and *roo-90* in all the populations analyzed. a: *FBti0019985* alellic frequency, b: *roo-90* alellic frequency, and c: absent alellic frequency. (XLSX 47 kb)

Additional file 3: **A.** Consensus target site duplication (TSD) sequence identified in Merenciano et al. (2016) (left panel) and consensus TSD identified with the data of this paper and Merenciano et al. (2016) (right panel). **B.** TSD sequences of the 20 *roo* insertions. Frequency represents the number of strains that harbor the TSD out of the number of strains with a complete sequenced region. (DOCX 332 kb)

Additional file 4: Consensus sequence of the transcription factor binding sites and matrix attachment regions identified in all the *roo* sequences identified in the *CG18446* promoter region. (DOCX 330 kb)

Additional file 5: The formation of *roo* insertional clusters in gene promoter regions is not a *roo* family characteristic. Scheme of the gene promoter regions containing *roo* elements present in the reference genome (left) and present in 177 DGRP inbred strains (right). (DOCX 29 kb)

Additional file 6: **A.** Coordinates (R6), length, recombination rates and PCR results of the 26 promoter regions with a *roo* insertion in the reference genome. **B.** PCR results and *de novo* TE information of the 28 promoter regions where > = 3 strains putatively have a roo insertion based on TIDAL software predictions. **C.** Promoter regions where < 3 strains putatively have a *roo* insertion based on TIDAL software predictions. (XLSX 61 kb)

Additional file 7: Genome-wide distribution of *de novo roo* elements found in 177 DGRP strains. Number of predicted *de novo roo* elements found in 177 DGRP strains inserted in 1 kb windows in chromosomes 2, 3, 4, and X. (DOCX 102 kb)

Additional file 8: **A.** PCR results of the five 1 kb regions with more *roo* insertions predicted by TIDAL software. **B.** 1 kb regions with at least 1 *roo* insertion predicted by TIDAL software. Regions checked by PCR are highlighted in yellow. (XLSX 166 kb)

Additional file 9: Schematic representation of the round-robin cross-design for outbred *FBti0019985* (+), *FBti0019985* (−), roo_{-90} (+), and roo_{-90} (−) generation. (DOCX 123 kb)

Additional file 10: Phylogenetic tree of the 20 *roo* solo-LTR found in *CG18446* promoter region and 115 other *roo* insertions annotated in the reference genome. (TXT 6 kb)

Additional file 11: List of primers used for insertional cluster validation. (XLSX 12 kb)

Abbreviations
DGRP: Drosophila Genetic Reference Panel; LTR: Long Terminal Repeat; TE: Transposable Element; TSD: Target Site Duplication

Acknowledgements
We thank members of the González lab for providing comments on the manuscript. We also thank members of the DrosEU consortium (droseu.net) for sharing isofemale lines from their lab collections.

Funding
This work was funded by the European Commission (H2020-ERC-2014-CoG-647900). C.I. was funded by an ERASMUS+ fellowship. We acknowledge the support of the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de laGeneralitat de Catalunya (GRC 2017 SGR 880). We also acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).
The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

Availability of data and materials
The datasets generated as part of the study are available as supplementary information.

Authors' contributions
MM designed the study, acquired, analyzed and interpreted the data, and wrote the manuscript. CI acquired and analyzed the data. JG designed the study, interpreted the data, and wrote the manuscript. All the authors approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Maritim de la Barceloneta 37,49, 08003 Barcelona, Spain. [2]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy.

Received: 21 December 2018 Accepted: 28 February 2019
Published online: 13 March 2019

References
1. Bergman CM, Quesneville H, Anxolabehere D, Ashburner M. Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol. 2006; 7(11):R112.
2. Makałowski W, Pande A, Gotea V, Makałowska I. Transposable elements and their identification. Methods Mol Biol. 2012;855:337–59.
3. Ewing AD. Transposable element detection from whole genome sequence data. Mob DNA. 2015;6:24.
4. Modolo L, Lerat E. In: Poptsova MS, editor. Identification and Analysis of transposable elements in genomic sequences. U.K: Caister Academic Press; 2014.
5. Cridland JM, Macdonald SJ, Long AD, Thornton KR. Abundance and distribution of transposable elements in two Drosophila QTL mapping resources. Mol Biol Evol. 2013;30(10):2311–27.
6. King EG, Macdonald SJ, Long AD. Properties and power of the Drosophila synthetic population resource for the routine dissection of complex traits. Genetics. 2012;191(3):935–49.
7. Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. Mol Biol Evol. 2003;20(1):135–44.
8. Walser JC, Chen B, Feder ME. Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. PLoS Genet. 2006;2(10):e165.
9. Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM. Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci U S A. 1995;92(24):10824–30.
10. Merenciano M, Ullastres A, de Cara MA, Barron MG, Gonzalez J. Multiple independent Retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in Drosophila. PLoS Genet. 2016;12(8):e1006249.
11. Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. Selection for cold resistance alters gene transcript levels in Drosophila melanogaster. J Insect Physiol. 2009;55(6):549–55.
12. Kong EC, Allouche L, Chapot PA, Vranizan K, Moore MS, Heberlein U, et al. Ethanol-regulated genes that contribute to ethanol sensitivity and rapid tolerance in Drosophila. Alcohol Clin Exp Res. 2010;34(2):302–16.
13. Gramates LS, Marygold SJ, Santos GD, Urbano JM, Antonazzo G, Matthews BB, et al. FlyBase at 25: looking to the future. Nucleic Acids Res. 2017;45(D1): D663–D71.
14. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, et al. The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. 2002;3(12):RESEARCH0084.
15. Rahman R, Chirn GW, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, et al. Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. Nucleic Acids Res. 2015;43(22):10655–72.
16. Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster genetic reference panel lines. Genome Res. 2014;24(7):1193–208.
17. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, et al. The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics. 2015;199(4):1229–41.
18. Roiha H, Rubin GM, O'Hare K. P element insertions and rearrangements at the singed locus of Drosophila melanogaster. Genetics. 1988;119(1):75–83.
19. Shilova VY, Garbuz DG, Myasyankina EN, Chen B, Evgen'ev MB, Feder ME, et al. Remarkable site specificity of local transposition into the Hsp70 promoter of Drosophila melanogaster. Genetics. 2006;173(2):809–20.
20. Fiston-Lavier AS, Barron MG, Petrov DA, Gonzalez J. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. Nucleic Acids Res. 2015;43(4):e22.
21. Linheiro RS, Bergman CM. Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. PLoS One. 2012;7(2):e30008.
22. Rech GE, Bogaerts-Márquez M, Barrón MG, Merenciano M, Villanueva-Cañas JL, Horváth V, et al. Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila. PLoS Genet. 2019;15(2):e1007900.
23. Papaceit M, Avila V, Aguade M, Garcia-Dorado A. The dynamics of the roo transposable element in mutation-accumulation lines and segregating populations of Drosophila melanogaster. Genetics. 2007;177(1):511–22.
24. Diaz-Gonzalez J, Vazquez JF, Albornoz J, Dominguez A. Long-term evolution of the roo transposable element copy number in mutation accumulation lines of Drosophila melanogaster. Genet Res (Camb). 2011;93(3):181–7.
25. Saunders A, Core LJ, Sutcliffe C, Lis JT, Ashe HL. Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Genes Dev. 2013;27(10):1146–58.
26. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol. 2010;339(2): 225–9.

27. Mamillapalli A, Pathak RU, Garapati HS, Mishra RK. Transposable element 'roo' attaches to nuclear matrix of the Drosophila melanogaster. J Insect Sci. 2013;13:111.
28. Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol. 2008;6(10):e251.
29. Hill T, Schlotterer C, Betancourt AJ. Hybrid dysgenesis in Drosophila simulans associated with a rapid invasion of the P-element. PLoS Genet. 2016;12(3):e1005920.
30. Bergman CM, Han S, Nelson MG, Bondarenko V, Kozeretska I. Genomic analysis of P elements in natural populations of Drosophila melanogaster. PeerJ. 2017;5:e3824.
31. Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour AB, Vieira C, et al. Population specific dynamics and selection patterns of transposable element insertions in European natural populations. Mol Ecol. 2018.
32. Fiston-Lavier AS, Carrigan M, Petrov DA, González J. T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. Nucleic Acids Res. 2011;39(6):e36.
33. Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in Drosophila melanogaster. PLoS Genet. 2012;8(10):e1002905.
34. Song J, Liu J, Schnakenberg SL, Ha H, Xing J, Chen KC. Variation in piRNA and transposable element content in strains of Drosophila melanogaster. Genome Biol Evol. 2014;6(10):2786–98.
35. Sytnikova YA, Rahman R, Chirn GW, Clark JP, Lau NC. Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in Drosophila ovarian cell cultures. Genome Res. 2014;24(12):1977–90.
36. Saint-Leandre B, Clavereau I, Hua-Van A, Capy P. Transcriptional polymorphism of piRNA regulatory genes underlies the mariner activity in Drosophila simulans testes. Mol Ecol. 2017;26(14):3715–31.
37. Lerman DN, Feder ME. Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster. Mol Biol Evol. 2005; 22(3):776–83.
38. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The Arabidopsis thaliana mobilome and its impact at the species level. Elife. 2016;5:e15716.
39. Mackay TF, Heinsohn SL, Lyman RF, Moehring AJ, Morgan TJ, Rollmann SM. Genetics and genomics of Drosophila mating behavior. Proc Natl Acad Sci U S A. 2005;102(Suppl 1):6622–9.
40. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188–90.
41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
42. Schou MF. Fast egg collection method greatly improves randomness of egg sampling in Drosophila melanogaster. Fly (Austin). 2013;7(1):44–6.

3.5. A versatile transposable elements affects the expression of a transcription factor depending on the developmental stage and the environmental conditions in *Drosophila melanogaster*

Resum

Els transposons son seqüències d'ADN repetitives amb l'habilitat de moure's pel genoma augmentant, en alguns casos, el seu nombre de copies. Els transposons han estat considerats fonts de regions reguladores capaços de regular l'expressió dels gens que tenen al seu voltant. Aquesta regulació podria dependre de les condicions ambientals, així com també del teixit i de l'estadi de desenvolupament. *FBti0019985* es un element *roo* solo-LTR de *Drosophila melanogaster* inserit en la regió promotora del factor de transcripció *CG18446*. Estudis previs han mostrat que *FBti0019985* pot actuar com a potenciador sota condicions d'estrès immune i també es va associar amb una sobreexpressió del gen *CG18446* en embrions. A més, mosques amb *FBti0019985* van ser associades amb un increment de la viabilitat en condicions normals i sota estrès per fred. En aquest treball, duem a terme assajos de potenciació *in vivo* i anàlisis d'expressió gènica per analitzar amb més profunditat els efectes de *FBti0019985* sobre el gen *CG18446* en diferents condicions d'estrès i en diferents estadis del desenvolupament. Hem trobat que *FBti0019985* afecta la expressió de *CG18446* depenent de les condicions ambientals i de l'estadi del desenvolupament. A part d'actuar com a potenciador sota condicions d'estrès immune, *FBti0019985* també mostra una activitat potenciadora en embrions sota condicions normals. En canvi, en estadis adults o sota estrès induït per fred o per la exposició a etanol, *FBti0019985* no actua com a potenciador provocant una reducció de la expressió de *CG18446*. Finalment, també hem observat un altre possible efecte fenotípic de *FBti0019985* associant la seva presència amb tolerància a infecció bacteriana. Els nostres resultats suggereixen que diferents estadis del desenvolupament i diferents condicions ambientals haurien de ser explorades per arribar a caracteritzar completament els efectes moleculars i fenotípics de qualsevol variant genètica.

A versatile transposable element affects the expression of a transcription factor depending on the developmental stage and the environmental conditions in *Drosophila melanogaster*

Miriam Merenciano[1] and Josefa González[1]

1 Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain.

Abstract

Transposable elements (TEs) are repetitive DNA sequences with the ability to move along the genome sometimes increasing their copy number. TEs have been considered a genome-wide source of regulatory elements capable of regulating the expression of nearby genes. This gene regulation might depend on the environmental context, as well as on the tissue and on the developmental stage. *FBti0019985* is a *Drosophila melanogaster roo* solo-LTR element inserted in the promoter region of the transcription factor *CG18446*. Previous studies found that *FBti0019985* acts as an enhancer under immune-stress conditions and it was also associated with *CG18446* up-regulation in embryos. Furthermore, flies with *FBti0019985* were associated with increased viability in nonstress and under cold-stress conditions. In this work, we performed *in vivo* enhancer assays and gene expression analysis to further explore the effects of *FBti0019985* on *CG18446* expression under different stress conditions and in different developmental stages. We found that *FBti0019985* affects the expression of its nearby gene *CG18446* depending on both environmental conditions and developmental stage. Besides acting as an enhancer under immune-stress conditions, *FBti0019985* also shows enhancer activity in embryos under nonstress conditions. However, in adult stages or under cold- and ethanol-stress conditions, *FBti0019985* does not have this enhancer activity causing *CG18446* down-regulation. Finally, we also found another possible phenotypic effect to *FBti0019985* associating its presence with tolerance to bacterial infection. Our results suggest that different developmental stages and environmental conditions should be tested in order to fully characterize the molecular and functional effects of a genetic variant.

Introduction

Understanding the link between genotype and phenotype is a relevant and challenging question in Biology. Genotype-phenotype mapping aim at understand the effect of DNA variations from changes in gene expression to all aspects of morphology, physiology and behaviour (1). In the field of evolutionary biology, linking genetic variants to adaptive phenotypes allows us to better understand the process of adaptation. Adaptive processes are pervasive in nature and its understanding is especially important in the era of rapid environmental change (2). However, pleiotropy, epistasis, and environmental interactions can challenge the study of genotype-phenotype associations (1-5). Pleiotropy is the phenomenon by which one single gene affects multiple traits (6, 7). Pleiotropic events have been reported in different organisms such as the threespine sticklebacks, where mutations in the *Eda* gene account for different phenotypes like the development of the lateral plates and schooling behaviour, among others (8, 9). Thus, to fully characterize the functional effects of a genetic variant, the analysis of different possible phenotypes is needed. The interaction between different loci, or epistasis, can also lead to misinterpretations in genotype-phenotype associations (10-12). In bar-headed geese and Tibetan mastiffs, different amino acid substitutions in the haemoglobin gene cause different O_2 binding affinities depending on the genetic background, likely resulting in different adaptive consequences under hypoxic conditions (13, 14). So, different genetic backgrounds should be considered when linking candidate adaptive variants to their phenotypic outcomes. Moreover, some adaptive traits have been found to be modulated by the environment. For instance, in *C. elegans*, a mutation in the *npr-1* gene affects nematode aggregation behaviour only at certain oxygen levels (15). Thus, different environmental conditions should be tested in order to find the appropriate one in which a candidate adaptive mutation acts.

To date, most studies that aim at characterizing adaptive mutations have been focused on SNP variants, that are easier to detect by the common-used short-read sequencing techniques. However, other types of mutations like copy number variants (CNVs) have been a source of adaptive variation in different species (16-18). Transposable elements (TEs) are DNA sequences with the ability to move along the genome. They are powerful mutagens that can modulate both gene structure and expression through myriad ways (19, 20). As such, TEs with adaptive effects have been found in different organisms (21-26). Specifically, in *Drosophila melanogaster*, TEs

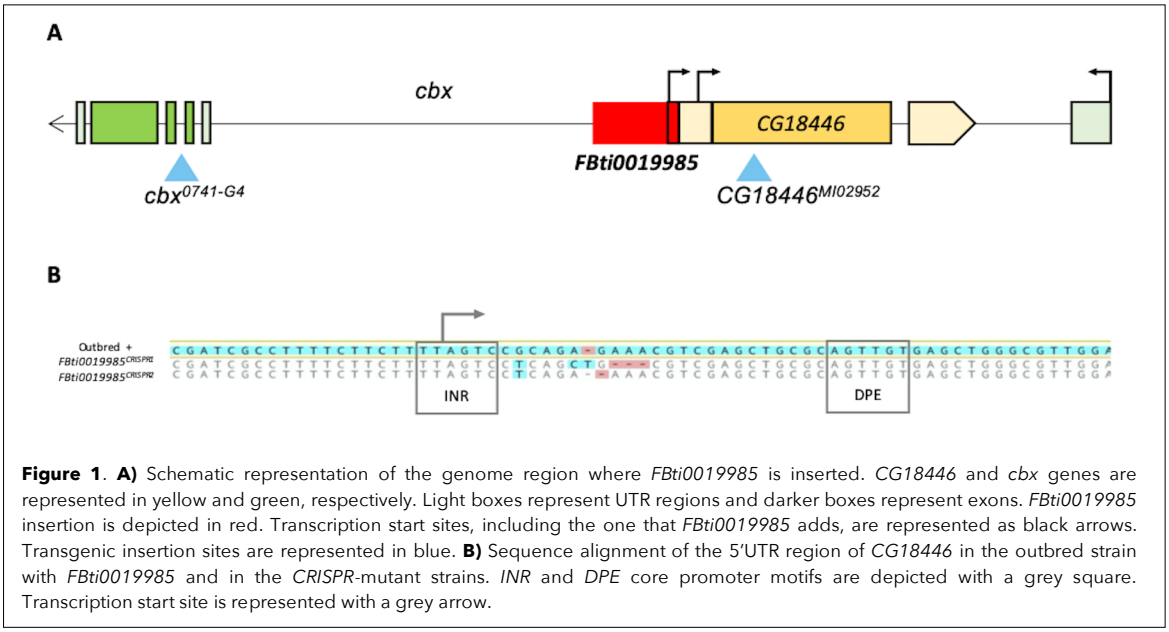


Figure 1. **A)** Schematic representation of the genome region where *FBti0019985* is inserted. *CG18446* and *cbx* genes are represented in yellow and green, respectively. Light boxes represent UTR regions and darker boxes represent exons. *FBti0019985* insertion is depicted in red. Transcription start sites, including the one that *FBti0019985* adds, are represented as black arrows. Transgenic insertion sites are represented in blue. **B)** Sequence alignment of the 5'UTR region of *CG18446* in the outbred strain with *FBti0019985* and in the *CRISPR*-mutant strains. *INR* and *DPE* core promoter motifs are depicted with a grey square. Transcription start site is represented with a grey arrow.

have been associated with resistance to insecticides, xenobiotics, oxidative stress, shorter developmental time, cold-stress, and tolerance to *P. entomophila* infection (27-34).

One of the identified candidate-adaptive TEs in *D. melanogaster* is *FBti0019985*, a *roo* solo-LTR element inserted in the promoter region of *CG18446*, which is located in the first intron of *cbx* gene. *FBti0019985* is one of the elements found in the unique *roo* insertional cluster located in the promoter region of *CG18446* (33, 35). Previous studies showed that *FBti0019985* provides an alternative transcript to the *CG18446* gene (33, 36). *FBti0019985* is also associated with *CG18446* up-regulation in embryos (33). Moreover, flies with this insertion have increased egg-to-adult viability in different genetic backgrounds under nonstress and cold-stress conditions (33).

CG18446 is a C_2H_2-type zinc finger transcription factor involved in chill-coma and immunity (37, 38). *CG18446* was up-regulated in flies artificially selected for increased resistance to chill-coma stress (37). Furthermore, *CG18446* has been recently associated with reduced development and number of immune cells in larvae due to systemic metabolic changes (38). The other gene in which *FBti0019985* is inserted, *cbx*, is a ubiquitin-conjugating enzyme that has also been associated with immune stress (34, 39). *cbx* mutant flies were found to be more sensitive to infection with the gram-positive bacteria *S. aureus,* an important pathogen in humans (39, 40). Moreover, a different *cbx* mutant was associated with increased tolerance to the gram-negative bacteria *P. entomophila* (34), a natural pathogen of *D. melanogaster* (41, 42).

Since *FBti0019985* nearby genes are involved in different stress responses, it is possible that *FBti0019985* is involved in different phenotypes, besides the already described cold-stress resistance. Furthermore, whether *FBti0019985* affects *CG18446* and *cbx* expression under stress conditions is still unknown.

In this work, we studied the molecular and phenotypic effects of *FBti0019985* under different stress conditions that are relevant for *D. melanogaster* in nature: immune-, cold-, and ethanol-stress. To do that, we performed expression analysis and *in vivo* reporter assays to investigate the role of *FBti0019985* in its nearby genes' expression. Furthermore, to check whether *FBti0019985* has an associated phenotypic effect in different stress responses, we also performed phenotypic experiments using laboratory outbred populations and CRISPR/Cas9-mutant strains.

Results

CG18446^{MI02952} mutant flies are more sensitive to immune- and cold-stress conditions

As mentioned above, *FBti0019985* is inserted in the promoter region of *CG18446*, which is nested in the first intron of *cbx* (Fig. 1). It has been shown that *CG18446* expression changes were associated with immunity and chill-coma (37, 38). Moreover, two *cbx* mutant strains were associated with different survival rates when exposed to different pathogens (34, 39). To further test whether changes in expression of *CG18446* and *cbx* genes are associated with changes in immune- and cold-stress responses, we exposed mutant flies with reduced/null expression levels of these genes to these stresses (Table S1 and S2). We first confirmed that *CG18446*^{MI02952} mutant down-regulates the expression of *CG18446* (Fig. 2A, Table S2). Because *CG18446* is nested in *cbx,* we also discarded that *CG18446*^{MI02952} affected the expression of *cbx* (Fig. 2A, Table S2). We found that *CG18446*^{MI02952} mutant flies were more sensitive to *P. entomophila* infection (Fig. 2B

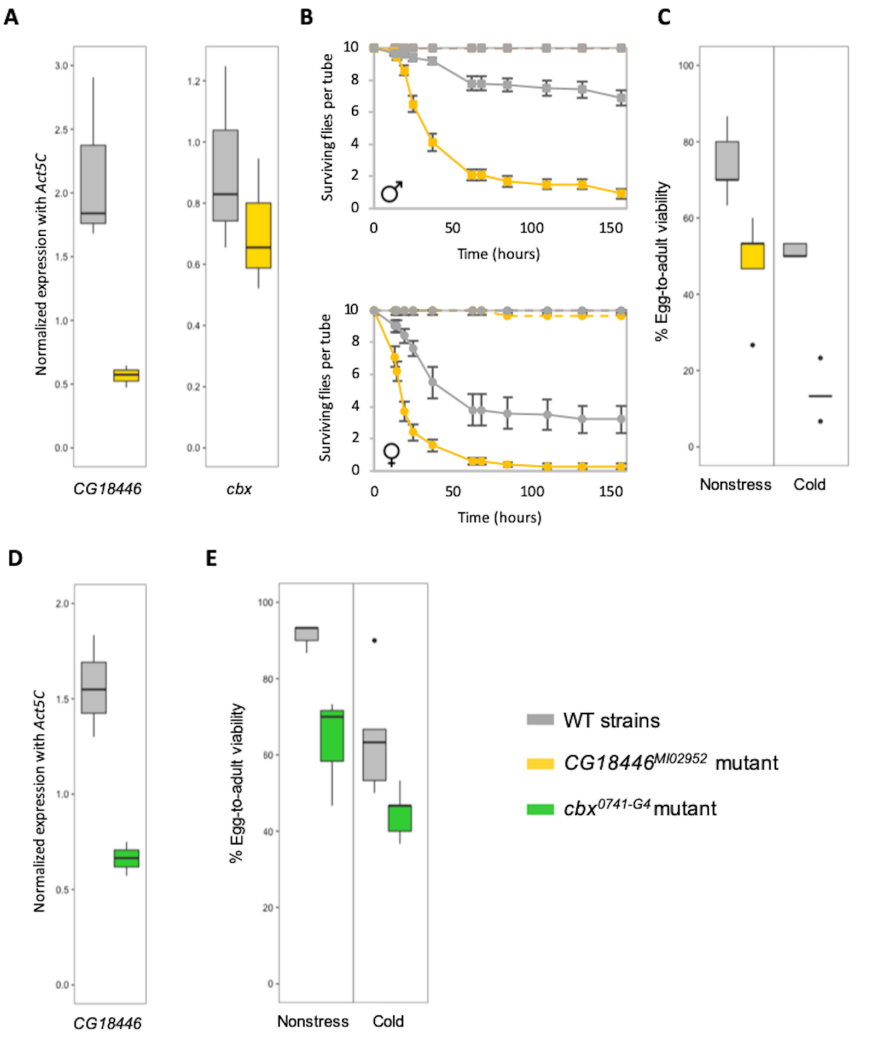


Figure 2. *CG18446*^{MI02952} mutants are more sensitive to *P. entomophila* infection and have reduced egg-to-adult viability under cold-stress. A) Normalized expression of *CG18446* and *cbx* in the *CG18446*^{MI02952} mutant strain compared with the WT strain. **B)** Survival curves of *CG18446*^{MI02952} mutant males and females compared with the WT strain. Survival curves in non-infected conditions are depicted as dotted lines and survival curves after *P. entomophila* infection are depicted as continuous lines. Error bars represent SEM. **C)** Egg-to-adult viability in nonstress and in cold-stress conditions of *CG18446*^{MI02952} mutants compared with the WT strain. **D)** Normalized expression of *CG18446* in *cbx*$^{0741-G4}$ mutant strain compared with the WT strain. **E)** Egg-to-adult viability in nonstress and in cold-stress conditions of *cbx*$^{0741-G4}$ mutants compared with the WT strain.

and Table S3). *CG18446*^{MI02952} mutant flies also showed reduced egg-to-adult viability in nonstress and in cold-stress conditions (Fig. 2C, Table S4). Thus, our results confirm that changes in *CG18446* expression are associated with changes in immune- and cold-stress responses (Fig. 2A-C) (37, 38).

Previous expression analysis of the *cbx*$^{0741-G4}$ mutant confirmed that it is a null mutant of its target gene (34). However, we found that *cbx*$^{0741-G4}$ also down-regulates the expression of *CG18446* (Fig. 2D, Table S2). *cbx*$^{0741-G4}$ mutant flies were previously described to be more tolerant to *P. entomophila* infection (34). We showed that *cbx*$^{0741-G4}$ mutants also have reduced egg-to-adult viability in nonstress and in cold-stress conditions (Fig. 2E and Table S4).

Overall, we found that *CG18446*^{MI02952} mutants, in which only *CG18446* is down-regulated, were more sensitive to *P. entomophila* infection. On the other hand, *cbx*$^{0741-G4}$ mutants, in which *cbx* and *CG18446* expression were knockout and down-regulated, respectively, were more tolerant to *P. entomophila* infection (34). These results suggested that *cbx* might also be involved in immune response as previously suggested (34, 39). On the other hand, we found that *CG18446*^{MI02952} and *cbx*$^{0741-G4}$ mutants were more sensitive to cold-stress conditions, suggesting that either *cbx* does not affect this stress responses or, if it does, it has the same effect as *CG18446*.

Our results add to the previous evidence suggesting that changes in expression in *CG18446* and *cbx* affect immune-

stress responses, while *CG18446* also affects cold-stress responses (34, 37-39).

***FBti0019985* is not associated with changes in *cbx* expression in nonstress, immune- or cold-stress conditions**

We have previously shown that *FBti0019985* up-regulates *CG18446* embryonic expression in nonstress conditions (33). As mention above, an artificial TE insertion located in the third intron of *cbx*, affects both *cbx* and *CG18446* expression (Fig. 1). Thus, we tested whether *FBti0019985*, that is inserted in the first intron of *cbx*, is also associated with expression changes of this gene. We measured *cbx* expression in outbred populations that differ in the presence/absence of *FBti0019985* in two experimental conditions: immune- and cold-stress (35). We found that the expression level of *cbx* is not significantly affected by the insertion genotype (presence/absence of *FBti0019985*), the experimental condition (nonstress vs stress) or the interaction between the genotype and the experimental condition (ANOVA p-value > 0.05 for all comparisons) (Fig. 3, Table 1, and Table S5). We thus focused on the effect of *FBti0019985* on *CG18446* for the rest of this work.

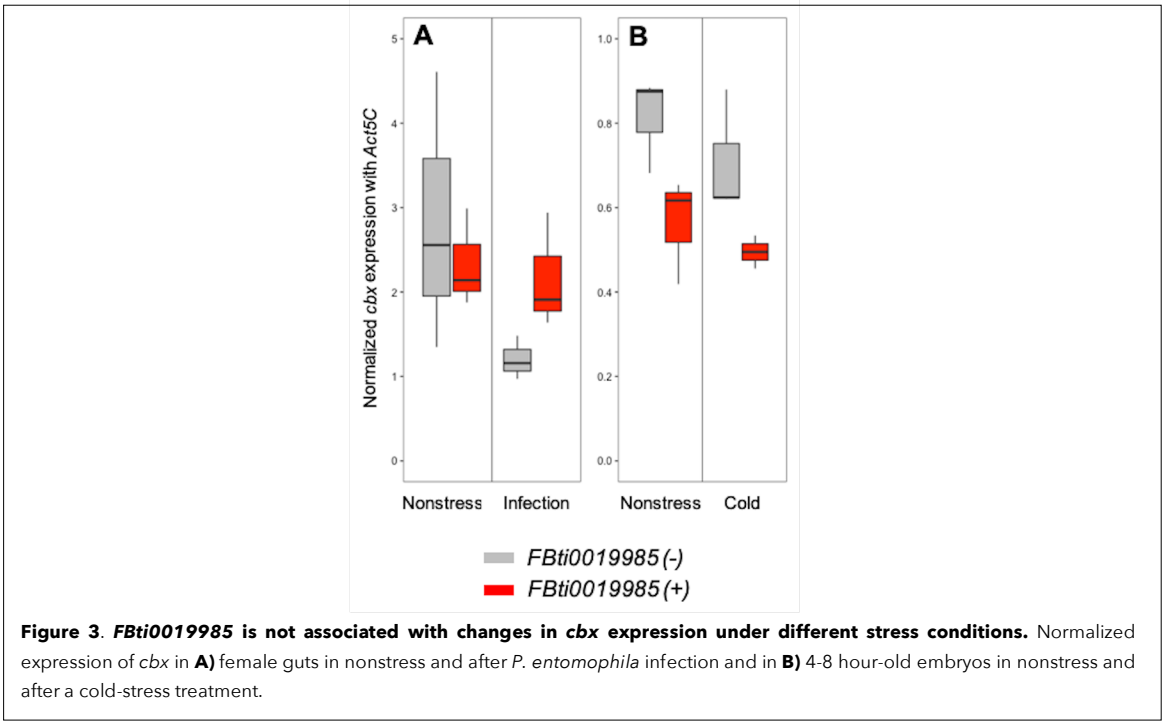


Figure 3. ***FBti0019985* is not associated with changes in *cbx* expression under different stress conditions.** Normalized expression of *cbx* in **A)** female guts in nonstress and after *P. entomophila* infection and in **B)** 4-8 hour-old embryos in nonstress and after a cold-stress treatment.

***FBti0019985* is associated with *CG18446* down-regulation in nonstress conditions and with *CG18446* up-regulation under immune-stress**

We next investigated whether outbred flies with *FBti0019985* were associated with *CG18446* expression changes in nonstress and after a *P. entomophila* oral infection. We found that in nonstress conditions, *FBti0019985* was associated with *CG18446* down-regulation in female guts (Fig. 4A, Table 1, and Table S6). However, under immune stress conditions, *FBti0019985* was associated with *CG18446* up-regulation (Fig. 4A, Table 1, and Table S6). These results are consistent with *FBti0019985* acting as an enhancer in immune stress conditions but not in nonstress as previously reported by Ullastres et al. (2019) (34).

To further study the role of *FBti0019985* in immune-stress response, we used the CRISPR/Cas9 genome engineering tool to delete *FBti0019985* in the outbred natural population. We generated two different CRISPR-mutant fly strains with a precise deletion of *FBti0019985* (see Material and Methods). However, these fly strains also have small indels in the 5'UTR of *CG18446* (Fig. 1B). Nevertheless, the core promoter elements *Initiator* (INR) and *Downstream Promoter Element* (DPE) were conserved in the CRISPR-mutant strains, suggesting that *CG18446* expression might not be affected by these small indels (Fig. 1B). Besides the two CRISPR mutants, we also measured *CG18446* expression in the outbred populations, which are the baseline of the experiment. As expected based on our previous experiment (Fig. 4A), we found that female guts with *FBti0019985* had decreased *CG18446* expression in nonstress conditions while this expression was increased under immune-stress conditions (Fig. 4B, Table 1, and Table S6). Thus, if *FBti0019985* is the cause of these expression changes we would expect the CRISPR mutants to show *CG18446* up-regulation in nonstress conditions and down-regulation in stress conditions. Indeed, this is what we found when analyzing $FBti0019985^{CRISPR1}$ mutant (Fig. 4B, Table 1, and Table S6). On the other hand, the $FBti0019985^{CRISPR2}$ mutant, showed down-regulation of *CG18446* in nonstress conditions and no differences in immune-stress conditions compared with the outbred strain with *FBti0019985* (Fig. 4B, Table 1, and Table

S6). Thus, in females only *FBti0019985^{CRISPR1}* mutant gave the expected results if *FBti0019985* is responsible for the changes in expression of *CG18446* in nonstress and in stress conditions (Fig. 4B).

Contrary to what we found in females, the level of expression of *CG18446* in guts from outbred males without *FBti0019985* was very low in nonstress conditions (Fig. 4C, Table 1, and Table S6). Moreover, similar to the results in females, guts from male flies with *FBti0019985* had increased *CG18446* expression levels after the infection (Fig. 4C, Table 1, and Table S6). These results suggest that *FBti0019985* also acts as an enhancer in males under immune-stress conditions. The two CRISPR mutants showed *CG18446* expression changes consistent with the deletion of *FBti0019985* in males: *CG18446* is down-regulated in nonstress and in immune stress conditions (Fig. 4C, Table 1, and Table S6).

Overall, under nonstress conditions, *FBti0019985* was associated with *CG18446* down-regulation only in female guts. Our results indicated that there are differences in *CG18446* expression between female and males without *FBti0019985* under nonstress conditions. However, we found that *FBti0019985* was associated with *CG18446* up-regulation under immune-stress conditions in both female and in male guts. While both CRISPR mutants showed *CG18446* expression changes consistent with *FBti0019985* being the causative mutation in males, only *FBti0019985^{CRISPR1}* showed consistent changes in *CG18446* expression in females (Fig. 4, Table 1).

Table 1. Two-way ANOVA results of the *cbx* and *CG18446* expression analysis. Statistically significant results are highlighted in bold.

Gene	Stress	Strain	Tissue / stage	Sex	Insertion genotype effect	Experimental condition effect	Interaction effect
cbx	Immunity	Outbred *FBti0019985* + vs. outbred *FBti0019985* -	Gut	Females	0.162	0.090	0.215
	Cold	Outbred *FBti0019985* + vs. outbred *FBti0019985* -	Embryo	-	0.136	0.873	0.350
CG18446	Immunity	Outbred *FBti0019985* + vs. outbred *FBti0019985* - (Rep. 1)	Gut	Females	0.505	0.380	**0.006**
		Outbred *FBti0019985* + vs. outbred *FBti0019985* - (Rep. 2)	Gut	Females	**0.001**	**0.001**	**<0.001**
		FBti0019985^{CRISPR1} vs. outbred *FBti0019985* +	Gut	Females	0.151	**0.019**	**<0.001**
		FBti0019985^{CRISPR2} vs. outbred *FBti0019985* +	Gut	Females	0.941	**0.002**	0.322
		Outbred *FBti0019985* + vs. outbred *FBti0019985* -	Gut	Males	**0.004**	**0.001**	**0.005**
		FBti0019985^{CRISPR1} vs. outbred *FBti0019985* +	Gut	Males	**0.002**	**0.003**	**0.003**
		FBti0019985^{CRISPR2} vs. outbred *FBti0019985* +	Gut	Males	**0.002**	**0.003**	**0.003**
	Cold	Outbred *FBti0019985* + vs. outbred *FBti0019985* -	Embryo	-	0.207	**0.001**	**0.011**
	Ethanol	Outbred *FBti0019985* + vs. outbred *FBti0019985* -	Adult	Females	**0.002**	0.456	0.945

FBti0019985 is associated with tolerance to *P. entomophila* infection

To test whether the up-regulation of *CG18446* in outbred flies with *FBti0019985* under immune-stress conditions affects fly survival, we performed infection tolerance assays with *P. entomophila*. In outbred populations, we found that both female and males flies with *FBti0019985* were more tolerant to infection than flies without the insertion (Fig. 4D and Table S7). Thus, up-regulation of *CG18446* is associated with infection tolerance in outbred populations.

We then repeated the experiment using the outbred populations as the baseline, and the two CRISPR-mutant strains. As expected, we observed that outbred flies with *FBti0019985* were more tolerant to infection compared with flies without the insertion (Fig. 4D and Table S7). *FBti0019985^{CRISPR1}* strain was more sensitive to infection compared to flies with *FBti0019985* in both females and males, as expected because *CG18446* is not up-regulated in this strain (Fig. 4B and 4D and Table S7). Finally, *FBti0019985^{CRISPR2}* females showed no differences, while males were more sensitive to infection compared to flies with *FBti0019985* (Fig. 4D and Table S7). These results are consistent with the *CG18446* expression results found in the *FBti0019985^{CRISPR2}* mutant (Fig. 4B-D).

Overall, our results with outbred and CRISPR-mutant strains suggested that *CG18446* up-regulation is associated with tolerance to *P. entomophila* infection in both sexes. Moreover, since *FBti0019985* is driving the expression of *CG18446* under

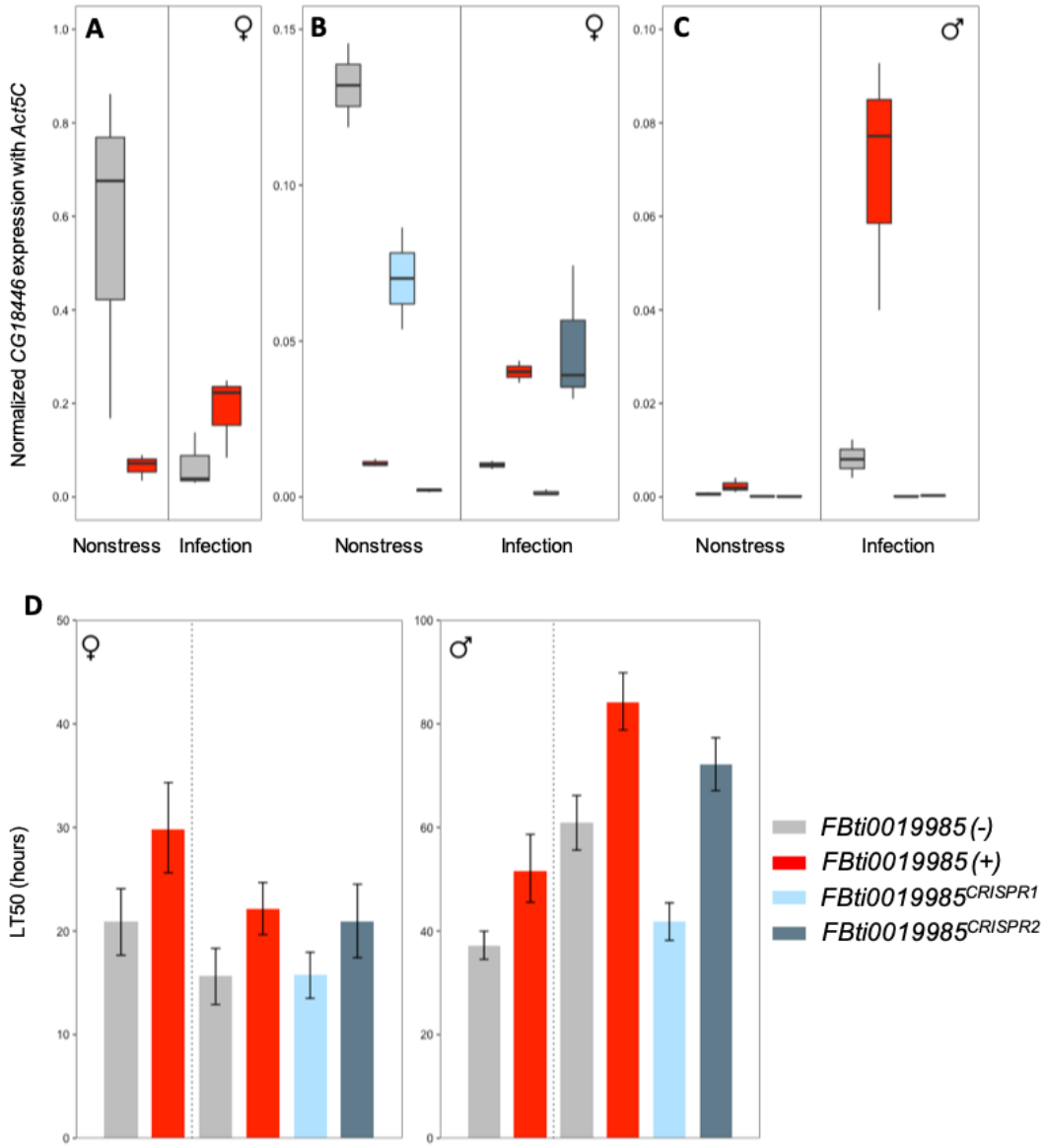


Figure 4. ***FBti0019985* is associated with *CG18446* up-regulation in guts under immune-stress conditions and with increased tolerance to *P. entomophila* infection. A)** Normalized expression of *CG18446* with *Act5C* in nonstress and immune-stress conditions of female guts from outbred populations with and without *FBti0019985*. **B)** Normalized expression of *CG18446* with *Act5C* in nonstress and immune-stress conditions of female guts from outbred and *CRISPR*-mutant strains. **C)** Normalized expression of *CG18446* with *Act5C* in nonstress and immune-stress conditions of male guts from outbred and *CRISPR*-mutant strains. **D)** LT50 values after an oral infection with *P. entomophila* in outbred and *CRISPR*-mutant strains. Different replicates are separated with a dotted line.

immune-stress conditions, we suggest that *FBti0019985* is associated with tolerance to *P. entomophila* infection.

***FBti0019985* acts as an enhancer in embryos under nonstress conditions while it is associated with *CG18446* down-regulation in cold-stress**

We have previously reported that *FBti0019985* is associated with up-regulation of *CG18446* in embryos in nonstress conditions (33). In this work, we confirmed these results by comparing outbred populations with and without *FBti0019985* (Fig. 5A, Table 1, and Table S6). Furthermore, we found that *CG18446* is down-regulated under cold-stress (Fig. 5A, Table 1, and Table S6).

We finally tested whether *FBti0019985* can act as an enhancer in embryos in nonstress conditions using the *lacZ* enhancer reporter gene assay. We found that transgenic embryos containing the *FBti0019985* sequence showed significant increased *lacZ* expression in nonstress conditions while no differences in expression were found under cold-stress compared with the empty vector (p-value = 0.013 and 0.155, respectively; Fig. 5B and Table S8).

Altogether, these results suggest that *FBti0019985* is acting as an enhancer promoting the expression of *CG18446* in embryos under nonstress conditions. However, under cold-stress, *FBti0019985* does not act as an enhancer causing a *CG18446* down-regulation (Fig. 5A and 5B).

***FBti0019985* is not associated with increased egg-to-adult viability under cold-stress in outbred populations**

FBti0019985 was previously associated with increased egg-to-adult viability in nonstress and in cold-stress conditions in three different genetic backgrounds (33). We then subjected the outbred strains with and without the element to the same cold-stress conditions and measured egg-to-adult viability. ANOVA analyses showed that the effect of the experimental condition (nonstress or cold-stress) was significant ($p = 0.046$) while the insertion genotype (presence/absence of *FBti0019985*) and the interaction between these two factors were not significant ($p = 0.572$ and $p = 0.411$, respectively) (Fig. 5C and Table S9). Although we observed that outbred flies with *FBti0019985* had high viability levels under nonstress and cold-stress conditions (82.7% and 66.0%, respectively), the outbred population without the insertion also showed high viability levels in both conditions (74.7% in nonstress and 62.2% in cold-stress). Indeed, viability levels in nonstress of flies without *FBti0019985* were even higher than the ones expected for flies without the element considering previous egg-to-adult viability results in three different natural populations (49.2% ± 22.2) (33). Thus, we could not find differences in egg-to-adult viability between outbred populations with and without *FBti0019985* probably due to the high viability of the outbred population without the element.

***FBti0019985* is not acting as an enhancer in adult flies and under ethanol-stress conditions**

As mentioned above, *FBti0019985* is acting as an enhancer in embryo stages under nonstress conditions and in guts under immune-stress (34) (Fig. 5B). However, we did not find this enhancer activity when we exposed embryos to cold-stress (Fig. 5B). We then performed *in vivo* assays to test whether *FBti0019985* acts as an enhancer in adult stages and under a different stress condition present in the fly natural environment: ethanol-stress. Flies feed and breed on rotten fruit that contain small concentrations of ethanol (43, 44). We found no *lacZ* expression in adult transgenic flies under nonstress conditions or after the ethanol exposure (Fig. 6A and Table S8), suggesting that *FBti0019985* is not acting as an enhancer neither in adult stages nor under ethanol-stress (Fig 6A). To confirm that *FBti0019985* is not driving the expression of *CG18446*, we also measured the expression of *CG18446* in the outbred populations under these specific contexts. As expected, we found that *FBti0019985* was associated with *CG18446* down-regulation in adults in both nonstress and ethanol-stress conditions (Fig. 6B, Table S6).

In all, these results suggest that *FBti0019985* is not acting as enhancer in adults under nonstress and ethanol-stress conditions. These findings correlate with the *CG18446* down-regulation found in outbred populations with *FBti0019985* under these conditions (Fig. 6).

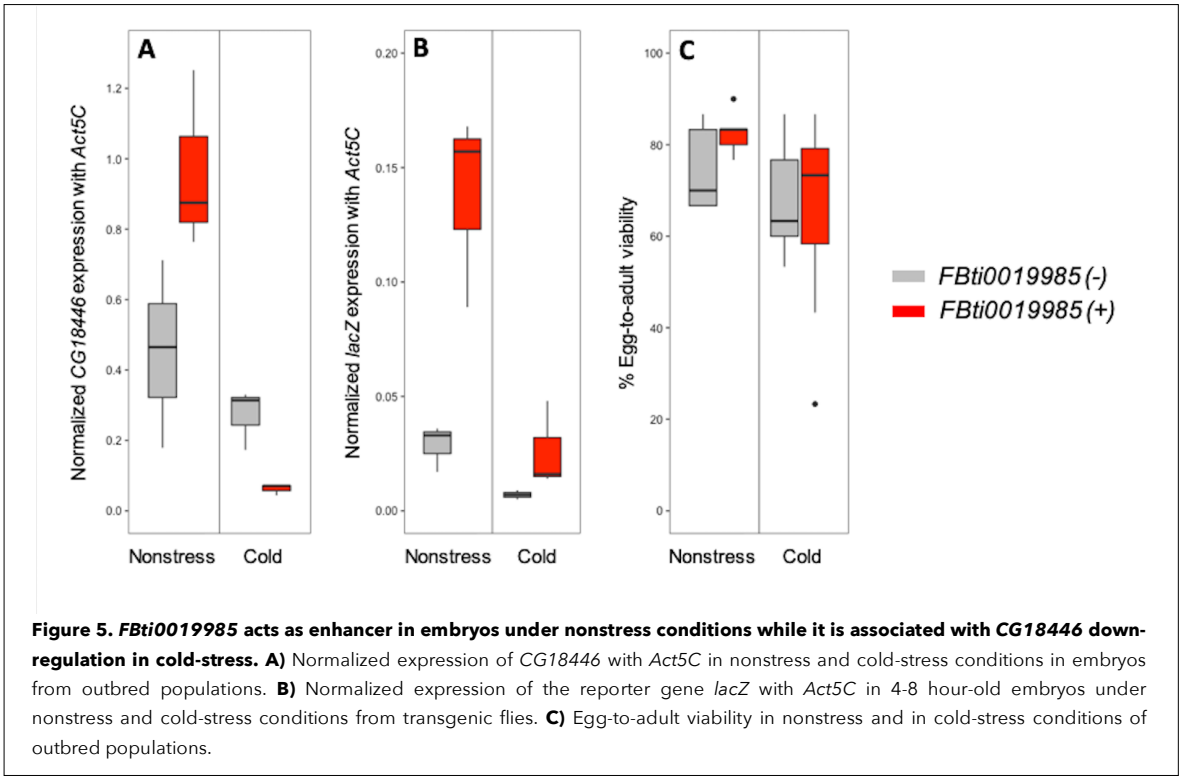


Figure 5. *FBti0019985* acts as enhancer in embryos under nonstress conditions while it is associated with *CG18446* down-regulation in cold-stress. A) Normalized expression of *CG18446* with *Act5C* in nonstress and cold-stress conditions in embryos from outbred populations. **B)** Normalized expression of the reporter gene *lacZ* with *Act5C* in 4-8 hour-old embryos under nonstress and cold-stress conditions from transgenic flies. **C)** Egg-to-adult viability in nonstress and in cold-stress conditions of outbred populations.

***CG18446* expression changes are not due to *FBti0019985* transcript start site**

FBti0019985 provides an alternative transcription start site (TSS) to *CG18446* in embryos under nonstress conditions (33). This alternative transcript includes 50 bp of the *FBti0019985* sequence and it is 75 bp longer compared with the transcript that does not start in the TE. We thus tested whether changes in *CG18446* expression in flies with *FBti0019985* could be due to the presence of this alternative transcript. We found that the alternative transcript was present in the gut under both nonstress and under immune-stress conditions (Table S6). However, most of the *CG18446* expression in both conditions was due to the shorter transcript (99.41% and 96.12%, respectively) (Table S6). We also found that the *CG18446* alternative transcript starting in *FBti0019985* was present in embryos under cold-stress. Nevertheless, our results suggest that the up-regulation of *CG18446* in embryos under nonstress is not due to the presence of the alternative transcript since the majority of the expression came from the short transcript (86.36%) (Table S6). Similar results were obtained in cold-stress conditions where 89.79% of the expression corresponds to the short transcript.

Overall, the observed *CG18446* expression changes are not caused by a preference towards the alternative TSS located in *FBti0019985* (Table S6).

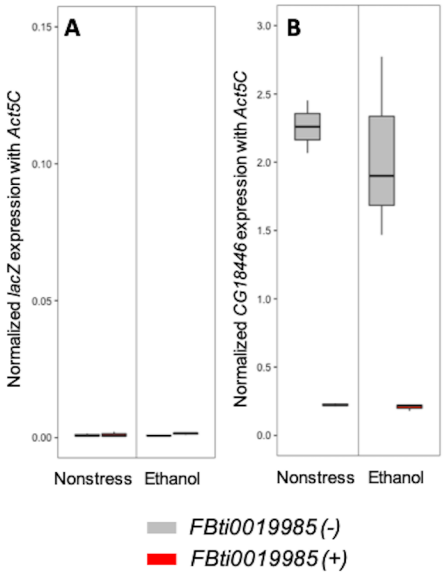


Figure 6. *FBti0019985* is not acting as enhancer under ethanol-stress. A) Normalized expression of the reporter gene *lacZ* with *Act5C* in transgenic adult female flies under nonstress conditions and after an ethanol exposure. **B)** Normalized expression of *CG18446* with *Act5C* in nonstress and ethanol-stress conditions of adult females from outbred populations with and without *FBti0019985*.

Discussion

In this work, we found that the effect of *FBti0019985* insertion on *CG18446* expression depends on the environmental conditions and the developmental stage. TEs have been repeatedly described to contain cis-regulatory elements capable of regulating the expression of nearby genes depending on the environmental context (20, 45, 46). Moreover, TEs can also act as enhancers in specific tissues or developmental stages (47-50). For instance, several retroelements were found to be spatially and temporally regulated during *D. melanogaster* embryogenesis, probably due to the interaction of transcription factors with cis-regulatory elements within them (51). Also, in the same species, the *hobo*-related sequence *hoboVA* has cis-regulatory elements of developmental genes, such as *hunchback* and *even-skipped*, that are mainly expressed during early embryonic stages (52, 53). However, to the best of our knowledge, a single transposable element insertion has not been shown to affect the expression of its nearby gene depending on both environmental conditions and the developmental stage. The versatile effects of *FBti0019985* could be due to an alteration of the spacing between the cis-regulatory elements present in the *CG18446* gene promoter region and by the presence of TFBSs in its sequence. On the one hand, in adults or after a cold and ethanol exposure, *FBti0019985* was associated with *CG18446* down-regulation, and we showed that it was not acting as an enhancer. Since *FBti0019985* is inserted in the 5'UTR region of *CG18446,* it could be possible that the insertion affects the spacing of existing cis-regulatory elements of the promoter, thus leading to a decrease *CG18446* expression. Indeed, this

mechanism was already described for different naturally occurring TEs inserted in close proximity to the *D. melanogaster hsp70* gene (54). The replacement of these TEs by intergenic sequences confirmed that the presence of the different TEs disrupted the *hsp70* promoter architecture by spacing cis-regulatory elements needed for *hsp70* transcription (55). On the other hand, in immune-stress conditions and in embryos under nonstress, *FBti0019985* was associated with *CG18446* up-regulation and it was shown to act as an enhancer (33, 34). Indeed, TFBSs related with both immune responses and development were predicted in the *FBti0019985* sequence (33, 46). Yet, functional validation of these predicted TFBSs will be necessary to specifically link *FBti0019985* with the context-specific *CG18446* up-regulation.

We also found that the *CG18446* up-regulation driven by *FBti0019985* in guts under immune-stress was associated with tolerance to *P. entomophila* infection. Thus, besides the already described increased viability in nonstress and in cold-stress (33), our data suggested that *FBti0019985* is also involved in tolerance to bacterial infection. *CG18446* is a transcription factor that has been associated with different processes. Recently, *CG18446* has been found to simultaneously regulate metabolic and immune processes, and it was renamed *Linking Immunity and Metabolism* (*Lime*) (38). Authors showed that *CG18446* mutants have low levels of glycogen and trehalose, the two main energy-storing molecules in the fly (38). During infection, *CG18446* mutants fail to increase glucose levels probably because they already had low levels of glycogen and trehalose (38). This glucose deprivation affects in turn, immune cell proliferation and activation (38). Indeed, *CG18446* mutants develop fewer immune cells in the lymph gland and in circulation (38). Moreover, changes in the metabolism of sugars could also be related with the observed increased in egg-to-adult viability after cold-stress in flies with *FBti0019985* (33). Cold-stress induces changes in the fly metabolite profile. Specifically, high levels of trehalose were found after cold-shock and rapid cold hardening treatments in adults as well as during long-term cold acclimation in larvae (56, 57). Further experiments will be needed to link the effect of *FBti0019985* on *CG18446* expression with changes in sugar metabolism and its subsequent effects on cold and immune tolerance. Finally, the high expression levels of *CG18446* in embryonic stages suggested that it may have an important role during development (58). In fact, *CG18446* interacts with several transcription factors involved in developmental processes such as *foxo* (59). *CG18446* has also been reported as a direct target of *Notch* activation, a receptor that controls different developmental processes and contributes to the maintenance of adult tissues (60, 61).

Overall, this paper identified *FBti0019985* as a TE that modulates the expression of *CG18446* depending on different developmental stages and environmental conditions. Besides conferring increased egg-to-adult viability in nonstress and cold-stress conditions, we have associated *FBti0019985* with tolerance to immune-stress. These results reflect the importance of considering the effect of a candidate adaptive insertion under different contexts to fully characterize its adaptive effects.

Material and methods

Fly stocks

Fly stocks were reared on fly food medium in a 12:12 hr light/dark cycle at 25C.

Laboratory mutant and RNAi knock-down strains

We used three laboratory mutant strains and one RNAi knock-down strain that were likely to affect *CG18446* and / or *cbx* genes. We used $CG18446^{MI02952}$ mutants (Bloomington Drosophila Stock Center (BDSC) stock number #36170) that contains a MiMIC insertion in the first exon of *CG18446* gene and a RNAi knock-down strain for the same gene (stock number #33735) (Table S1). We also used $cbx^{0741\text{-}G4}$ mutants (stock number #63767) and cbx^{c00428} mutants (stock number #10067) with a *PiggyBac* insertion in the first and third intron of *cbx*, respectively. Insertional mutant strains were compared to WT strains with similar genetic backgrounds in all the molecular and phenotypic assays (Table S1). The RNAi knock-down strain was crossed with a strain containing *GAL4* in heterozygosis under the control of an *Actin* promoter. Offspring containing *GAL4* were compared to the parental line with the *Actin* promoter in all the molecular and phenotypic assays (Table S1).

CG18446 and *cbx* expression levels were measured in all the strains mentioned above and only those strains performing changes in expression were used for phenotypic assays ($CG18446^{MI02952}$ and $cbx^{0741\text{-}G4}$).

Outbred strains

We used the laboratory outbred populations with and without *FBti0019985* generated in Merenciano et al. (2019) (35). Briefly, they were generated by a round-robin cross-design of inbred lines from the Drosophila Genetic Reference Panel (DGRP) and isofemale lines from different European populations (35). Outbred populations were maintained by random mating with a large population size for over five generation before starting the experiments.

CRISPR/Cas9 mutant strains

Guide RNAs (gRNAs) were designed in the *FBti0019985* flanking region and cloned into pCFD5 (62) plasmid following the pCFD5 cloning protocol (www.crisprflydesign.com) using the primers 5'-gcggcccgggttcgattcccggccgatgcaagctagacttatttgagatagttttaga gctagaaatagcaag-3' and 5'-attttaacttgctatttctagctctaaaaccaga gaaacgtcgagctgcgtgcaccagccgggaatcgaaccc-3'. A donor DNA containing two homology arms flanking the DsRed sequence for homology repair were cloned into the pHD-ScarlessDsRed plasmid. Homology arm 1 contained the sequence in 2R:9870299-9871095 (Release 6) from the outbred population with *FBti0019985*. Homology arm 2 contained the sequence in 2R:9871529-9872365 (Release 6) from the outbred population with *FBti0019985*. The pCFD5 plasmid containing the gRNAs, the donor pHD-ScarlessDsRed plasmid containing the homology arms, and a plasmid containing Cas9 endonuclease were co-

injected as a unique mix into approximately 550 embryos from the outbred population with *FBti0019985*. All the injections were performed using the following mix concentrations: pCFD5 plasmid at 100ng/ul, donor plasmid at 500ng/ul, and Cas9 plasmid at 250ng/ul. Offspring was screened for eye fluorescence. Flies with the desired mutation were backcrossed with the parental line for a minimum of five generations. Then, two homozygous strains containing the deletion of *FBti0019985* were established. The deletion was checked by PCR with two primer pairs: 5'-aacaatgcaagtccgtgctc-3', 5'-gtggttcctccacccttgtg-3' and 5'-ggccgcgactctagatcataatc-3' and 5'-gtggttcctccacccttgtg-3'. PCR bands were confirmed by Sanger sequencing.

Transgenic strains

We used transgenic flies for *lacZ* enhancer reporter assays we used transgenic with the *FBti0019985* sequence cloned in front of the reporter gene generated in Ullastres et al. (2019) (34) and we compared them with transgenic strains with the *placZ.attB* empty vector to control for possible *lacZ* expression driven by the vector sequence itself.

Expression analysis

Sample collection

P. entomophila infection: 5-7 day-old flies from every strain were separated by sex and placed in six vials with fresh food in groups of 25-35. We allowed flies to recover from CO_2 anesthesia for 24 h at 25C. To expose the flies to the gram-negative bacteria *P. entomophila* infection, we followed the protocol described in Neyen et al. (2014) (63). Briefly, after two hours of starvation, we transfer 75-105 flies in groups of 25-35 into three vials with fresh food and a filter paper soaked with 120ul of a solution containing 1.25% sucrose and bacterial preparation adjusted to a final OD_{600} = 50 for females and OD_{600} = 150 for males. Flies were kept at 29C, the optimal temperature condition for *P. entomophila* infection. Simultaneously, a total of 75-105 flies were also transferred in groups of 25-35 into three vials with fresh food and a filter paper soaked with 120ul of a solution containing sterile LB with 1.25% sucrose as a control. Guts were dissected after 12 and 19.5 hours, for females and males, respectively. Samples were then flash-frozen in liquid nitrogen and stored at -80C until sample processing.

Cold stress treatment: 5-7 day-old flies were allowed to lay eggs at 25C in a fly cage with egg-laying medium (2% agar with apple juice and a piece of fresh yeast) for four hours. After these four hours, adults were removed and the plate containing embryos was kept at 1C for four additional hours. Simultaneously, another plate with embryos was kept at 25C for four additional hours as a control. 4-8 hour-old embryos were then collected from both plates using the method described in Schou et al. (2013) (64) adding an additional step of dechorionation during 10 minutes with 50% bleach. Finally, samples were flash-frozen in liquid nitrogen and stored at -80C until sample processing.

RNA extraction and cDNA synthesis

RNA was extracted using the GeneEluteTM Mammalian Total RNA Miniprep Kit following manufacturer´s instructions (Sigma). RNA was then treated with DNase I (Thermo). cDNA was synthesized from a total of 250-1,000 ng of RNA using the NZY Fisrt-Strand cDNA synthesis kit (NZYTech).

qRT-PCR analysis

CG18446 expression was measured using the forward primer 5' -gagcagttggaatcgggttttac -3' and the reverse primer 5'-gtatgaatcgcagtccagccata-3' spanning 99 bp cDNA in the exon 1/exon 2 junction. *cbx* expression was measured with the forward primer 5'-gggaaaacgatctgggagca -3' and the reverse primer 5'- gtcggagaagttgagtggga -3' spanning 233 bp cDNA in the exon 2/exon 3 junction. *lacZ* reporter gene expression was measured using the forward primer 5'-cctgctgatgaagcagaacaact-3' and the reverse primer 5'-gctacggcctgtatgtggtg-3'. Gene expression was normalized with Act5C (5' -gcgcccttactctttcacca-3' and 5' -atgtcacggacgatttcacg-3' primers). We performed the qRT-PCR analysis with SYBR Green (BioRad) or with the qPCRBIO SyGreen Mix Lo-Rox (PCRBiosystems) on iQ5 and CFX384 Thermal cyclers, respectively. Results were analyzed using the dCT method (65).

Transcript start site detection

To detect whether *FBti0019985* is adding an alternative TSS to the *CG18446* gene in outbred flies carrying *FBti0019985* after different stress conditions, we performed RT-PCRs. We used the forward primer 5'- aaaactcaacgagtaaagtcttc -3' and the reverse primer 5'- tataaagttccaacgcccagc -3' to detect the *CG18446* transcript starting in the TE. The forward primer 5'-cgcagagaaacgtcgagctg -3' and the reverse primer 5'-cacgttaaattcactagggtggc -3' were used to detect *CG18446* total transcript. Outbred population without *FBti0019985* was used as control sample.

Phenotypic assays

P. entomophila infection

100 5-7 day-old male flies and 100 5-7 day-old female from the different strains were infected with the gram-negative bacteria *P. entomophila* infection as described before. Simultaneously, a total of 30 were tested as controls. We counted the number of dead flies in every vial at different time points until a maximum of 157 hours post infection.

For *CG18446* mutants, log-rank tests were performed to analyze survival curves with SPSS v21 software. For outbred and CRISPR-mutant populations, we calculated lethal time 50 (LT50), the timepoint at which mortality was 50%, using Probit analysis (66, 67).

Egg-to-adult viability under cold stress

5-7 day-old flies from every strain were allowed to lay eggs for 4 h at 25C in a fly cage with egg-laying medium (2% agar with apple juice and a piece of fresh yeast) for four hours. Then, adults were removed from the cage and plates with were kept four additional hours at 25C. After that, 4-8 hour-old embryos were collected using the method described in Schou et al. (2013) (64) and placed in vials with fresh food in groups of 30.

In total, 240-300 embryos were tested for each strain. Cold-stressed vials were kept at 1C for 15 h and then maintained at 25C until adult emergence. Simultaneously, control vials were kept at 25C and never exposed to cold. Percentage egg-to-adult viability was calculated based on the number of emerged flies to the total number of embryos placed in each vial. Statistical significance was calculated performing ANOVA using SPSS v.21 combining all the data into a full model: experimental condition (stress and nonstress), insertion genotype (presence/absence of the insertion) and interaction between these two factors.

Acknowledgements

We thank Fillip Port from the Division of Signaling and Functional Genomic led by Prof. Dr. Michael Boutros for his advice in the generation of the CRISPR mutants.

References

1. Mackay TFC, Huang W. Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. Wiley Interdiscip Rev Dev Biol. 2018;7(1).
2. Nelson TC, Jones MR, Velotta JP, Dhawanjewar AS, Schweizer RM. UNVEILing connections between genotype, phenotype, and fitness in natural populations. Mol Ecol. 2019;28(8):1866-76.
3. Mackay TF. Mutations and quantitative genetic variation: lessons from Drosophila. Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1229-39.
4. Barrett RD, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet. 2011;12(11):767-80.
5. Orgogozo V, Morizot B, Martin A. The differential view of genotype-phenotype relationships. Front Genet. 2015;6:179.
6. Grüneberg H. An analysis of the "pleiotropic" efects of a new lethal mutation in the rat *(Mus norvegicus)*. Proc R Soc Lond B; 1938. p. 123-44.
7. Paaby AB, Rockman MV. The many faces of pleiotropy. Trends Genet. 2013;29(2):66-73.
8. Mills MG, Greenwood AK, Peichel CL. Pleiotropic effects of a single gene on skeletal development and sensory system patterning in sticklebacks. Evodevo. 2014;5(1):5.
9. Greenwood AK, Mills MG, Wark AR, Archambeault SL, Peichel CL. Evolution of Schooling Behavior in Threespine Sticklebacks Is Shaped by the Eda Gene. Genetics. 2016;203(2):677-81.
10. Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9(11):855-67.
11. Sackton TB, Hartl DL. Genotypic Context and Epistasis in Individuals and Populations. Cell. 2016;166(2):279-87.
12. Evangelou A, Ignatiou A, Antoniou C, Kalanidou S, Chatzimatthaiou S, Shianiou G, et al. Unpredictable Effects of the Genetic Background of Transgenic Lines in Physiological Quantitative Traits. G3 (Bethesda). 2019;9(11):3877-90.
13. Natarajan C, Jendroszek A, Kumar A, Weber RE, Tame JRH, Fago A, et al. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose. PLoS Genet. 2018;14(4):e1007331.
14. Signore AV, Yang YZ, Yang QY, Qin G, Moriyama H, Ge RL, et al. Adaptive Changes in Hemoglobin Function in High-Altitude Tibetan Canids Were Derived via Gene Conversion and Introgression. Mol Biol Evol. 2019;36(10):2227-37.
15. Andersen EC, Bloom JS, Gerke JP, Kruglyak L. A variant in the neuropeptide receptor npr-1 is a major determinant of Caenorhabditis elegans growth and physiology. PLoS Genet. 2014;10(2):e1004156.
16. Tigano A. A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird. In: Reiertsen TK, Walters, J. R., Friesen, V. L., editor.: *bioRxiv*; 2018.
17. Vickrey AI, Bruders R, Kronenberg Z, Mackey E, Bohlender RJ, Maclary ET, et al. Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. Elife. 2018;7.
18. Nelson TC, Monnahan PJ, McIntosh MK, Anderson K, MacArthur-Waltz E, Finseth FR, et al. Extreme copy number variation at a tRNA ligase gene affecting phenology and fitness in yellow monkeyflowers. Mol Ecol. 2019;28(6):1460-75.
19. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013;22(6):1503-17.
20. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18(2):71-86.
21. Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-5.
22. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083-7.
23. Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. Nature. 2016;536(7616):329-32.
24. Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, et al. ZmCCT9 enhances maize adaptation to higher latitudes. Proc Natl Acad Sci U S A. 2018;115(2):E334-E41.
25. Li ZW, Hou XH, Chen JF, Xu YC, Wu Q, González J, et al. Transposable Elements Contribute to the Adaptation of Arabidopsis thaliana. Genome Biol Evol. 2018;10(8):2140-50.
26. Esnault C, Lee M, Ham C, Levin HL. Transposable element insertions in fission yeast drive adaptation to environmental stress. Genome Res. 2019;29(1):85-95.
27. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A single p450 allele associated with insecticide resistance in Drosophila. Science. 2002;297(5590):2253-6.
28. Aminetzach YT, Macpherson JM, Petrov DA. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science. 2005;309(5735):764-7.
29. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication. PLoS Genet. 2011;7(10):e1002337.
30. Mateo L, Ullastres A, González J. A transposable element insertion confers xenobiotic resistance in Drosophila. PLoS Genet. 2014;10(8):e1004560.
31. Guio L, Barrón MG, González J. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Mol Ecol. 2014;23(8):2020-30.
32. Ullastres A, Petit N, González J. Exploring the Phenotypic Space and the Evolutionary History of a Natural Mutation in Drosophila melanogaster. Mol Biol Evol. 2015;32(7):1800-14.
33. Merenciano M, Ullastres A, de Cara MA, Barrón MG, González J. Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in Drosophila. PLoS Genet. 2016;12(8):e1006249.
34. Ullastres A. Natural transposable element insertions drive expression changes in genes underlying Drosophila immune response. In: Merenciano M, González, J., editor.: *bioRxiv*; 2019.
35. Merenciano M, Iacometti C, González J. A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila Melanogaster*. Mob DNA. 2019;10:10.
36. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 2013;23(1):169-80.
37. Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. Selection for cold resistance alters gene transcript levels in Drosophila melanogaster. J Insect Physiol. 2009;55(6):549-55.
38. Mihajlovic Z, Tanasic D, Bajgar A, Perez-Gomez R, Steffal P, Krejci A. Lime is a new protein linking immunity and metabolism in Drosophila. Dev Biol. 2019;452(2):83-94.

39. Ayres JS, Freitag N, Schneider DS. Identification of Drosophila mutants altering defense of and endurance to Listeria monocytogenes infection. Genetics. 2008;178(3):1807-15.
40. Needham AJ, Kibart M, Crossley H, Ingham PW, Foster SJ. Drosophila melanogaster as a model host for Staphylococcus aureus infection. Microbiology. 2004;150(Pt 7):2347-55.
41. Bou Sleiman MS, Osman D, Massouras A, Hoffmann AA, Lemaitre B, Deplancke B. Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. Nat Commun. 2015;6:7829.
42. Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, et al. Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. Proc Natl Acad Sci U S A. 2005;102(32):11414-9.
43. McKenzie JA. A comparative study of resource utilization in natural populations of *Drosophila melanogaster* and *D. simulans.* In: McKechnie SW, editor.: Oecologia; 1979. p. 299-309.
44. Gibson JB. Genetic variation at the alcohol dehydrogenase locus in *Drosophila melanogaster* in relation to environmental variation: ethanol levels in breeding sites and allozyme frequencies. In: May TW, Wilks, A. V., editor.: Oecologia; 1981. p. 191-8.
45. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet. 2015;11(1):e1004915.
46. Villanueva-Cañas JL, Horvath V, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in Drosophila melanogaster. Nucleic Acids Res. 2019;47(13):6842-57.
47. Garcia-Perez JL, Widmann TJ, Adams IR. The impact of transposable elements on mammalian development. Development. 2016;143(22):4101-14.
48. Simonti CN, Pavlicev M, Capra JA. Transposable Element Exaptation into Regulatory Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints. Mol Biol Evol. 2017;34(11):2856-69.
49. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, et al. Transposable elements are the primary source of novelty in primate gene regulation. Genome Res. 2017;27(10):1623-33.
50. Todd CD, Deniz Ö, Taylor D, Branco MR. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife. 2019;8.
51. Ding D, Lipshitz HD. Spatially regulated expression of retrovirus-like transposons during Drosophila melanogaster embryogenesis. Genet Res. 1994;64(3):167-81.
52. Deprá M, Valente VL, Margis R, Loreto EL. The hobo transposon and hobo-related elements are expressed as developmental genes in Drosophila. Gene. 2009;448(1):57-63.
53. Loreto ELS, Deprá M, Diesel JF, Panzera Y, Valente-Gaiesky VLS. Drosophila relics hobo and hobo-MITEs transposons as raw material for new regulatory networks. Genet Mol Biol. 2018;41(1 suppl 1):198-205.
54. Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. Mol Biol Evol. 2003;20(1):135-44.
55. Lerman DN, Feder ME. Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster. Mol Biol Evol. 2005;22(3):776-83.
56. Overgaard J, Malmendal A, Sørensen JG, Bundy JG, Loeschcke V, Nielsen NC, et al. Metabolomic profiling of rapid cold hardening and cold shock in Drosophila melanogaster. J Insect Physiol. 2007;53(12):1218-32.
57. Koštál V, Korbelová J, Rozsypal J, Zahradníčková H, Cimlová J, Tomčala A, et al. Long-term cold acclimation extends survival time at 0°C and modifies the metabolomic profiles of the larvae of the fruit fly Drosophila melanogaster. PLoS One. 2011;6(9):e25025.
58. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al. FlyBase 2.0: the next generation. Nucleic Acids Res. 2019;47(D1):D759-D65.
59. Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, et al. A Comprehensive Drosophila melanogaster Transcription Factor Interactome. Cell Rep. 2019;27(3):955-70.e7.
60. Krejcí A, Bernard F, Housden BE, Collins S, Bray SJ. Direct response to Notch activation: signaling crosstalk and incoherent logic. Sci Signal. 2009;2(55):ra1.
61. Housden BE, Fu AQ, Krejci A, Bernard F, Fischer B, Tavaré S, et al. Transcriptional dynamics elicited by a short pulse of notch activation involves feed-forward regulation by E(spl)/Hes genes. PLoS Genet. 2013;9(1):e1003162.
62. Port F, Bullock SL. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. Nat Methods. 2016;13(10):852-4.
63. Neyen C, Bretscher AJ, Binggeli O, Lemaitre B. Methods to study Drosophila immunity. Methods. 2014;68(1):116-28.
64. Schou MF. Fast egg collection method greatly improves randomness of egg sampling in Drosophila melanogaster. Fly (Austin). 2013;7(1):44-6.
65. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29(9):e45.
66. Finney DJ. Probit analysis. Cambridge: Cambridge University Press; 1971.
67. Johnson RM, Dahlgren L, Siegfried BD, Ellis MD. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS One. 2013;8(1):e54092.

SECTION 4
DISCUSSION

4. DISCUSSION

In this thesis, we have discovered a unique *roo* insertional cluster in the promoter region of the stress-response gene *CG18446*. We have also contributed to a better understanding of the role of TEs in adaptation and stress responses by deciphering the molecular and fitness-related effects of the *D. melanogaster FBti0019985* TE, which is located in the newly discovered *roo* insertional cluster. We have performed different molecular and phenotypic experiments that suggested that *FBti0019985* modulates *CG18446* expression in a versatile way, thus having an impact in egg-to-adult viability and in response to cold and immune stresses.

4.1. The *roo* solo-LTR insertional cluster in *CG18446* promoter region

We have demonstrated the presence of a unique *roo* insertional cluster in the *CG18446* promoter region of *D. melanogaster* (Merenciano et al., 2016, 2019). We have found that this *roo* insertional cluster is formed by independent insertions, which appeared to be the result of several bursts of transposition (Merenciano et al., 2019).

We have identified the target site duplication (TSD) for 17 of the 20 *roo* insertions located in the *CG18448* promoter region (Merenciano et al., 2019). TSDs are characteristic marks in the flanking regions of TEs that are generated as a result of the double-strand breaks that took place at the insertion site (Linheiro and Bergman, 2012). The 17 *roo* insertions had different TSD sequences suggesting that they are independent insertions (Merenciano et al., 2019). Thus, the different *roo* elements in the *CG18446* promoter region could have been the result of different transposition events rather than small rearrangements, which would not change TSD sequences.

TE transposition rates in natural and laboratory *D. melanogaster* populations are usually low (10^{-4} and 10^{-6}, respectively) (García Guerreiro, 2012). However, compared to other TE types, *roo* is one of the most active TE families with high transposition rates (Kaminker et al., 2002; Papaceit et al., 2007; Diaz-Gonzalez et al., 2011; Rahman et al., 2015). A rapid increase in TE copy number can be due to transposition bursts (Le Rouzic and Capy, 2005; García Guerreiro, 2012). These events are massive TE outbreaks that account for a rapid multiplication of one or several TEs (Le Rouzic and Capy, 2005; Belyayev et al., 2014). Some spontaneous transposition bursts have been described in *D. melanogaster* although there are often produced in response

to environmental or genomic stress (Le Rouzic and Capy, 2005; García Guerreiro, 2012; Kofler et al., 2015). It has been proposed that the *D. melanogaster* recent out-of-Africa expansion may have triggered transposition burst events thus promoting genomic variability that could ultimately favor adaptation and speciation processes (Vieira et al., 1999; 2004; 2012; Kofler et al., 2015). Our PCR screening of the *CG18446* promoter region has revealed that the ancestral range population from Africa is one of the populations with a smaller number of *roo* solo-LTRs: there are only three private insertions not found in the out-of-Africa populations (Merenciano et al., 2019). Moreover, this African population has the highest frequency of alleles without any insertion (77%) (Merenciano et al., 2019). Hence, these findings could indicate that the increased *roo* solo-LTR copy number in the *CG18446* promoter is the result of several bursts of transposition and could be related to a habitat expansion.

Why *roo* solo-LTR elements recurrently insert in the *CG18446* promoter region is not yet completely understood. We suggest that both the genomic region where the cluster is present and the already mentioned high transposition rates of *roo* elements could account for the recurrent insertions in that particular promoter region. The presence of the *roo* insertional cluster in the *CG18446* promoter could be a region-specific phenomenon since we have not detected any other *roo* insertional cluster with similar characteristics along the *D. melanogaster* genome (Merenciano et al., 2019). Indeed, similar to *hsp* genes with recurrent insertions of *P-elements* in their promoter regions (Walser et al., 2006), *CG18446* is regulated by polymerase pausing (Saunders et al., 2013), and is highly transcribed in the germline (Gramates et al., 2017). Hence, the location of *CG18446* in an open chromatin region could be one of the factors explaining the high TE density in its promoter region (Merenciano et al., 2019).

Overall, the *roo* insertional cluster in *CG18446* promoter is formed by independent insertions that could be a result of several burst of transposition. Moreover, the specificities of the *CG18446* promoter region could have favored the recurrent insertions in this unique region. Why other genomic regions with similar properties do not harbor such number of *roo* insertions is still unknown.

4.2. *FBti0019985* affects *CG18446* expression depending on the environmental conditions and the developmental stage

We have shown that *FBti0019985* is associated with *CG18446* down-regulation. However, it acts as an enhancer in guts under immune-stress conditions and in embryo stages under nonstress conditions (Ullastres et al., 2019; Section 3.5) (Figure 4.1). We have associated the enhancer activity of *FBti0019985* with an up-regulation of the nearby gene *CG18446* under the aforementioned specific contexts (Section 3.5). Specifically, *FBti0019985* has been associated with increased *CG18446* expression in embryos under nonstress conditions in four different backgrounds (three natural populations and one outbred population) (Merenciano et al., 2016, Section 3.5).

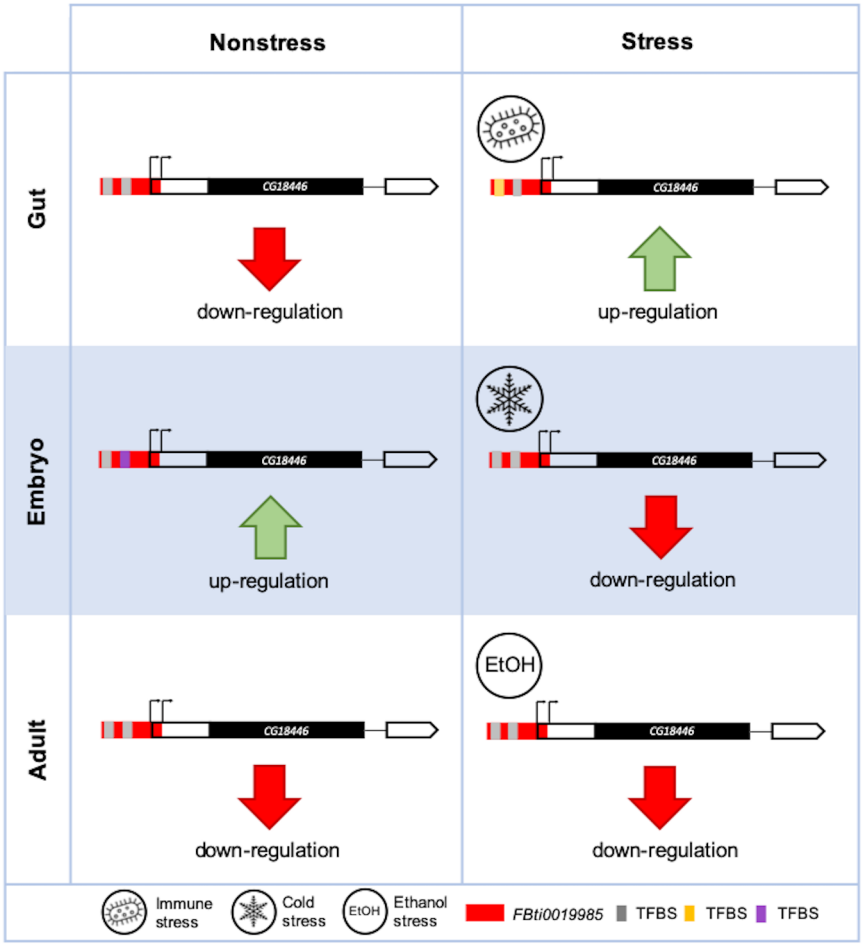


Figure 4.1. *FBti0019985* modulates *CG18446* expression depending on the environmental conditions and the developmental stage. *FBti0019985* is associated with *CG18446* down-regulation. However, in guts under immune-stress and in embryos under nonstress conditions, *FBti0019985* acts as an enhancer driving *CG18446* expression. Black boxes represent exons and white boxes represent UTR regions. TSSs are represented with black arrows.

We have suggested that *FBti0019985* harbours cis-regulatory elements responsible for its enhancer activity in embryos and under immune-stress. Indeed, TFBSs related with both developmental and immune processes have been predicted in the *FBti0019985* sequence (Merenciano et al., 2016; Villanueva-Cañas et al., 2019). Since *FBti0019985* shows very high sequence conservation with the other *roo* elements present in the *CG18446 roo* insertional cluster (Merenciano et al., 2019), we would expect them to also contain the predicted TFBSs related with development and immunity. However, neither roo_{+7}, roo_{-90}, nor roo_{-44} was associated with increased *CG18446* expression in embryos under nonstress conditions (Merenciano et al., 2016). There are different hypotheses that could explain these findings. First, the regulatory function of the TFBSs within the *roo* sequences can be affected by its location, orientation, and spacing in relation to the nearby gene (Beer and Tavazoie, 2004; Nguyen and D'haeseleer, 2006; Ng et al., 2014; Lis and Walther, 2016). Indeed, a statistical analysis of non-palindromic TFBSs and core promoter motifs in Arabidopsis revealed that regions closer to gene transcription start sites showed increased frequencies of motifs exhibiting orientation preferences (Lis and Walther, 2016). Since roo_{+7}, roo_{-90}, and roo_{-44}, are not inserted in the same position as *FBti0019985,* and roo_{-44} is not inserted in the same orientation neither, differences in spacing and orientation between the TFBSs and the gene could explain differences in *CG18446* expression in embryos. Second, although *roo* elements were acting as enhancers despite its location and orientation, epistatic interactions within each specific genetic background could affect *CG18446* expression. In fact, we found that *roo* elements inserted in the same position were associated with different effects on embryonic *CG18446* expression depending on the genetic background (Merenciano et al., 2016). roo_{+7} and roo_{-90} were both associated with significant *CG18446* expression changes only in one out of the three genetic backgrounds analysed for each insertion (Merenciano et al., 2016).

Altogether, we have found that although *FBti0019985* acts as an enhancer in embryos resulting in *CG18446* up-regulation, other *roo* elements inserted in the *roo* cluster had different molecular effects. The insertion site and the orientation of the *roo* elements could be crucial for the proper function of the cis-regulatory elements within their sequence. Moreover, we cannot discard the effect of the genetic background on the molecular effect of the different *roo* elements.

4.3. *FBti0019985* is involved in different stress-response phenotypes

Our results have shown that *FBti0019985* is associated with increased egg-to-adult viability in nonstress and cold-stress conditions, and with tolerance to *P. entomophila* infection (Merenciano et al., 2016, Section 3.5).

In *D. melanogaster*, there are a few examples of TEs being involved in different phenotypes. The *Doc* element *Doc1420* inserted in the second exon of the *CHKov1* gene produces four types of altered transcripts (Aminetzach et al., 2005). Authors found that flies with this insertion were more resistant to the organophosphate pesticide AZM (Aminetzach et al., 2005). *Doc1420* was inserted approximately 90,000 years ago, but its frequency expansion took place in the last 25-240 years (Aminetzach et al., 2005). Therefore, the authors suggested that *Doc1420* might be related to some other beneficial phenotype (Aminetzach et al., 2005). Indeed, a few years later, Magwire et al. (2011) associated *Doc1420* with resistance to viral infection. Moreover, they also found a rearrangement consisting in partial duplications of the *Doc1420, CHKov1*, and its paralog *CHKov2* that conferred even more resistance to viral infection (Magwire et al., 2011). Thus, *Doc1420* would have initially contributed to the resistance against viral infection and later, when the use of organophosphate pesticides was introduced, the insertion would have been rapidly selected for its beneficial role in facing those compounds (Magwire et al., 2011). In this case, the two *Doc1420* associated phenotypes appeared to be beneficial for the fly (Aminetzach et al., 2005; Magwire et al., 2011) (Table 4.1).

However, there are TEs that have been associated with both positive and negative fitness effects. This is the case of the *Bari-Jheh* element, a *Bari1* element inserted upstream of *Jheh1* and *Jheh2* genes and downstream of *Jheh3* gene. Under oxidative stress, *Bari-Jheh* was associated with changes in the local chromatin state leading to an up-regulation of *Jheh1* and *Jheh2* (Guio et al., 2014, 2018). This up-regulation is likely due to the presence of antioxidant response elements (AREs) within the *Bari-Jheh* element (Guio et al., 2014). Flies with this insertion were then associated with resistance to two different oxidative stress-inducing agents: paraquat and malathion (Guio et al., 2014). Oxidative stress can also be induced by an excessive production of reactive oxygen species (ROS) after a bacterial infection (Buchon et al., 2014). Indeed, *Bari-Jheh* was later associated with increased *P. entomophila* infection (Guio et al., 2016). However, besides the already mentioned phenotypes, *Bari-Jheh* was also associated with reduced viability and extended developmental time, both effects related with reduced fitness (González et al., 2009) (Table 4.1). Another adaptive TE involved in different phenotypes

was *FBti0019386*, an *invader4* element inserted in the 5'UTR region of *Bin1* gene (González et al., 2008, 2010; Ullastres et al., 2015, 2019). Authors found that flies with *FBti0019386* showed shorter developmental time and increased tolerance to bacterial infection (Ullastres et al., 2015, 2019). *FBti0019386* was also associated with reduced egg-to-adult viability under cold-stress, which has negative fitness consequences (Ullastres et al., 2015) (Table 4.1).

TE	Phenotypic effect	Fitness effect
Doc1420	Insecticide resistance	Positive
	Viral infection resistance	Positive
Bari-Jheh	Oxidative stress resistance	Positive
	Bacterial infection resistance	Positive
	Reduced viability	Negative
	Extended developmental time	Negative
FBti0019386	Short developmental time	Positive
	Bacterial infection tolerance	Positive
	Reduced viability under cold-stress	Negative
FBti0019985	Increased viability under nonstress	Positive
	Increased viability under cold stress	Positive
	Bacterial infection tolerance	Positive

Table 4.1. Adaptive TE insertions of D. melanogaster involved in different phenotypes.

These examples point out how TEs can have an impact on different phenotypes, sometimes with opposite fitness effects (Table 4.1). Hence, to explore all the possible phenotypic outcomes that a candidate mutation may have is important to fully understand its role in adaptation (Mackay et al., 2010; Olson-Manning et al., 2012; Mackay and Huang, 2018). Besides the described increase viability in nonstress and cold-stress conditions, and tolerance to infection, we cannot discard the effect of *FBti0019985* on other phenotypes.

4.4. Trade-off effects might be responsible for the low population frequency of *FBti0019985*

Based on our population screenings, *FBti0019985* is present at different allelic frequencies in out-of-Africa population, ranging from 0% to 22% (Merenciano et al., 2019). Considering the effect of this TE on the adaptive phenotypes that we have described in this work, we would expect this mutation to be positively selected and present at higher frequencies in all out-of-Africa natural populations. Thus, it could be possible that the low *FBti0019985* frequencies observed are due to its role in other phenotypes not identified yet with a related negative fitness effect.

It is known that a single mutation can have widespread phenotypic effects due to pleiotropic interactions (Grüneberg, 1938; Paaby and Rockman, 2013). In that way, a genetic variant can have different beneficial effects or present trade-offs. Antagonistic pleiotropy was first proposed in the field of senescence by Williams (1957) and occurs when a beneficial allele is at the same time deleterious for at least another trait (Williams, 1957). In *D. melanogaster*, it is well documented that female flies that are more tolerant to infection or cold-stress, reduce their number of offspring and are more sensitive to starvation (Watson and Hoffmann, 1996; Lazzaro et al., 2008; Marshall and Sinclair, 2010). Fecundity and starvation resistance experiments could be performed to test whether flies with *FBti0019985* show reduced number of offspring or are more sensitive to starvation than flies without the insertion.

In this work, we have found that, in adults, or under cold- and ethaol-stress, *FBti0019985* is associated with *CG18446* down-regulation. *CG18446* mutants have been shown to contain reduced glycogen and trehalose levels, the two main energy-storing carbohydrates in the fly (Mihajlovic et al., 2019). Trehalose is a nonreducing disaccharide synthesize in the fat body and released into the hemolymph (Matsuda et al., 2015). Availability of high levels of circulating trehalose are necessary to provide energy for insect flight muscle, immune defence, brain function, and to protect against environmental stress such as low humidity levels (Becker et al., 1996; Mattila and Hietakangas, 2017; Mihajlovic et al., 2019). Hence, *CG18446* down-regulation could impair the metabolism homeostasis causing a wide range of negative-fitness effects. Considering that *CG18446* is involved in metabolic processes, carrying *FBti0019985* might be a cost under certain circumstances. Further analysis could be performed to test whether the presence of *FBti0019985* is responsible for systemic metabolism changes.

4.5. Deletion of *FBti0019985* using the CRISPR/Cas9 technique in *D. melanogaster* natural populations

In this work, we used the CRISPR/Cas9 system to delete *FBti0019985* in *D. melanogaster* natural populations. We consider that this technique is a useful complementary approach to decipher molecular and phenotypic effects of naturally occurring TEs.

After its discovery as a bacterial defence mechanism (Mojica et al., 2000; Jansen et al., 2002), CRISPR/Cas9 technology has emerged as a powerful genome-editing tool in numerous species (Zhang et al., 2014; Manghwar et al., 2019). In *D. melanogaster*, Gratz and colleagues applied for the first time the CRISPR/Cas9 system targeting the *yellow* marker gene (Gratz et al., 2013). Authors injected the Cas9 endonuclease and a guide RNA (gRNA) that guides Cas9 to cut at specific location as plasmids into fly embryos (Gratz et al., 2013). Indels produced by imprecise repair of the Cas9 double-strand breaks in the *yellow* gene were transmitted to the offspring and resulted in the loss of the yellow phenotype (Gratz et al., 2013). Thus, authors confirmed CRISPR/Cas9 technique as a suitable tool to target genomic DNA in Drosophila although the efficiency of the process was very low (0.25%-1.37%) (Gratz et al., 2013). Later, several improvements allowed reaching higher efficiency levels. The injection of the Cas9 and the gRNA as *in vitro*-transcribed RNAs provided a 10-fold increase in efficiency (Bassett et al., 2013; Yu et al., 2013). However, with the generation of transgenic fly strains expressing Cas9, the CRISPR/Cas9 system reached up to 100% efficiency in some cases (Kondo and Ueda, 2013; Ren et al., 2013; Gratz et al., 2014; Port et al., 2014; Sebo et al., 2014). In these transgenic flies, there is no need to inject the Cas9 endonuclease since it is already expressed under the control of different promoters. Thus, only the injection of the gRNA is required. So far, CRISPR/Cas9 genome engineering has typically involved the use of transgenic strains expressing Cas9 due to its significantly increased efficiency and its reduction in cost. Here, we aimed to remove *FBti0019985* from an outbred population in which molecular and phenotypic differences were previously associated to the presence of the TE. Since we thought that the effect of *FBti0019985* depends on the genetic background, we avoided the use of the Cas9 expressing strains as they have been generated in a different genetic background, which might influence the phenotypes under investigation.

Nowadays CRISPR/Cas9 technique has evolved to allow the generation of precise genomic modifications with the addition of a donor template that contains a DNA sequence of interest (e.g. selection markers or specific alleles) flanked by homology arms. This donor template is

introduced in the genome through homology-directed repair system (Gratz et al., 2014). In this work, we use a DNA donor template to achieve a precise deletion of *FBti0019985* and to add the *DsRed* selection marker that allowed us to screen for successful deletion events. Recently, the possibility to produce conditional and tissue-specific knockouts has also been incorporated to the broad spectrum of genomic modifications that the CRISPR/Cas9 system offers (Gratz et al., 2014; Xue et al., 2014; Port et al., 2019). With this, we could further investigate the role of *FBti0019985* in specific tissues.

Although CRISPR/Cas9 systems is considered a powerful tool for genome engineering due to its high specificity rates, the presence of off-target mutations is still a major concern. In *D. melanogaster*, approximately 87% of gRNAs are specific to the target sequence (Ren et al., 2014). Yet, an appropriate gRNA design that ensures the target of a unique genome region is crucial. Some tools have been developed to allow researchers to find specific gRNAs (Gratz et al., 2014; Heigwer et al., 2014). In this work, gRNAs for *FBti0019985* deletion were designed using the CRISPR Optimal Target Finder (Gratz et al., 2014) and no off-targets were predicted by the software. Nevertheless, we cannot discard the presence of off-target mutations not considered by the gRNAs design tool.

To date, there are few examples of the use of CRISPR/Cas9 targeted to TEs. In human cells, the deletion of an ERV element allowed to discover the regulatory functions of this TE in mammalian immune responses (Chuong et al., 2016). Also, the deletion of the transposon-like human element 1B (THE1B) in transgenic mice permitted to identify its role in controlling the expression of corticotropin-releasing hormone (CRH) that, in turn, influences gestational length and birth timing (Dunn-Fletcher et al., 2018). In mice, TE deletions proved their role in gene regulation (Du et al., 2016; Todd et al., 2019). Moreover, CRISPR/Cas9 technique has been recently applied to delete a TE in plants for the first time (Saika et al., 2019). The deletion of the retrotransposon *Tos17* in rice opened a new avenue to elucidate the contribution of TEs in plant evolution and to produce novel variants without crossing (Saika et al., 2019). Finally, in *D. simulans*, CRISPR/Cas9 has been successfully employed to delete the *Shedller* element. This element was then associated with variations in the fly courtship song (Ding et al., 2016). Overall, all these studies evidence that CRISPR/Cas9 is a good tool to delete TE sequences.

In this thesis, we have generated two CRISPR-mutant strains with a deletion of *FBti0019985* (Section 5.3). In females, we have observed differences in *CG18446* expression between the two mutant strains under immune-stress conditions. Thus, we cannot discard the presence of

off-target deletion events that could influence *CG18446* expression. We then suggest analysing several CRISPR-mutant strains, when possible, to avoid misinterpretations caused by off-target events.

4.6. Future perspectives for functional validation of candidate adaptive TEs

As mentioned above, to decipher the molecular and phenotypic effects of *FBti0019985*, we have used for the first time the CRISPR/Cas9 technique to delete a TE insertion in a *D. melanogaster* outbred population.

We now propose other approaches to delete candidate adaptive-TE insertions using the CRISPR/Cas9 technique that could improve the efficiency of the system. In this work, we injected the Cas9 nuclease, the gRNAs, and the donor sequence for homology directed repair as plasmids in an outbred population. As such, we got a very low efficiency (1.5%) (Section 5.3). So, to considerably increase the efficiency of deletion events, we first propose to delete the candidate adaptive-TEs in Cas9 expressing transgenic flies (Gratz et al., 2014). However, this approach would only be possible when the studied TE is present in the Cas9 transgenic strain and there are previous evidences showing an effect of the TE in that specific background. Since Cas9 transgenic flies were generated in the laboratory, complementary experiments in natural populations could be useful to unravel the effect of the mutation in a natural genetic background. Second, we suggest designing different pairs of gRNAs flanking the candidate TE, and inject them together to ensure deletion events. In *D. melanogaster*, there are available plasmids that allow the cloning of several gRNAs (Port and Bullock, 2016). However, it should be considered that the introduction of many gRNAs would increase the frequency of off-target events.

So far, the validation of candidate adaptive-TE insertions in *D. melanogaster* has been done considering one insertion at a time (*e.g.* Mateo et al., 2014; Guio et al., 2014; Ullastres et al., 2015). However, the availability of whole genome sequences of fly populations from different geographic locations has allowed to significantly increase the number of candidate adaptive-TE insertions (Rech et al., 2019). In Rech et al. (2019), different approaches were used to look for signatures of positive selection in TEs present in high recombination regions and at high population frequencies considering 60 worldwide natural populations (Rech et al., 2019). A

total of 300 polymorphic and 21 fixed TEs were identified as putatively adaptive insertions (Rech et al., 2019). Moreover, authors also suggested that these candidate TEs could contribute to stress-response, developmental, and behavioral traits (Rech et al., 2019). However, this list is expected to grow thanks to the emergence of long-read sequencing techniques and the subsequent identification of TE insertions not present in the reference genome. Thus, in order to have a comprehensive picture of the role of TEs in adaptation processes, we need to move from the study of single TE candidates to the study of the general effects of multiple TEs. The injection of several gRNAs targeting multiple candidate TEs at the same time, will also allow studying the role of several TEs in a specific trait.

The binding specificities of the CRISPR/Cas9 system has permitted to repurpose CRISPR/Cas9 as a method to control gene expression at a sequence-specific manner (Qi et al., 2013). CRISPRi is a powerful and reversible tool to study gene regulation that consists in a catalytically inactive version of Cas9 lacking the endonuclease activity that, when coupled with a gRNA, generates a recognition complex that can silence genes through a physical transcription blockage of RNA polymerase (Qi et al., 2013). Gilbert et al. (2013) later proposed the fusion of the inactivated Cas9 to a transcriptional repressor domain (Krüppel associated box, KRAB) allowing the silence of proximal regulatory elements within a promoter by recruiting chromatin modifiers (Gilbert et al., 2013). Considering that, this system can be used to simultaneously decipher the effect of multiple TEs on gene regulation with the coexpression of different gRNAs. Since TEs are repetitive sequences and CRISPRi is design to target possible cis-regulatory elements within them, this system allows the study of different TEs from the same family rather than single insertions. Indeed, some TE families such as SVA, LTR5H and RLTR13D6 elements have already been targeted with the CRISPRi technology demonstrating their role in the evolutionary turnover of transcriptional networks in mice and humans specific cell types (Fuentes et al., 2018; Todd et al., 2019; Pontis et al., 2019).

Overall, there are several CRISPR-based technologies that can be used to validate the effects of candidate adaptive-TEs. Because each one of these technologies has advantages and disadvantages, the selection of the best approach has to consider both the scientific question that needs to be answered and the specificities of the targeted loci.

SECTION 5
CONCLUSIONS

5. CONCLUSIONS

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

1. The *D. melanogaster CG18446* promoter region contains a *roo* solo-LTR insertional cluster with at least 20 independent TE insertions, ranging in frequency from 0.2% to 16.5% on average. All these insertions have the same size and show high sequence conservation.

2. The presence of the recurrent *roo* insertions in the *CG18446* promoter region is likely to be the result of several bursts of transposition. We suggest that chromatin accessibility could be one of the factors explaining the recurrent insertions.

3. We suggest that the *roo* insertional cluster found in the *CG18446* promoter region is unique, as we could not detect any other *roo* insertional cluster in the 177 *D. melanogaster* genomes analyzed considering both reference and non-reference insertions.

4. Different *roo* elements of the insertional cluster have different effects on *CG18446* structure. Besides the already described alternative transcription start site in *FBti0019985*, we found that roo_{+7} also adds a transcription start site modifying the *CG18446* transcript length. However, roo_{-44} and roo_{-90} do not affect *CG18446* structure.

5. We found that *FBti0019985* affects *CG18446* expression depending on the developmental stage and the environmental conditions. In embryos, *FBti0019985* is associated with *CG18446* up-regulation while in adults it is associated with down-regulation. Moreover, *FBti0019985* up-regulates *CG18446* after infection while it down-regulates this gene under cold-, and ethanol-stress. We suggest that the increased expression of *CG18446* in embryos and under immune-stress conditions is likely due to the presence of transcription factor binding sites in its sequence related with development and immune processes. We also suggest that in adults or after a cold and ethanol exposure, *FBti0019985* could be affecting the spacing of the existing

cis-regulatory elements of the promoter region, thus leading to a decrease *CG18446* expression.

6. *FBti0019985* is associated with increased viability under nonstress and cold-stress conditions, and with increased tolerance to *P. entomophila* infection in *D. melanogaster* natural populations. These phenotypes could be related with changes in sugar metabolism performed by the *CG18446* transcription factor.

SECTION 6

REFERENCES

6. REFERENCES

Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science, 309*(5735), 764-767. doi:10.1126/science.1112699

Arguello, J. R., Laurent, S., & Clark, A. G. (2019). Demographic History of the Human Commensal *Drosophila melanogaster*. *Genome Biol Evol, 11*(3), 844-854. doi:10.1093/gbe/evz022

Arkhipova, I. R. (2017). Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mob DNA, 8*, 19. doi:10.1186/s13100-017-0103-2

Arkhipova, I. R. (2018). Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. *Mol Biol Evol, 35*(6), 1332-1337. doi:10.1093/molbev/msy083

Barrón, M. G., Fiston-Lavier, A. S., Petrov, D. A., & González, J. (2014). Population genomics of transposable elements in *Drosophila*. *Annu Rev Genet, 48*, 561-581. doi:10.1146/annurev-genet-120213-092359

Bassett, A. R., Tibbit, C., Ponting, C. P., & Liu, J. L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep, 4*(1), 220-228. doi:10.1016/j.celrep.2013.06.020

Batut, P., Dobin, A., Plessy, C., Carninci, P., & Gingeras, T. R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res, 23*(1), 169-180. doi:10.1101/gr.139618.112

Becker, A., Schlöder, P., Steele, J. E., & Wegener, G. (1996). The regulation of trehalose metabolism in insects. *Experientia, 52*(5), 433-439. doi:10.1007/bf01919312

Beer, M. A., & Tavazoie, S. (2004). Predicting gene expression from sequence. *Cell, 117*(2), 185-198. doi:10.1016/s0092-8674(04)00304-6

Beinert, N., Werner, M., Dowe, G., Chung, H. R., Jäckle, H., & Schäfer, U. (2004). Systematic gene targeting on the X chromosome of *Drosophila melanogaster*. *Chromosoma, 113*(6), 271-275. doi:10.1007/s00412-004-0313-5

Bellen, H. J., Levis, R. W., Liao, G., He, Y., Carlson, J. W., Tsang, G., . . . Spradling, A. C. (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics, 167*(2), 761-781. doi:10.1534/genetics.104.026427

Belyayev, A. (2014). Bursts of transposable elements as an evolutionary driving force. *J Evol Biol, 27*(12), 2573-2584. doi:10.1111/jeb.12513

Benoit, M., Drost, H. G., Catoni, M., Gouil, Q., Lopez-Gomollon, S., Baulcombe, D., & Paszkowski, J. (2019). Environmental and epigenetic regulation of *Rider* retrotransposons in tomato. *PLoS Genet, 15*(9), e1008370. doi:10.1371/journal.pgen.1008370

Bergman, C. M., Quesneville, H., Anxolabéhère, D., & Ashburner, M. (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol, 7*(11), R112. doi:10.1186/gb-2006-7-11-r112

Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R. W., & Pardue, M. L. (1992). HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol Cell Biol, 12*(9), 3910-3918. doi:10.1128/mcb.12.9.3910

Biémont, C. (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics, 186*(4), 1085-1093. doi:10.1534/genetics.110.124180

Biémont, C., & Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. *Nature, 443*(7111), 521-524. doi:10.1038/443521a

Blumenstiel, J. P., Chen, X., He, M., & Bergman, C. M. (2014). An age-of-allele test of neutrality for transposable element insertions. *Genetics, 196*(2), 523-538. doi:10.1534/genetics.113.158147

Bourgeois, Y., & Boissinot, S. (2019). On the Population Dynamics of Junk: A Review on the population genomics of transposable elements. *Genes (Basel), 10*(6). doi:10.3390/genes10060419

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., . . . Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biol, 19*(1), 199. doi:10.1186/s13059-018-1577-z

Bouttier, M., Laperriere, D., Memari, B., Mangiapane, J., Fiore, A., Mitchell, E., . . . White, J. H. (2016). *Alu* repeats as transcriptional regulatory platforms in macrophage responses to *M. tuberculosis* infection. *Nucleic Acids Res, 44*(22), 10571-10587. doi:10.1093/nar/gkw782

Brandt, J., Schrauth, S., Veith, A. M., Froschauer, A., Haneke, T., Schultheis, C., . . . Volff, J. N. (2005). Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene, 345*(1), 101-111. doi:10.1016/j.gene.2004.11.022

Britten, R. J., & Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol, 46*(2), 111-138. doi:10.1086/406830

Brönner, G., Taubert, H., & Jäckle, H. (1995). Mesoderm-specific B104 expression in the *Drosophila* embryo is mediated by internal cis-acting elements of the transposon. *Chromosoma, 103*(10), 669-675. doi:10.1007/bf00344227

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

Béguin, P., Charpin, N., Koonin, E. V., Forterre, P., & Krupovic, M. (2016). Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res, 44*(21), 10367-10376. doi:10.1093/nar/gkw821

Cam, H. P., Noma, K., Ebina, H., Levin, H. L., & Grewal, S. I. (2008). Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature, 451*(7177), 431-436. doi:10.1038/nature06499

Cappucci, U., Noro, F., Casale, A. M., Fanti, L., Berloco, M., Alagia, A. A., . . . Pimpinelli, S. (2019). The Hsp70 chaperone is a major player in stress-induced transposable element activation. *Proc Natl Acad Sci U.S.A, 116*(36), 17943-17950. doi:10.1073/pnas.1903936116

Capy, P., Gasperi, G., Biémont, C., & Bazin, C. (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity (Edinb), 85 (Pt 2)*, 101-106. doi:10.1046/j.1365-2540.2000.00751.x

Carareto, C. M., Hernandez, E. H., & Vieira, C. (2014). Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling Drosophila species. *Gene, 537*(1), 93-99. doi:10.1016/j.gene.2013.11.080

Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Mol Ecol, 22*(6), 1503-1517. doi:10.1111/mec.12170

Cheung, S., Manhas, S., & Measday, V. (2018). Retrotransposon targeting to RNA polymerase III-transcribed genes. *Mob DNA, 9*, 14. doi:10.1186/s13100-018-0119-2

Chou, H. H., Berthet, J., & Marx, C. J. (2009). Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genet, 5*(9), e1000652. doi:10.1371/journal.pgen.1000652

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science, 351*(6277), 1083-1087. doi:10.1126/science.aad5497

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet, 18*(2), 71-86. doi:10.1038/nrg.2016.139

Chénais, 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*(1), 7-15. doi:10.1016/j.gene.2012.07.042

Cowley, M., & Oakey, R. J. (2013). Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet, 9*(1), e1003234. doi:10.1371/journal.pgen.1003234

Cridland, J. M., Macdonald, S. J., Long, A. D., & Thornton, K. R. (2013). Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol Biol Evol, 30*(10), 2311-2327. doi:10.1093/molbev/mst129

Curcio, M. J., & Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol, 4*(11), 865-877. doi:10.1038/nrm1241

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., . . . ffrench-Constant, R. H. (2002). A single p450 allele associated with insecticide resistance in *Drosophila*. *Science, 297*(5590), 2253-2256. doi:10.1126/science.1074170

David, J. R., & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet, 4*(4), 106-111.

de Souza, F. S., Franchini, L. F., & Rubinstein, M. (2013). Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol, 30*(6), 1239-1251. doi:10.1093/molbev/mst045

Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nat Rev Genet, 20*(7), 417-431. doi:10.1038/s41576-019-0106-6

Ding, Y., Berrocal, A., Morita, T., Longden, K. D., & Stern, D. L. (2016). Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature, 536*(7616), 329-332. doi:10.1038/nature19093

Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature, 284*(5757), 601-603. doi:10.1038/284601a0

Du, J., Leung, A., Trac, C., Lee, M., Parks, B. W., Lusis, A. J., . . . Schones, D. E. (2016). Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics Chromatin, 9*, 28. doi:10.1186/s13072-016-0078-0

Dunn-Fletcher, C. E., Muglia, L. M., Pavlicev, M., Wolf, G., Sun, M. A., Hu, Y. C., . . . Muglia, L. J. (2018). Anthropoid primate-specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length. *PLoS Biol, 16*(9), e2006337. doi:10.1371/journal.pbio.2006337

Duret, L., Marais, G., & Biémont, C. (2000). Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics, 156*(4), 1661-1669.

Díaz-González, J., Vázquez, J. F., Albornoz, J., & Domínguez, A. (2011). Long-term evolution of the *roo* transposable element copy number in mutation accumulation lines of *Drosophila melanogaster*. *Genet Res (Camb), 93*(3), 181-187. doi:10.1017/S0016672311000103

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science, 351*(6274), aac7247. doi:10.1126/science.aac7247

Ellison, C. E., & Bachtrog, D. (2013). Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science, 342*(6160), 846-850. doi:10.1126/science.1239552

Esnault, C., Lee, M., Ham, C., & Levin, H. L. (2019). Transposable element insertions in fission yeast drive adaptation to environmental stress. *Genome Res, 29*(1), 85-95. doi:10.1101/gr.239699.118

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

Faulkner, G. J., & Carninci, P. (2009). Altruistic functions for selfish DNA. *Cell Cycle, 8*(18), 2895-2900. doi:10.4161/cc.8.18.9536

Feiner, N. (2016). Accumulation of transposable elements in *Hox* gene clusters during adaptive radiation of *Anolis* lizards. *Proc Biol Sci, 283*(1840). doi:10.1098/rspb.2016.1555

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

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet, 5*(4), 103-107. doi:10.1016/0168-9525(89)90039-5

Forestan, C., Aiese Cigliano, R., Farinati, S., Lunardon, A., Sanseverino, W., & Varotto, S. (2016). Stress-induced and epigenetic-mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis. *Sci Rep, 6*, 30446. doi:10.1038/srep30446

Fuentes, D. R., Swigut, T., & Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *Elife, 7*. doi:10.7554/eLife.35989

Gaffé, J., McKenzie, C., Maharjan, R. P., Coursange, E., Ferenci, T., & Schneider, D. (2011). Insertion sequence-driven evolution of *Escherichia coli* in chemostats. *J Mol Evol, 72*(4), 398-412. doi:10.1007/s00239-011-9439-2

Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., & Song, C. (2016). The contribution of transposable elements to size variations between four teleost genomes. *Mob DNA, 7*, 4. doi:10.1186/s13100-016-0059-7

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

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., . . . Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature, 419*(6906), 498-511. doi:10.1038/nature01097

Gerdes, P., Richardson, S. R., Mager, D. L., & Faulkner, G. J. (2016). Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol, 17*, 100. doi:10.1186/s13059-016-0965-5

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., . . . Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell, 154*(2), 442-451. doi:10.1016/j.cell.2013.06.044

Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature, 470*(7333), 284-288. doi:10.1038/nature09701

González, J., Karasov, T. L., Messer, P. W., & Petrov, D. A. (2010). Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet, 6*(4), e1000905. doi:10.1371/journal.pgen.1000905

González, J., Lenkov, K., Lipatov, M., Macpherson, J. M., & Petrov, D. A. (2008). High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol, 6*(10), e251. doi:10.1371/journal.pbio.0060251

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

Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., . . . the FlyBase Consortium. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res, 45*(D1), D663-D671. doi:10.1093/nar/gkw1016

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., . . . O'Connor-Giles, K. M. (2013). Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics, 194*(4), 1029-1035. doi:10.1534/genetics.113.152710

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics, 196*(4), 961-971. doi:10.1534/genetics.113.160713

Grüneberg, H. (1938). An analysis of the "pleiotropic" efects of a new lethal mutation in the rat *(Mus norvegicus). Proc R Soc Lond B*, 125, 123-144. doi:10.1098/rspb.1938.0017

Guio, L. (2016). The transposon *Bari-Jheh* in *Drosophila melanogaster*: adaptive phenotypes, molecular mecanisms, and genetic inheritance (Unpublished doctoral dissertation). In: University of Barcelona.

Guio, L., Barrón, M. G., & González, J. (2014). The transposable element *Bari-Jheh* mediates oxidative stress response in *Drosophila*. *Mol Ecol, 23*(8), 2020-2030. doi:10.1111/mec.12711

Guio, L., & González, J. (2019). New insights on the evolution of genome content: population dynamics of transposable elements in flies and humans. *Methods Mol Biol, 1910*, 505-530. doi:10.1007/978-1-4939-9074-0_16

Guio, L., Vieira, C., & González, J. (2018). Stress affects the epigenetic marks added by natural transposable element insertions in *Drosophila melanogaster*. *Sci Rep, 8*(1), 12197. doi:10.1038/s41598-018-30491-w

Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat Methods, 11*(2), 122-123. doi:10.1038/nmeth.2812

Hickey, D. A. (1982). Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics, 101*(3-4), 519-531.

Horváth, V., Merenciano, M., & González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet, 33*(11), 832-841. doi:10.1016/j.tig.2017.08.007

Hu, S. B., Xiang, J. F., Li, X., Xu, Y., Xue, W., Huang, M., . . . Chen, L. L. (2015). Protein arginine methyltransferase CARM1 attenuates the paraspeckle-mediated nuclear retention of mRNAs containing IRAlus. *Genes Dev, 29*(6), 630-645. doi:10.1101/gad.257048.114

Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filée, J., & Capy, P. (2011). The struggle for life of the genome's selfish architects. *Biol Direct, 6*, 19. doi:10.1186/1745-6150-6-19

Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H. A., . . . Xu, A. (2016). Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell, 166*(1), 102-114. doi:10.1016/j.cell.2016.05.032

Hummel, B., Hansen, E. C., Yoveva, A., Aprile-Garcia, F., Hussong, R., & Sawarkar, R. (2017). The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. *Nat Struct Mol Biol, 24*(3), 234-242. doi:10.1038/nsmb.3368

Ikeda, Y., Pélissier, T., Bourguet, P., Becker, C., Pouch-Pélissier, M. N., Pogorelcnik, R., . . . Mathieu, O. (2017). *Arabidopsis* proteins with a transposon-related domain act in gene silencing. *Nat Commun, 8*, 15122. doi:10.1038/ncomms15122

Jacques, P., Jeyakani, J., & Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet, 9*(5), e1003504. doi:10.1371/journal.pgen.1003504

Jangam, D., Feschotte, C., & Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet, 33*(11), 817-831. doi:10.1016/j.tig.2017.07.011

Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol, 43*(6), 1565-1575. doi:10.1046/j.1365-2958.2002.02839.x

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., . . . Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature, 546*(7659), 524-527. doi:10.1038/nature22971

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., . . . Celniker, S. E. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol, 3*(12), RESEARCH0084.

Kanazawa, A., Liu, B., Kong, F., Arase, S., & Abe, J. (2009). Adaptive evolution involving gene duplication and insertion of a novel Ty1/copia-like retrotransposon in soybean. *J Mol Evol, 69*(2), 164-175. doi:10.1007/s00239-009-9262-1

Kapitonov, V. V., & Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol, 3*(6), e181. doi:10.1371/journal.pbio.0030181

Kapitonov, V. V., & Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet, 9*(5), 411-412; author reply 414. doi:10.1038/nrg2165-c1

Kawakami, T., Mugal, C. F., Suh, A., Nater, A., Burri, R., Smeds, L., & Ellegren, H. (2017). Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol Ecol, 26*(16), 4158-4172. doi:10.1111/mec.14197

Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., & Antonarakis, S. E. (1988). Haemophilia A resulting from *de novo* insertion of *L1* sequences represents a novel mechanism for mutation in man. *Nature, 332*(6160), 164-166. doi:10.1038/332164a0

Kissinger, J. C., & DeBarry, J. (2011). Genome cartography: charting the apicomplexan genome. *Trends Parasitol, 27*(8), 345-354. doi:10.1016/j.pt.2011.03.006

Klein, S. J., & O'Neill, R. J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res, 26*(1-2), 5-23. doi:10.1007/s10577-017-9569-5

Kofler, R., Betancourt, A. J., & Schlötterer, C. (2012). Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet, 8*(1), e1002487. doi:10.1371/journal.pgen.1002487

Kofler, R., Hill, T., Nolte, V., Betancourt, A. J., & Schlötterer, C. (2015). The recent invasion of natural *Drosophila simulans* populations by the *P-element*. *Proc Natl Acad Sci U S A, 112*(21), 6659-6663. doi:10.1073/pnas.1500758112

Kondo, S., & Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics, 195*(3), 715-721. doi:10.1534/genetics.113.156737

Krupovic, M., Béguin, P., & Koonin, E. V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr Opin Microbiol, 38*, 36-43. doi:10.1016/j.mib.2017.04.004

Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D., & Koonin, E. V. (2014). Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol, 12*, 36. doi:10.1186/1741-7007-12-36

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:10.1038/35057062

Laricchia, K. M., Zdraljevic, S., Cook, D. E., & Andersen, E. C. (2017). Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol Biol Evol, 34*(9), 2187-2202. doi:10.1093/molbev/msx155

Lazzaro, B. P., Flores, H. A., Lorigan, J. G., & Yourth, C. P. (2008). Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in *Drosophila melanogaster*. *PLoS Pathog, 4*(3), e1000025. doi:10.1371/journal.ppat.1000025

Le Manh, H., Guio, L., Merenciano, M., Rovira, Q., Barrón, M. G., & González, J. (2017). Natural and laboratory mutations in *kuzbanian* are associated with zinc stress phenotypes in *Drosophila melanogaster*. *Sci Rep, 7*, 42663. doi:10.1038/srep42663

Le Rouzic, A., & Capy, P. (2005). The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics, 169*(2), 1033-1043. doi:10.1534/genetics.104.031211

Lee, J. Y., Ji, Z., & Tian, B. (2008). Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res, 36*(17), 5581-5590. doi:10.1093/nar/gkn540

Leem, Y. E., Ripmaster, T. L., Kelly, F. D., Ebina, H., Heincelman, M. E., Zhang, K., . . . Levin, H. L. (2008). Retrotransposon Tf1 is targeted to Pol II promoters by transcription activators. *Mol Cell, 30*(1), 98-107. doi:10.1016/j.molcel.2008.02.016

Lerman, D. N., & Feder, M. E. (2005). Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster. *Mol Biol Evol, 22*(3), 776-783. doi:10.1093/molbev/msi063

Li, H., & Stephan, W. (2006). Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet, 2*(10), e166. doi:10.1371/journal.pgen.0020166

Liao, G. C., Rehm, E. J., & Rubin, G. M. (2000). Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A, 97*(7), 3347-3351. doi:10.1073/pnas.050017397

Lin, R., Ding, L., Casola, C., Ripoll, D. R., Feschotte, C., & Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science, 318*(5854), 1302-1305. doi:10.1126/science.1146281

Linheiro, R. S., & Bergman, C. M. (2012). Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS One, 7*(2), e30008. doi:10.1371/journal.pone.0030008

Lis, M., & Walther, D. (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics, 17*, 185. doi:10.1186/s12864-016-2549-x

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., & Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics, 180*(2), 995-1007. doi:10.1534/genetics.108.092742

Liu, S., Yeh, C. T., Ji, T., Ying, K., Wu, H., Tang, H. M., . . . Schnable, P. S. (2009). Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet, 5*(11), e1000733. doi:10.1371/journal.pgen.1000733

Lorenz, D. R., Mikheyeva, I. V., Johansen, P., Meyer, L., Berg, A., Grewal, S. I., & Cam, H. P. (2012). CENP-B cooperates with Set1 in bidirectional transcriptional silencing and genome organization of retrotransposons. *Mol Cell Biol, 32*(20), 4215-4225. doi:10.1128/MCB.00395-12

Loreto, E. L. S., Deprá, M., Diesel, J. F., Panzera, Y., & Valente-Gaiesky, V. L. S. (2018). *Drosophila* relics *hobo* and *hobo*-MITEs transposons as raw material for new regulatory networks. *Genet Mol Biol, 41*(1 suppl 1), 198-205. doi:10.1590/1678-4685-GMB-2017-0068

Mackay, T. F. (2010). Mutations and quantitative genetic variation: lessons from *Drosophila*. *Philos Trans R Soc Lond B Biol Sci, 365*(1544), 1229-1239. doi:10.1098/rstb.2009.0315

Mackay, T. F. C., & Huang, W. (2018). Charting the genotype-phenotype map: lessons from the *Drosophila melanogaster* Genetic Reference Panel. *Wiley Interdiscip Rev Dev Biol, 7*(1). doi:10.1002/wdev.289

Magwire, M. M., Bayer, F., Webster, C. L., Cao, C., & Jiggins, F. M. (2011). Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication. *PLoS Genet, 7*(10), e1002337. doi:10.1371/journal.pgen.1002337

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., & Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet, 11*(1), e1004915. doi:10.1371/journal.pgen.1004915

Makałowski, W., Mitchell, G. A., & Labuda, D. (1994). Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet, 10*(6), 188-193. doi:10.1016/0168-9525(94)90254-2

Mamillapalli, A., Pathak, R. U., Garapati, H. S., & Mishra, R. K. (2013). Transposable element 'roo' attaches to nuclear matrix of the *Drosophila melanogaster*. *J Insect Sci, 13*, 111. doi:10.1673/031.013.11101

Manghwar, H., Lindsey, K., Zhang, X., & Jin, S. (2019). CRISPR/Cas system: recent advances and future prospects for genome editing. *Trends Plant Sci, 24*(12), 1102-1125. doi:10.1016/j.tplants.2019.09.006

Marshall, K. E., & Sinclair, B. J. (2010). Repeated stress exposure results in a survival-reproduction trade-off in *Drosophila melanogaster*. *Proc Biol Sci, 277*(1683), 963-969. doi:10.1098/rspb.2009.1807

Mason, J. M., & Biessmann, H. (1995). The unusual telomeres of *Drosophila*. *Trends Genet, 11*(2), 58-62. doi:10.1016/s0168-9525(00)88998-2

Mateo, L., Ullastres, A., & González, J. (2014). A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLoS Genet, 10*(8), e1004560. doi:10.1371/journal.pgen.1004560

Matsuda, H., Yamada, T., Yoshida, M., & Nishimura, T. (2015). Flies without trehalose. *J Biol Chem, 290*(2), 1244-1255. doi:10.1074/jbc.M114.619411

Mattila, J., & Hietakangas, V. (2017). Regulation of carbohydrate energy metabolism in *Drosophila melanogaster*. *Genetics, 207*(4), 1231-1253. doi:10.1534/genetics.117.199885

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A, 36*(6), 344-355. doi:10.1073/pnas.36.6.344

McClintock, B. (1953). Induction of instability at selected loci in maize. *Genetics, 38*(6), 579-599.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol, 21*, 197-216. doi:10.1101/sqb.1956.021.01.017

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science, 226*(4676), 792-801. doi:10.1126/science.15739260

McLaughlin, R. N., Young, J. M., Yang, L., Neme, R., Wichman, H. A., & Malik, H. S. (2014). Positive selection and multiple losses of the LINE-1-derived *L1TD1* gene in mammals suggest a dual role in genome defense and pluripotency. *PLoS Genet, 10*(9), e1004531. doi:10.1371/journal.pgen.1004531

Menees, T. M., & Sandmeyer, S. B. (1996). Cellular stress inhibits transposition of the yeast retrovirus-like element Ty3 by a ubiquitin-dependent block of virus-like particle formation. *Proc Natl Acad Sci U S A, 93*(11), 5629-5634. doi:10.1073/pnas.93.11.5629

Merenciano, M., Iacometti, C., & González, J. (2019). A unique cluster of *roo* insertions in the promoter region of a stress response gene in *Drosophila melanogaster*. *Mob DNA, 10*, 10. doi:10.1186/s13100-019-0152-9

Merenciano, M., Ullastres, A., de Cara, M. A., Barrón, M. G., & González, J. (2016). Multiple independent retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in *Drosophila*. *PLoS Genet, 12*(8), e1006249. doi:10.1371/journal.pgen.1006249

Metaxakis, A., Oehler, S., Klinakis, A., & Savakis, C. (2005). Minos as a genetic and genomic tool in *Drosophila melanogaster*. *Genetics, 171*(2), 571-581. doi:10.1534/genetics.105.041848

Mihajlovic, Z., Tanasic, D., Bajgar, A., Perez-Gomez, R., Steffal, P., & Krejci, A. (2019). Lime is a new protein linking immunity and metabolism in *Drosophila*. *Dev Biol, 452*(2), 83-94. doi:10.1016/j.ydbio.2019.05.005

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., . . . Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res, 52*(3), 643-645.

Miller, W. J., McDonald, J. F., & Pinsker, W. (1997). Molecular domestication of mobile elements. *Genetica, 100*(1-3), 261-270.

Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol, 36*(1), 244-246. doi:10.1046/j.1365-2958.2000.01838.x

Myers, S., Freeman, C., Auton, A., Donnelly, P., & McVean, G. (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet, 40*(9), 1124-1129. doi:10.1038/ng.213

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., . . . Wessler, S. R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature, 461*(7267), 1130-1134. doi:10.1038/nature08479

Negi, P., Rai, A. N., & Suprasanna, P. (2016). : Emerging Regulatory Roles for Transposons in Plant Stress Response. *Front Plant Sci, 7*, 1448. doi:10.3389/fpls.2016.01448

Nekrutenko, A., & Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends Genet, 17*(11), 619-621. doi:10.1016/s0168-9525(01)02445-3

Ng, F. S., Schütte, J., Ruau, D., Diamanti, E., Hannah, R., Kinston, S. J., & Göttgens, B. (2014). Constrained transcription factor spacing is prevalent and important for transcriptional control of mouse blood cells. *Nucleic Acids Res, 42*(22), 13513-13524. doi:10.1093/nar/gku1254

Nguyen, D. H., & D'haeseleer, P. (2006). Deciphering principles of transcription regulation in eukaryotic genomes. *Mol Syst Biol, 2*, 2006.0012. doi:10.1038/msb4100054

Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Genet, 13*(12), 867-877. doi:10.1038/nrg3322

Orgel, L. E., & Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature, 284*(5757), 604-607. doi:10.1038/284604a0

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends Genet, 29*(2), 66-73. doi:10.1016/j.tig.2012.10.010

Palazzo, A., Lorusso, P., Miskey, C., Walisko, O., Gerbino, A., Marobbio, C. M. T., . . . Marsano, R. M. (2019). Transcriptionally promiscuous "blurry" promoters in Tc1/*mariner* transposons

allow transcription in distantly related genomes. *Mob DNA, 10*, 13. doi:10.1186/s13100-019-0155-6

Papaceit, M., Avila, V., Aguadé, M., & García-Dorado, A. (2007). The dynamics of the roo transposable element in mutation-accumulation lines and segregating populations of *Drosophila melanogaster*. *Genetics, 177*(1), 511-522. doi:10.1534/genetics.107.076174

Persson, J., Steglich, B., Smialowska, A., Boyd, M., Bornholdt, J., Andersson, R., . . . Ekwall, K. (2016). Regulating retrotransposon activity through the use of alternative transcription start sites. *EMBO Rep, 17*(5), 753-768. doi:10.15252/embr.201541866

Peter, A., Schöttler, P., Werner, M., Beinert, N., Dowe, G., Burkert, P., . . . Schäfer, U. (2002). Mapping and identification of essential gene functions on the X chromosome of *Drosophila*. *EMBO Rep, 3*(1), 34-38. doi:10.1093/embo-reports/kvf012

Piégu, 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. doi:10.1016/j.ympev.2015.03.009

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., . . . Trono, D. (2019). Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell, 24*(5), 724-735.e725. doi:10.1016/j.stem.2019.03.012

Port, F., Strein, C., Stricker, M., Rauscher, B., Heigwer, F., Zhou, J., … Boutros, M. (2019). A large-scale resource for tissue-specific CRISPR mutagenesis in *Drosophila*. *bioRxiv. doi:*10.1101/636076

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nat Methods, 13*(10), 852-854. doi:10.1038/nmeth.3972

Port, F., Chen, H. M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci U S A, 111*(29), E2967-2976. doi:10.1073/pnas.1405500111

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell, 152*(5), 1173-1183. doi:10.1016/j.cell.2013.02.022

Rahman, R., Chirn, G. W., Kanodia, A., Sytnikova, Y. A., Brembs, B., Bergman, C. M., & Lau, N. C. (2015). Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. *Nucleic Acids Res, 43*(22), 10655-10672. doi:10.1093/nar/gkv1193

Rech, G. E., Bogaerts-Márquez, M., Barrón, M. G., Merenciano, M., Villanueva-Cañas, J. L., Horváth, V., . . . González, J. (2019). Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet, 15*(2), e1007900. doi:10.1371/journal.pgen.1007900

Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., . . . Ni, J. Q. (2013). Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. *Proc Natl Acad Sci U S A, 110*(47), 19012-19017. doi:10.1073/pnas.1318481110

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., . . . Ni, J. Q. (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell Rep, 9*(3), 1151-1162. doi:10.1016/j.celrep.2014.09.044

Rey, O., Danchin, E., Mirouze, M., Loot, C., & Blanchet, S. (2016). Adaptation to global change: a transposable element-epigenetics perspective. *Trends Ecol Evol, 31*(7), 514-526. doi:10.1016/j.tree.2016.03.013

Rishishwar, L., Wang, L., Clayton, E. A., Mariño-Ramírez, L., McDonald, J. F., & Jordan, I. K. (2017). Population and clinical genetics of human transposable elements in the (post) genomic era. *Mob Genet Elements, 7*(1), 1-20. doi:10.1080/2159256X.2017.1280116

Rubin, G. M., Kidwell, M. G., & Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell, 29*(3), 987-994. doi:10.1016/0092-8674(82)90462-7

Saika, H., Mori, A., Endo, M., & Toki, S. (2019). Targeted deletion of rice retrotransposon *Tos17* via CRISPR/Cas9. *Plant Cell Rep, 38*(4), 455-458. doi:10.1007/s00299-018-2357-7

Samuelson, L. C., Wiebauer, K., Snow, C. M., & Meisler, M. H. (1990). Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol, 10*(6), 2513-2520. doi:10.1128/mcb.10.6.2513

Sarkar, A., Sim, C., Hong, Y. S., Hogan, J. R., Fraser, M. J., Robertson, H. M., & Collins, F. H. (2003). Molecular evolutionary analysis of the widespread *piggyBac* transposon family and related "domesticated" sequences. *Mol Genet Genomics, 270*(2), 173-180. doi:10.1007/s00438-003-0909-0

Sasaki, M., Lange, J., & Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol, 11*(3), 182-195. doi:10.1038/nrm2849

Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T., & Ashe, H. L. (2013). Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription. *Genes Dev, 27*(10), 1146-1158. doi:10.1101/gad.215459.113

Schmidt, J. M., Good, R. T., Appleton, B., Sherrard, J., Raymant, G. C., Bogwitz, M. R., . . . Robin, C. (2010). Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genet, 6*(6), e1000998. doi:10.1371/journal.pgen.1000998

Schmitz, J., & Brosius, J. (2011). Exonization of transposed elements: A challenge and opportunity for evolution. *Biochimie, 93*(11), 1928-1934. doi:10.1016/j.biochi.2011.07.014

Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Mol Ecol, 28*(6), 1537-1549. doi:10.1111/mec.14794

Sebo, Z. L., Lee, H. B., Peng, Y., & Guo, Y. (2014). A simplified and efficient germline-specific CRISPR/Cas9 system for *Drosophila* genomic engineering. *Fly (Austin), 8*(1), 52-57. doi:10.4161/fly.26828

Sela, N., Kim, E., & Ast, G. (2010). The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biol, 11*(6), R59. doi:10.1186/gb-2010-11-6-r59

Sessegolo, C., Burlet, N., & Haudry, A. (2016). Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. *Biol Lett, 12*(8). doi:10.1098/rsbl.2016.0407

Shapiro, J. A. (1969). Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J Mol Biol, 40*(1), 93-105. doi:10.1016/0022-2836(69)90298-8

Shen, S., Lin, L., Cai, J. J., Jiang, P., Kenkel, E. J., Stroik, M. R., . . . Xing, Y. (2011). Widespread establishment and regulatory impact of Alu exons in human genes. *Proc Natl Acad Sci U S A, 108*(7), 2837-2842. doi:10.1073/pnas.1012834108

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., . . . Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature, 423*(6942), 825-837. doi:10.1038/nature01722

Sorek, R., Ast, G., & Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res, 12*(7), 1060-1067. doi:10.1101/gr.229302

Sotero-Caio, C. G., Platt, R. N., Suh, A., & Ray, D. A. (2017). Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol Evol, 9*(1), 161-177. doi:10.1093/gbe/evw264

Spaller, T., Groth, M., Glöckner, G., & Winckler, T. (2017). TRE5-A retrotransposition profiling reveals putative RNA polymerase III transcription complex binding sites on the *Dictyostelium* extrachromosomal rDNA element. *PLoS One, 12*(4), e0175729. doi:10.1371/journal.pone.0175729

Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., & Rubin, G. M. (1995). Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. *Proc Natl Acad Sci U S A, 92*(24), 10824-10830.

Staudt, N., Molitor, A., Somogyi, K., Mata, J., Curado, S., Eulenberg, K., . . . Vorbrüggen, G. (2005). Gain-of-function screen for genes that affect *Drosophila* muscle pattern formation. *PLoS Genet, 1*(4), e55. doi:10.1371/journal.pgen.0010055

Stoebel, D. M., & Dorman, C. J. (2010). The effect of mobile element IS10 on experimental regulatory evolution in *Escherichia coli*. *Mol Biol Evol, 27*(9), 2105-2112. doi:10.1093/molbev/msq101

Stoebel, D. M., Hokamp, K., Last, M. S., & Dorman, C. J. (2009). Compensatory evolution of gene regulation in response to stress by *Escherichia coli* lacking RpoS. *PLoS Genet, 5*(10), e1000671. doi:10.1371/journal.pgen.1000671

Strobel, E., Dunsmuir, P., & Rubin, G. M. (1979). Polymorphisms in the chromosomal locations of elements of the 412, *copia* and 297 dispersed repeated gene families in *Drosophila*. *Cell, 17*(2), 429-439. doi:10.1016/0092-8674(79)90169-7

Sultana, T., Zamborlini, A., Cristofari, G., & Lesage, P. (2017). Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet, 18*(5), 292-308. doi:10.1038/nrg.2017.7

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., . . . Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res, 24*(12), 1963-1976. doi:10.1101/gr.168872.113

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., . . . Margolis, J. (2004). A complementary transposon tool kit for *Drosophila melanogaster* using *P* and *piggyBac*. *Nat Genet, 36*(3), 283-287. doi:10.1038/ng1314

Thompson, P. J., Macfarlan, T. S., & Lorincz, M. C. (2016). Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol Cell, 62*(5), 766-776. doi:10.1016/j.molcel.2016.03.029

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., . . . Consortium, F. (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res, 47*(D1), D759-D765. doi:10.1093/nar/gky1003

Todd, C. D., Deniz, Ö., Taylor, D., & Branco, M. R. (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *Elife, 8*. doi:10.7554/eLife.44344

Touchon, M., & Rocha, E. P. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol, 24*(4), 969-981. doi:10.1093/molbev/msm014

Trivedi, M., Shah, J., Hodgson, N., Byun, H. M., & Deth, R. (2014). Morphine induces redox-based changes in global DNA methylation and retrotransposon transcription by inhibition of excitatory amino acid transporter type 3-mediated cysteine uptake. *Mol Pharmacol, 85*(5), 747-757. doi:10.1124/mol.114.091728

Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., . . . Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res, 27*(10), 1623-1633. doi:10.1101/gr.218149.116

Ullastres, A. , Merenciano, M., & González, J. (2019). Natural transposable element insertions drive expression changes in genes underlying *Drosophila* immune response. *bioRxiv* doi:10.1101/655225

Ullastres, A., Petit, N., & González, J. (2015). Exploring the Phenotypic Space and the Evolutionary History of a Natural Mutation in *Drosophila melanogaster*. *Mol Biol Evol, 32*(7), 1800-1814. doi:10.1093/molbev/msv061

Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., & Gorbunova, V. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun, 5*, 5011. doi:10.1038/ncomms6011

Van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., . . . Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature, 534*(7605), 102-105. doi:10.1038/nature17951

Vieira, C., & Biémont, C. (2004). Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica, 120*(1-3), 115-123. doi:10.1023/b:gene.0000017635.34955.b5

Vieira, C., Fablet, M., Lerat, E., Boulesteix, M., Rebollo, R., Burlet, N., . . . Biémont, C. (2012). A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *J Environ Radioact, 113*, 83-86. doi:10.1016/j.jenvrad.2012.04.001

Vieira, C., Lepetit, D., Dumont, S., & Biémont, C. (1999). Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol, 16*(9), 1251-1255. doi:10.1093/oxfordjournals.molbev.a026215

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

Walser, J. C., Chen, B., & Feder, M. E. (2006). Heat-shock promoters: targets for evolution by *P* transposable elements in *Drosophila*. *PLoS Genet, 2*(10), e165. doi:10.1371/journal.pgen.0020165

Warren, I. A., Naville, M., Chalopin, D., Levin, P., Berger, C. S., Galiana, D., & Volff, J. N. (2015). Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Res, 23*(3), 505-531. doi:10.1007/s10577-015-9493-5

Watson, M. J. O., & Hoffmann, A. A. (1996). Acclimation, cross-generation effects, and the response to selection for increased cold resistance in *Drosophila*. *Evolution, 50*(3), 1182-1192. doi:10.1111/j.1558-5646.1996.tb02359.x

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., . . . Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet, 8*(12), 973-982. doi:10.1038/nrg2165

Williams, G. C. (1957). Pleiotropy, natural selection, and the evolution of senescence. *Evolution,* 11(4), 398-411. doi:10.2307/2406060

Xue, Z., Wu, M., Wen, K., Ren, M., Long, L., Zhang, X., & Gao, G. (2014). CRISPR/Cas9 mediates efficient conditional mutagenesis in *Drosophila*. *G3 (Bethesda), 4*(11), 2167-2173. doi:10.1534/g3.114.014159

Yu, Z., Ren, M., Wang, Z., Zhang, B., Rong, Y. S., Jiao, R., & Gao, G. (2013). Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. *Genetics, 195*(1), 289-291. doi:10.1534/genetics.113.153825

Zhang, F., Wen, Y., & Guo, X. (2014). CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet, 23*(R1), R40-46. doi:10.1093/hmg/ddu125

Zhao, Y., Xu, T., Shen, C. Y., Xu, G. H., Chen, S. X., Song, L. Z., . . . Deng, X. (2014). Identification of a retroelement from the resurrection plant Boea hygrometrica that confers osmotic and alkaline tolerance in *Arabidopsis thaliana*. *PLoS One, 9*(5), e98098. doi:10.1371/journal.pone.0098098

Zovoilis, A., Cifuentes-Rojas, C., Chu, H. P., Hernandez, A. J., & Lee, J. T. (2016). Destabilization of B2 RNA by EZH2 activates the stress response. *Cell, 167*(7), 1788-1802.e1713. doi:10.1016/j.cell.2016.11.041

SECTION 7

ANNEXES

7. ANNEXES

7.1. Supplementary material: Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*

Figure S1. *roo* consensus Target Site Duplication (TSD).

Figure S2. Phylogenetic tree including the nine *roo* elements analyzed in this work and 115 *roo* elements annotated in the *D. melanogaster* reference genome.

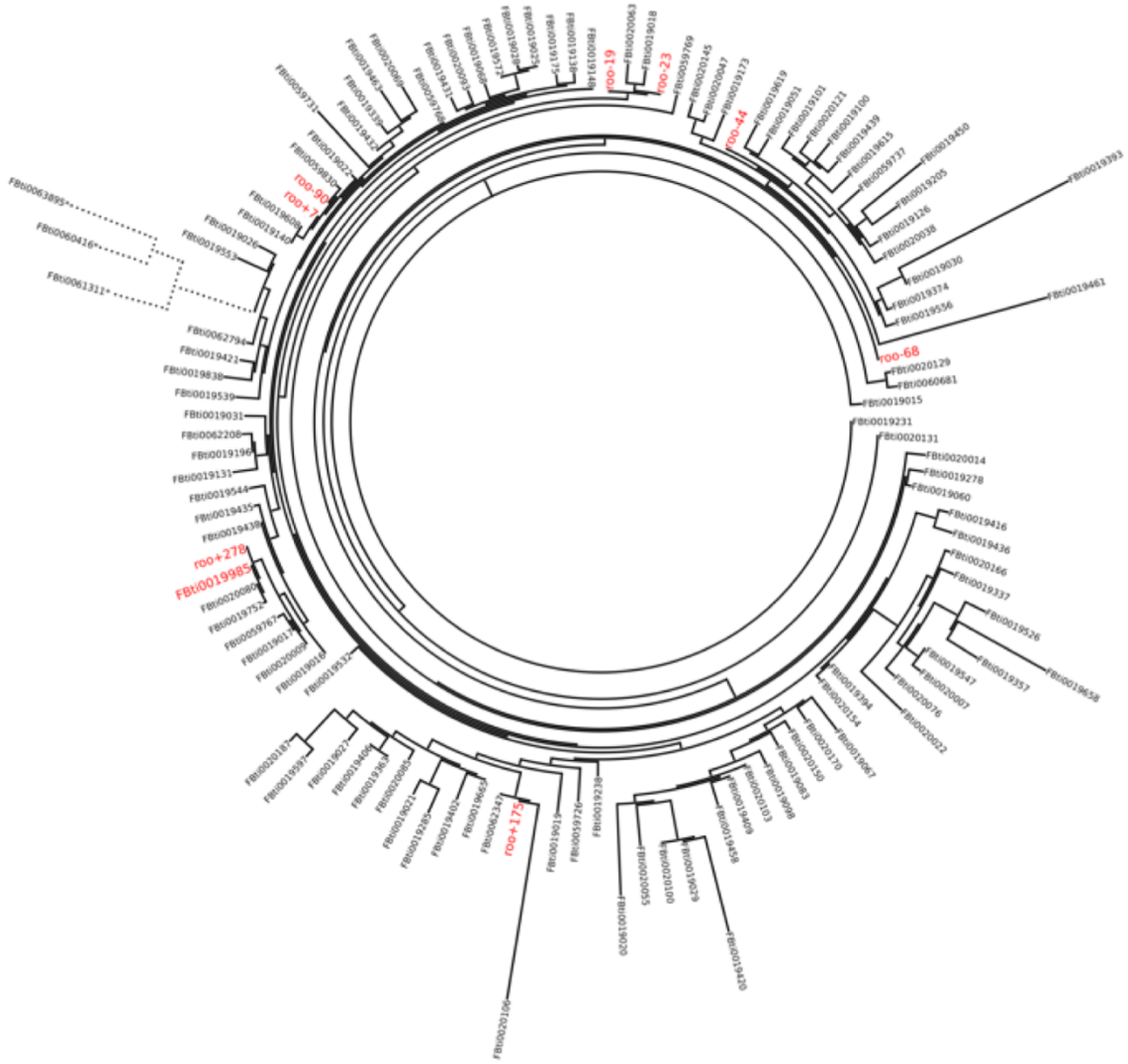


Figure S3. Sequence alignments of the regulatory regions identified in *roo* insertions and in the *CG18446* promoter region.

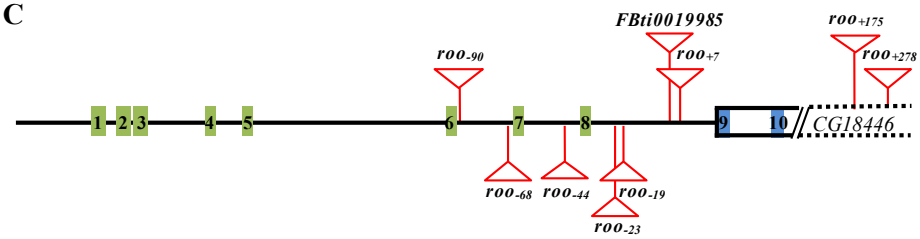
A

	1 Deaf1	2 Nub	3 ara mirr caup	4 ara mirr	5 Bap	6 Tin	7 Vnd	8 Btd	9 INR
FBti0019985 (RAL-639)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
FBti0019985 (RAL-802)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
FBti0019985 (RAL-810)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
FBti0019985 (IV68)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	CGAAGGCGCG	ATCAGTT
roo$_{+7}$ (RAL- 405)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{+7}$ (RAL- 887)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{+7}$ (RAL-911)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{+7}$ (RAL-441)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{+7}$ (RAL-801)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{+175}$ (IV145)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATTAGTT
roo$_{+278}$ (RAL-502)									ATCAGTT
roo$_{-19}$ (IV42)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-19}$ (IV127)	CTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-28}$ (IV40)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-44}$ (RAL-195)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-44}$ (RAL-383)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-68}$ (RAL-75)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ACCAGTT
roo$_{-68}$ (RAL-716)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-68}$ (IV69)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-21)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-88)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-177)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-737)	TTGCAT	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-820)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
roo$_{-90}$ (RAL-857)									ATCAGTT
roo$_{-90}$ (IV50)	TTCGTG	TATGTAAATGAA	TAACA	AAACA	TTAAGTG	CTCAAGTG	TCTCAAGTG	AGGAGGCGGG	ATCAGTT
	*	**********	*****	*****	*******	********	*********	* ***** *	* ****

B

	1	2	3
FBti0019985 (RAL-639)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
FBti0019985 (RAL-802)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
FBti0019985 (RAL-810)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
FBti0019985 (IV68)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+7}$ (RAL- 405)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+7}$ (RAL- 887)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+7}$ (RAL-911)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+7}$ (RAL-441)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+7}$ (RAL-801)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{+175}$ (IV145)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATTA	TTCTCATTTGGGATTTTACA
roo$_{+278}$ (RAL-502)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-19}$ (IV42)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-19}$ (IV127)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-28}$ (IV40)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-44}$ (RAL-195)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-44}$ (RAL-383)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-68}$ (RAL-75)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAACCA	TTCTCATTTGGGATTTTACA
roo$_{-68}$ (RAL-716)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-68}$ (IV69)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (RAL-21)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (RAL-88)	GTAGGTCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (RAL-177)	GTATGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTCACA
roo$_{-90}$ (RAL-737)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (RAL-820)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (RAL-857)	A- - GGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA
roo$_{-90}$ (IV50)	GTAGGCCATTTACTTTAAGA	ATGTCACCTATTTAAACCGAAGATATTTCCAAATAAAATCA	TTCTTATTTGGGATTTTACA

C



	1 Abd-B	2 ct	3 eve zen	4 Optix	5 Optix	6 ara mirr	7 ara mirr caup	8 ara mirr	9 INR	10 DPE
FBti0019985 (RAL-639)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	ATTTGT
FBti0019985 (RAL-802)	TTTTATG	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
FBti0019985 (RAL-810)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
FBti0019985 (IV68)	----------	----------	----------	--------	--------	--------	-------	--------	TTAGTC	AGTTGT
roo+7 (RAL- 405)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo+7 (RAL- 887)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo+7 (RAL-911)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo+7 (RAL-441)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AACCC	TAACA	AAACA	TCAGTC	AGTTGT
roo+7 (RAL-801)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo+175 (IV145)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-19 (IV42)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-19 (IV127)	----------	----------	----------	--------	--------	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-28 (IV40)	----------	----------	----------	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-44 (RAL-195)	TTTTATG	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-44 (RAL-383)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-68 (RAL-75)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-68 (RAL-716)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-68 (IV69)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-90 (RAL-21)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-90 (RAL-88)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-90 (RAL-177)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AATTA	TTAGTC	AGTTGT
roo-90 (RAL-737)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	ATTTGT
roo-90 (RAL-820)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TTAGTC	AGTTGT
roo-90 (RAL-857)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo-90 (IV50)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo- (RAL-371)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo- (RAL-391)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo- (RAL-783)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo- (RAL-908)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
roo- (IV22)	TTTATGA	TTGAAC	CTAATGA	TGATA	TGATA	AAACA	TAACA	AAACA	TCAGTC	AGTTGT
	***	******	*******	*****	*****	** *	*****	** *	* ****	* ****

Figure S4. From top to bottom: Tajima's D in the 23 strains with one of the nine solo-LTR insertions, Tajima's D in the 10 strains with the *FBti0019985* insertion, and Tajima's D in the 15 strains without any of the nine insertions.

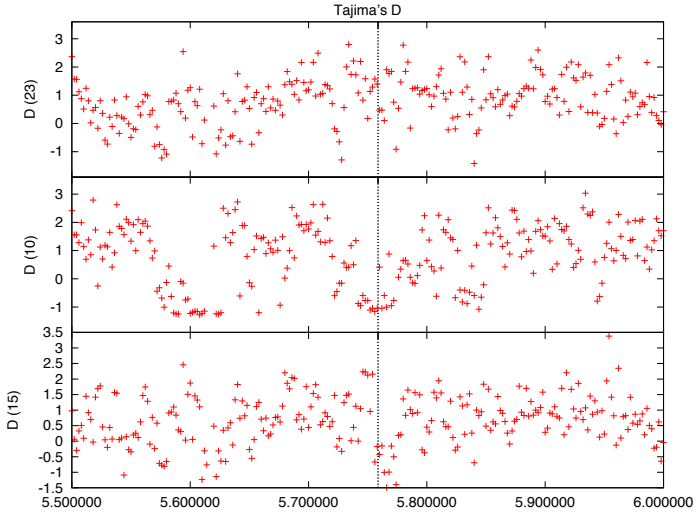


Figure S5. From top to bottom, results for *XP-EHH* , H_{12}, *nSL* , and *iHS* .

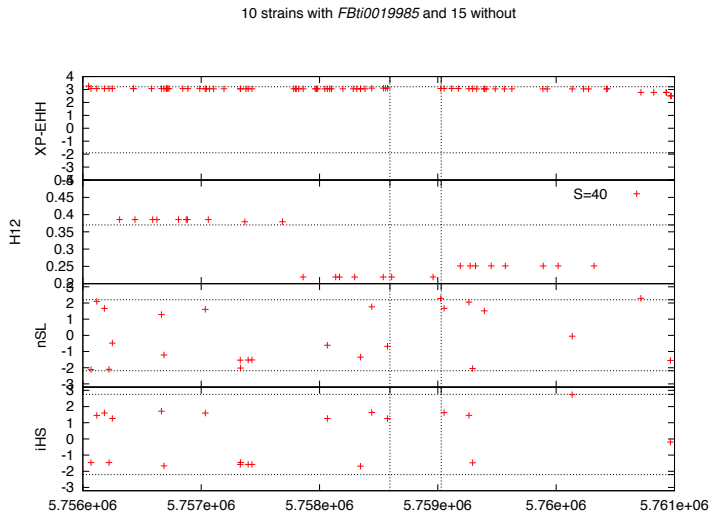


Figure S6. Results for *XP-EHH* , H_{12}, *nSL* , and *iHS* from top to bottom calculated with the 23 strains that contain one of the nine *roo* insertions and the 15 strains without any of the *roo* insertions.

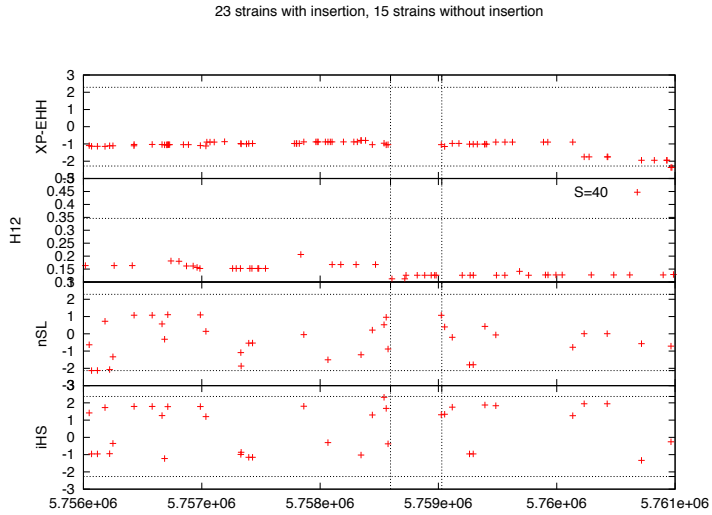


Table S1. Allele frequency estimates using *T-lex2* for the nine *roo* solo-LTR insertions analyzed.

Supplementary table S1A. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain.

	Population							
	North Carolina (DGRP)			Italy	Sweden	Zambia		
***roo* insertion**	**Allele frequency (% PCR)**	**Allele frequency (% T-lex2)**	**Total frequency**	**Allele frequency (% PCR)**	**Allele frequency (% T-lex2)**	**Allele frequency (% T-lex2)**	**% Allele frequency Out of Africa**	**% Allele frequency Total**
FBti0019985	11	19	14	8	4	1	10	8
roo $_{+7}$	19	3	13	0	0	2	7	6
roo $_{+175}$	0	0	0	4	0	2	1	1
roo $_{+278}$	4	0	2	0	0	1	1	1
roo $_{-19}$	0	0	0	17	0	1	3	2
roo $_{-28}$	0	0	0	4	0	1	1	1
roo $_{-44}$	7	25	14	0	0	2	8	6
roo $_{-68}$	7	13	10	4	2	13	7	8
roo $_{-90}$	20	11	16	4	11	4	13	10
9 insertions	69	57	64	42	17	23	47	39

Supplementary table S1B. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain in DGRP population.

	T-lex2 results									
Strain	***FBti0019985***	***roo+7***	***roo+175***	***roo+278***	***roo-19***	***roo-28***	***roo-44***	***roo-68***	***roo-90***	**TOTAL**
RAL142	no data	no data	absent	absent	no data	no data	no data	no data	no data	absent
RAL235	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL31	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL382	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL566	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL57	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL634	absent	absent	absent	absent	no data	no data	no data	no data	absent	absent
RAL892	absent	absent	absent	absent	absent	absent	absent	no data	no data	absent
RAL149	present	no data	absent	absent	no data	no data	no data	absent	absent	FBti0019985
RAL42	present	no data	absent	absent	no data	no data	absent	no data	absent	FBti0019985
RAL491	present	no data	absent	absent	absent	no data	absent	absent	absent	FBti0019985
RAL85	polymorphic	absent	absent	absent	absent	absent	absent	no data	absent	FBti0019985 (polymorphic)
RAL129	absent	absent	absent	polymorphic	absent	absent	absent	no data	present	no data
RAL158	polymorphic	absent	absent	absent	absent	absent	absent	absent	present	no data
RAL161	absent	absent	absent	polymorphic	absent	absent	no data	present	no data	no data
RAL176	present	present	polymorphic	polymorphic	no data	no data	absent	absent	absent	no data
RAL181	polymorphic	absent	polymorphic	polymorphic	absent	absent	absent	polymorphic	present	no data
RAL228	absent	polymorphic	absent	absent	absent	absent	absent	no data	present	no data
RAL229	absent	absent	polymorphic	absent	absent	absent	absent	absent	polymorphic	no data
RAL239	absent	absent	polymorphic	absent	absent	absent	absent	present	absent	no data
RAL256	polymorphic	absent	absent	absent	polymorphic	absent	absent	present	absent	no data
RAL332	polymorphic	polymorphic	polymorphic	absent	absent	absent	polymorphic	absent	present	no data
RAL350	absent	absent	absent	absent	polymorphic	absent	absent	absent	present	no data
RAL45	absent	absent	absent	absent	absent	polymorphic	absent	absent	present	no data
RAL49	absent	polymorphic	absent	absent	absent	absent	absent	absent	polymorphic	no data
RAL517	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	present	polymorphic	no data
RAL555	present	no data	absent	polymorphic	present	no data	absent	absent	absent	no data
RAL59	present	absent	absent	polymorphic	absent	absent	absent	absent	absent	no data
RAL642	absent	polymorphic	absent	absent	absent	polymorphic	polymorphic	absent	absent	no data
RAL69	absent	polymorphic	absent	absent	absent	absent	absent	polymorphic	present	no data
RAL707	absent	absent	absent	absent	absent	absent	absent	present	present	no data
RAL727	absent	polymorphic	polymorphic	absent	absent	polymorphic	polymorphic	present	polymorphic	no data
RAL73	absent	absent	polymorphic	absent	absent	absent	absent	absent	present	no data
RAL738	absent	absent	absent	polymorphic	absent	absent	present	polymorphic	absent	no data
RAL757	polymorphic	absent	absent	absent	polymorphic	polymorphic	absent	absent	polymorphic	no data
RAL786	present	present	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	polymorphic	no data
RAL799	absent	no data	absent	polymorphic	polymorphic	polymorphic	absent	polymorphic	absent	no data
RAL808	polymorphic	absent	absent	absent	polymorphic	absent	polymorphic	absent	polymorphic	no data
RAL852	absent	absent	absent	absent	polymorphic	absent	present	no data	absent	no data
RAL861	present	present	absent	absent	absent	absent	absent	absent	polymorphic	no data
RAL91	no data	absent	absent	absent	absent	absent	absent	present	present	no data
RAL138	absent	absent	absent	absent	absent	absent	present	absent	absent	roo-44
RAL321	absent	absent	absent	absent	absent	absent	present	absent	absent	roo-44
RAL370	absent	absent	absent	absent	absent	absent	present	absent	absent	roo-44
RAL38	absent	absent	absent	absent	absent	absent	present	no data	absent	roo-44
RAL317	absent	absent	no data	absent	no data	absent	no data	present	absent	roo-68
RAL381	absent	absent	absent	absent	absent	absent	absent	present	absent	roo-68
RAL93	no data	polymorphic	absent	absent	absent	absent	absent	absent	absent	roo+7 (polymorphic)
RAL100	absent	absent	absent	absent	absent	absent	absent	absent	present	roo-90
RAL336	no data	no data	absent	no data	no data	no data	no data	absent	present	roo-90

Supplementary table S1C. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain in Sweden population.

	T-lex2 results									
Strain	***FBti0019985***	***roo+7***	***roo+175***	***roo+278***	***roo-19***	***roo-28***	***roo-44***	***roo-68***	***roo-90***	**TOTAL**
B14	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
B16	polymorphic	absent	absent	absent	absent	absent	absent	absent	no data	FBti0019985 (polymorphic)
B17	absent	absent	absent	absent	no data	absent	absent	absent	no data	absent
B18	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
B21	absent	absent	absent	absent	absent	absent	polymorphic	absent	polymorphic	no data
B22	absent	absent	absent	absent	absent	absent	polymorphic	no data	absent	roo44 (polymorphic)
B24	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B25	absent	no data	absent	absent	absent	absent	absent	absent	absent	absent
B26	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B27	no data	absent	absent	absent	absent	absent	absent	no data	polymorphic	roo-90 (polymorphic)
B29	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B32	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B33	absent	absent	absent	absent	absent	absent	absent	no data	polymorphic	roo-90 (polymorphic)
B36	absent	no data	absent	absent	absent	absent	absent	absent	absent	absent
B38	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B39	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
B4	polymorphic	no data	absent	absent	absent	absent	absent	absent	absent	FBti0019985 (polymorphic)
B41	absent	absent	absent	absent	absent	absent	no data	absent	absent	absent
B42	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B45	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B46	absent	no data	absent	absent	absent	absent	no data	no data	absent	absent
B47	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B6	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B7	absent	absent	absent	absent	absent	absent	no data	absent	no data	absent
B8	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
B9	absent	absent	absent	absent	absent	absent	absent	no data	polymorphic	roo-90 (polymorphic)
S3	absent	no data	absent	no data	absent	absent	absent	absent	no data	absent

Supplementary table S1D. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain in Zambia population.

	T-lex2 results									
Strain	***FBti0019985***	***roo+7***	***roo+175***	***roo+278***	***roo-19***	***roo-28***	***roo-44***	***roo-68***	***roo-90***	**TOTAL**
ZI10	absent	absent	absent	polymorphic	absent	absent	polymorphic	polymorphic	absent	no data
ZI114N	absent	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI117	absent	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI161	no data	no data	absent	absent	absent	absent	absent	absent	absent	absent
ZI184	absent	absent	absent	absent	polymorphic	absent	polymorphic	polymorphic	polymorphic	no data
ZI194	absent	absent	absent	absent	absent	polymorphic	no data	present	polymorphic	no data
ZI206	absent	absent	absent	absent	polymorphic	no data	polymorphic	absent	polymorphic	no data
ZI207	absent	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI210	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI213	polymorphic	absent	absent	polymorphic	polymorphic	absent	absent	absent	polymorphic	no data
ZI214	absent	absent	absent	absent	polymorphic	absent	absent	polymorphic	absent	no data
ZI219	polymorphic	absent	absent	absent	absent	absent	absent	polymorphic	absent	no data
ZI228	absent	no data	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
ZI230	absent	no data	absent	absent	absent	absent	no data	absent	absent	absent
ZI232	no data	no data	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
ZI235	no data	no data	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI237	polymorphic	absent	absent	absent	absent	absent	absent	polymorphic	absent	no data
ZI239	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI250	absent	absent	absent	no data	absent	polymorphic	absent	polymorphic	absent	no data
ZI252	absent	absent	absent	absent	absent	absent	no data	no data	absent	absent
ZI253	no data	no data	absent	polymorphic	absent	absent	absent	absent	absent	roo+278 (polymorphic)
ZI255	absent	absent	polymorphic	absent	absent	absent	polymorphic	absent	absent	no data
ZI264	polymorphic	absent	absent	absent	absent	polymorphic	present	no data	absent	no data
ZI265	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI27	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI271	no data	no data	no data	absent	absent	no data	absent	no data	absent	absent
ZI284	absent	absent	polymorphic	absent	absent	absent	absent	polymorphic	absent	no data
ZI292	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI296	absent	absent	absent	absent	absent	polymorphic	absent	polymorphic	absent	no data
ZI303	absent	absent	absent	absent	absent	absent	no data	no data	absent	absent
ZI311N	absent	polymorphic	absent	absent	absent	polymorphic	no data	no data	absent	no data
ZI320	absent	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI321	absent	absent	polymorphic	absent	no data	absent	no data	no data	absent	roo+175 (polymorphic)
ZI324	absent	absent	absent	absent	absent	polymorphic	absent	absent	absent	roo-28 (polymorphic)
ZI332	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI339	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
ZI341	absent	absent	absent	absent	absent	polymorphic	present	no data	absent	no data
ZI344	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI348	polymorphic	absent	absent	polymorphic	absent	absent	polymorphic	polymorphic	polymorphic	no data
ZI357N	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI364	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI365	absent	absent	absent	no data	absent	absent	absent	absent	absent	absent
ZI378	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	polymorphic	no data
ZI379	absent	absent	absent	absent	absent	absent	no data	no data	no data	absent
ZI384	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI386	absent	polymorphic	absent	absent	absent	absent	no data	no data	absent	roo+7 (polymorphic)
ZI398	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI400	absent	absent	absent	absent	absent	no data	absent	absent	absent	absent
ZI402	absent	absent	absent	absent	no data	absent	no data	no data	absent	absent
ZI418N	polymorphic	absent	absent	absent	absent	absent	no data	no data	absent	FBti0019985 (polymorphic)
ZI420	absent	absent	polymorphic	absent	no data	no data	no data	absent	absent	roo+175 (polymorphic)
ZI437	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI443	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI445	absent	absent	absent	absent	absent	absent	no data	no data	absent	absent
ZI447	absent	absent	absent	absent	absent	absent	polymorphic	absent	absent	roo-44 (polymorphic)
ZI455N	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
ZI456	absent	absent	absent	absent	polymorphic	absent	polymorphic	absent	absent	no data
ZI457	absent	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI460	polymorphic	absent	absent	absent	absent	absent	polymorphic	absent	absent	no data
ZI476	no data	no data	absent	absent	absent	absent	no data	absent	absent	absent
ZI477	absent	absent	absent	absent	no data	absent	absent	no data	absent	absent
ZI486	polymorphic	absent	absent	absent	absent	absent	absent	polymorphic	polymorphic	no data
ZI517	absent	absent	absent	absent	absent	no data	absent	polymorphic	absent	roo-68 (polymorphic)
ZI76	no data	absent	absent	absent	absent	absent	absent	absent	no data	absent
ZI85	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)
ZI90	absent	absent	absent	absent	polymorphic	absent	absent	absent	absent	roo-19 (polymorphic)
ZI99	no data	no data	absent	absent	absent	absent	absent	polymorphic	absent	roo-68 (polymorphic)

Supplementary table S1E. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain in Italian population.

	T-lex2 results									
Strain	***FBti0019985***	***roo+7***	***roo+175***	***roo+278***	***roo-19***	***roo-28***	***roo-44***	***roo-68***	***roo-90***	**TOTAL**
IV145	absent	absent	polymorphic	absent	absent	absent	absent	absent	absent	roo+175 (polymorphic)
IV148	absent	absent	absent	absent	absent	absent	absent	no_data	absent	absent
IV22	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
IV33	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
IV40	absent	absent	absent	absent	absent	polymorphic	absent	absent	absent	roo-28 (polymorphic)
IV49	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
IV52	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
IV66	polymorphic	absent	absent	absent	absent	absent	absent	absent	absent	FBti0019985 (polymorphic)
IV68	polymorphic	absent	absent	absent	absent	absent	absent	absent	absent	FBti0019985 (polymorphic)
IV69	absent	absent	absent	absent	absent	absent	absent	polymorphic	absent	roo-68
IV72	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
IV75	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
IV125	polymorphic	absent	absent	absent	absent	absent	absent	absent	polymorphic	no data
IV127	absent	absent	absent	absent	polymorphic	polymorphic	absent	absent	polymorphic	no data
IV42	absent	absent	absent	absent	present	present	absent	absent	polymorphic	no data
IV50	absent	no data	absent	absent	polymorphic	absent	absent	absent	polymorphic	no data

***Note that these strains where also estimated with PCR**

Supplementary table S1F. Allele frequency of all the nine *roo* solo-LTR insertions analyzed. T-lex2 results per strain in DGRP population.

	T-lex2 results									
Strain	***FBti0019985***	***roo+7***	***roo+175***	***roo+278***	***roo-19***	***roo-28***	***roo-44***	***roo-68***	***roo-90***	**TOTAL**
RAL776	absent	absent	absent	polymorphic	absent	absent	absent	present	absent	roo+278 (polymorphic)
RAL911	no data	no data	absent	no data	absent	absent	absent	no data	absent	absent
RAL405	no data	no data	absent	absent	absent	absent	absent	no data	absent	absent
RAL822	absent	no data	absent	absent	absent	absent	no data	no data	absent	absent
RAL195	absent	absent	absent	absent	no data	no data	present	no data	absent	roo-44
RAL855	absent	absent	absent	no data	absent	no data	absent	absent	absent	absent
RAL502	absent	absent	absent	present	absent	absent	absent	absent	absent	roo+278
RAL802	present	no data	absent	absent	absent	absent	absent	absent	absent	FBti0019985
RAL908	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL508	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL40	absent	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL801	no data	absent	absent	absent	absent	absent	absent	absent	absent	absent
RAL783	absent	absent	polymorphic	absent	absent	absent	absent	absent	absent	roo+175 (polymorphic)
RAL383	absent	absent	absent	absent	absent	absent	present	absent	absent	roo-44
RAL75	absent	no data	absent	no data	absent	absent	no data	present	no data	roo-68
RAL857	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
RAL441	absent	absent	absent	absent	absent	absent	absent	absent	polymorphic	roo-90 (polymorphic)
RAL21	absent	absent	absent	absent	absent	absent	absent	absent	present	roo-90 (polymorphic)
RAL177	absent	absent	absent	absent	absent	absent	absent	absent	present	roo-90 (polymorphic)
RAL639	present	no data	absent	absent	no data	no data	absent	absent	absent	FBti0019985
RAL887	no data	no data	absent	absent	absent	absent	absent	absent	absent	absent

***Note that these strains where also estimated with PCR**

Table S2. Sequence alignments of the cis-regulatory motifs located in *roo* solo-LTR insertions and in the *CG18446* promoter region.

Table S2A. Transcription factor binding sites and promoter motifs found inside *FBti0019985.*

Model name	Relative score	Chromosome	Start	End	Strand	Predicted site sequence
Deaf1	1.000	2R	9871110	9871115	-1	ttcgtg
mirr	0.996	2R	9871197	9871201	-1	taaca
ara	1.000	2R	9871197	9871201	-1	taaca
caup	1.000	2R	9871197	9871201	-1	taaca
ara	0.995	2R	9871230	9871234	1	aaaca
mirr	0.999	2R	9871230	9871234	1	aaaca
bap	0.999	2R	9871241	9871247	1	ttaagtg
vnd	0.995	2R	9871266	9871274	-1	tctcaagtg
Nub	NA[(a)]	2R	9871156	9871167	1	tatgtaaatgaa
Tin	NA[(a)]	2R	9871266	9871273	-1	ctcaagtg
Btd	NA[(a)]	2R	9871281	9871290	-1	aggaggcggg
INR	NA[(b)]	2R	9871470	9871476	1	atcagtt

Table S2B. Genomic regions with matrix association potential (MARs) found inside *FBti0019985.*

Model name	Chromosome	Start	End	Strand	Predicted site sequence
MAR	2R	9871403	9871422	1	gtaggccatttactttaaga
MAR	2R	9871433	9871473	1	atgtcacctatttaaaccgaagatatttccaaataaaatca
MAR	2R	9871504	9871523	1	ttcttatttgggattttaca

Table S2C. Transcription factor binding sites and promoter motifs found in *CG18446* promoter region.

Model name	Relative score	Chromosome	Start	End	Strand	Predicted site sequence
Abd-B	1.000	2R	9870850	9870856	-1	tttatga
ct	0.999	2R	9870861	9870866	1	ttgaac
eve	0.999	2R	9870868	9870874	1	ctaatga
zen	1.000	2R	9870868	9870874	1	ctaatga
Optix	1.000	2R	9870898	9870902	-1	tgata
Optix	1.000	2R	9870914	9870918	-1	tgata
ara	0.995	2R	9871999	9871003	-1	aaaca
mirr	0.999	2R	9871999	9871003	-1	aaaca
mirr	0.996	2R	9871028	9871032	-1	taaca
ara	1.000	2R	9871028	9871032	-1	taaca
caup	1.000	2R	9871028	9871032	-1	taaca
ara	0.995	2R	9871056	9871060	1	aaaca
mirr	0.999	2R	9871056	9871060	1	aaaca
INR	NA[(b)]	2R	9871547	9871552	1	tcagtc
DPE	NA[(b)]	2R	9871576	9871581	1	agttgt

[(a)]TFBS described in Batut et al 2013. These TFBS are not included in the JASPAR database.[(b)]Core promoter motifs described in Juven-Gershon and Kadonaga 2010. These TFBS are not included in the JASPAR database.

Table S3. Number of piRNA reads (A) and HP1a reads (B) mapping to each one of the nine *roo* insertions analyzed in this work.

Supplementary table S3A. Mapping of piRNA reads in all the *roo* insertions sequenced in this work.

	FBti0019985		*FBti0019985*		*FBti0019985*		*FBti0019985*		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+175}$		*roo* $_{+278}$		*roo* $_{-19}$		*roo* $_{-19}$		*roo* $_{-23}$	
	Reference genome		*RAL-802*		*RAL-810*		*IV68*		*RAL-405*		*RAL-911*		*RAL-441*		*RAL-801*		*IV145*		*RAL-502*		*IV42*		*IV127*		*IV40*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
1	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
2	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
3	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
4	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
5	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
6	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
7	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
8	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
9	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
10	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
11	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
12	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
13	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
14	0	0	0	0	0	0			0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0
26	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1			0	1	0	1	0	1
27	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1			1	1	1	1	1	1
28	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1			1	1	1	1	1	1
29	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1			1	1	1	1	1	1
30	1	2	1	2	1	2	1	2	0	0	1	2	1	2	1	2	1	2			1	2	1	2	1	2
31	1	2	1	2	1	2	1	2	0	0	1	2	1	2	1	2	1	2			1	2	1	2	1	2
32	1	2	1	2	1	2	1	2	0	0	1	2	1	2	1	2	1	2			1	2	1	2	1	2
33	1	3	1	3	1	3	1	3	0	1	1	3	1	3	1	3	1	3			1	3	1	3	1	3
34	1	3	1	3	1	3	1	3	0	1	1	3	1	3	1	3	1	3			1	3	1	3	1	3
35	1	3	1	3	1	3	1	3	0	1	1	3	1	3	1	3	1	3			1	3	1	3	1	3
36	1	3	1	3	1	3	1	3	0	1	1	3	1	3	1	3	1	3			1	3	1	3	1	3
37	1	4	1	4	1	4	1	4	0	2	1	4	1	4	1	4	1	4			1	4	1	4	1	4
38	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
39	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
40	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
41	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
42	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
43	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
44	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
45	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
46	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
47	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
48	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
49	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
50	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11

	FBti0019985		FBti0019985		FBti0019985		rod +7		rod +7		rod +7		rod +175		rod +278		rod -19		rod -19		rod -23	
	Reference genome		RAL-802		RAL-810		IV68		RAL-405		RAL-911		RAL-441		RAL-801		IV145		RAL-502		IV42		IV127		IV40	
Position	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense
51	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
52	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
53	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
54	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
55	1	11	1	11	1	11	1	11	0	9	1	11	1	11	1	11	1	11			1	11	1	11	1	11
56	1	10	1	10	1	10	1	10	0	9	1	10	1	10	1	10	1	10			1	10	1	10	1	10
57	0	12	0	12	0	12	0	12	0	11	0	12	0	12	0	12	0	12			0	12	0	12	0	12
58	1	12	1	12	1	12	1	12	1	11	1	12	1	12	1	12	1	12			1	12	1	12	1	12
59	1	12	1	12	1	12	1	12	1	11	1	12	1	12	1	12	1	12			1	12	1	12	1	12
60	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
61	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
62	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
63	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
64	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
65	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
66	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
67	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11	1	11	1	11
68	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3	1	3	1	3
69	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
70	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
71	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
72	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
73	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
74	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
75	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
76	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2	1	2	1	2
77	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
78	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
79	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
80	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
81	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
82	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
83	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
84	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
85	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
86	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
87	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7			1	7	1	7	1	7
88	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7	0	7	0	7
89	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7	0	7	0	7
90	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7	0	7	0	7
91	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7	0	7	0	7
92	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7	2	7	2	7
93	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7	2	7	2	7
94	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7	2	7	2	7
95	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7	2	7	2	7
96	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7	2	7	2	7
97	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
98	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
99	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
100	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8

	FBti0019985		*FBti0019985*		*FBti0019985*		*rno -7*		*rno -7*		*rno -7*		*rno -125*		*rno -228*		*rno -10*		*rno -10*		*rno -33*	
	Reference genome		*RAL-802*		*RAL-810*		*RAL-405*		*RAL-911*		*RAL-441*		*RAL-801*		*IVI145*		*RAL-502*		*IVI42*		*IVI27*		*IVI40*	
Position	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense
101	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
102	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
103	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
104	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
105	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8	2	8	2	8
106	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10			2	10	2	10	2	10
107	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
108	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
109	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
110	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
111	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
112	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
113	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3			2	3	2	3	2	3
114	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
115	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
116	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
117	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
118	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
119	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
120	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
121	2	16	2	16	2	16	2	16	2	16	2	16	2	16	2	16			2	16	2	16	2	16
122	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16	0	16	0	16
123	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16	0	16	0	16
124	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16	0	16	0	16
125	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16	0	16	0	16
126	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16	0	16	0	16
127	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
128	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
129	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
130	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
131	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
132	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
133	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
134	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
135	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15	0	15	0	15
136	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
137	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
138	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
139	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
140	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
141	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
142	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
143	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22			0	22	0	22	0	22
144	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13	0	13	0	13
145	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9			0	9	0	9	0	9
146	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
147	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
148	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
149	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
150	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
151	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9	1	9	1	9
152	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9	8	9	8	9
153	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9	8	9	8	9
154	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9	8	9	8	9
155	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9	8	9	8	9
156	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9	9	9	9	9
157	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9	9	9	9	9
158	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9	9	9	9	9
159	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9	9	9	9	9
160	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9	9	9	9	9

	FBti0019985		FBti0019985		FBti0019985		FBti0019985		rod_7		rod_7		rod_7		rod_7		mo_t75		rod_278		rod_19		rod_19		rod_23	
	Reference genome		RAL-802		RAL-810		IV68		RAL-405		RAL-911		RAL-441		RAL-801		IV745		RAL-502		IV42		IV727		IV40	
Position	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense
161	9		9		9		9		9		9		9		9		9				9		9		9	
162	9		9		9		9		9		9		9		9		9				9		9		9	
163	10		9		10		9		10		9		10		9		10				10		10		10	
164	10		9		10		9		10		9		10		9		10				10		10		10	
165	10		9		10		9		10		9		10		9		10				10		10		10	
166	10		9		10		9		10		9		10		9		10				10		10		10	
167	10		9		10		9		10		9		10		9		10				10		10		10	
168	10		9		10		9		10		9		10		9		10				10		10		10	
169	10		9		10		9		10		9		10		9		10				10		10		10	
170	10		9		10		9		10		9		10		9		10				10		10		10	
171	10		9		10		9		10		9		10		9		10				10		10		10	
172	10		10		10		9		10		10		10		10		10				10		10		10	
173	10		1		1		1		10		1		10		1		10				10		1		10	
174	10		1		10		1		10		1		10		1		10				10		1		10	
175	11		1		11		1		11		1		11		1		11				11		1		11	
176	10		1		10		1		10		1		10		1		10				10		1		10	
177	10		1		10		1		10		1		10		1		10				10		1		10	
178	10		1		10		1		10		1		10		1		10				10		1		10	
179	10		1		10		1		10		1		10		1		10				10		1		10	
180	10		1		10		1		10		1		10		1		10				10		1		10	
181	10		1		10		1		10		1		10		1		10				10		1		10	
182	10		1		10		1		10		1		10		1		10				10		1		10	
183	10		1		10		1		10		1		10		1		10				10		1		10	
184	4		1		4		1		4		1		4		1		4				4		4		4	
185	3		1		3		1		3		1		3		1		3				3		3		3	
186	2		1		2		1		2		1		2		1		2				2		2		2	
187	2		1		2		1		2		1		2		1		2				2		2		2	
188	2		1		2		1		2		1		2		1		2				2		2		2	
189	2		1		2		1		2		1		2		1		2				2		2		2	
190	2		1		2		1		2		1		2		1		2				2		2		2	
191	2		1		2		1		2		1		2		1		2				2		2		2	
192	2		1		2		1		2		1		2		1		2				2		2		2	
193	2		1		2		1		2		1		2		1		2				2		2		2	
194	1		1		1		1		1		1		1		1		1				1		1		1	
195	1		1		1		1		1		1		1		1		1				1		1		1	
196	1		1		1		1		1		1		1		1		1				1		1		1	
197	1		1		1		1		1		1		1		1		1				1		1		1	
198	1		1		1		1		1		1		1		1		1				1		1		1	
199	1		1		1		1		1		1		1		1		1				1		1		1	
200	1		1		1		1		1		1		1		1		1				1		1		1	
201	1	2	1	2	1	1	1	1	1	2	1	1	1	2	1	2	1	2			1	2	1	2	1	2
202	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1			1	1	1	1	1	1
203	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1			1	1	1	1	1	1
204	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1			1	1	1	1	1	1
205	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
206	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
207	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
208	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
209	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
210	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
211	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
212	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
213	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1	0	1			0	1	0	1	0	1
214	0	21	0	21	0	0	21	0	21	0	21	0	1	0	21	0	21	0			0	21	0	21	0	21
215	0	21	0	21	0	0	21	0	21	0	21	0	1	0	21	0	21	0			0	21	0	21	0	21
216	0	21	0	21	0	0	21	0	21	0	21	0	1	0	21	0	21	0			0	21	0	21	0	21
217	2	21	2	21	0	0	21	2	21	2	21	2	1	2	21	2	21	2			2	21	2	21	2	21
218	2	21	2	21	0	0	21	2	21	2	21	2	1	2	21	2	21	2			2	21	2	21	2	21
219	2	21	2	21	0	0	21	2	21	2	21	2	1	2	21	2	21	2			2	21	2	21	2	21
220	2	21	2	21	0	0	21	2	21	2	21	2	1	2	21	2	21	2			2	21	2	21	2	21

	FBti0019985		*FBti0019985*		*FBti0019985*		*FBti0019985*		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+175}$		*roo* $_{+278}$		*roo* $_{-19}$		*roo* $_{-19}$		*roo* $_{-23}$	
	Reference genome		*RAL-802*		*RAL-810*		*IV68*		*RAL-405*		*RAL-911*		*RAL-441*		*RAL-801*		*IV145*		*RAL-502*		*IV42*		*IV127*		*IV40*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
221	2	21	2	21	0	0	2	21	2	21	2	21	2	1	2	21	2	21			2	21	2	21	2	21
222	2	21	2	21	0	0	2	21	2	21	2	21	2	1	2	21	2	21			2	21	2	21	2	21
223	2	23	2	23	0	0	2	23	2	23	2	23	2	3	2	23	2	23			2	23	2	23	2	23
224	2	25	2	25	0	4	2	25	2	25	2	25	2	5	2	25	2	25			2	25	2	25	2	25
225	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
226	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
227	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
228	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
229	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
230	15	25	15	25	13	4	15	25	15	25	15	25	15	5	15	25	15	25			15	25	15	25	15	25
231	15	24	15	24	13	4	15	24	15	24	15	24	15	5	15	24	15	24			15	24	15	24	15	24
232	15	24	15	24	13	4	15	24	15	24	15	24	15	5	15	24	15	24			15	24	15	24	15	24
233	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
234	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
235	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
236	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
237	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
238	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
239	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
240	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
241	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
242	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
243	17	24	17	24	15	4	17	24	17	24	17	24	17	5	17	24	17	24			17	24	17	24	17	24
244	27	4	27	4	25	4	27	4	27	4	27	4	27	4	27	4	27	4			27	4	27	4	27	4
245	31	4	31	4	29	4	31	4	31	4	31	4	31	4	31	4	31	4	14	0	31	4	31	4	31	4
246	440	4	440	4	438	4	440	4	440	4	440	4	440	4	440	4	440	4	423	0	440	4	440	4	440	4
247	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4	438	4	438	4
248	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4	438	4	438	4
249	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4	438	4	438	4
250	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4	438	4	438	4
251	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4	438	4	438	4
252	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	423	3	438	7	438	7	438	7
253	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	423	3	438	7	438	7	438	7
254	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	3	438	4	438	4	438	4
255	425	5	425	5	425	5	425	5	425	5	425	5	425	5	425	5	425	5	423	4	425	5	425	5	425	5
256	427	7	427	7	427	7	427	7	427	7	427	7	427	7	427	7	427	7	425	7	427	7	427	7	427	7
257	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
258	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
259	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
260	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
261	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
262	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7	428	7	428	7
263	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
264	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
265	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
266	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
267	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
268	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
269	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
270	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
271	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
272	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
273	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
274	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
275	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12	426	12
276	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12	4	12
277	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12
278	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12
279	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12
280	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12	3	12

	FBti0019985		*FBti0019985*		*FBti0019985*		*FBti0019985*		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+7}$		*roo* $_{+175}$		*roo* $_{+278}$		*roo* $_{-19}$		*roo* $_{-19}$		*roo* $_{-23}$	
	Reference genome		*RAL-802*		*RAL-810*		*IV68*		*RAL-405*		*RAL-911*		*RAL-441*		*RAL-801*		*IV145*		*RAL-502*		*IV42*		*IV127*		*IV40*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
281	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14	3	14
282	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11
283	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11
284	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11
285	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11	3	11
286	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7
287	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
288	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
289	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
290	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
291	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
292	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
293	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
294	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
295	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
296	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
297	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
298	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
299	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3
300	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3
301	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3
302	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4
303	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4
304	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
305	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
306	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
307	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
308	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
309	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
310	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7
311	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
312	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
313	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
314	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
315	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
316	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5	2	5
317	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5
318	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5
319	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5	3	5
320	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
321	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4
322	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4
323	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4
324	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4
325	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4	10	4
326	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5
327	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5
328	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5
329	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5	10	5
330	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
331	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
332	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
333	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9
334	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
335	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
336	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
337	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11	9	11
338	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12
339	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12	9	12
340	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

	FBti0019985		FBti0019985		FBti0019985		r00 +7		r00 +7		r00 +7		r00 +175		r00 +278		r00 -19		r00 -19		r00 -23	
	Reference genome		RAL-802		RAL-810		IV68		RAL-405		RAL-911		RAL-441		RAL-801		IV145		RAL-502		IV42		IV127		IV40	
Position	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense
341	9		9		9		9		9		9		9		9		9		9		9		9		9	
342	9		9		9		9		9		9		9		9		9		9		9		9		9	
343	9		9		9		9		9		9		9		9		9		9		9		9		9	
344	9		9		9		9		9		9		9		9		9		9		9		9		9	
345	9		9		9		9		9		9		9		9		9		9		9		9		9	
346	9		9		9		9		9		9		9		9		9		9		9		9		9	
347	8		8		8		8		8		8		8		8		8		8		8		8		8	
348	8		9		8		9		8		9		8		9		9		9		8		9		8	
349	8		8		8		8		8		8		8		8		8		8		8		8		8	
350	8		8		8		8		8		8		8		8		8		8		8		8		8	
351	1		9		1		1		1		1		1		1		1		1		1		1		1	
352	1		9		1		1		1		1		1		1		1		1		1		1		1	
353	1		9		1		1		1		1		1		1		1		1		1		1		1	
354	1		9		1		1		1		1		1		1		1		1		1		1		1	
355	1		9		1		1		1		1		1		1		1		1		1		1		1	
356	1		8		1		1		1		1		1		1		1		1		1		1		1	
357	1		8		1		1		1		1		1		1		1		1		1		1		1	
358	1		8		1		1		1		1		1		1		1		1		1		1		1	
359	1		12		1		1		1		1		1		1		1		1		1		1		1	
360	1		12		1		1		1		1		1		1		1		1		1		1		1	
361	1		7		1		7		1		1		7		1		1		7		1		7		1	
362	1		8		1		1		1		1		1		1		1		1		1		1		1	
363	1		8		1		8		1		1		1		1		1		1		1		1		1	
364	1		8		1		8		1		1		1		1		1		1		1		1		1	
365	1		8		1		8		1		1		1		1		1		1		1		1		1	
366	0		8		0		8		0		0		0		0		0		0		0		0		0	
367	0		6		0		6		0		0		0		0		0		0		0		0		0	
368	0		5		0		5		0		0		0		0		0		0		0		0		0	
369	0		5		0		5		0		0		0		0		0		0		0		0		0	
370	0		5		0		5		0		0		0		0		0		0		0		0		0	
371	0		5		0		5		0		0		0		0		0		0		0		0		0	
372	0		5		0		5		0		0		0		0		0		0		0		0		0	
373	0		5		0		5		0		0		0		0		0		0		0		0		0	
374	0		5		0		5		0		0		0		0		0		0		0		0		0	
375	0		5		0		5		0		0		0		0		0		0		0		0		0	
376	0		5		0		5		0		0		0		0		0		0		0		0		0	
377	0		5		0		5		0		0		0		0		0		0		0		0		0	
378	0		5		0		5		0		0		0		0		0		0		0		0		0	
379	1		5		1		5		0		1		1		1		1		1		1		1		1	
380	5		5		5		5		0		5		5		5		5		5		5		5		5	
381	19		19		19		19		0		19		19		19		19		19		19		19		19	
382	19		19		19		19		0		19		19		19		19		19		19		19		19	
383	19		19		19		19		0		19		19		19		19		19		19		19		19	
384	19		19		19		19		0		19		19		19		19		19		19		19		19	
385	19		19		19		19		0		19		19		19		19		19		19		19		19	
386	19		19		19		19		0		19		19		19		19		19		19		19		19	
387	19		19		19		19		0		19		19		19		19		19		19		19		19	
388	19		19		19		19		0		19		19		19		19		19		19		19		19	
389	19		19		19		19		0		19		19		19		19		19		19		19		19	
390	19		19		19		19		0		19		19		19		19		19		19		19		19	

	FBti0019985		FBti0019985		FBti0019985		rod +7		rod +7		rod -7		rod +7		rod +171		rod +278		rod -19		rod -19		rod -23	
	Reference genome		R4L-802		R4L-810		I1168		R4L-405		R4L-911		R4L-441		R4L-801		R4L-502		I1142		I1127		I1140	
Position	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense	Sense	Antisense
391	19	1	19	1	0	1	19	1	0	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1
392	19	1	19	1	0	1	19	1	0	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
393	19	1	19	1	0	1	19	1	0	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
394	19	0	19	0	0	0	19	0	0	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0
395	19	0	19	0	0	0	19	0	0	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0
396	19	0	19	0	0	0	19	0	0	0	19	0	19	1	19	0	19	0	19	0	19	0	19	0
397	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
398	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
399	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
400	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
401	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
402	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
403	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
404	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
405	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
406	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
407	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
408	19	1	19	1	0	0	19	1	0	0	19	1	19	1	19	1	19	1	19	1	19	1	19	1
409	18	1	18	1	0	0	18	1	0	0	18	1	18	1	18	1	18	1	18	1	18	1	18	1
410	17	1	17	1	0	0	17	1	0	0	17	1	17	1	17	1	17	1	17	1	17	1	17	1
411	16	1	16	1	0	0	16	1	0	0	16	1	16	1	16	1	16	1	16	1	16	1	16	1
412	4	1	4	1	0	0	4	1	0	0	4	1	4	1	4	1	4	1	4	1	4	1	4	1
413	3	1	3	1	0	0	3	1	0	0	3	1	3	1	3	1	3	1	3	1	3	1	3	1
414	3	1	3	1	0	0	3	1	0	0	3	1	3	1	3	1	3	1	3	1	3	1	3	1
415	3	1	3	1	0	0	3	1	0	0	3	1	3	1	3	1	3	1	3	1	3	1	3	1
416	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
417	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
418	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
419	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
420	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
421	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
422	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
423	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
424	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
425	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
426	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
427	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
428	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
average	34	7	34	7	33	6	35	8	33	7	34	7	34	6	34	7	34	10	75	6	34	7	34	7

	TOO -41		*TOO -41*		*TOO -48*		*TOO -48*		*TOO -68*		*TOO -90*		*TOO -90*		*TOO -90*		*TOO -90*		*TOO -90*		*TOO -90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			0	0
26	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
30	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
31	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
32	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
33	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3
34	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3
35	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3
36	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3
37	1	4	1	4	1	4	1	4	1	4	1	4	1	4	1	4	1	4	1	4			1	4
38	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
39	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
40	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
41	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
42	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
43	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
44	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
45	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
46	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
47	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
48	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
49	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
50	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11

	r00_44		*r00_44*		*r00_48*		*r00_48*		*r00_50*		*r00_50*		*r00_50*		*r00_90*		*r00_90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
51	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
52	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
53	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
54	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
55	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
56	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10			1	10
57	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12			0	12
58	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12			1	12
59	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12			1	12
60	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
61	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
62	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
63	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
64	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
65	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
66	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
67	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11			1	11
68	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3			1	3
69	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
70	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
71	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
72	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
73	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
74	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
75	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
76	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
77	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
78	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
79	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
80	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
81	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
82	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
83	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
84	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
85	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
86	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
87	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7			1	7
88	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7
89	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7
90	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7
91	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7			0	7
92	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7
93	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7
94	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7
95	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7
96	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7	2	7			2	7
97	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8
98	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8
99	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8
100	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8	2	8			2	8

	YOO -44		*YOO -44*		*YOO -48*		*YOO -48*		*YOO -48*		*YOO -90*		*YOO -90*		*YOO -90*		*YOO -90*		*YOO -90*		*YOO -90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
101	2	8	2	8	2	8	2	8	0	1	2	8	2	8	2	8	2	8	2	8			2	8
102	2	8	2	8	2	8	2	8	0	1	2	8	2	8	2	8	2	8	2	8			2	8
103	2	8	2	8	2	8	2	8	0	1	2	8	2	8	2	8	2	8	2	8			2	8
104	2	8	2	8	2	8	2	8	0	1	2	8	2	8	2	8	2	8	2	8			2	8
105	2	8	2	8	2	8	2	8	0	1	2	8	2	8	2	8	2	8	2	8			2	8
106	2	10	2	10	2	10	2	10	0	3	2	10	2	10	2	10	2	10	2	10			2	10
107	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
108	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
109	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
110	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
111	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
112	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
113	2	3	2	3	2	3	2	3	0	3	2	3	2	3	2	3	2	3	2	3			2	3
114	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
115	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
116	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
117	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
118	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
119	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
120	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
121	2	16	2	16	2	16	2	16	0	16	2	16	2	16	2	16	2	16	2	16			2	16
122	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16
123	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16
124	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16
125	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16
126	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16	0	16			0	16
127	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
128	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
129	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
130	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
131	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
132	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
133	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
134	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
135	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15			0	15
136	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
137	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
138	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
139	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
140	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
141	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
142	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
143	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22			0	22
144	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13			0	13
145	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9			0	9
146	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
147	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
148	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
149	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
150	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
151	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9			1	9
152	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9
153	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9
154	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9
155	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9			8	9
156	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
157	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
158	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
159	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
160	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9

	VOO -44		*VOO -44*		*VOO -48*		*VOO -48*		*VOO -48*		*VOO -90*		*VOO -90*		*VOO -90*		*VOO -90*		*VOO -90*		*VOO -90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
161	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
162	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9			9	9
163	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
164	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
165	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
166	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
167	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
168	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
169	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
170	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
171	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9	10	9			10	9
172	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10			10	10
173	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
174	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
175	11	1	10	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1	11	1			11	1
176	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
177	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
178	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
179	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
180	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
181	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
182	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
183	10	1	9	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1	10	1			10	1
184	4	1	3	1	4	1	4	1	4	1	4	1	4	1	4	1	4	1	4	1			4	1
185	3	1	2	1	3	1	2	1	3	1	3	1	3	1	3	1	3	1	3	1			3	1
186	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
187	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
188	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
189	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
190	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
191	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
192	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
193	2	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1			2	1
194	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
195	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
196	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
197	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
198	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
199	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
200	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
201	1	2	0	1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2			1	2
202	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
203	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
204	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1
205	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
206	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
207	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
208	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
209	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
210	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
211	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
212	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
213	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1			0	1
214	0	21	0	20	0	21	0	21	0	21	0	21	0	21	0	21	0	21	0	21			0	21
215	0	21	0	20	0	21	0	21	0	21	0	21	0	21	0	21	0	21	0	21			0	21
216	0	21	0	20	0	21	0	21	0	21	0	21	0	21	0	21	0	21	0	21			0	21
217	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21
218	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21
219	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21
220	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21

	rno -44		*rno -44*		*rno -48*		*rno -48*		*rno -48*		*rno -50*		*rno -50*		*rno -50*		*rno -50*		*rno -50*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
221	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21
222	2	21	2	20	2	21	2	21	2	21	2	21	2	21	2	21	2	21	2	21			2	21
223	2	23	2	22	2	23	2	23	2	23	2	23	2	23	2	23	2	23	2	23			2	23
224	2	25	2	24	2	25	2	25	2	25	2	25	2	25	2	25	2	25	2	25			2	25
225	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
226	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
227	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
228	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
229	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
230	15	25	15	24	15	25	15	25	15	25	15	25	15	25	15	25	15	25	15	25			15	25
231	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24			15	24
232	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24	15	24			15	24
233	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
234	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
235	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
236	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
237	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
238	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
239	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
240	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
241	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
242	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
243	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24	17	24			17	24
244	27	4	27	4	27	4	27	4	27	4	27	4	27	4	27	4	27	4	27	4			27	4
245	31	4	31	4	31	4	31	4	31	4	31	4	31	4	31	4	31	4	31	4	0	4	31	4
246	440	4	440	4	440	4	440	4	440	4	440	4	440	4	440	4	440	4	440	4	10	4	440	4
247	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	14	0	438	4
248	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4
249	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4
250	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4
251	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	0	438	4
252	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	423	0	438	7
253	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	438	7	423	0	438	7
254	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	438	4	423	3	438	4
255	425	5	425	5	425	5	425	5	425	5	425	5	425	5	425	5	425	5	425	5	423	3	425	5
256	427	7	427	7	427	7	427	7	427	7	427	7	427	7	427	7	427	7	427	7	423	3	427	7
257	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	423	4	428	7
258	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	425	7	428	7
259	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7
260	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7
261	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7
262	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	428	7	426	7	428	7
263	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
264	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
265	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
266	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
267	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
268	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7	426	7
269	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	7	426	12
270	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	7	426	12
271	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	12	426	12
272	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	12	426	12
273	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	12	426	12
274	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	12	426	12
275	426	7	426	7	426	12	426	7	426	7	426	12	426	12	426	12	426	12	426	12	426	12	426	12
276	4	7	4	7	4	12	4	7	4	7	4	12	4	12	4	12	4	12	4	12	426	12	4	12
277	3	7	3	7	3	12	3	7	3	7	3	12	3	12	3	12	3	12	3	12	426	12	3	12
278	3	7	3	7	3	12	3	7	3	7	3	12	3	12	3	12	3	12	3	12	4	12	3	12
279	3	7	3	7	3	12	3	7	3	7	3	12	3	12	3	12	3	12	3	12	3	12	3	12
280	3	7	3	7	3	12	3	7	3	7	3	12	3	12	3	12	3	12	3	12	3	12	3	12

	roo -44		*roo -44*		*roo -68*		*roo -68*		*roo -68*		*roo -90*		*roo -90*		*roo -90*		*roo -90*		*roo -90*		*roo -90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*1V69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*1V50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
281	3	7	3	7	3	14	3	7	3	7	3	14	3	12	3	14	3	14	3	14	3	12	3	14
282	3	4	3	4	3	11	3	4	3	4	3	11	3	9	3	11	3	11	3	11	3	12	3	11
283	3	4	3	4	3	11	3	4	3	4	3	11	3	9	3	11	3	11	3	11	3	12	3	11
284	3	4	3	4	3	11	3	4	3	4	3	11	3	9	3	11	3	11	3	11	3	9	3	11
285	3	4	3	4	3	11	3	4	3	4	3	11	3	9	3	11	3	11	3	11	3	9	3	11
286	1	0	1	0	1	7	1	0	1	0	1	7	1	5	1	7	1	7	1	7	3	9	1	7
287	0	0	0	0	0	7	0	0	0	0	0	7	0	5	0	7	0	7	0	7	3	9	0	7
288	0	0	0	0	0	7	0	0	0	0	0	7	0	5	0	7	0	7	0	7	1	5	0	7
289	0	0	0	0	0	7	0	0	0	0	0	7	0	5	0	7	0	7	0	7	0	5	0	7
290	0	0	0	0	0	8	0	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
291	0	0	0	0	0	8	0	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
292	0	0	0	0	0	8	0	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
293	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
294	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
295	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
296	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
297	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
298	1	0	1	0	0	8	1	0	0	0	0	8	0	5	0	7	0	8	0	8	0	5	0	8
299	1	0	1	0	0	3	1	0	0	0	0	3	0	0	0	2	0	3	0	3	0	5	0	3
300	1	0	1	0	0	3	1	0	0	0	0	3	0	0	0	2	0	3	0	3	0	5	0	3
301	1	0	1	0	0	3	1	0	0	0	0	3	0	0	0	2	0	3	0	3	0	5	0	3
302	1	1	1	1	0	4	1	1	0	0	0	4	0	0	0	2	0	4	0	4	0	5	0	4
303	1	1	1	1	0	4	1	1	0	0	0	4	0	0	0	2	0	4	0	4	0	5	0	4
304	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
305	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
306	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
307	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
308	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
309	3	1	3	1	2	4	3	1	2	4	2	4	0	0	0	2	2	4	2	4	0	5	0	4
310	3	4	3	4	2	7	3	4	2	7	2	7	0	0	0	2	2	7	2	7	0	3	2	7
311	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
312	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
313	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
314	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
315	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
316	3	4	3	4	2	5	3	4	2	5	2	5	0	0	0	2	2	5	2	5	0	3	2	5
317	4	4	4	4	3	5	4	4	3	5	3	5	1	0	1	3	3	5	3	5	1	3	3	5
318	4	4	4	4	3	5	4	4	3	5	3	5	1	0	1	3	3	5	3	5	1	3	3	5
319	4	4	4	4	3	5	4	4	3	5	3	5	1	0	1	3	3	5	3	5	1	3	3	5
320	4	4	4	4	3	4	4	4	3	4	3	4	1	0	1	3	3	4	3	4	1	3	3	4
321	11	4	11	4	10	4	11	4	10	4	10	4	8	0	8	10	10	4	10	4	8	3	10	4
322	11	4	11	4	10	4	11	4	10	4	10	4	8	0	8	10	10	4	10	4	8	3	10	4
323	10	4	10	4	10	4	10	4	10	4	10	4	8	0	8	10	10	4	10	4	8	3	10	4
324	10	4	10	4	10	4	10	4	10	4	10	4	8	0	8	10	10	4	10	4	8	3	10	4
325	10	4	10	4	10	4	10	4	10	4	10	4	8	0	8	10	10	4	10	4	8	3	10	4
326	10	5	10	5	10	5	10	5	10	5	10	5	8	1	8	10	10	5	10	5	8	4	10	5
327	10	5	10	5	10	5	10	5	10	5	10	5	8	1	8	10	10	5	10	5	8	4	10	5
328	10	5	10	5	10	5	10	5	10	5	10	5	8	1	8	10	10	5	10	5	8	4	10	5
329	10	5	10	5	10	5	10	5	10	5	10	5	8	1	8	10	10	5	10	5	8	4	10	5
330	10	10	10	10	10	10	10	10	10	10	10	10	8	6	8	10	10	10	10	10	8	9	10	10
331	10	10	10	10	10	10	10	10	10	10	10	10	8	6	8	10	10	10	10	10	8	9	10	10
332	10	10	10	10	10	10	10	10	10	10	10	10	8	6	8	10	10	10	10	10	8	9	10	10
333	10	9	10	9	10	9	10	9	10	9	10	9	8	6	8	10	9	9	10	9	8	9	10	9
334	8	9	8	9	8	9	8	9	8	9	8	9	8	6	8	9	8	9	8	9	8	9	8	9
335	9	9	9	9	9	9	9	9	9	9	9	9	9	6	9	9	9	9	9	9	9	9	9	9
336	9	9	9	9	9	9	9	9	9	9	9	9	9	6	9	9	9	9	9	9	9	9	9	9
337	9	11	9	11	9	11	9	11	9	11	9	11	9	8	9	9	9	11	9	11	9	11	9	11
338	9	12	9	12	9	12	9	12	9	12	9	12	9	9	9	9	9	12	9	12	9	12	9	12
339	9	12	9	12	9	12	9	12	9	12	9	12	9	9	9	9	9	12	9	12	9	12	9	12
340	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

	rod -44		*rod -44*		*rod -48*		*rod -48*		*rod -48*		*rod -49*		*rod -49*		*rod -49*		*rod -50*		*rod -50*		*rod -50*	
	RAL-795		***RAL-383***		***RAL-75***		***RAL-716***		***IV69***		***RAL-21***		***RAL-88***		***RAL-177***		***RAL-737***		***RAL-820***		***RAL-857***		***IV50***	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
341	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
342	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
343	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
344	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
345	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
346	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
347	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
348	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
349	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
350	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9	8	9
351	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9
352	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9
353	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9
354	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9
355	1	9	1	9	1	12	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9	1	9
356	1	8	1	8	1	11	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
357	1	8	1	8	1	12	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
358	1	8	1	8	1	13	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
359	1	12	1	12	1	21	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12
360	1	12	1	12	1	21	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12	1	12
361	1	7	1	7	1	16	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7	1	7
362	1	8	1	8	1	16	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
363	1	8	1	8	1	16	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
364	1	8	1	8	1	16	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
365	1	8	1	8	1	16	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8	1	8
366	0	8	0	8	0	16	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8
367	0	6	0	6	0	14	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6
368	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
369	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
370	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
371	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
372	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
373	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
374	0	5	0	5	0	13	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
375	0	5	0	5	0	16	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
376	0	5	0	5	0	16	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
377	0	5	0	5	0	16	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
378	0	5	0	5	0	16	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
379	1	5	1	5	1	16	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5	1	5
380	5	5	5	5	5	16	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
381	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
382	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
383	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
384	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
385	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
386	19	5	19	5	19	16	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
387	19	5	19	5	19	12	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
388	19	5	19	5	19	12	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
389	19	5	19	5	19	12	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5	19	5
390	19	2	19	2	19	4	19	2	19	2	19	2	19	2	19	2	19	2	19	2	19	2	19	2

	r00 -44		*r00 -44*		*r00 -48*		*r00 -48*		*r00 -48*		*r00 -50*		*r00 -50*		*r00 -50*		*r00 -90*		*r00 -90*		*r00 -90*	
	RAL-195		*RAL-383*		*RAL-75*		*RAL-716*		*IV69*		*RAL-21*		*RAL-88*		*RAL-177*		*RAL-737*		*RAL-820*		*RAL-857*		*IV50*	
Position	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**	**Sense**	**Antisense**
391	19	1	19	1	19	3	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1
392	19	1	19	1	19	3	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1
393	19	1	19	1	19	3	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1	19	1
394	19	0	19	0	19	3	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0
395	19	0	19	0	19	3	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0
396	19	0	19	0	19	4	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0	19	0
397	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
398	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
399	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
400	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
401	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
402	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
403	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
404	19	1	19	1	19	46	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
405	19	1	19	1	19	44	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
406	19	1	19	1	19	43	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
407	19	1	19	1	19	43	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
408	19	1	19	1	19	43	19	1	19	1	19	1	19	1	19	0	19	1	19	1	19	1	19	1
409	18	1	18	1	18	43	18	1	18	1	18	1	18	1	18	0	18	1	18	1	18	1	18	1
410	17	1	17	1	17	43	17	1	17	1	17	1	17	1	17	0	17	1	17	1	17	1	17	1
411	16	1	16	1	16	43	16	1	16	1	16	1	16	1	16	0	16	1	16	1	16	1	16	1
412	4	1	4	1	1	43	4	1	4	1	4	1	4	1	4	0	4	1	4	1	4	1	4	1
413	3	1	3	1	1	43	3	1	3	1	3	1	3	1	3	0	3	1	3	1	3	1	3	1
414	3	1	3	1	0	43	3	1	3	1	3	1	3	1	3	0	3	1	3	1	3	1	3	1
415	3	1	3	1	0	43	3	1	3	1	3	1	3	1	3	0	3	1	3	1	3	1	3	1
416	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
417	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
418	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
419	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
420	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
421	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
422	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
423	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
424	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
425	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
426	0	1	0	1	0	43	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1
427	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
428	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
average	34	7	34	7	34	11	34	7	34	6	34	7	34	7	34	7	34	7	34	7	74	5	34	7

Supplementary table S3B. Mapping of HP1a reads in all the *roo* insertions sequenced in this work.

	FBti0019985	*FBti0019985*	*FBti0019985*	*FBti0019985*	*roo* $_{+7}$	*roo* $_{+7}$	*roo* $_{+7}$	*roo* $_{+7}$	*roo* $_{+175}$	*roo* $_{+278}$	*roo* $_{-19}$	*roo* $_{-19}$	*roo* $_{-23}$	*roo* $_{-44}$	*roo* $_{-44}$	*roo* $_{-68}$	*roo* $_{-68}$	*roo* $_{-68}$	*roo-90*	*roo* $_{-90}$	*roo* $_{-90}$	*roo* $_{-90}$	*roo* $_{-90}$	*roo* $_{-90}$	*roo* $_{-90}$
Position	*Reference genome*	*RAL-802*	*RAL-810*	*IV68*	*RAL-405*	*RAL-911*	*RAL-441*	*RAL-801*	*IV145*	*RAL-502*	*IV42*	*IV127*	*IV40*	*RAL-195*	*RAL-383*	*RAL-75*	*RAL-716*	*IV69*	*RAL-21*	*RAL-88*	*RAL-177*	*RAL-737*	*RAL-820*	*RAL-857*	*IV50*
1	2631	2631	2631		0	2631	2631	2631	2631		2631	1460	2631	2631	2631	2631	2631	2631	2631	2631	2631	2631	0	2631		2631
2	2842	2842	2842		0	2842	2842	2842	2842		2842	1550	2842	2842	2842	2842	2842	2842	2842	2842	2842	2842	0	2842		2842
3	3029	3029	3029		0	3029	3029	3029	3029		3029	1643	3029	3029	3029	3029	3029	3029	3029	3029	3029	3029	0	3029		3029
4	3169	3169	3169		0	3169	3169	3169	3169		3169	1773	3169	3169	3169	3169	3169	3169	3169	3169	3169	3169	0	3169		3169
5	3378	3378	3378		0	3378	3378	3378	3378		3378	1967	3378	3378	3378	3378	3378	3378	3378	3378	3378	3378	0	3378		3378
6	3538	3538	3538		0	3538	3538	3538	3538		3538	2111	3538	3538	3538	3538	3538	3538	3538	3538	3538	3538	0	3538		3538
7	3702	3702	3702		0	3702	3702	3702	3702		3702	2261	3702	3702	3702	3702	3702	3702	3702	3702	3702	3702	0	3702		3702
8	3867	3867	3867		0	3867	3867	3867	3867		3867	2415	3867	3867	3867	3867	3867	3867	3867	3867	3867	3867	0	3867		3867
9	4064	4064	4064		0	4064	4064	4064	4064		4064	2600	4064	4064	4064	4064	4064	4064	4064	4064	4064	4064	0	4064		4064
10	4288	4288	4288		0	4288	4288	4288	4288		4288	2818	4288	4288	4288	4288	4288	4288	4288	4288	4288	4288	0	4288		4288
11	4491	4491	4491		0	4491	4491	4491	4491		4491	3010	4491	4491	4491	4491	4491	4491	4491	4491	4491	4491	0	4491		4491
12	4659	4659	4659		0	4659	4659	4659	4659		4659	3168	4659	4659	4659	4659	4659	4659	4659	4659	4659	4659	0	4659		4659
13	4856	4856	4856		0	4856	4856	4856	4856		4856	3363	4856	4856	4856	4856	4856	4856	4856	4856	4856	4856	0	4856		4856
14	5037	5037	5037		0	5037	5037	5037	5037		5037	3539	5037	5037	5037	5037	5037	5037	5037	5037	5037	5037	0	5037		5037
15	5271	5271	5271	2391	0	5271	5271	5271	5271		5271	3770	5271	5271	5271	5271	5271	5271	5271	5271	5271	5271	0	5271		5271
16	5452	5452	5452	2581	0	5452	5452	5452	5452		5452	3942	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	0	5452		5452
17	5628	5628	5628	2776	0	5628	5628	5628	5628		5628	4107	5628	5628	5628	5628	5628	5628	5628	5628	5628	5628	0	5628		5628
18	5804	5804	5804	2959	0	5804	5804	5804	5804		5804	4272	5804	5804	5804	5804	5804	5804	5804	5804	5804	5804	0	5804		5804
19	5989	5989	5989	3147	0	5989	5989	5989	5989		5989	4443	5989	5989	5989	5989	5989	5989	5989	5989	5989	5989	0	5989		5989
20	6165	6165	6165	3326	0	6165	6165	6165	6165		6165	4616	6165	6165	6165	6165	6165	6165	6165	6165	6165	6165	0	6165		6165
21	6352	6352	6352	3513	0	6352	6352	6352	6352		6352	4815	6352	6352	6352	6352	6352	6352	6352	6352	6352	6352	0	6352		6352
22	6550	6550	6550	3711	0	6550	6550	6550	6550		6550	5014	6550	6550	6550	6550	6550	6550	6550	6550	6550	6550	0	6550		6550
23	6695	6695	6695	3856	146	6695	6695	6695	6695		6695	5161	6695	6695	6695	6695	6695	6695	6695	6695	6695	6695	146	6695		6695
24	6928	6928	6928	4089	2338	6928	6928	6928	6928		6928	5404	6928	6928	6928	6928	6928	6928	6928	6928	6928	6928	2338	6928		6928
25	7071	7071	7071	4232	2483	7071	7071	7071	7071		7071	5548	7071	7071	7071	7071	7071	7071	7071	7071	7071	7071	2483	7071		7071
26	7231	7231	7231	4392	2647	7231	7231	7231	7231		7231	5708	7231	7231	7231	7231	7231	7231	7231	7231	7231	7231	2647	7231		7231
27	7374	7374	7374	4535	2791	7374	7374	7374	7374		7374	5851	7374	7374	7374	7374	7374	7374	7374	7374	7374	7374	2791	7374		7374
28	7564	7564	7564	4725	2983	7564	7564	7564	7564		7564	6041	7564	7564	7564	7564	7564	7564	7564	7564	7564	7564	2983	7564		7564
29	7787	7787	7787	4948	3263	7787	7787	7787	7787		7787	6264	7787	7787	7787	7787	7787	7787	7787	7787	7787	7787	3263	7787		7787
30	7985	7985	7985	5146	3461	7985	7985	7985	7985		7985	6462	7985	7985	7985	7985	7985	7985	7985	7985	7985	7985	3461	7985		7985
31	8002	8002	8002	5369	3684	8002	8002	8002	8002		8002	6685	8002	8002	8002	8002	8002	8002	8002	8002	8002	8002	3684	8002		8002
32	7953	7953	7953	5565	3880	7953	7953	7953	7953		7953	6881	7953	7953	7953	7953	7953	7953	7953	7953	7953	7953	3880	7953		7953
33	7953	7953	7953	5758	4073	7953	7953	7953	7953		7953	7074	7953	7953	7953	7953	7953	7953	7953	7953	7953	7953	4073	7953		7953
34	7874	7874	7874	5922	4237	7874	7874	7874	7874		7874	7238	7874	7874	7874	7874	7874	7874	7874	7874	7874	7874	4237	7874		7874
35	7821	7821	7821	6095	4410	7821	7821	7821	7821		7821	7411	7821	7821	7821	7821	7821	7821	7821	7821	7821	7821	4410	7821		7821
36	7804	7804	7804	6256	4571	7804	7804	7804	7804		7804	7410	7804	7804	7804	7804	7804	7804	7804	7804	7804	7804	4571	7804		7804
37	7786	7786	7786	6413	4728	7786	7786	7786	7786		7786	7404	7786	7786	7786	7786	7786	7786	7786	7786	7786	7786	4728	7786		7786
38	7782	7782	7782	6620	4935	7782	7782	7782	7782		7782	7521	7782	7782	7782	7782	7782	7782	7782	7782	7782	7782	4935	7782		7782
39	7755	7755	7755	6784	5099	7755	7755	7755	7755		7755	7599	7755	7755	7755	7755	7755	7755	7755	7755	7755	7755	5099	7755		7755
40	7811	7811	7811	6977	5292	7811	7811	7811	7811		7811	7665	7811	7811	7811	7811	7811	7811	7811	7811	7811	7811	5292	7811		7811
41	7773	7773	7773	7123	5438	7773	7773	7773	7773		7773	7632	7773	7773	7773	7773	7773	7773	7773	7773	7773	7773	5438	7773		7773
42	7847	7847	7847	7325	5640	7847	7847	7847	7847		7847	7713	7847	7847	7847	7847	7847	7847	7847	7847	7847	7847	5640	7847		7847
43	7845	7845	7845	7482	5797	7845	7845	7845	7845		7845	7718	7845	7845	7845	7845	7845	7845	7845	7845	7845	7845	5797	7845		7845
44	7847	7847	7847	7669	5984	7847	7847	7847	7847		7847	7732	7847	7847	7847	7847	7847	7847	7847	7847	7847	7847	5984	7847		7847
45	7851	7851	7851	7695	6170	7851	7851	7851	7851		7851	7750	7851	7851	7851	7851	7851	7851	7851	7851	7851	7851	6170	7851		7851
46	7794	7794	7794	7657	6355	7794	7794	7794	7794		7794	7696	7794	7794	7794	7794	7794	7794	7794	7794	7794	7794	6355	7794		7794
47	7795	7795	7795	7685	6579	7795	7795	7795	7795		7795	7708	7795	7795	7795	7795	7795	7794	7795	7795	7795	6579	7795		7795
48	7846	7846	7846	7762	6810	7846	7846	7846	7846		7846	7764	7846	7846	7846	7846	7846	7845	7846	7846	7846	6810	7846		7846
49	7797	7797	7797	7732	6974	7797	7797	7797	7797		7797	7722	7797	7797	7797	7797	7797	7795	7797	7797	7797	6974	7797		7797
50	7814	7814	7814	7759	7146	7814	7814	7814	7814		7814	7744	7814	7814	7814	7814	7814	7812	7814	7814	7814	7146	7814		7814

Position	FBti0019985 Reference genome	FBti0019985 RAL-802	FBti0019985 RAL-810	FBti0019985 IV68	roo +7 RAL-405	roo +7 RAL-911	roo +7 RAL-441	roo +7 RAL-801	roo +175 IV145	roo +278 RAL-502	roo -19 IV42	roo -19 IV127	roo -23 IV40	roo -44 RAL-195	roo -44 RAL-383	roo -68 RAL-75	roo -68 RAL-716	roo -68 IV69	roo-90 RAL-21	roo -90 RAL-88	roo -90 RAL-177	roo -90 RAL-737	roo -90 RAL-820	roo -90 RAL-857	roo -90 IV50
51	7782	7782	7782	7740	7313	7782	7782	7782	7782		7782	7715	7782	7782	7782	7782	7782	7780	7782	7782	7782	7313	7782		7782
52	7791	7791	7791	7756	7496	7791	7791	7791	7791		7791	7735	7791	7791	7791	7791	7791	7789	7791	7791	7791	7496	7791		7791
53	7785	7785	7785	7753	7655	7785	7785	7785	7785		7785	7739	7785	7785	7785	7785	7785	7783	7785	7785	7785	7655	7785		7785
54	7785	7785	7785	7758	7668	7785	7785	7785	7785		7785	7743	7785	7785	7785	7785	7785	7783	7785	7785	7785	7668	7785		7785
55	7832	7832	7832	7808	7736	7832	7832	7832	7832		7832	7800	7832	7832	7832	7832	7832	7829	7832	7832	7832	7736	7832		7832
56	7830	7830	7830	7812	7749	7830	7830	7830	7830		7830	7805	7830	7830	7830	7830	7830	7826	7830	7830	7830	7749	7830		7830
57	7855	7855	7855	7847	7795	7855	7855	7855	7855		7855	7841	7855	7855	7855	7855	7855	7841	7855	7855	7855	7795	7855		7855
58	7881	7881	7881	7878	7840	7881	7881	7881	7881		7881	7873	7881	7881	7881	7881	7881	7856	7881	7881	7881	7840	7881		7881
59	7902	7902	7902	7899	7869	7902	7902	7902	7902		7902	7894	7902	7902	7902	7902	7902	7866	7902	7902	7902	7869	7902		7902
60	7848	7848	7848	7845	7821	7848	7848	7848	7848		7848	7841	7848	7848	7848	7848	7848	7803	7848	7848	7848	7821	7848		7848
61	7923	7923	7923	7921	7903	7923	7923	7923	7923		7923	7917	7923	7923	7923	7923	7923	7864	7923	7923	7923	7903	7923		7923
62	8003	8003	8003	8002	7988	8003	8003	8003	8003		8003	7998	8003	8003	8003	8003	8003	7926	8003	8003	8003	7988	8003		8003
63	8045	8045	8045	8045	8036	8045	8045	8045	8045		8045	8041	8045	8045	8045	8045	8045	7948	8045	8045	8045	8036	8045		8045
64	8107	8107	8107	8107	8099	8107	8107	8107	8107		8107	8103	8107	8107	8107	8107	8107	8002	8107	8107	8107	8099	8107		8107
65	8059	8059	8059	8059	8055	8059	8059	8059	8059		8059	8056	8059	8059	8059	8059	8059	7939	8059	8059	8059	8055	8059		8059
66	8095	8095	8095	8095	8091	8095	8095	8095	8095		8095	8092	8095	8095	8095	8095	8095	7971	8095	8095	8095	8091	8095		8095
67	8037	8037	8037	8037	8034	8037	8037	8037	8037		8037	8035	8037	8037	8037	8037	8037	7907	8037	8037	8037	8034	8037		8037
68	8088	8088	8088	8088	8086	8088	8088	8088	8088		8088	8086	8088	8088	8088	8088	8088	7954	8088	8088	8088	8086	8088		8088
69	8036	8036	8036	8036	8034	8036	8036	8036	8036		8036	8034	8036	8036	8036	8036	8036	7900	8036	8036	8036	8034	8036		8036
70	8019	8019	8019	8019	8019	8019	8019	8019	8019		8019	8019	8019	8019	8019	8019	8019	7880	8019	8019	8019	8019	8019		8019
71	8071	8071	8071	8071	8071	8071	8071	8071	8071		8071	8071	8071	8071	8071	8071	8071	7928	8071	8071	8071	8071	8071		8071
72	8137	8137	8137	8137	8137	8137	8137	8137	8137		8137	8137	8137	8137	8137	8137	8137	7980	8137	8137	8137	8137	8137		8137
73	8165	8165	8165	8165	8165	8165	8165	8165	8165		8165	8165	8165	8165	8165	8165	8165	8004	8165	8165	8165	8165	8165		8165
74	8152	8152	8152	8152	8152	8152	8152	8152	8152		8152	8152	8152	8152	8152	8152	8152	7985	8152	8152	8152	8152	8152		8152
75	8203	8203	8203	8203	8203	8203	8203	8203	8203		8203	8203	8203	8203	8203	8203	8203	8031	8203	8203	8203	8203	8203		8203
76	8178	8178	8178	8178	8178	8178	8178	8178	8178		8178	8178	8178	8178	8178	8178	8178	8002	8178	8178	8178	8178	8178		8178
77	8262	8262	8262	8262	8262	8262	8262	8262	8262		8262	8262	8262	8262	8262	8262	8262	8073	8262	8262	8262	8262	8262		8262
78	8209	8209	8209	8209	8209	8209	8209	8209	8209		8209	8209	8209	8209	8209	8209	8209	7922	8209	8209	8209	8209	8209		8209
79	8200	8200	8200	8200	8200	8200	8200	8200	8200		8200	8200	8200	8200	8200	8200	8200	7832	8200	8200	8200	8200	8200		8200
80	8232	8232	8232	8232	8232	8232	8232	8232	8232		8232	8232	8232	8232	8232	8232	8232	7849	8232	8232	8232	8232	8232		8232
81	8231	8231	8231	8231	8231	8231	8231	8231	8231		8231	8231	8231	8231	8231	8231	8231	7843	8231	8231	8231	8231	8231		8231
82	8299	8299	8299	8299	8299	8299	8299	8299	8299		8299	8299	8299	8299	8299	8299	8299	7894	8299	8299	8299	8299	8299		8299
83	8240	8240	8240	8240	8240	8240	8240	8240	8240		8240	8240	8240	8240	8240	8240	8240	7829	8240	8240	8240	8240	8240		8240
84	8286	8286	8286	8286	8286	8286	8286	8286	8286		8286	8286	8286	8286	8286	8286	8286	7864	8286	8286	8286	8286	8286		8286
85	8249	8249	8249	8249	8249	8249	8249	8249	8249		8249	8249	8249	8249	8249	8249	8249	7823	8249	8249	8249	8249	8249		8249
86	8235	8235	8235	8235	8235	8235	8235	8235	8235		8235	8235	8235	8235	8235	8235	8235	7803	8235	8235	8235	8235	8235		8235
87	8205	8205	8205	8205	8205	8205	8205	8205	8205		8205	8205	8205	8205	8205	8205	8205	7765	8205	8205	8205	8205	8205		8205
88	8169	8169	8169	8169	8169	8169	8169	8169	8169		8169	8169	8169	8169	8169	8169	8169	7720	8169	8169	8169	8169	8169		8169
89	8175	8175	8175	8175	8175	8175	8175	8175	8175		8175	8175	8175	8175	8175	8175	8175	7721	8175	8175	8175	8175	8175		8175
90	8181	8181	8181	8181	8181	8181	8181	8181	8181		8181	8181	8181	8181	8181	8181	8181	7716	8181	8181	8181	8181	8181		8181
91	8184	8184	8184	8184	8184	8184	8184	8184	8184		8184	8184	8184	8184	8184	8184	8184	7706	8184	8184	8184	8184	8184		8184
92	8249	8249	8249	8249	8249	8249	8249	8249	8249		8249	8249	8249	8249	8249	8249	8249	7762	8249	8249	8249	8249	8249		8249
93	8193	8193	8193	8193	8193	8193	8193	8193	8193		8193	8193	8193	8193	8193	8193	8193	7697	8193	8193	8193	8193	8193		8193
94	8227	8227	8227	8227	8227	8227	8227	8227	8227		8227	8227	8227	8227	8227	8227	8227	7725	8227	8227	8227	8227	8227		8227
95	8150	8150	8150	8150	8150	8150	8150	8150	8150		8150	8150	8150	8150	8150	8150	8150	7636	8150	8150	8150	8150	8150		8150
96	8165	8165	8165	8165	8165	8165	8165	8165	8165		8165	8165	8165	8165	8165	8165	8165	7680	8165	8165	8165	8165	8165		8165
97	8114	8114	8114	8114	8114	8114	8114	8114	8114		8114	8114	8114	8114	8114	8114	8114	7601	8114	8114	8114	8114	8114		8114
98	8029	8029	8029	8029	8029	8029	8029	8029	8029		8029	8029	8029	8029	8029	8029	8029	7522	8029	8029	8029	8029	8029		8029
99	8001	8001	8001	8001	8001	8001	8001	8001	8001		8001	8001	8001	8001	8001	8001	8001	7502	8001	8001	8001	8001	8001		8001
100	7975	7975	7975	7975	7975	7975	7975	7975	7975		7975	7975	7975	7975	7975	7975	7975	7475	7975	7975	7975	7975	7975		7975
101	7997	7997	7997	7997	7997	7997	7997	7997	7997		7997	7997	7997	7997	7997	7997	7997	7501	7997	7997	7997	7997	7997		7997
102	7975	7975	7975	7975	7975	7975	7975	7975	7975		7975	7975	7975	7975	7975	7975	7975	7484	7975	7975	7975	7975	7975		7975
103	8032	8032	8032	8032	8032	8032	8032	8032	8032		8032	8032	8032	8032	8032	8032	8032	7548	8032	8032	8032	8032	8032		8032
104	8032	8032	8032	8032	8032	8032	8032	8032	8032		8032	8032	8032	8032	8032	8032	8032	7552	8032	8032	8032	8032	8032		8032
105	8062	8062	8062	8062	8062	8062	8062	8062	8062		8062	8062	8062	8062	8062	8062	8062	7588	8062	8062	8062	8062	8062		8062
106	8117	8117	8117	8117	8117	8117	8117	8117	8117		8117	8117	8117	8117	8117	8117	8117	7647	8117	8117	8117	8117	8117		8117
107	8071	8071	8071	8071	8071	8071	8071	8071	8071		8071	8071	8071	8071	8071	8071	8071	7613	8071	8071	8071	8071	8071		8071
108	8130	8130	8130	8130	8130	8130	8130	8130	8130		8130	8130	8130	8130	8130	8130	8130	7691	8130	8130	8130	8130	8130		8130
109	8083	8083	8083	8083	8083	8083	8083	8083	8083		8083	8083	8083	8083	8083	8083	8083	7661	8083	8083	8083	8083	8083		8083
110	8063	8063	8063	8063	8063	8063	8063	8063	8063		8063	8063	8063	8063	8063	8063	8063	7649	8063	8063	8063	8063	8063		8063

Position	FBti0019985 Reference genome	FBti0019985 RAL-802	FBti0019985 RAL-810	FBti0019985 IV68	roo +7 RAL-405	roo +7 RAL-911	roo +7 RAL-441	roo +7 RAL-801	roo +175 IV145	roo +278 RAL-502	roo -19 IV42	roo -19 IV127	roo -23 IV40	roo -44 RAL-195	roo -44 RAL-383	roo -68 RAL-75	roo -68 RAL-716	roo -68 IV69	roo-90 RAL-21	roo -90 RAL-88	roo -90 RAL-177	roo -90 RAL-737	roo -90 RAL-820	roo -90 RAL-857	roo -90 IV50
111	8041	8041	8041	8041	8041	8041	8041	8041	8041		8041	8041	8041	8041	8041	8041	8041	7644	8041	8041	8041	8041	8041		8041
112	8014	8014	8014	8014	8014	8014	8014	8014	8014		8014	8014	8014	8014	8014	8014	8014	7635	8014	8014	8014	8014	8014		8014
113	7927	7927	7927	7927	7927	7927	7927	7927	7927		7927	7927	7927	7927	7927	7927	7927	7573	7927	7927	7927	7927	7927		7927
114	7845	7845	7845	7845	7845	7845	7845	7845	7845		7845	7845	7845	7845	7845	7845	7845	7590	7845	7845	7845	7845	7845		7845
115	7891	7891	7891	7891	7891	7891	7891	7891	7891		7891	7891	7891	7891	7891	7891	7891	7731	7891	7891	7891	7891	7891		7891
116	7825	7825	7825	7825	7825	7825	7825	7825	7825		7825	7825	7825	7825	7825	7825	7825	7675	7825	7825	7825	7825	7825		7825
117	7851	7851	7851	7851	7851	7851	7851	7851	7851		7851	7851	7851	7851	7851	7851	7851	7701	7851	7851	7851	7851	7851		7851
118	7796	7796	7796	7796	7796	7796	7796	7796	7796		7796	7796	7796	7796	7796	7796	7796	7662	7796	7796	7796	7796	7796		7796
119	7832	7832	7832	7832	7832	7832	7832	7832	7832		7832	7832	7832	7832	7832	7832	7832	7705	7832	7832	7832	7832	7832		7832
120	7912	7912	7912	7912	7912	7912	7912	7912	7912		7912	7912	7912	7912	7912	7912	7912	7792	7912	7912	7912	7912	7912		7912
121	7910	7910	7910	7910	7910	7910	7910	7910	7910		7910	7910	7910	7910	7910	7910	7910	7797	7910	7910	7910	7910	7910		7910
122	8078	8078	8078	8078	8078	8078	8078	8078	8078		8078	8078	8078	8078	8078	8078	8078	7967	8078	8078	8078	8078	8078		8078
123	8174	8174	8174	8174	8174	8174	8174	8174	8174		8174	8174	8174	8174	8174	8174	8174	8067	8174	8174	8174	8174	8174		8174
124	8179	8179	8179	8179	8179	8179	8179	8179	8179		8179	8179	8179	8179	8179	8179	8179	8079	8179	8179	8179	8179	8179		8179
125	8220	8220	8220	8220	8220	8220	8220	8220	8220		8220	8220	8220	8220	8220	8220	8220	8124	8220	8220	8220	8220	8220		8220
126	8261	8261	8261	8261	8261	8261	8261	8261	8261		8261	8261	8261	8261	8261	8261	8261	8171	8261	8261	8261	8261	8261		8261
127	8237	8237	8237	8237	8237	8237	8237	8237	8237		8237	8237	8237	8237	8237	8237	8237	8154	8237	8237	8237	8237	8237		8237
128	8209	8209	8209	8209	8209	8209	8209	8209	8209		8209	8209	8209	8209	8209	8209	8209	8138	8209	8209	8209	8209	8209		8209
129	8225	8225	8225	8225	8225	8225	8225	8225	8225		8225	8225	8225	8225	8225	8225	8225	8158	8225	8225	8225	8225	8225		8225
130	8274	8274	8274	8274	8274	8274	8274	8274	8274		8274	8274	8274	8274	8274	8274	8274	8217	8274	8274	8274	8274	8274		8274
131	8279	8279	8279	8279	8279	8279	8279	8279	8279		8279	8279	8279	8279	8279	8279	8279	8228	8279	8279	8279	8279	8279		8279
132	8258	8258	8258	8258	8258	8258	8258	8258	8258		8258	8258	8258	8258	8258	8258	8258	8212	8258	8258	8258	8258	8258		8258
133	8277	8277	8277	8277	8277	8277	8277	8277	8277		8277	8277	8277	8277	8277	8277	8277	8231	8277	8277	8277	8277	8277		8277
134	8288	8288	8288	8288	8288	8288	8288	8288	8288		8288	8288	8288	8288	8288	8288	8288	8246	8288	8288	8288	8288	8288		8288
135	8341	8341	8341	8341	8341	8341	8341	8341	8341		8341	8341	8341	8341	8341	8341	8341	8303	8341	8341	8341	8341	8341		8341
136	8442	8442	8442	8442	8442	8442	8442	8442	8442		8442	8442	8442	8442	8442	8442	8442	8407	8442	8442	8442	8442	8442		8442
137	8469	8469	8469	8469	8469	8469	8469	8469	8469		8469	8469	8469	8469	8469	8469	8469	8439	8469	8469	8469	8469	8469		8469
138	8550	8550	8550	8550	8550	8550	8550	8550	8550		8550	8550	8550	8550	8550	8550	8550	8522	8550	8550	8550	8550	8550		8550
139	8511	8511	8511	8511	8511	8511	8511	8511	8511		8511	8511	8511	8511	8511	8511	8511	8483	8511	8511	8511	8511	8511		8511
140	8562	8562	8562	8561	8562	8562	8562	8562	8562		8562	8562	8562	8562	8562	8562	8562	8539	8562	8562	8562	8562	8562		8562
141	8601	8601	8601	8599	8601	8601	8601	8601	8601		8601	8601	8601	8601	8601	8601	8601	8586	8601	8601	8601	8601	8601		8601
142	8567	8567	8567	8564	8567	8567	8567	8567	8567		8567	8567	8567	8567	8567	8567	8567	8557	8567	8567	8567	8567	8567		8567
143	8648	8648	8648	8644	8648	8648	8648	8648	8648		8648	8648	8648	8648	8648	8648	8648	8639	8648	8648	8648	8648	8648		8648
144	8551	8551	8551	8547	8551	8551	8551	8551	8551		8551	8551	8551	8551	8551	8551	8551	8544	8551	8551	8551	8551	8551		8551
145	8568	8568	8568	8563	8568	8568	8568	8568	8568		8568	8568	8568	8568	8568	8568	8568	8563	8568	8568	8568	8568	8568		8568
146	8543	8543	8543	8535	8543	8543	8543	8543	8543		8543	8543	8543	8543	8543	8543	8543	8540	8543	8543	8543	8543	8543		8543
147	8524	8524	8524	8515	8524	8524	8524	8524	8524		8524	8524	8524	8524	8524	8524	8524	8521	8524	8524	8524	8524	8524		8524
148	8505	8505	8505	8490	8505	8505	8505	8505	8505		8505	8505	8505	8505	8505	8505	8505	8502	8505	8505	8505	8505	8505		8505
149	8521	8521	8521	8499	8521	8521	8521	8521	8521		8521	8521	8521	8521	8521	8521	8521	8521	8521	8521	8521	8521	8521		8521
150	8645	8645	8645	8611	8645	8645	8645	8645	8645		8645	8645	8645	8645	8645	8645	8645	8645	8645	8645	8645	8645	8645		8645
151	8640	8640	8640	8596	8640	8640	8640	8640	8640		8640	8640	8640	8640	8640	8640	8640	8640	8640	8640	8640	8640	8640		8640
152	8662	8662	8662	8608	8662	8662	8662	8662	8662		8662	8662	8662	8662	8662	8662	8662	8662	8662	8662	8662	8662	8662		8662
153	8745	8745	8745	8676	8745	8745	8745	8745	8745		8745	8745	8745	8745	8745	8745	8745	8745	8745	8745	8745	8745	8745		8745
154	8810	8810	8810	8707	8810	8810	8810	8810	8810		8810	8810	8810	8810	8809	8810	8810	8810	8810	8810	8810	8810	8810		8810
155	8821	8821	8821	8692	8821	8821	8821	8821	8821		8821	8821	8821	8821	8819	8821	8821	8821	8821	8821	8821	8821	8821		8821
156	8858	8858	8858	8677	8858	8858	8858	8858	8858		8858	8858	8858	8858	8856	8858	8858	8858	8858	8858	8858	8858	8858		8858
157	8914	8914	8914	8697	8914	8914	8914	8914	8914		8914	8914	8914	8914	8912	8914	8914	8914	8914	8914	8914	8914	8914		8914
158	8769	8769	8769	8522	8769	8769	8769	8769	8769		8769	8769	8769	8769	8766	8769	8769	8769	8769	8769	8769	8769	8769		8769
159	8828	8828	8828	8349	8828	8828	8828	8828	8828		8828	8828	8828	8828	8822	8828	8828	8828	8828	8828	8828	8828	8828		8828
160	8888	8888	8888	8208	8888	8888	8888	8888	8888		8888	8888	8888	8888	8878	8888	8888	8888	8888	8888	8888	8888	8888		8888
161	8894	8894	8894	8019	8894	8894	8894	8894	8894		8894	8894	8894	8894	8881	8894	8894	8894	8894	8894	8894	8894	8894		8894
162	8855	8855	8855	7829	8855	8855	8855	8855	8855		8855	8855	8855	8855	8841	8855	8855	8855	8855	8855	8855	8855	8855		8855
163	8880	8880	8880	7712	8880	8880	8878	8880	8880		8880	8880	8880	8880	8862	8880	8880	8880	8880	8880	8880	8880	8880		8880
164	8928	8928	8928	7563	8928	8928	8925	8928	8928		8928	8928	8928	8928	8909	8928	8928	8928	8928	8928	8928	8928	8928		8928
165	9080	9080	9080	7434	9080	9080	9075	9080	9080		9080	9080	9080	9080	9056	9080	9080	9080	9080	9080	9080	9080	9080		9080
166	8985	8985	8985	7199	8985	8985	8976	8985	8985		8985	8985	8985	8985	8956	8985	8985	8985	8985	8985	8985	8985	8985		8985
167	9131	9131	9131	7062	9131	9131	9121	9131	9131		9131	9131	9131	9131	9095	9131	9131	9131	9131	9131	9131	9131	9131		9131
168	9125	9125	9125	6865	9125	9125	9107	9125	9125		9125	9125	9125	9125	9079	9125	9125	9125	9125	9125	9125	9125	9125		9125
169	9237	9237	9237	6720	9237	9237	9214	9237	9237		9237	9237	9237	9237	9165	9237	9237	9237	9237	9237	9237	9237	9237		9237
170	9226	9226	9226	6504	9226	9226	9195	9226	9226		9226	9226	9226	9226	9140	9226	9226	9226	9226	9226	9226	9226	9226		9226

	FBti0019985	FBti0019985	FBti0019985	FBti0019985	roo +7	roo +7	roo +7	roo +7	roo +175	roo +278	roo -19	roo -19	roo -23	roo -44	roo -44	roo -68	roo -68	roo -68	roo-90	roo -90	roo -90	roo -90	roo -90	roo -90	roo -90
Position	Reference genome	RAL-802	RAL-810	IV68	RAL-405	RAL-911	RAL-441	RAL-801	IV145	RAL-502	IV42	IV127	IV40	RAL-195	RAL-383	RAL-75	RAL-716	IV69	RAL-21	RAL-88	RAL-177	RAL-737	RAL-820	RAL-857	IV50
171	9267	9267	9267	6337	9267	9267	9221	9267	9267		9267	9267	9267	9267	9166	9267	9267	9267	9267	9267	9267	9267	9267		9267
172	9221	9221	9221	6085	9221	9221	9161	9221	9221		9221	9221	9221	9221	9097	9221	9221	9221	9221	9221	9221	9221	9221		9221
173	9261	9261	9261	5889	9261	9261	9164	9261	9261		9261	9261	9261	9261	9119	9261	9261	9261	9261	9261	9261	9261	9261		9261
174	9228	9228	9226	5707	9228	9228	9093	9228	9228		9228	9228	9228	9228	9073	9228	9228	9228	9228	9228	9228	9228	9228		9228
175	9280	9280	9278	5505	9280	9280	9089	9280	9280		9280	9280	9280	9280	9103	9280	9280	9280	9280	9280	9280	9280	9280		9280
176	9262	9262	9255	5316	9262	9262	9038	9262	9262		9262	9262	9262	9262	9080	9262	9262	9262	9262	9262	9262	9262	9262		9262
177	9218	9218	9207	5081	9218	9218	8954	9218	9218		9218	9218	9218	9218	9026	9218	9218	9218	9218	9218	9218	9218	9218		9218
178	9187	9187	9172	4894	9187	9187	8880	9187	9187		9187	9187	9187	9187	8984	9187	9187	9187	9187	9187	9187	9187	9187		9187
179	9172	9172	9150	4667	9172	9172	8823	9172	9172		9172	9172	9172	9172	8953	9172	9172	9172	9172	9172	9172	9172	9172		9172
180	9091	9091	9068	4454	9091	9091	8710	9091	9091		9091	9091	9091	9091	8867	9091	9091	9091	9091	9091	9091	9091	9091		9091
181	9137	9137	9111	4243	9137	9137	8564	9137	9137		9137	9137	9137	9137	8899	9137	9137	9137	9137	9137	9137	9137	9137		9137
182	9118	9118	9086	4037	9118	9118	8440	9118	9118		9118	9118	9118	9118	8875	9118	9118	9118	9118	9118	9118	9118	9118		9118
183	9172	9172	9131	3820	9172	9172	8324	9172	9172		9172	9172	9172	9172	8915	9172	9172	9172	9172	9172	9172	9172	9172		9172
184	9206	9206	9160	3641	9206	9206	8264	9206	9206		9206	9206	9206	9206	8942	9206	9206	9206	9206	9206	9206	9206	9206		9206
185	9223	9223	9175	3444	9223	9223	8171	9223	9223		9223	9223	9223	9223	8884	9223	9223	9223	9223	9223	9223	9223	9223		9223
186	9183	9183	9134	3177	9183	9183	7984	9183	9183		9183	9183	9183	9183	8714	9183	9183	9183	9183	9183	9183	9183	9183		9183
187	9165	9165	9113	2815	9165	9165	7839	9165	9165		9165	9165	9165	9165	8685	9165	9165	9165	9165	9165	9165	9165	9165		9165
188	9091	9091	9029	2554	9091	9091	7657	9091	9091		9091	9091	9091	9091	8598	9091	9091	9091	9091	9091	9091	9091	9091		9091
189	9011	9011	8942	2348	9011	9011	7502	9011	9011		9011	9011	9011	9011	8507	9011	9011	9011	9011	9011	9011	9011	9011		9011
190	8954	8954	8870	2128	8954	8954	7339	8954	8954		8954	8954	8954	8954	8442	8954	8954	8954	8954	8954	8954	8954	8954		8954
191	9024	9024	8933	1957	9024	9024	7264	9024	9024		9024	9024	9024	9024	8493	9024	9024	9024	9024	9024	9024	9024	9024		9024
192	8863	8863	8763	1681	8863	8863	7015	8863	8863		8863	8863	8863	8863	8318	8863	8863	8863	8863	8863	8863	8863	8863		8863
193	8809	8809	8699	284	8809	8809	6840	8809	8809		8809	8809	8809	8809	8245	8809	8809	8809	8809	8809	8809	8809	8809		8809
194	8832	8832	8708	107	8832	8832	6750	8832	8832		8832	8832	8832	8832	8247	8832	8832	8832	8832	8832	8832	8832	8832		8832
195	8814	8814	8687	4	8814	8814	6657	8814	8814		8814	8814	8814	8814	8223	8814	8814	8814	8814	8814	8814	8814	8814		8814
196	8820	8820	8689	0	8820	8820	6579	8820	8820		8820	8820	8820	8820	8226	8820	8820	8820	8820	8820	8820	8820	8820		8820
197	8876	8876	8736	0	8876	8876	6535	8876	8876		8876	8876	8876	8876	8274	8876	8876	8876	8876	8876	8876	8876	8876		8876
198	8901	8901	8758	0	8901	8901	6483	8901	8901		8901	8901	8901	8901	8293	8901	8901	8901	8901	8901	8901	8901	8901		8901
199	8965	8965	8816	202	8965	8965	6457	8965	8965		8965	8965	8965	8965	8330	8965	8965	8965	8965	8965	8965	8965	8965		8965
200	8842	8842	8691	2314	8842	8842	6133	8842	8842		8842	8842	8842	8842	8018	8842	8842	8842	8842	8842	8842	8842	8842		8842
201	8729	8729	8571	2841	8729	8729	5907	8729	8729		8729	8729	8729	8729	7884	8729	8729	8729	8729	8729	8729	8729	8729		8729
202	9016	9016	8843	3390	9016	9016	5896	9016	9016		9016	9016	9016	9016	8058	9016	9016	9016	9016	9016	9016	9016	9016		9016
203	8954	8954	8769	3648	8954	8954	5741	8954	8954		8954	8954	8954	8954	7976	8954	8954	8954	8954	8954	8954	8954	8954		8954
204	8928	8928	8741	3834	8928	8928	5648	8928	8928		8928	8928	8928	8928	8164	8928	8928	8928	8928	8928	8928	8928	8928		8928
205	8911	8911	8638	4064	8911	8911	5402	8911	8911		8911	8911	8911	8911	8193	8911	8911	8911	8911	8911	8911	8911	8911		8911
206	8944	8944	8545	4325	8944	8944	5173	8944	8944		8944	8944	8944	8944	8270	8944	8944	8944	8944	8944	8944	8944	8944		8944
207	8900	8900	8495	4481	8900	8900	4973	8900	8900		8900	8900	8900	8900	8251	8900	8900	8900	8900	8900	8900	8900	8900		8900
208	8899	8899	8487	4676	8899	8899	4778	8899	8899		8899	8899	8899	8899	8275	8899	8899	8899	8899	8899	8899	8899	8899		8899
209	8918	8918	8495	4931	8918	8918	4544	8918	8918		8918	8918	8918	8918	8317	8918	8918	8918	8918	8918	8918	8918	8918		8918
210	9041	9041	8611	5222	9041	9041	3593	9041	9041		9041	9041	9041	9041	8462	9041	9041	9041	9041	9041	9041	9041	9041		9041
211	9048	9048	8610	5489	9048	9048	3358	9048	9048		9048	9048	9048	9048	8505	9048	9048	9048	9048	9048	9048	9048	9048		9048
212	9118	9118	8670	5719	9118	9118	3215	9118	9118		9118	9118	9118	9118	8593	9118	9118	9118	9118	9118	9118	9118	9118		9118
213	9144	9144	8684	5946	9144	9144	3044	9144	9144		9144	9144	9144	9144	8651	9144	9144	9144	9144	9144	9144	9144	9144		9144
214	9187	9187	8721	6129	9187	9187	3230	9187	9187		9187	9187	9187	9187	8704	9187	9187	9187	9187	9187	9187	9187	9187		9187
215	9212	9212	8737	6333	9212	9212	5189	9212	9212		9212	9212	9212	9212	8740	9212	9212	9212	9212	9212	9212	9212	9212		9212
216	9296	9296	8816	6523	9296	9296	5392	9296	9296		9296	9296	9296	9296	8835	9296	9296	9296	9296	9296	9296	9296	9296		9296
217	9246	9246	8757	6769	9246	9246	5638	9246	9246		9246	9246	9246	9246	8815	9246	9246	9246	9246	9246	9246	9246	9246		9246
218	9290	9290	8792	6983	9290	9290	5868	9290	9290		9290	9290	9290	9290	8866	9290	9290	9290	9290	9290	9290	9290	9290		9290
219	9296	9296	8783	7245	9296	9296	6132	9296	9296		9296	9296	9296	9296	8897	9296	9296	9296	9296	9296	9296	9296	9296		9296
220	9322	9322	8801	7436	9322	9322	6326	9322	9322		9322	9322	9322	9322	8944	9322	9322	9322	9322	9322	9322	9322	9322		9322
221	9337	9337	8810	7652	9337	9337	6440	9337	9337		9337	9337	9337	9337	9045	9337	9337	9337	9337	9337	9337	9337	9337		9337
222	9320	9320	8780	7877	9320	9320	6573	9320	9320		9320	9320	9320	9320	9166	9320	9320	9320	9320	9320	9320	9320	9320		9320
223	9291	9291	8741	8051	9291	9291	6651	9291	9291		9291	9291	9291	9291	9145	9291	9291	9291	9291	9291	9291	9291	9291		9291
224	9345	9345	8810	8300	9345	9345	6846	9345	9345		9345	9345	9345	9345	9207	9345	9345	9345	9345	9345	9345	9345	9345		9345
225	9358	9358	8837	8473	9358	9358	6956	9358	9358		9358	9358	9358	9358	9230	9358	9358	9358	9358	9358	9358	9358	9358		9358
226	9393	9393	8880	8672	9393	9393	7097	9393	9393		9393	9393	9393	9393	9273	9393	9393	9393	9393	9393	9393	9393	9393		9393
227	9312	9312	8806	8796	9312	9312	7156	9312	9312		9312	9312	9312	9312	9198	9312	9312	9312	9312	9312	9312	9312	9312		9312
228	9363	9363	8866	9001	9363	9363	7296	9363	9363		9363	9363	9363	9363	9254	9363	9363	9363	9363	9363	9363	9363	9363		9363
229	9381	9381	8897	9250	9381	9381	7452	9381	9381		9381	9381	9381	9381	9285	9381	9381	9381	9381	9381	9381	9381	9381		9381
230	9348	9348	8873	9248	9348	9348	7529	9348	9348		9348	9348	9348	9348	9261	9348	9348	9348	9348	9348	9348	9348	9348		9348

Position	FBti0019985 Reference genome	FBti0019985 RAL-802	FBti0019985 RAL-810	FBti0019985 IV68	roo +7 RAL-405	roo +7 RAL-911	roo +7 RAL-441	roo +7 RAL-801	roo +175 IV145	roo +278 RAL-502	roo -19 IV42	roo -19 IV127	roo -23 IV40	roo -44 RAL-195	roo -44 RAL-383	roo -68 RAL-75	roo -68 RAL-716	roo -68 IV69	roo-90 RAL-21	roo -90 RAL-88	roo -90 RAL-177	roo -90 RAL-737	roo -90 RAL-820	roo -90 RAL-857	roo -90 IV50
231	9421	9421	8959	9336	9421	9421	7703	9421	9421		9421	9421	9421	9421	9346	9421	9421	9421	9421	9421	9421	9421	9421		9421
232	9490	9490	9037	9419	9490	9490	7893	9490	9490		9490	9490	9490	9490	9421	9490	9490	9490	9490	9490	9490	9490	9490		9490
233	9424	9424	8984	9364	9424	9424	7943	9424	9424		9424	9424	9424	9424	9368	9424	9424	9424	9424	9424	9424	9424	9424		9424
234	9532	9532	9103	9485	9532	9532	8139	9532	9532		9532	9532	9532	9532	9481	9532	9532	9532	9532	9532	9532	9532	9532		9532
235	9518	9518	9097	9477	9518	9518	8237	9518	9518		9518	9518	9518	9518	9470	9518	9518	9518	9518	9518	9518	9518	9518		9518
236	9628	9628	9218	9594	9628	9628	8453	9628	9628		9628	9628	9628	9628	9587	9628	9628	9628	9628	9628	9628	9628	9628		9628
237	9668	9668	9268	9638	9668	9668	8609	9668	9668		9668	9668	9668	9668	9634	9668	9668	9668	9668	9668	9668	9668	9668		9668
238	9542	9542	9148	9519	9542	9542	8704	9542	9542		9542	9542	9542	9542	9515	9542	9542	9542	9542	9542	9542	9542	9542		9542
239	9578	9578	9205	9559	9578	9578	8844	9578	9578		9578	9578	9578	9578	9555	9578	9578	9578	9578	9578	9578	9578	9578		9578
240	9624	9624	9255	9606	9624	9624	8971	9624	9624		9624	9624	9624	9624	9603	9624	9624	9624	9624	9624	9624	9624	9624		9624
241	9592	9592	9317	9580	9592	9592	9062	9592	9592		9592	9592	9592	9592	9576	9592	9592	9592	9592	9592	9592	9592	9592		9592
242	9607	9607	9457	9599	9607	9607	9235	9607	9607		9607	9607	9607	9607	9595	9607	9607	9607	9607	9607	9607	9607	9607		9607
243	9559	9559	9413	9553	9559	9559	9293	9559	9559		9559	9559	9559	9559	9550	9559	9559	9559	9559	9559	9559	9559	9559		9559
244	9455	9455	9317	9453	9455	9455	9300	9455	9455		9455	9455	9455	9455	9450	9455	9455	9455	9455	9455	9455	9455	9455		9455
245	9415	9415	9289	9414	9415	9415	9291	9415	9415	2243	9415	9415	9415	9415	9412	9415	9415	9415	9415	9415	9415	9415	9415	2251	9415
246	9266	9266	9150	9265	9266	9266	9154	9266	9266	2393	9266	9266	9266	9266	9263	9266	9266	9266	9266	9266	9266	9266	9266	2349	9266
247	9085	9085	8983	9085	9085	9085	9002	9085	9085	2544	9085	9085	9085	9085	9084	9085	9085	9085	9085	9085	9085	9085	9085	2569	9085
248	9053	9053	8960	9053	9053	9053	8991	9053	9053	2752	9053	9053	9053	9041	9041	9053	9041	9041	9053	9053	9053	9053	9053	2719	9053
249	8921	8921	8838	8921	8921	8921	8880	8921	8921	2906	8921	8921	8921	8895	8895	8921	8895	8895	8921	8921	8921	8921	8921	2868	8921
250	8995	8995	8920	8995	8995	8995	8963	8995	8995	3186	8995	8995	8995	8944	8944	8995	8944	8944	8995	8995	8995	8995	8995	3076	8995
251	9015	9015	8949	9015	9015	9015	8987	9015	9015	3416	9015	9015	9015	8942	8942	9015	8942	8942	9015	9015	9015	9015	9015	3229	9015
252	8940	8940	8882	8940	8940	8940	8915	8940	8940	3585	8940	8940	8940	8844	8844	8940	8844	8844	8940	8940	8940	8940	8940	3505	8940
253	8892	8892	8840	8892	8892	8892	8871	8892	8892	3853	8892	8892	8892	8782	8782	8892	8782	8782	8892	8892	8892	8892	8892	3735	8892
254	8827	8827	8782	8827	8827	8827	8809	8827	8827	3994	8827	8827	8827	8706	8706	8827	8706	8706	8827	8827	8827	8827	8827	3903	8827
255	8760	8760	8722	8760	8760	8760	8745	8760	8760	4189	8760	8760	8760	8627	8627	8760	8627	8627	8760	8760	8760	8760	8760	4168	8760
256	8669	8669	8635	8669	8669	8669	8656	8669	8669	4311	8669	8669	8669	8522	8522	8669	8522	8522	8669	8669	8669	8669	8669	4307	8669
257	8655	8655	8625	8655	8655	8655	8646	8655	8655	4474	8655	8655	8655	8487	8487	8655	8487	8487	8655	8655	8655	8655	8655	4501	8655
258	8656	8656	8629	8656	8656	8656	8648	8656	8656	4686	8656	8656	8656	8470	8470	8656	8470	8470	8656	8656	8656	8656	8656	4622	8656
259	8672	8672	8647	8672	8672	8672	8665	8672	8672	4882	8672	8672	8672	8470	8470	8672	8470	8470	8672	8672	8672	8672	8672	4782	8672
260	8588	8588	8564	8588	8588	8588	8582	8588	8588	5030	8588	8588	8588	8382	8382	8588	8382	8382	8588	8588	8588	8588	8588	4982	8588
261	8557	8557	8535	8557	8557	8557	8551	8557	8557	5176	8557	8557	8557	8337	8337	8557	8337	8337	8557	8554	8557	8557	8557	5172	8557
262	8582	8582	8565	8582	8582	8582	8580	8582	8582	5418	8582	8582	8582	8335	8335	8582	8335	8335	8582	8574	8577	8582	8582	5320	8582
263	8598	8598	8584	8598	8598	8598	8598	8598	8598	5564	8598	8598	8598	8325	8325	8598	8325	8325	8598	8579	8592	8598	8598	5457	8598
264	8680	8680	8666	8680	8680	8680	8680	8680	8680	5849	8680	8680	8680	8371	8371	8680	8371	8371	8680	8651	8670	8680	8680	5670	8680
265	8737	8737	8724	8737	8737	8737	8737	8737	8737	6142	8737	8737	8737	8395	8395	8737	8395	8395	8737	8698	8722	8737	8737	5790	8737
266	8774	8774	8762	8774	8774	8774	8774	8774	8774	6322	8774	8774	8774	8423	8423	8774	8423	8423	8774	8729	8758	8774	8774	6037	8774
267	8800	8800	8789	8800	8800	8800	8800	8800	8800	6524	8800	8800	8800	8436	8436	8800	8436	8436	8800	8744	8779	8800	8800	6301	8800
268	8704	8704	8697	8704	8704	8704	8704	8704	8704	6682	8704	8704	8704	8335	8335	8704	8335	8335	8704	8626	8677	8704	8704	6464	8704
269	8705	8705	8700	8705	8705	8705	8705	8705	8705	6884	8705	8705	8705	8327	8327	8705	8327	8327	8705	8599	8665	8705	8705	6585	8705
270	8616	8616	8612	8616	8616	8616	8616	8616	8616	7048	8616	8616	8616	8234	8234	8616	8234	8234	8616	8491	8571	8616	8616	6688	8616
271	8566	8566	8566	8566	8566	8566	8566	8566	8566	7205	8566	8566	8566	8178	8178	8566	8178	8178	8566	8414	8511	8566	8566	6820	8566
272	8526	8526	8526	8526	8526	8526	8526	8526	8526	7447	8526	8526	8526	8129	8129	8526	8129	8129	8526	8343	8460	8526	8526	6925	8526
273	8443	8443	8443	8443	8443	8443	8443	8443	8443	7611	8443	8443	8443	8041	8041	8443	8041	8041	8443	8222	8377	8443	8443	7022	8443
274	8463	8463	8463	8463	8463	8463	8463	8463	8463	7822	8463	8463	8463	8044	8044	8463	8044	8044	8463	8186	8378	8463	8463	7022	8463
275	8427	8427	8427	8427	8427	8427	8427	8427	8427	7833	8427	8427	8427	7997	7997	8427	7997	7997	8427	8106	8336	8427	8427	6872	8427
276	8443	8443	8443	8443	8443	8443	8443	8443	8443	7930	8443	8443	8443	7988	7988	8443	7988	7988	8443	8043	8349	8443	8443	6730	8443
277	8431	8431	8431	8431	8431	8431	8431	8431	8431	7957	8431	8431	8431	7962	7962	8431	7962	7962	8431	7986	8310	8431	8431	6610	8431
278	8349	8349	8349	8349	8349	8349	8349	8349	8349	7923	8349	8349	8349	7854	7854	8349	7854	7854	8349	7822	8203	8349	8349	6488	8349
279	8370	8370	8370	8370	8370	8370	8370	8370	8370	8004	8370	8370	8370	7832	7832	8370	7832	7832	8370	7728	8197	8370	8370	6367	8370
280	8437	8437	8437	8437	8437	8437	8437	8437	8437	8101	8437	8437	8437	7883	7883	8437	7883	7883	8437	7691	8227	8437	8437	6164	8437
281	8341	8341	8341	8341	8341	8341	8341	8341	8341	8056	8341	8341	8341	7784	7784	8341	7784	7784	8341	7467	8089	8341	8341	6039	8341
282	8390	8390	8390	8390	8390	8390	8390	8390	8390	8146	8390	8390	8390	7808	7808	8390	7808	7808	8390	7361	8077	8390	8390	5944	8390
283	8373	8373	8373	8373	8373	8373	8373	8373	8373	8147	8373	8373	8373	7781	7781	8373	7781	7781	8373	7266	8035	8373	8373	5721	8373
284	8284	8284	8284	8284	8284	8284	8284	8284	8284	8091	8284	8284	8284	7684	7684	8284	7684	7684	8284	7079	7927	8284	8284	5563	8284
285	8351	8351	8351	8351	8351	8351	8351	8351	8351	8185	8351	8351	8351	7737	7737	8351	7737	7737	8351	7003	7937	8351	8351	5418	8351
286	8271	8271	8271	8271	8271	8271	8271	8271	8271	8135	8271	8271	8271	7655	7655	8271	7655	7655	8271	6827	7832	8271	8271	5187	8271
287	8287	8287	8287	8287	8287	8287	8287	8287	8287	8174	8287	8287	8287	7666	7666	8287	7666	7666	8287	6715	7812	8287	8287	5041	8287
288	8374	8374	8374	8374	8374	8374	8374	8374	8374	8286	8374	8374	8374	7741	7741	8374	7741	7741	8374	6643	7840	8374	8374	4822	8374
289	8229	8229	8229	8229	8229	8229	8229	8229	8229	8173	8229	8229	8229	7584	7584	8229	7584	7584	8229	6406	7669	8229	8229	4641	8229
290	8244	8244	8244	8244	8244	8244	8244	8244	8244	8215	8244	8244	8244	7594	7594	8244	7594	7594	8244	6314	7655	8244	8244	4493	8244

Position	FBti0019985 Reference genome	FBti0019985 RAL-802	FBti0019985 RAL-810	FBti0019985 1V68	TOO_+7 RAL-405	TOO_+7 RAL-911	TOO_+7 RAL-441	TOO_+7 RAL-801	TOO_+175 1V745	TOO_+278 RAL-502	TOO_-9 1V42	TOO_-19 1V127	TOO_-23 1V40	TOO_-44 RAL-195	TOO_-44 RAL-383	TOO_-48 RAL-75	TOO_-48 RAL-716	TOO_48 1V69	TOO-90 RAL-21	TOO_-90 RAL-88	TOO_90 RAL-177	TOO_90 RAL-737	TOO_-90 RAL-820	TOO_90 RAL-857	TOO_90 1V50
291	8242	8242	8242	8242	8242	8242	8242	8242	8242	8233	8242	8242	8242	7583	7583	8242	7583	7583	8242	6169	7607	8242	8242	4210	8242
292	8242	8242	8242	8242	8242	8242	8242	8242	8242	8234	8242	8242	8242	7579	7579	8242	7579	7579	8242	6112	7593	8242	8242	4055	8242
293	8297	8297	8297	8297	8297	8297	8297	8297	8297	8289	8297	8297	8297	7627	7627	8297	7627	7627	8297	6065	7563	8297	8297	3856	8297
294	8367	8367	8367	8367	8367	8367	8367	8367	8367	8361	8367	8367	8367	7680	7680	8367	7680	7680	8367	6032	7547	8367	8367	3721	8367
295	8367	8367	8367	8367	8367	8367	8367	8367	8367	8363	8367	8367	8367	7667	7667	8367	7667	7667	8367	5922	7520	8367	8367	3595	8367
296	8415	8415	8415	8415	8415	8415	8415	8411	8415	8415	8415	8415	8415	7708	7708	8415	7708	7708	8415	5825	7554	8415	8415	3431	8415
297	8433	8433	8433	8433	8433	8433	8433	8433	8433	8429	8433	8433	8433	7734	7734	8433	7734	7734	8433	5748	7560	8433	8433	3237	8433
298	8473	8473	8473	8473	8473	8473	8473	8473	8473	8472	8473	8473	8473	7789	7789	8473	7789	7789	8473	5650	7568	8473	8473	2873	8473
299	8516	8516	8516	8516	8516	8516	8516	8516	8516	8516	8516	8516	8516	7873	7873	8516	7873	7873	8516	5611	7607	8516	8516	2707	8516
300	8421	8421	8421	8421	8421	8421	8421	8421	8421	8421	8421	8421	8421	7806	7806	8421	7806	7806	8421	5404	7504	8421	8421	2529	8421
301	8304	8304	8304	8304	8304	8304	8304	8304	8304	8304	8304	8304	8304	7719	7719	8304	7719	7719	8304	5191	7378	8304	8304	2376	8304
302	8439	8439	8439	8439	8439	8439	8439	8439	8439	8439	8439	8439	8439	7883	7883	8439	7883	7883	8439	5129	7497	8439	8439	2069	8439
303	8370	8370	8370	8370	8370	8370	8370	8370	8370	8370	8370	8370	8370	7836	7836	8370	7836	7836	8370	4967	7414	8370	8370	199	8370
304	8537	8537	8537	8537	8537	8537	8537	8537	8537	8537	8537	8537	8537	8012	8012	8537	8012	8012	8537	4827	7571	8537	8537	0	8537
305	8539	8539	8539	8539	8539	8539	8539	8539	8539	8539	8539	8539	8539	8029	8029	8539	8029	8029	8539	4643	7566	8539	8539	0	8539
306	8564	8564	8564	8564	8564	8564	8564	8564	8564	8564	8564	8564	8564	8068	8068	8564	8068	8068	8564	4488	7585	8564	8564	0	8564
307	8711	8711	8711	8711	8711	8711	8711	8711	8711	8711	8711	8711	8711	8237	8237	8711	8237	8237	8711	3584	7693	8711	8711	1	8711
308	8749	8749	8749	8749	8749	8749	8749	8749	8749	8749	8749	8749	8749	8294	8294	8749	8294	8294	8749	3599	7711	8749	8749	240	8749
309	8805	8805	8805	8805	8805	8805	8805	8805	8805	8805	8805	8805	8805	8371	8371	8805	8371	8371	8805	3428	7740	8805	8805	482	8805
310	8904	8904	8904	8904	8904	8904	8904	8904	8904	8904	8904	8904	8904	8478	8478	8904	8478	8478	8904	3305	7821	8904	8904	2931	8904
311	8964	8964	8964	8964	8964	8964	8964	8964	8964	8964	8964	8964	8964	8560	8560	8964	8560	8560	8964	3480	7842	8964	8964	3126	8964
312	8840	8840	8840	8840	8840	8840	8840	8840	8840	8840	8840	8840	8840	8485	8485	8840	8485	8485	8840	3652	7756	8840	8840	3327	8840
313	8905	8905	8905	8905	8905	8905	8905	8905	8905	8905	8905	8905	8905	8583	8583	8905	8583	8583	8905	3747	7848	8905	8905	3547	8905
314	9008	9008	9008	9008	9008	9008	9008	9008	9008	9008	9008	9008	9008	8739	8739	9008	8739	8739	9008	5002	7965	9008	9008	3837	9008
315	9030	9030	9030	9030	9030	9030	9030	9030	9030	9030	9030	9030	9030	8822	8822	9030	8822	8822	9030	5206	7999	9030	9030	4039	9030
316	9082	9082	9082	9082	9082	9082	9082	9082	9082	9082	9082	9082	9082	8882	8882	9082	8882	8882	9082	5436	8080	9082	9082	4248	9082
317	9098	9098	9098	9098	9098	9098	9098	9098	9098	9098	9098	9098	9098	8905	8905	9098	8905	8905	9098	5650	8133	9098	9098	4459	9098
318	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	8932	8932	9110	8932	8932	9110	5786	8188	9110	9110	4657	9110
319	9167	9167	9167	9167	9167	9167	9167	9167	9167	9167	9167	9167	9167	9000	9000	9167	9000	9000	9167	5939	8275	9167	9167	4889	9167
320	9192	9192	9192	9192	9192	9192	9192	9192	9192	9192	9192	9192	9192	9035	9035	9192	9035	9035	9192	6074	8321	9192	9192	5099	9192
321	9121	9121	9121	9121	9121	9121	9121	9121	9121	9121	9121	9121	9121	8983	8983	9121	8983	8983	9121	6164	8315	9121	9121	5247	9121
322	9071	9071	9071	9071	9071	9071	9071	9071	9071	9071	9071	9071	9071	8946	8946	9071	8946	8946	9071	6238	8300	9071	9071	5413	9071
323	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	8963	8963	9080	8963	8963	9080	6374	8340	9080	9080	5596	9080
324	9009	9009	9009	9009	9009	9009	9009	9009	9009	9009	9009	9009	9009	8918	8918	9009	8918	8918	9009	6484	8339	9009	9009	5804	9009
325	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9080	9003	9003	9080	9003	9003	9080	6646	8431	9080	9080	6011	9080
326	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9110	9043	9043	9110	9043	9043	9110	6774	8495	9110	9110	6204	9110
327	9120	9120	9120	9120	9120	9120	9120	9120	9120	9120	9120	9120	9120	9069	9069	9120	9069	9069	9120	6929	8558	9120	9120	6424	9120
328	9132	9132	9132	9132	9132	9132	9132	9132	9132	9132	9132	9132	9132	9084	9084	9132	9084	9084	9132	7022	8608	9132	9132	6596	9132
329	9091	9091	9091	9091	9091	9091	9091	9091	9091	9091	9091	9091	9091	9052	9052	9090	9052	9052	9091	7104	8676	9091	9091	6764	9091
330	9112	9112	9112	9112	9112	9112	9112	9112	9111	9112	9112	9112	9112	9080	9080	9109	9080	9080	9112	7241	8803	9112	9112	6989	9112
331	9163	9163	9163	9163	9163	9163	9163	9163	9162	9163	9163	9163	9163	9133	9133	9160	9133	9133	9163	7405	8883	9163	9163	7203	9163
332	9119	9119	9119	9119	9119	9119	9119	9119	9118	9119	9119	9119	9119	9100	9100	9115	9100	9100	9119	7526	8847	9119	9119	7416	9119
333	9130	9130	9130	9130	9130	9130	9130	9130	9129	9130	9130	9130	9130	9114	9114	9125	9114	9114	9130	7621	8871	9130	9130	7574	9130
334	9100	9100	9100	9100	9100	9100	9100	9100	9099	9100	9100	9100	9100	9089	9089	9094	9089	9089	9100	7697	8853	9100	9100	7721	9100
335	9084	9084	9084	9084	9084	9084	9084	9083	9084	9084	9084	9084	9084	9075	9075	9077	9075	9075	9084	7780	8849	9084	9084	7901	9084
336	9061	9061	9061	9061	9061	9061	9061	9061	9060	9061	9061	9061	9061	9051	9051	9052	9051	9051	9061	7861	8835	9061	9061	8054	9061
337	9004	9004	9004	9004	9004	9004	9004	9004	9001	9004	9004	9004	9004	8995	8995	8993	8995	8995	9004	7927	8802	9004	9004	8211	9004
338	8849	8849	8849	8849	8849	8849	8849	8849	8845	8849	8849	8849	8849	8840	8840	8837	8840	8840	8849	7970	8669	8849	8849	8355	8849
339	8845	8845	8845	8845	8845	8845	8845	8845	8839	8845	8845	8845	8845	8838	8838	8832	8838	8838	8845	8059	8679	8845	8845	8522	8845
340	8747	8747	8747	8747	8747	8747	8747	8747	8747	8747	8747	8747	8747	8741	8741	8733	8741	8741	8747	8145	8591	8747	8747	8471	8747
341	8737	8737	8737	8737	8737	8737	8737	8737	8731	8737	8737	8737	8737	8732	8732	8722	8732	8732	8737	8259	8592	8737	8737	8497	8737
342	8835	8835	8835	8835	8835	8835	8835	8835	8826	8835	8835	8835	8835	8830	8830	8805	8830	8830	8835	8439	8694	8835	8835	8633	8835
343	8745	8745	8745	8745	8745	8745	8745	8745	8728	8745	8745	8745	8745	8740	8740	8707	8740	8740	8745	8522	8628	8745	8745	8593	8745
344	8564	8564	8564	8564	8564	8564	8564	8564	8549	8564	8564	8564	8564	8561	8561	8528	8561	8561	8564	8405	8481	8564	8564	8495	8564
345	8528	8528	8528	8528	8528	8528	8528	8528	8508	8528	8528	8528	8528	8526	8526	8487	8526	8526	8528	8407	8469	8528	8528	8475	8528
346	8513	8513	8513	8513	8513	8513	8513	8513	8484	8513	8513	8513	8513	8512	8512	8466	8512	8512	8513	8417	8475	8513	8513	8472	8513
347	8552	8552	8552	8552	8552	8552	8552	8552	8512	8552	8552	8552	8552	8552	8552	8475	8552	8552	8552	8483	8520	8552	8552	8520	8552
348	8556	8556	8556	8556	8556	8556	8556	8556	8556	8503	8556	8556	8556	8556	8556	8449	8556	8556	8556	8516	8539	8556	8556	8537	8556
349	8505	8505	8505	8505	8505	8505	8505	8505	8426	8505	8505	8505	8505	8505	8505	8374	8505	8505	8505	8471	8492	8505	8505	8491	8505
350	8430	8430	8430	8430	8430	8430	8430	8430	8329	8430	8430	8430	8430	8430	8430	8277	8430	8430	8430	8402	8421	8430	8430	8420	8430

	FBti0019985	FBti0019985	FBti0019985	FBti0019985	roo +7	roo +7	roo +7	roo +7	roo +175	roo +278	roo -19	roo -19	roo -23	roo -44	roo -44	roo -68	roo -68	roo -68	roo-90	roo -90	roo -90	roo -90	roo -90	roo -90	roo -90
Position	*Reference genome*	*RAL-802*	*RAL-810*	*IV68*	*RAL-405*	*RAL-911*	*RAL-441*	*RAL-801*	*IV145*	*RAL-502*	*IV42*	*IV127*	*IV40*	*RAL-195*	*RAL-383*	*RAL-75*	*RAL-716*	*IV69*	*RAL-21*	*RAL-88*	*RAL-177*	*RAL-737*	*RAL-820*	*RAL-857*	*IV50*
351	8393	8393	8393	8393	8393	8393	8393	8393	8285	8393	8393	8393	8393	8393	8393	8231	8393	8393	8393	8371	8384	8393	8393	8384	8393
352	8363	8363	8363	8363	8363	8363	8363	8363	8230	8363	8363	8363	8363	8363	8363	8170	8363	8363	8363	8344	8356	8363	8363	8356	8363
353	8373	8373	8373	8373	8373	8373	8373	8373	8228	8373	8373	8373	8373	8373	8373	8156	8373	8373	8373	8360	8367	8373	8373	8368	8373
354	8316	8316	8316	8316	8316	8316	8316	8316	8167	8316	8316	8316	8316	8316	8316	8092	8316	8316	8316	8307	8311	8316	8316	8314	8316
355	8324	8324	8324	8324	8324	8324	8324	8324	8157	8324	8324	8324	8324	8324	8324	8086	8324	8324	8324	8317	8320	8324	8324	8323	8324
356	8287	8287	8287	8287	8287	8287	8287	8287	8109	8287	8287	8287	8287	8287	8287	8038	8287	8287	8287	8281	8284	8287	8287	8286	8287
357	8251	8251	8251	8251	8251	8251	8251	8251	8064	8251	8251	8251	8251	8251	8251	7993	8251	8251	8251	8246	8248	8251	8251	8250	8251
358	8325	8325	8325	8325	8325	8325	8325	8325	8132	8325	8325	8325	8325	8325	8325	7984	8325	8325	8325	8321	8322	8325	8325	8324	8325
359	8267	8267	8267	8267	8267	8267	8267	8267	8042	8267	8267	8267	8267	8267	8267	7879	8267	8267	8267	8264	8265	8267	8267	8267	8267
360	8188	8188	8186	8188	8188	8188	8188	8188	7914	8188	8188	8188	8188	8188	8188	7799	8188	8188	8188	8187	8188	8188	8188	8188	8188
361	8212	8212	8205	8212	8212	8212	8212	8212	7935	8212	8212	8212	8212	8212	8212	7809	8212	8212	8212	8211	8212	8212	8212	8212	8212
362	8181	8181	8170	8181	8181	8181	8181	8181	7892	8181	8181	8181	8181	8181	8181	7771	8181	8181	8181	8180	8181	8181	8181	8181	8181
363	8218	8218	8205	8218	8218	8218	8218	8218	7910	8218	8218	8218	8218	8218	8218	7787	8218	8218	8218	8218	8218	8218	8218	8218	8218
364	8165	8165	8151	8165	8165	8165	8165	8165	7852	8165	8165	8165	8165	8165	8165	7720	8165	8165	8165	8165	8165	8165	8165	8165	8165
365	8137	8137	8116	8137	8137	8137	8137	8137	7820	8137	8137	8137	8137	8137	8137	7684	8137	8137	8137	8137	8137	8137	8137	8137	8137
366	8103	8103	8075	8103	8103	8103	8103	8103	7781	8103	8103	8103	8103	8103	8103	7629	8103	8103	8103	8103	8103	8103	8103	8103	8103
367	8037	8037	8004	8037	8037	8037	8037	8037	7710	8037	8037	8037	8037	8037	8037	7547	8037	8037	8037	8037	8037	8037	8037	8037	8037
368	8058	8058	8015	8058	8058	8058	8058	8058	7720	8058	8058	8058	8058	8058	8058	7550	8058	8058	8058	8058	8058	8058	8058	8058	8058
369	8049	8049	7999	8049	8049	8049	8049	8049	7705	8049	8049	8049	8049	8049	8049	7530	8049	8049	8049	8049	8049	8049	8049	8049	8049
370	8038	8038	7977	8038	8038	8038	8038	8038	7688	8038	8038	8038	8038	8038	8038	7514	8038	8038	8038	8038	8038	8038	8038	8038	8038
371	8058	8058	7994	8058	8058	8058	8058	8058	7708	8058	8058	8058	8058	8058	8058	7515	8058	8058	8058	8058	8058	8058	8058	8058	8058
372	8022	8022	7954	8022	8022	8022	8022	8022	7673	8022	8022	8022	8022	8022	8022	7477	8022	8022	8022	8022	8022	8022	8022	8022	8022
373	8010	8010	7932	8010	8010	8010	8010	8010	7645	8010	8010	8010	8010	8010	8010	7441	8010	8010	8010	8010	8010	8010	8010	8010	8010
374	8037	8037	7946	8037	8037	8037	8037	8037	7655	8037	8037	8037	8037	8037	8037	7440	8037	8037	8037	8037	8037	8037	8037	8037	8037
375	8020	8020	7921	8020	8020	8020	8020	8020	7611	8020	8020	8020	8020	8020	8020	7391	8020	8020	8020	8020	8020	8020	8020	8020	8020
376	7993	7993	7873	7993	7993	7993	7993	7993	7479	7993	7993	7993	7993	7993	7993	7164	7993	7993	7993	7993	7993	7993	7993	7993	7993
377	7945	7945	7811	7945	7945	7945	7945	7945	7418	7945	7945	7945	7945	7945	7945	7136	7945	7945	7945	7945	7944	7945	7945	7945	7945
378	7883	7883	7734	7883	7883	7883	7883	7883	7443	7883	7883	7883	7883	7883	7883	7211	7883	7883	7883	7883	7882	7883	7883	7883	7883
379	7892	7892	7713	7892	7892	7892	7892	7892	7476	7892	7892	7892	7892	7892	7892	7264	7892	7892	7892	7892	7890	7892	7892	7892	7892
380	7839	7839	7663	7839	7839	7839	7839	7839	7518	7839	7839	7839	7839	7839	7839	7303	7839	7839	7839	7839	7872	7839	7839	7839	7839
381	7741	7741	7580	7741	7741	7741	7741	7741	7463	7741	7741	7741	7741	7741	7741	7272	7741	7741	7741	7741	7802	7741	7741	7741	7741
382	7670	7670	7523	7670	7670	7670	7670	7670	7421	7670	7670	7670	7670	7670	7670	7233	7670	7670	7670	7670	7762	7670	7670	7670	7670
383	7546	7546	7418	7546	7546	7546	7546	7546	7368	7546	7546	7546	7546	7546	7546	7208	7546	7546	7546	7546	7675	7546	7546	7546	7546
384	7525	7525	7399	7525	7525	7525	7525	7525	7380	7525	7525	7525	7525	7525	7525	7241	7525	7525	7525	7525	7681	7525	7525	7525	7525
385	7498	7498	7376	7498	7498	7498	7498	7498	7395	7498	7498	7498	7498	7498	7498	7254	7498	7498	7498	7498	7681	7498	7498	7498	7498
386	7453	7453	7345	7453	7453	7453	7453	7453	7395	7453	7453	7453	7453	7453	7453	7259	7453	7453	7453	7453	7665	7453	7453	7453	7453
387	7445	7445	7337	7445	7445	7445	7445	7445	7452	7445	7445	7445	7445	7445	7445	7313	7445	7445	7445	7445	7682	7445	7445	7445	7445
388	7370	7370	7270	7370	7370	7370	7370	7370	7421	7370	7370	7370	7370	7370	7370	7289	7370	7370	7370	7370	7631	7370	7370	7370	7370
389	7309	7309	7121	7309	7309	7309	7309	7309	7391	7309	7309	7309	7309	7309	7309	7270	7309	7309	7309	7309	7595	7309	7309	7309	7309
390	7328	7328	7027	7328	7328	7328	7328	7328	7439	7328	7328	7328	7328	7328	7328	7323	7328	7328	7328	7328	7639	7328	7328	7328	7328
391	7256	7256	6915	7256	7256	7256	7256	7256	7420	7256	7256	7256	7256	7256	7256	7300	7256	7256	7256	7256	7566	7256	7256	7256	7256
392	7257	7257	6865	7257	7257	7257	7257	7257	7472	7257	7257	7257	7257	7257	7257	7346	7257	7257	7257	7257	7544	7257	7257	7257	7257
393	7250	7250	6802	7250	7250	7250	7250	7250	7504	7250	7250	7250	7250	7250	7250	7376	7250	7250	7250	7250	7512	7250	7250	7250	7250
394	7172	7172	6693	7172	7172	7172	7172	7172	7479	7172	7172	7172	7172	7172	7172	7424	7172	7172	7172	7172	7452	7172	7172	7172	7172
395	7039	7039	6505	7039	7039	7039	7039	7039	7348	7039	7039	7039	7039	7039	7039	7311	7039	7039	7039	7039	7364	7039	7039	7039	7039
396	6935	6935	6347	6935	6935	6935	6935	6935	7267	6935	6935	6935	6935	6935	6935	7181	6935	6935	6935	6935	7206	6935	6935	6935	6935
397	6795	6795	6163	6795	6795	6795	6795	6795	7116	6795	6795	6795	6795	6795	6795	7042	6795	6795	6795	6795	7022	6795	6795	6795	6795
398	6653	6653	5948	6653	6653	6653	6653	6653	6942	6653	6653	6653	6653	6653	6653	6864	6653	6653	6653	6653	6811	6653	6653	6653	6653
399	6495	6495	5743	6495	6495	6495	6495	6495	6773	6495	6495	6495	6495	6495	6495	6697	6495	6495	6495	6495	6606	6495	6495	6495	6495
400	6316	6316	5564	6316	6316	6316	6316	6316	6600	6316	6316	6316	6316	6316	6316	6534	6316	6316	6316	6316	6427	6316	6316	6316	6316
401	6197	6197	5445	6197	6197	6197	6197	6197	6482	6197	6197	6197	6197	6197	6197	6420	6197	6197	6197	6197	6308	6197	6197	6197	6197
402	6013	6013	5260	6013	6013	6013	6013	6013	6299	6013	6013	6013	6013	6013	6013	6254	6013	6013	6013	6013	6124	6013	6013	6013	6013
403	5870	5870	5117	5870	5870	5870	5870	5870	6160	5870	5870	5870	5870	5870	5870	6123	5870	5870	5870	5870	5981	5870	5870	5870	5870
404	5683	5683	4925	5683	5683	5683	5683	5683	5979	5683	5683	5683	5683	5683	5683	5948	5683	5683	5683	5683	5794	5683	5683	5683	5683
405	5496	5496	4729	5496	5496	5496	5496	5496	5801	5496	5496	5496	5496	5496	5496	5774	5496	5496	5496	5496	5607	5496	5496	5496	5496
406	5347	5347	4568	5347	5347	5347	5347	5347	5656	5347	5347	5347	5347	5347	5347	5627	5347	5347	5347	5347	5458	5347	5347	5347	5347
407	5137	5137	4315	5137	5137	5137	5137	5137	5451	5137	5137	5137	5137	5137	5137	5440	5137	5137	5137	5137	5247	5137	5137	5137	5137
408	4973	4973	4161	4973	4973	4973	4973	4973	5288	4973	4973	4973	4973	4973	4973	5283	4973	4973	4973	4973	5083	4973	4973	4973	4973
409	4816	4816	4040	4816	4816	4816	4816	4816	5130	4816	4816	4816	4816	4816	4816	5126	4816	4816	4816	4816	4919	4816	4816	4816	4816
410	4643	4643	3888	4643	4643	4643	4643	4643	4958	4643	4643	4643	4643	4643	4643	4954	4643	4643	4643	4643	4746	4643	4643	4643	4643

Position	*FBti0019985* *Reference genome*	*FBti0019985* *RAL-802*	*FBti0019985* *RAL-810*	*FBti0019985* *IV68*	*TOO +7* *RAL-405*	*TOO +7* *RAL-911*	*TOO +7* *RAL-441*	*TOO +7* *RAL-801*	*TOO +175* *IV145*	*TOO +278* *RAL-502*	*TOO -9* *IV42*	*TOO -9* *IV127*	*TOO -23* *IV40*	*TOO -44* *RAL-195*	*TOO -44* *RAL-383*	*TOO -48* *RAL-75*	*TOO -48* *RAL-716*	*TOO -48* *IV69*	*TOO-90* *RAL-21*	*TOO -90* *RAL-88*	*TOO -90* *RAL-177*	*TOO -90* *RAL-737*	*TOO -90* *RAL-820*	*TOO -90* *RAL-857*	*TOO -90* *IV50*
411	4482	4482	3748	4482	4482	4482	4482	4482	4801	4482	4482	4482	4482	4482	4482	4798	4482	4482	4482	4482	4585	4482	4482	4482	4482
412	4301	4301	3596	4301	4301	4301	4301	4301	4622	4301	4301	4301	4301	4301	4301	4619	4301	4301	4301	4301	4404	4301	4301	4301	4301
413	4082	4082	3408	4082	4082	4082	4082	4082	4414	4082	4082	4082	4082	4082	4082	4415	4082	4082	4082	4082	4180	4082	4082	4082	4082
414	3924	3924	3264	3924	3924	3924	3924	3924	4255	3924	3924	3924	3924	3924	3924	4257	3924	3924	3924	3924	4022	3924	3924	3924	3924
415	3716	3716	3081	3716	3716	3716	3716	3716	4057	3716	3716	3716	3716	3716	3716	4059	3716	3716	3716	3716	3814	3716	3716	3716	3716
416	3531	3531	2917	3531	3531	3531	3531	3531	3879	3531	3531	3531	3531	3531	3531	3881	3531	3531	3531	3531	3629	3531	3531	3531	3531
417	3363	3363	2770	3363	3363	3363	3363	3363	3724	3363	3363	3363	3363	3363	3363	3725	3363	3363	3363	3363	3461	3363	3363	3363	3363
418	3140	3140	2566	3140	3140	3140	3140	3140	3506	3140	3140	3140	3140	3140	3140	3507	3140	3140	3140	3140	3238	3140	3140	3140	3140
419	2943	2943	2389	2943	2943	2943	2943	2943	3310	2943	2943	2943	2943	2943	2943	3313	2943	2943	2943	2943	3041	2943	2943	2943	2943
420	2762	2762	2239	2762	2762	2762	2762	2762	3128	2762	2762	2762	2762	2762	2762	3130	2762	2762	2762	2762	2859	2762	2762	2762	2762
421	2598	2598	2097	2598	2598	2598	2598	2598	2963	2598	2598	2598	2598	2598	2598	2966	2598	2598	2598	2598	2690	2598	2598	2598	2598
422	2414	2414	1949	2414	2414	2414	2414	2414	2775	2414	2414	2414	2414	2414	2414	2778	2414	2414	2414	2414	2486	2414	2414	2414	2414
423	2228	2228	1782	2228	2228	2228	2228	2228	2588	2228	2228	2228	2228	2228	2228	2591	2228	2228	2228	2228	2295	2228	2228	2228	2228
424	2080	2080	1658	2080	2080	2080	2080	2080	2421	2080	2080	2080	2080	2080	2080	2423	2080	2080	2080	2080	2137	2080	2080	2080	2080
425	1922	1922	1563	1922	1922	1922	1922	1922	2269	1922	1922	1922	1922	1922	1922	2271	1922	1922	1922	1922	704	1922	1922	1922	1922
426	1745	1745	1476	1745	1745	1745	1745	1745	2089	1745	1745	1745	1745	1745	1745	2089	1745	1745	1745	1745	538	1745	1745	1745	1745
427	1574	1574	1322	1574	1574	1574	1574	1574	1914	1574	1574	1574	1574	1574	1574	1914	1574	1574	1574	1574	391	1574	1574	1574	1574
428	1286	1286	1077	1286	1286	1286	1286	1286	1604	1286	1286	1286	1286	1286	1286	1604	1286	1286	1286	1286	162	1286	1286	1286	1286
average	7775	7775	7690	7691	7260	7775	7737	7775	7780	7070	7775	7636	7775	7692	7690	7759	7692	7635	7775	7327	7664	7260	7775	5655	7775

Table S4. Results of the different statistics used to infer positive selection in the region flanking the nine solo-LTR insertions.

Region analyzed	10 strains with *FBti0019985* insertion	15 strains without insertion	23 strains with one of the nine *roo* insertions
S in 2R: 5,758,000-5,760,000	5	12	10
S/L in 2R:5,758,000-5,760,000	**0.0025**	0.006	0.0050
S/L in 2R	0.0149	0.0211	0.0240
S/L in whole genome	0.0135	0.0185	0.0206
Tajima's D	**-1.0446**	-0.1724	1.3970
iHS	**3.3470**	NA	**2.3200**
nSL	2.2730	NA	-1.7970
H_{12}	**0.16**	NA	0.023

Text S1. Phylogenetic tree containing the nine *roo* elements sequenced in this work and the 115 *roo* elements annotated in the *D. melanogaster* reference genome.

(FBti0019231:0.00000094674890955829,((FBti0020131:0.00000094674890955829
,(((FBti0020014:0.00046705729892767634,((FBti0019278:0.00034910171019756
432,FBti0019060:0.00011647222618549342):0.00000094674890955829,(((((FBti
0019416:0.00023589557753226124,FBti0019436:0.00047274191532750774):0.000
47295867215318093,(((FBti0020166:0.00011650078264661326,(FBti0019337:0.0
0349622065315623 69,(((FBti0019526:0.00055068784639003337,(FBti0019658:0
.00254363643921472942,FBti0019357:0.00088380144693645489):0.000000946748
90955829):0.00063203211838770246,(FBti0019547:0.00000094674890955829,FBt
i0020007:0.00023311291390921458):0.00034914144010396111):0.0000009467489
0955829,FBti0020076:0.00027530527528500529):0.00011647125085568128):0.00
000094674890955829):0.00011470909640405142,FBti0020022:0.001405024972338
49916):0.00070114820814951332,(FBti0019394:0.00000094674890955829,FBti00
20154:0.00012246847025752179):0.00000094674890955829):0.0000009467489095
5829):0.00000094674890955829,(((((FBti0019067:0.00102321006638694443,FBt
i0020170:0.00034997064535262774):0.00000094674890955829,(FBti0020150:0.0
0233335925341388893,(FBti0019083:0.00037032401638158354,(FBti0019098:0.0
0355891399745167 03,(FBti0020103:0.00000094674890955829,((FBti0019409:0.
00000094674890955829,FBti0019458:0.00058550282702813402):0.0000009467489
0955829,((((FBti0019420:0.00246333460744282626,FBti0019029:0.00000094674
890955829):0.00000094674890955829,FBti0020100:0.00000094674890955829):0.
00096037383233113364,FBti0020055:0.00100676243750998594):0.0000009467489
0955829,FBti0019020:0.00229179886670175057):0.00014059349275914320):0.00
000094674890955829):0.00011544715921784761):0.00023355624993562710):0.00
000094674890955829):0.00011647070169532554):0.00011645663267179169,(FBti
0019238:0.00034967198262425000,(FBti0059726:0.00083105394741331111,(FBti
0019019:0.00096585656848167023,(((FBti0020106:0.00498337992216761096,roo
+175:0.00000094674890955829):0.00000094674890955829,FBti0062347:0.000000
94674890955829):0.00059230297737466813,(((FBti0019665:0.0000009467489095
5829,FBti0019402:0.00070013989630343784):0.00000094674890955829,(FBti001
9285:0.00063942258338534584,FBti0019021:0.00038321049730827159):0.000640
45910257451519):0.00000094674890955829,(FBti0020085:0.000245622962029754
06,((FBti0019363:0.00000094674890955829,FBti0019406:0.000000946748909558
29):0.00034908240583596820,(FBti0019027:0.00035027127949199512,(FBti0019
597:0.00023308544117736079,FBti0020187:0.00023303758620180919):0.0010541
5689959827640):0.00023247554741548013):0.00000094674890955829):0.0001165
5726716069866):0.00023314045691881540):0.00008694019110376380):0.0001170
3578503542120):0.00011639650995372900):0.00000094674890955829):0.0001164
0340915149504,FBti0019532:0.00024334692488410758):0.00000094674890955829
,((((((FBti0019016:0.00000094674890955829,((FBti0020009:0.000000946748909
55829,(FBti0019017:0.00000094674890955829,FBti0059767:0.000160812140060
14038):0.00000094674890955829):0.00000094674890955829,(((FBti0019752:0.0
0058764956850761 57,FBti0020080:0.00000094674890955829):0.00000188471819
235699,FBti0019985:0.00000094674890955829):0.00000094674890955829,roo+27
8:0.00000094674890955829):0.00017521447173615061):0.00010009962955183680
):0.00008529966824280773,FBti0019438:0.00000094674890955829):0.000002354
89214169335,FBti0019435:0.00011663322309375492):0.00017502266315656346,F
Bti0019544:0.00038308306553177081):0.00017458759449437975,((FBti0019131:
0.00060213234277013651,(FBti0019196:0.00000094674890955829,FBti0062208:0
.00023300418546467672):0.00000094674890955829):0.00023203875373026364,FB
ti0019031:0.00051016023326125532):0.00000094674890955829):0.000000946748
90955829,(((((FBti0019539:0.00058307364617038687,((FBti0019838:0.0006891
4656870463272,FBti0019421:0.00034970032526873094):0.00000094674890955829
,(FBti0062794:0.00014664431224641232,((FBti0061311:0.0157057942850776788
3,(FBti0060416:0.00760100351227275227,FBti0063895:0.00234836887426894585
):0.00710324500758199081):0.00782437109130388643,(FBti0019553:0.00081887
607198007377,FBti0019026:0.00046823093382644242):0.00000094674890955829)
:0.00011496491355045804):0.00011670222712300768):0.00011648375495678730)

:0.00000094674890955829,(((((FBti0019140:0.00000094674890955829,FBti0019
608:0.00000094674890955829):0.00014183575346046612,roo+7:0.0000009467489
0955829):0.00000094674890955829,(roo-
90:0.00000094674890955829,FBti0059830:0.00011368545117857174):0.00000094
674890955829):0.00000094674890955829,FBti0019022:0.00012766241572181899)
:0.00000094674890955829,(FBti0059731:0.00188998558752436082,(FBti0019432
:0.00000094674890955829,(FBti0019339:0.00023329468776995760,(FBti0019463
:0.00090743874594468330,FBti0020069:0.00070023036405809621):0.0000009467
4890955829):0.00011643780030084637):0.00011643356050822003):0.0000009467
4890955829):0.00000094674890955829):0.00000094674890955829,FBti0059768:0
.00000094674890955829):0.00000094674890955829,(((FBti0019431:0.000349733
26705781685,FBti0020093:0.00034970726847986962):0.00000094674890955829,(
FBti0019068:0.00034989425564840165,(FBti0019572:0.00000094674890955829,(
FBti0019028:0.00011686551313038240,FBti0019025:0.00000094674890955829):0
.00000094674890955829):0.00035002137650958241):0.00000094674890955829):0
.00000094674890955829,(FBti0019175:0.00023304891497400555,FBti0019138:0.
00023316842656060640):0.00000094674890955829):0.00000094674890955829):0.
00000094674890955829,FBti0019148:0.00000094674890955829):0.0000009467489
0955829):0.00000094674890955829):0.00000094674890955829):0.0000009467489
0955829,(roo-
19:0.00000094674890955829,((FBti0020063:0.00046669903916543375,FBti00190
18:0.00035006271361163143):0.00011545750272533375,roo-
23:0.00000094674890955829):0.00000094674890955829):0.0001165032367971929
0):0.00011645855838157035,FBti0059769:0.00046635405217756309):0.00011646
220241308808):0.00000094674890955829):0.00046575648124465887,((((FBti002
0145:0.00034967720727706661,FBti0020047:0.00016091406379169885):0.000311
11281904480734,FBti0019173:0.00069620148365664254):0.00012027985713808095
,((roo-
44:0.00000094674890955829,((FBti0019619:0.00062894919272469998,FBti00190
51:0.00046571129321417717):0.00000094674890955829,(((FBti0019101:0.00035
180008134408196,(FBti0020121:0.00034940957368561113,(FBti0019100:0.00055
770415172411148,FBti0019439:0.00035004255314180033):0.000009467489095582
9):0.00011641402756658853):0.00000094674890955829,FBti0019615:0.0005825
8274967293864):0.00023295436224067640,((FBti0059737:0.000709916735699674
79,((FBti0019450:0.00191564085519543826,FBti0019205:0.000754267607756412
45):0.00029947998423395666,FBti0019126:0.00023632182075776495):0.0000009
4674890955829):0.00000094674890955829,FBti0020038:0.00023295307908169428
):0.00000094674890955829):0.00011646212305469723):0.00011547980131795422
):0.00000094674890955829,(((FBti0019393:0.00482091246787618322,FBti00190
30:0.00000094674890955829):0.00064553400584700031,FBti0019374:0.00011655
665979014553):0.00011632580859900959,FBti0019556:0.00023289689166103858)
:0.00000094674890955829):0.00000094674890955829):0.00011634831534742763,
FBti0019461:0.00440852981502604536):0.00000094674890955829):0.0000009467
4890955829,roo-
68:0.00000094674890955829):0.00011541761253000161):0.00014499380490196709
3,(FBti0020129:0.00011632654074328896,FBti0060681:0.00023325141455608287
):0.00043819243282028938):0.00041365948668212454,FBti0019015:0.000275262
29925931977):0.0;

7.2. Supplementary material: A unique cluster of *roo* insertions in the promoter region of a stress response

Additional file 1. Populations used for the analysis.

Population	Type of strain	Latitude (° N)	Longitude (° E)	Climate (Köppen-Geiger)	Year of the collection	Collector name / Reference
Akaa, FI	isofemale	61.1	23.5	Cold (Dfc)	2015	Maaria Kankare (Kankare Lab)
Stockholm, SE	isofemale	59.9	17.6	Cold (Dfb)	2011	Josefa Gonzalez (Gonzalez Lab)
Karensminde, DK	isofemale	56.0	10.2	Temperate (Cfb)	2015	Mads Fristrup (Loeschcke Lab)
Lund, SE	isofemale	55.7	13.2	Temperate (Cfb)	2015	Jessica Abbott (Abbott Lab)
Market Harborough, UK	isofemale	52.5	-0.9	Temperate (Cfb)	2015	Mirko Pegoraro
Munich, DE	isofemale	48.2	11.6	Temperate (Cfb)	2015	Eliza Argyridou (Parsch Lab)
Gotheron, FR	isofemale	45.0	4.9	Temperate (Cfb)	2015	Cristina Vieira (Vieira Lab)
Gimenells, ES	isofemale	41.7	0.4	Temperate (Csa)	2015	Marta Pascual (Pascual Lab)
Bari, IT	isofemale	40.9	17.2	Temperate (Csa)	2011	Josefa Gonzalez (Gonzalez Lab)
Tomelloso, ES	isofemale	39.2	-3.0	Arid (Bsk)	2015	Josefa Gonzalez (Gonzalez Lab)
Cortes de Baza, ES	isofemale	37.7	-2.8	Arid (Bsk)	2015	Lain Guio and Anna Ullastres (Gonzalez Lab)
Guadix, ES	isofemale	37.3	-3.2	Temperate (Csa)	2015	Anan Ullastres and Miriam Merenciano (Gonzalez Lab)
Raleigh, USA	inbred	35.7	-78.7	Temperate (Cfa)	2003	Huang et al., 2014
San Cristóbal de la Laguna, ES	isofemale	28.5	-16.3	Temperate (Csb)	2015	Quirze Rovira (Gonzalez Lab)
Siavonga, ZI	isofemale / inbred	-16.5	28.7	Arid (Bsh)	2010	Lack et al., 2015

Additional file 2

Additional file 2A. PCR results for the 277 strains analyzed in this work and in Merenciano et al. (2016). Strains used in Merenciano et al. (2016) are highlighted in blue.

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Akaa	Finland	FI_Aka_15_4	Polymorphic	roo-90						
		FI_Aka_15_9	Absent	Absent						
		FI_Aka_15_11	Absent	Absent						
		FI_Aka_15_12_2	Absent	Absent						
		FI_Aka_15_13	Present	roo-64						
		FI_Aka_15_14	Absent	Absent						
		FI_Aka_15_15	Absent	Absent						
		FI_Aka_15_18	No data	No data						
		FI_Aka_15_16	Present	roo-64						
		FI_Aka_15_23	Present	roo-64						
		FI_Aka_15_17	Polymorphic	roo-291						
		FI_Aka_15_19	Polymorphic	roo-64						
		FI_Aka_15_20	No data	No data	No data	No data	No data	No data		
		FI_Aka_15_21	Absent	Absent						
		FI_Aka_15_22	Polymorphic	roo-64						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Gimenells	Spain	ES_Gim_15_1	Polymorphic	roo-90						
		ES_Gim_15_2	Polymorphic	roo+175						
		ES_Gim_15_4	Polymorphic	No data						
		ES_Gim_15_5	Polymorphic	roo-90						
		ES_Gim_15_8	Absent	Absent						
		ES_Gim_15_12	Polymorphic	No data						
		ES_Gim_15_13	Polymorphic	roo-90						
		ES_Gim_15_14	Present	roo-90						
		ES_Gim_15_15	Absent	Absent						
		ES_Gim_15_16	Polymorphic	roo-44						
		ES_Gim_15_18	Polymorphic	roo-44						
		ES_Gim_15_20	Present	FBti0019985						
		ES_Gim_15_21	Present	roo-44						
		ES_Gim_15_25	Polymorphic	FBti0019985						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Munich	Germany	DE_Mun_15_6	Polymorphic	roo-90						
		DE_Mun_15_7	Polymorphic	roo-90						
		DE_Mun_15_8	Present	roo-90						
		DE_Mun_15_9	Absent	Absent						
		DE_Mun_15_10	Absent	Absent						
		DE_Mun_15_11	Present	roo+175						
		DE_Mun_15_12	Present	roo-68						
		DE_Mun_15_13	Present	roo-90						
		DE_Mun_15_14	No data	No data	No data	No data	No data	No data		
		DE_Mun_15_15	Polymorphic	roo+175						
		DE_Mun_15_16	Polymorphic	roo-90						
		DE_Mun_15_17	Present	roo-378						
		DE_Mun_15_18	Absent	Absent						
		DE_Mun_15_20	Polymorphic	roo-378						
		DE_Mun_15_21	Present	roo-90						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Gotheron	France	FR_Got 15_1	Absent	Absent						
		FR_Got 15_2	Absent	Absent						
		FR_Got 15_3	Present	roo-90						
		FR_Got 15_4	Polymorphic	roo-64						
		FR_Got 15_5	Absent	Absent						
		FR_Got 15_6	Absent	Absent						
		FR_Got 15_7	Absent	Absent						
		FR_Got 15_8	Polymorphic	roo-378						
		FR_Got 15_9	Present	roo-64						
		FR_Got 15_10	Present	roo-68						
		FR_Got 15_11	Absent	Absent						
		FR_Got 15_12	Absent	Absent						
		FR_Got 15_13	Absent	Absent						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Cortes de Baza	Spain	ES_Cor_15_13_1	Absent	Absent						
		ES_Cor_15_14	Polymorphic	roo-90						
		ES_Cor_15_15_1	Polymorphic	roo-90						
		ES_Cor_15_16_1	No data	No data	No data	No data	No data	No data		
		ES_Cor_15_17_1	Polymorphic	roo-90						
		ES_Cor_15_18	Polymorphic	roo-90						
		ES_Cor_15_19	Absent	Absent						
		ES_Cor_15_20	Polymorphic	roo-90						
		ES_Cor_15_21_1	Polymorphic	roo-90						
		ES_Cor_15_22_1	Polymorphic	roo-90						
		ES_Cor_15_23_1	Absent	Absent						
		ES_Cor_15_24_1	Polymorphic	roo-44						
		ES_Cor_15_25_1	Absent	Absent						
		ES_Cor_15_27_1	Polymorphic	roo-90 (5'-3')						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Lund	Sweden	SE_Lun_15_1	Polymorphic	roo-68						
		SE_Lun_15_2	Absent	Absent						
		SE_Lun_15_4	Present	roo-64						
		SE_Lun_15_5	Absent	Absent						
		SE_Lun_15_6	Present	No data						
		SE_Lun_15_7	Present	No data						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Tomelloso	Spain	ES_Tom_15_1	No data	No data	Absent	Absent				
		ES_Tom_15_2	No data	No data					Present	roo-90
		ES_Tom_15_3	Polymorphic	No data	Polymorphic	No data				
		ES_Tom_15_4	Polymorphic	No data			Polymorphic	No data		
		ES_Tom_15_5	Present	roo-90						
		ES_Tom_15_6	No data	No data	Absent	Absent				
		ES_Tom_15_7	Polymorphic	roo-90					Polymorphic	roo-90
		ES_Tom_15_8	Present	roo-90						
		ES_Tom_15_9	No data	No data	Polymorphic	roo-44				
		ES_Tom_15_10	Polymorphic	roo-90						
		ES_Tom_15_11	Polymorphic	roo-90						
		ES_Tom_15_12	Polymorphic	No data						
		ES_Tom_15_13	Polymorphic	roo-90						
		ES_Tom_15_14	Polymorphic	roo-44						
		ES_Tom_15_15	No data	No data	No data	No data	Polymorphic	roo+ 291		

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Karensminde	Denmark	DK_Jut_15 1	Present	roo-90 (5'-3')						
		DK_Jut_15 2	Present	roo-64						
		DK_Jut_15 3	Absent	Absent						
		DK_Jut_15 4	Present	roo-68						
		DK_Jut_15 5	Present	roo-19						
		DK_Jut_15 6	Absent	Absent						
		DK_Jut_15 7	Absent	Absent						
		DK_Jut_15 8	Polymorphic	No data						
		DK_Jut_15 9	Present	roo-90 (5'-3')						
		DK_Jut_15 10	Absent	Absent						
		DK_Jut_15 11	Absent	Absent						
		DK_Jut_15 12	Polymorphic	FBti0019985						
		DK_Jut_15 13	No data	No data	No data	No data				

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
San Cristobal de la Laguna	Spain	ES_Ten_15_1	Present	roo-90						
		ES_Ten_15_2	Absent	Absent						
		ES_Ten_15_3	Absent	Absent						
		ES_Ten_15_4	Present	roo-90						
		ES_Ten_15_5	Absent	Absent						
		ES_Ten_15_6	Present	roo-291						
		ES_Ten_15_7	Present	No data						
		ES_Ten_15_8	Present	roo-90						
		ES_Ten_15_9	No data	No data	No data	No data	No data	No data		
		ES_Ten_15_10	Polymorphic	roo-291						
		ES_Ten_15_11	Present	FBti0019985						
		ES_Ten_15_12	Polymorphic	roo-90						
		ES_Ten_15_13	No data	No data	No data	No data	No data	No data		
		ES_Ten_15_14	No data	No data	No data	No data	No data	No data		
		ES_Ten_15_15	Absent	Absent						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Market Harborough	UK	UK_Mar_15_5	Absent	Absent						
		UK_Mar_15_39	Present	roo-90 (5'-3')						
		UK_Mar_15_58	Absent	Absent						
		UK_Mar_15_44	Absent	Absent						
		UK_Mar_15_64	Polymorphic	roo-90						
		UK_Mar_15_46	No data	No data	No data	No data	Absent	Absent		
		UK_Mar_15_11	Absent	Absent						
		UK_Mar_15_54	No data	No data	No data	No data	Absent	Absent		
		UK_Mar_15_66	Polymorphic	roo-42						
		UK_Mar_15_7	Present	roo-90 (5'-3')						
		UK_Mar_15_40	No data	No data	Polymorphic	roo-68				
		UK_Mar_15_63	Absent	Absent						
		UK_Mar_15_70	No data	No data	No data	No data	Polymorphic	roo-90		
		UK_Mar_15_45	No data	No data	No data	No data	Absent	Absent		
		UK_Mar_15_42	Present	roo-42						
		UK_Mar_15_51	Present	roo-42						
		UK_Mar_15_60	Present	roo-90						
		UK_Mar_15_19	Plymorphic	roo+37						
		UK_Mar_15_14	Polymorphic	FBti0019985						
		UK_Mar_15_57	No data	No data	No data	No data	Polymorphic	roo-291		

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Stockholm	Sweden	SE_Sto_11_14	Polymorphic	roo-68						
		SE_Sto_11_16	No data	No data	No data	No data	No data	No data		
		SE_Sto_11_17	Present	roo-42						
		SE_Sto_11_18	Absent	Absent						
		SE_Sto_11_21	Present	roo-90						
		SE_Sto_11_22	Polymorphic	FBti0019985 (5'-3')						
		SE_Sto_11_24	Present	roo-393						
		SE_Sto_11_25	Absent	Absent						
		SE_Sto_11_26	Absent	Absent						
		SE_Sto_11_27	Present	roo-90						
		SE_Sto_11_29	Present	FBti0019985 (5'-3')						
		SE_Sto_11_32	Polymorphic	roo-42						
		SE_Sto_11_33	Present	roo-90						
		SE_Sto_11_36	No data	No data	No data	No data				
		SE_Sto_11_38	Absent	Absent						
		SE_Sto_11_39	Present	roo-90						
		SE_Sto_11_4	Absent	Absent						
		SE_Sto_11_41	Present	roo-42						
		SE_Sto_11_42	No data	No data	No data	No data	No data	No data		
		SE_Sto_11_45	Present	FBti0019985 (5'-3')						
		SE_Sto_11_46	Polymorphic	roo-64						
		SE_Sto_11_47	Absent	Absent						
		SE_Sto_11_6	Absent	Absent						
		SE_Sto_11_7	No data	No data	No data	No data	No data	No data		
		SE_Sto_11_8	Polymorphic	roo-64						
		SE_Sto_11_9	Absent	Absent						
		SE_Sto_11_53	Polymorphic	roo-44						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Raleig, NC (DGRP)	USA	RAL-100	Present	roo-90						
		RAL-142	Present	FBti0019985 (5'-3')						
		RAL-149	Present	FBti0019985						
		RAL-176	Present	FBti0019985						
		RAL-177			Present	roo-90				
		RAL-195			Present	roo-44				
		RAL-21			Present	roo-90				
		RAL-31	Absent	Absent						
		RAL-317	Present	roo-68						
		RAL-321	Present	roo-44						
		RAL-336	Present	roo-90						
		RAL-350	Present	roo-90						
		RAL-370	Present	roo-44						
		RAL-371			Absent					
		RAL-382	Absent	Absent						
		RAL-383			Present	roo-44				
		RAL-386	No data	No data	No data	No data				
		RAL-391			Absent	Absent				
		RAL-40			Absent	Absent				
		RAL-405			Present	roo+7				
		RAL-42	Present	FBti0019985						
		RAL-441			Present	roo+7				
		RAL-491	Present	FBti0019985						
		RAL-502			Present	roo+278				
		RAL-508			Absent					

Raleig, NC (DGRP)	USA	RAL-517	No data	No data						
		RAL-555	Present	FBti0019985						
		RAL-566	Absent	Absent						
		RAL-59	Present	FBti0019985						
		RAL-634	Present	roo-28						
		RAL-639	Present	FBti0019985						
		RAL-707	Present	roo-90						
		RAL-716			Present	roo-68				
		RAL-737			Present	roo-90				
		RAL-75			Present	roo-68				
		RAL-757	Absent	Absent						
		RAL-776			No data	No data				
		RAL-783			Absent					
		RAL-799	Present	FBti0019985						
		RAL-801			Present	roo+7				
		RAL-802			Present	FBti0019985				
		RAL-808	Absent	Absent						
		RAL-810			Present	FBti0019985				
		RAL-820			Present	roo-90				
		RAL-822			Absent					
		RAL-855			Absent					
		RAL-857			Present	roo-90				
		RAL-861	Present	FBti0019985						
		RAL-88			Present	roo-90				
		RAL-887	Present	roo+7						
		RAL-892	No data	No data	No data	No data				
		RAL-908			Absent					
		RAL-911			Present	roo+7				

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Purullena y el Bejarin (Guadix)	Spain	ES_Gua_15_1	No data	No data	Polymorphic	roo-90				
		ES_Gua_15_2	No data	No data					Polymorphic	roo-64
		ES_Gua_15_3	Polymorphic	No data						
		ES_Gua_15_4	No data	No data					Polymorphic	roo-90
		ES_Gua_15_5	No data	No data					Polymorphic	roo-90
		ES_Gua_15_6	No data	No data					Absent	Absent
		ES_Gua_15_7	No data	No data					Polymorphic	roo-90
		ES_Gua_15_8	No data	No data	Absent	Absent				
		ES_Gua_15_9	No data	No data	Polymorphic	roo-68				
		ES_Gua_15_10	No data	No data	Polymorphic	roo-68				
		ES_Gua_15_11	No data	No data	No data	No data	Polymorphic	roo-90		
		ES_Gua_15_12	No data	No data	Absent	Absent				
		ES_Gua_15_13	No data	No data	No data	No data	Polymorphic	FBti0019985 (5'-3')		
		ES_Gua_15_14	Polymorphic	roo-90						

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)	PCR result (FL6-R2, L-R2)	Seq result (FL6-R2, L-R2)	PCR result (FL2-R, L-R)	Seq result (FL2-R, L-R)
Bari	Italy	IT_Cas 11_125			No data	No data				
		IT_Cas 11_127			Present	roo-19				
		IT_Cas 11_145			Polymorphic	roo+175				
		IT_Cas 11_148			No data	No data				
		IT_Cas11_22			Absent					
		IT_Cas11_33-5			No data	No data				
		IT_Cas11_40			Polymorphic	roo-28				
		IT_Cas11_42-5			Present	roo-19				
		IT_Cas11_49-5			Absent					
		IT_Cas11_50-5			Polymorphic	roo-90				
		IT_Cas11_52-5			Absent					
		IT_Cas 11_68-5-8			Present	FBti0019985				
		IT_Cas 11_69-5-10			Polymorphic	roo-68				
		IT_Cas11_72			Absent					
		IT_Cas11_75			Absent					

Population	Country	Stock ID	PCR result (FL6-R, L-R)	Seq result (FL6-R, L-R)	PCR result (FL-R, L-R)	Seq result (FL-R, L-R)
Siavonga	Zambia	ZI114N	Absent	Absent		
		ZI161	Present	roo+7 (5'-3')		
		ZI177	Absent	Absent		
		ZI184	Absent	Absent		
		ZI207	No data	No data		
		ZI210	Absent	Absent		
		ZI214	Polymorphic	roo+7 (5'-3')		
		ZI228	Absent	Absent		
		ZI230	Absent	Absent		
		ZI232	Polymorphic	No data		
		ZI253	Absent	Absent		
		ZI264	Present	roo+7 (5'-3')		
		ZI271	Absent	Absent		
		ZI284	Absent	No data		
		ZI296	Polymorphic	No data	Polymorphic	roo+7 (5'-3')
		ZI303	Polymorphic	No data	Polymorphic	roo-90
		ZI324	Absent	Absent		
		ZI339	Absent	Absent		
		ZI348	Absent	Absent		
		ZI357N	Polymorphic	roo+192		
		ZI400	Polymorphic	No data	Polymorphic	roo+7 (5'-3')
		ZI437	Absent	Absent		
		ZI445	Polymorphic	No data	Polymorphic	roo+7 (5'-3')
		ZI447	Polymorphic	roo-56		
		ZI455	Polymorphic	roo-56		
		ZI457	Absent	Absent		
		ZI517	Absent	Absent		
		ZI76	Polymorphic	No data	Polymorphic	roo+7 (5'-3')

Additional file 2B. *Tlex-2* predictions in Merenciano et al. (2016) compared to PCR results in this work. Correct predictions are highlighted in green. Strains with *roo* insertions not identified in Merenciano et al. (2016) are highlighted in orange. Strains for which no results were obtained either by *Tlex-2* or by PCR are highlighted in grey.

Strain	*Tlex-2* prediction Merenciano *et al* ., 2016	PCR result This work
RAL-31	absent	absent
RAL-382	absent	absent
RAL-566	absent	absent
RAL-149	FBti0019985	FBti0019985
RAL-491	FBti0019985	FBti0019985
RAL-42	FBti0019985	FBti0019985
RAL-321	roo-44	roo-44
RAL-317	roo-68	roo-68
RAL-336	roo-90	roo-90
RAL-100	roo-90	roo-90
RAL-370	roo-44	roo-44
ZI210	absent	absent
ZI230	absent	absent
ZI271	absent	absent
ZI437	absent	absent
RAL-142	absent	FBti0019985 (5'-3')
RAL-892	absent	No data
RAL-634	absent	roo-28
ZI455N	absent	roo-56
ZI161	absent	roo+7 (5'-3')
ZI400	absent	roo+7 (5'-3')
ZI445	absent	roo+7 (5'-3')
ZI76	absent	roo+7 (5'-3')
ZI214	no data	roo+7 (5'-3')
ZI264	no data	roo+7 (5'-3')
ZI296	no data	roo+7 (5'-3')
ZI447	roo-44 (polymorphic)	roo-56
ZI357N	roo-68 (polymorphic)	roo+192
ZI303	absent	roo-90
ZI114N	no data	absent
ZI184	no data	absent
ZI348	no data	absent
ZI457	no data	absent
ZI207	no data	no data
ZI284	no data	no data
ZI324	roo-28 (polymorphic)	absent
ZI517	roo-68 (polymorphic)	absent
ZI228	roo-90 (polymorphic)	absent
ZI339	roo-90 (polymorphic)	absent
ZI232	roo-90 (polymorphic)	no data
ZI253	roo+278 (polymorphic)	absent
B25	absent	Absent
B26	absent	Absent
B38	absent	Absent
B47	absent	Absent
B6	absent	Absent
B14	roo-68 (polymorphic)	roo-68 (polymorphic)
B27	roo-90 (polymorphic)	roo-90
B33	roo-90 (polymorphic)	roo-90
B39	roo-90 (polymorphic)	roo-90
B29	absent	FBti0019985 (5'-3')
B45	absent	FBti0019985 (5'-3')
B24	absent	roo-393
B17	absent	roo-42
B41	absent	roo-42
B32	absent	roo-42 (polymorphic)
S3	absent	roo-44 (polymorphic)
B46	absent	roo-64 (polymorphic)
B8	absent	roo-64 (polymorphic)
B22	roo44 (polymorphic)	FBti0019985 (5'-3') (polymorphic)
B36	absent	No data
B42	absent	No data
B7	absent	No data
B4	FBti0019985 (polymorphic)	Absent
B16	FBti0019985 (polymorphic)	No data
B21	no data	roo-90
B18	roo-90 (polymorphic)	Absent
B9	roo-90 (polymorphic)	Absent

Additional file 2C. Allelic frequencies of all the 20 *roo* insertions in all the populations analyzed. EU: Europe, NA: North America and ZI: Zambia.

EU + NA + ZI populations	Present	Polymorphic	Absent	No data	Tested	Total strains result	Allelic freq. (%)
FBti0019985	14	3	228	32	277	245	**6,3**
roo $_{+7}$	5	0	240	32	277	245	**2,0**
roo $_{+37}$	0	1	244	32	277	245	**0,2**
roo $_{+175}$	1	2	242	32	277	245	**0,8**
roo $_{+278}$	1	0	244	32	277	245	**0,4**
roo $_{-19}$	3	0	242	32	277	245	**1,2**
roo $_{-28}$	1	1	243	32	277	245	**0,6**
roo $_{-44}$	5	6	234	32	277	245	**3,3**
roo $_{-68}$	6	6	233	32	277	245	**3,7**
roo $_{-90}$	26	29	190	32	277	245	**16,5**
roo $_{-393}$	1	0	244	32	277	245	**0,4**
roo $_{-378}$	1	2	242	32	277	245	**0,8**
roo $_{-291}$	1	4	240	32	277	245	**1,2**
roo $_{-64}$	6	6	233	32	277	245	**3,7**
roo $_{-42}$	4	2	239	32	277	245	**2,0**
FBti0019985 (3'-5')	3	2	240	32	277	245	**1,6**
roo $_{-90}$ ***(3'-5')***	4	1	240	32	277	245	**1,8**
roo $_{+7}$ ***(3'-5')***	2	5	238	32	277	245	**1,8**
roo $_{-56}$	0	2	243	32	277	245	**0,4**
roo $_{+192}$	0	1	244	32	277	245	**0,2**

Additional file 2D. Allelic frequencies (%) of the 20 *roo* insertions in the 15 different populations analyzed. Elements only present in one population are highlighted in red.

	Raleigh (DGRP)		Stockholm, SE		Bari, IT		Akaa, FI		Gimenells, ES		Gotheron, FR		Munich, DE		Baza, ES		Iund (SE)	
	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains
FBti0019985	22	49	0	23	8	12	0	13	14	11	0	13	0	14	0	13	0	4
roo $_{+7}$	10	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{+37}$	0	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{+175}$	0	49	0	23	4	12	0	13	0	11	0	13	11	14	0	13	0	4
roo $_{+278}$	2	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-19}$	0	49	0	23	17	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-28}$	2	49	0	23	4	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-44}$	8	49	2	23	0	12	0	13	18	11	0	13	0	14	4	13	0	4
roo $_{-68}$	6	49	2	23	4	12	0	13	0	11	8	13	7	14	0	13	13	4
roo $_{-90}$	20	49	17	23	4	12	4	13	23	11	8	13	32	14	27	13	0	4
roo $_{-393}$	0	49	4	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-378}$	0	49	0	23	0	12	0	13	0	11	4	13	11	14	0	13	0	4
roo $_{-291}$	0	49	0	23	0	12	4	13	0	11	0	13	0	14	0	13	0	4
roo $_{-64}$	0	49	4	23	0	12	31	13	0	11	12	13	0	14	0	13	25	4
roo $_{-42}$	0	49	11	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
FBti0019985 (3'-5')	2	49	11	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-90}$ ***(3'-5')***	0	49	0	23	0	12	0	13	0	11	0	13	0	14	4	13	0	4
roo $_{+7}$ ***(3'-5')***	0	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{-56}$	0	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
roo $_{+192}$	0	49	0	23	0	12	0	13	0	11	0	13	0	14	0	13	0	4
Absent	27	49	48	23	58	12	62	13	45	11	69	13	39	14	65	13	63	4

	Karensminde, DK		Tenerife, ES		Tomelloso, ES		Harborough, UK		Guadix, ES		Siavonga, ZI	
	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains	Allelic freq. (%)	Number of strains
FBti0019985	5	11	9	11	0	12	3	20	0	13	0	26
roo $_{+7}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{+37}$	0	11	0	11	0	12	3	20	0	13	0	26
roo $_{+175}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{+278}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{-19}$	9	11	0	11	0	12	0	20	0	13	0	26
roo $_{-28}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{-44}$	0	11	0	11	8	12	0	20	0	13	0	26
roo $_{-68}$	9	11	0	11	0	12	3	20	8	13	0	26
roo $_{-90}$	0	11	32	11	42	12	10	20	23	13	2	26
roo $_{-393}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{-378}$	0	11	0	11	0	12	0	20	0	13	0	26
roo $_{-291}$	0	11	14	11	4	12	3	20	0	13	0	26
roo $_{-64}$	9	11	0	11	0	12	0	20	4	13	0	26
roo $_{-42}$	0	11	0	11	0	12	13	20	0	13	0	26
FBti0019985 (3'-5')	0	11	0	11	0	12	0	20	4	13	0	26
roo $_{-90}$ ***(3'-5')***	18	11	0	11	0	12	10	20	0	13	0	26
roo $_{+7}$ ***(3'-5')***	0	11	0	11	0	12	0	20	0	13	15	26
roo $_{-56}$	0	11	0	11	0	12	0	20	0	13	4	26
roo $_{+192}$	0	11	0	11	0	12	0	20	0	13	2	26
Absent	50	11	45	11	46	12	58	20	62	13	77	26

Additional file 2E. Expected genotype frequency of heterozygous flies with the two most common insertions, *FBti0019985* and *roo-90* in all the populations analyzed. a: *FBti0019985* alellic frequency, b: *roo-90* alellic frequency, and c: absent alellic frequency.

Raleigh (DGRP)

	a	b	c
	0,22	**0,2**	**0,58**
a **0,22**	0,0484	0,044	0,1276
b **0,2**	0,044	0,04	0,116
c **0,58**	0,1276	0,116	0,3364

Expected genotype freq (%)	
aa	4,84
ab	8,8
bb	4

Stockholm

	a	b	c
	0	**0,17**	**0,83**
a **0**	0	0	0
b **0,17**	0	0,0289	0,1411
c **0,83**	0	0,1411	0,6889

Expected genotype freq (%)	
aa	0
ab	0
bb	2,89

Bari

	a	b	c
	0,08	**0,04**	**0,88**
a **0,08**	0,0064	0,0032	0,0704
b **0,4**	0,032	0,016	0,352
c **0,52**	0,0416	0,0208	0,4576

Expected genotype freq (%)	
aa	0,64
ab	0,64
bb	1,6

Akaa

	a	b	c
	0	**0,04**	**0,96**
a **0**	0	0	0
b **0,04**	0	0,0016	0,0384
c **0,96**	0	0,0384	0,9216

Expected genotype freq (%)	
aa	0
ab	0
bb	0,16

Gimenells

	a	b	c
	0,14	**0,23**	**0,63**
a **0,14**	0,0196	0,0322	0,0882
b **0,23**	0,0322	0,0529	0,1449
c **0,63**	0,0882	0,1449	0,3969

Expected genotype freq (%)	
aa	1,96
ab	6,44
bb	5,29

Gotheron

	a	b	c
	0	**0,08**	**0,92**
a **0**	0	0	0
b **0,08**	0	0,0064	0,0736
c **0,92**	0	0,0736	0,8464

Expected genotype freq (%)	
aa	0
ab	0
bb	0,64

Munich

	a	b	c
	0	**0,32**	**0,68**
a **0**	0	0	0
b **0,32**	0	0,1024	0,2176
c **0,68**	0	0,2176	0,4624

Expected genotype freq (%)	
aa	0
ab	0
bb	10,24

Cortes de Baza

		a	b	c
		0	**0,27**	**0,73**
a	**0**	0	0	0
b	**0,27**	0	0,0729	0,1971
c	**0,73**	0	0,1971	0,5329

Expected genotype freq (%)	
aa	0
ab	0
bb	7,29

Lund

		a	b	c
		0	**0**	**1**
a	**0**	0	0	0
b	**0**	0	0	0
c	**1**	0	0	1

Expected genotype freq (%)	
aa	0
ab	0
bb	0

Karensminde

		a	b	c
		0,05	**0**	**0,95**
a	**0,05**	0,0025	0	0,0475
b	**0**	0	0	0
c	**0,95**	0,0475	0	0,9025

Expected genotype freq (%)	
aa	0,25
ab	0
bb	0

Tenerife

		a	b	c
		0,09	**0,32**	**0,59**
a	**0,09**	0,0081	0,0288	0,0531
b	**0,32**	0,0288	0,1024	0,1888
c	**0,59**	0,0531	0,1888	0,3481

Expected genotype freq (%)	
aa	0,81
ab	5,76
bb	10,24

Tomelloso

		a	b	c
		0	**0,42**	**0,58**
a	**0**	0	0	0
b	**0,42**	0	0,1764	0,2436
c	**0,58**	0	0,2436	0,3364

Expected genotype freq (%)	
aa	0
ab	0
bb	17,64

Market Harborough

		a	b	c
		0,03	**0,1**	**0,87**
a	**0,03**	0,0009	0,003	0,0261
b	**0,1**	0,003	0,01	0,087
c	**0,87**	0,0261	0,087	0,7569

Expected genotype freq (%)	
aa	0,09
ab	0,6
bb	1

Guadix

		a	b	c
		0	**0,23**	**0,77**
a	**0**	0	0	0
b	**0,23**	0	0,0529	0,1771
c	**0,77**	0	0,1771	0,5929

Expected genotype freq (%)	
aa	0
ab	0
bb	5,29

Siavonga

		a	b	c
		0	**0,02**	**0,98**
a	**0**	0	0	0
b	**0,02**	0	0,0004	0,0196
c	**0,98**	0	0,0196	0,9604

Expected genotype freq (%)	
aa	0
ab	0
bb	0,04

Additional file 3

Additional file 3. A. Consensus target site duplication (TSD) sequence identified in Merenciano et al. (2016) (left panel) and consensus TSD identified with the data of this paper and Merenciano et al. (2016) (right panel). **B.** TSD sequences of the 20 *roo* insertions. Frequency represents the number of strains that harbor the TSD out of the number of strains with a complete sequenced region.

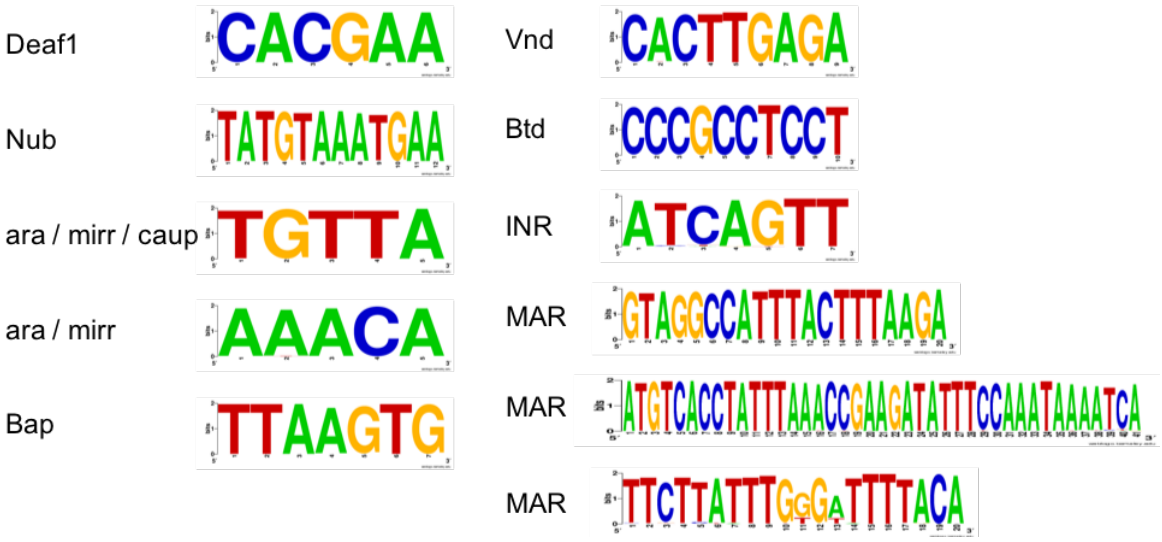
A

B

roo element	TSD sequence	Frequency
FBti0019985	GAATG	14/15
***roo*$_{+7}$**	ATCGC	4/7
***roo*$_{+37}$**	AACGT	1/1
***roo*$_{+175}$**	GCGAT	2/2
***roo*$_{+278}$**	-	-
***roo*$_{-19}$**	GCGT	3/3
***roo*$_{-28}$**	CGAGA	1/1
***roo*$_{-44}$**	CGTAC	11/11
***roo*$_{-68}$**	CTAGC	9/9
***roo*$_{-90}$**	TTTAG	43/45
***roo*$_{-393}$**	-	-
***roo*$_{-378}$**	-	-
***roo*$_{-291}$**	CTATT	3/3
***roo*$_{-64}$**	CTGTT	5/6
***roo*$_{-42}$**	TACT	7/7
FBti0019985 (3'-5')	GAATG	5/5
roo*$_{-90}$ *(3'-5')	TTTAG	4/6
roo*$_{+7}$ *(3'-5')	ATCGC	2/2
***roo*$_{-56}$**	CTGAT	2/2
***roo*$_{+192}$**	GTCAT	1/1

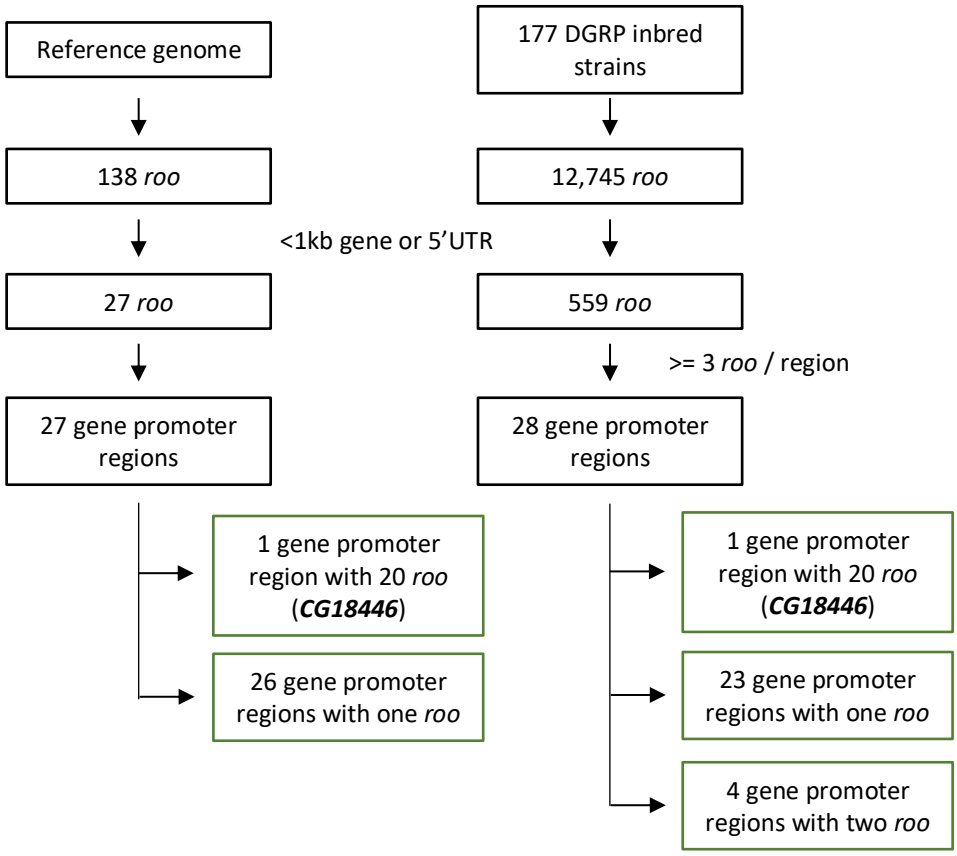
Additional file 4

Additional file 4. Consensus sequence of the transcription factor binding sites and matrix attachment regions identified in all the *roo* sequences identified in the *CG18446* promoter region.



Additional file 5

Additional file 5. The formation of *roo* insertional clusters in gene promoter regions is not a *roo* family characteristic. Scheme of the gene promoter regions containing *roo* elements present in the reference genome (left) and present in 177 DGRP inbred strains (right).



Additional file 6

Additional file 6A. Coordinates (R6), length, recombination rates and PCR results of the 26 promoter regions with a *roo* insertion in the reference genome.

							Insertion presence													
	Chr	Location max.	Location min.	Length	RRC Midpoint Rate	Comeron Midpoint Rate	*RAL-21*	*RAL-31*	*RAL-317*	*RAL-321*	*RAL-383*	*RAL-386*	*RAL-707*	*RAL-911*	*RAL-40*	*RAL-776*	*RAL-855*	*RAL-857*	*RAL-88*	*RAL-405*	*RAL-177*
FBti0020129	3L	16954877	16945783	9094	1.07	0.76	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019432	3R	25841869	25832746	9123	3.18	2.18	No data	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019996	2R	17827837	17818916	8921	3.75	4.01	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present					
FBti0019436	3R	26519559	26510557	9002	3.15	0.87	Absent	Absent	Absent	No data	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019556	X	3502247	3493147	9100	2.81	7.50	Absent	No data	No data	Absent	No data	Absent	No data	No data	Absent	Absent					
FBti0019278	2R	9361370	9352280	9090	1.58	2.49	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019017	2R	9213176	9212749	427	1.52	0.65	No data	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019394	3R	17763645	17763218	427	2.72	3.26	No data	No data	Absent	Absent	Absent	No data	Absent	No data	Absent	Absent					
FBti0019439	3R	27016536	27007460	9076	3.12	0.33	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0020022	3L	2456657	2447560	9097	3.30	1.25	Absent	Absent	Absent	Absent	Absent	No data	Absent	Absent	Absent	Absent					
FBti0020093	3L	11508212	11499112	9100	2.78	3.18	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019068	X	16349821	16340723	9098	3.02	2.71	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019597	X	7134401	7125301	9100	2.85	2.26	Absent	Absent	Absent	No data	Absent	Absent	No data	Absent	Absent	Absent					
FBti0019051	X	14114170	14105078	9092	3.67	4.43	Absent	Absent	Absent	Absent	Absent	No data	Absent	Absent	Absent	Absent					
FBti0019532	X	1072546	1063811	8735	0.00	0.00	No data	Absent	No data	Absent	Absent	No data	Absent	Absent	Absent	Absent					
FBti0020047	3L	6230063	6220970	9093	3.44	0.90	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0059710	X	4997254	4990099	7155	3.33	5.24	Absent	Absent	Absent	Absent	Absent	Present	Present	Absent	Absent	Absent					
FBti0059714	3L	13814138	13804601	9537	2.19	1.25	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present					
FBti0062208	2L	21450230	21441096	9134	0.00	0.00	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019285	3R	4375137	4366813	8324	0.00	0.00	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019393	3R	17638800	17638367	433	2.70	1.63	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019463	3R	30404480	30396273	8207	2.74	1.31	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0019665	X	21062346	21061919	427	0.70	0.36	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent					
FBti0059664	3L	23081420	23080432	988	0.00	0.00	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present					
FBti0059662	2R	9868272	9867900	372	1.81	1.08	Present	Present	Present	Present	Present	Present	Present	Present	Present	Present					
FBti0019450	3R	28763899	28758341	5558	2.96	3.16								Absent	Absent	Absent	No data	Absent	No data	Absent	No data

Additional file 6B. PCR results and *de novo* TE information of the 28 promoter regions where >=3 strains putatively have a roo insertion based on TIDAL software predictions.

Promoter region number	Gene	Chr	Strain	Chr_coord_5p (TIDAL)	Chr_coord_3p (TIDAL)	TE_start (TIDAL)	TE_end (TIDAL)	TE_length (TIDAL) +1	TE_length (TIDAL) -1	Absence / Presence	TE Insertion Site	TE Length	Orientation	TSD	>1 TE insertion
1	CR45023	chr3R	RAL-321	20359815	20359854	8664	8703	39		Present	20359890		1		NO
			RAL-584	20359805	20359855	15	70	55		Present	20359890	424 bp	1	ACACAGACT	
			RAL-737	20359837	20359867	8666	8695	29		Present	20359890	424 bp	1	ACACAGACT	
			RAL-748	20359789	20359854	4	8733	8729		-	-	-	-	-	
			RAL-799	20359831	20359868	8660	33		8627	Present	-	-	-	-	
			RAL-801	20359789	20359853	8668	66		8602	Present	20359890	424 bp	1	ACACAGACT	
			RAL-802	20359790	20359853	5	68	63		Absent	-	-	-	-	
			RAL-907	20359817	20359861	8666	46		8620	Absent	-	-	-	-	
			RAL-913	20359796	20359856	8670	71		8599	-	-	-	-	-	
2	cbx, CG18446	chr2R	RAL-142	9871532	9871580	8716	8675		41	Present	9871096	428 bp	-1	GAATG	YES
			RAL-208	9871524	9871557	8695	8663		32	Present	9871096	428 bp	-1	GAATG	
			RAL-223	9871532	9871593	74	8676	8602		-	-	-	-	-	
			RAL-229	9871526	9871569	22	8659	8637		Present	9871053	428 bp	-1	TACT	
			RAL-304	9871554	9871689	8700	398		8302	-	-	-	-	-	
			RAL-502	9871734	9871845	1	402	401		Present	9871802	428 bp	1		
			RAL-642	9871706	9871760	22	9071	9049		-	-	-	-	-	
			RAL-714	9871568	9871628	8664	9050	386		Present	9871622	428bp	1	TATG	
3	CR44779, CG3655	chrX	RAL-105	1087536	1087616	8665	412		8253	Absent	-	-	-	-	NO
			RAL-380	1087547	1087611	8665	405		8260	-	-	-	-	-	
			RAL-392	1087521	1087619	8671	404		8267	Present	1087588	-	1		
			RAL-810	1087534	1087604	8664	398		8266	Absent	-	-	-	-	
			RAL-812	1087535	1087618	8664	412		8252	Present	1087588	-	1		
			RAL-832	1087523	1087630	9	404	395		-	-	-	-	-	
			RAL-861	1087538	1087617	8667	411		8256	Present	1087588	-	1	GTGGC	
4	CR44940	chr3R	RAL-149	11085594	11085632	8665	385		8280	-	-	-	-	-	NO
			RAL-28	11085779	11085865	8661	411		8250	-	-	-	-	-	
			RAL-301	11085732	11085911	8664	407		8257	-	-	-	-	-	
			RAL-340	11085648	11086044	8751	375		8376	Present	110085836	-	1	TTGAC	
			RAL-774	11085545	11085695	20	398	378		-	-	-	-	-	
			RAL-850	11085738	11085904	6	400	394		Present	110085836	-	1	TTGAC	
			RAL-853	11085767	11085898	8664	409		8255	Present	110085836	-	1	TTGAC	
5	CR44895	chrX	RAL-748	6746270	6746436	10	404	394		-	-	-	-	-	NO
			RAL-801	6746275	6746420	15	9051	9036		Present	6746364	-	1	TGACT	
			RAL-802	6746269	6746435	8673	403		8270	Present	6746364	-	1	TGACT	
			RAL-810	6746314	6746391	8668	409		8259	Absent	-	-	-	-	
			RAL-850	6746286	6746437	8690	405		8285	Present	6746364	-	1	TGACT	
			RAL-859	6746316	6746392	6	411	405		Present	6746364	-	1		

6	CG34436	chr3R	RAL-391	28282358	28282480	408	8662	8254		Present	28282432	-	-1	ATTACT	NO
			RAL-45	28282377	28282457	409	8665	8256		-	-	-	-	-	
			RAL-85	28282396	28282445	390	13		377	-	-	-	-	-	
			RAL-907	28282369	28282472	397	8670	8273		Absent	-	-	-	-	
			RAL-911	28282387	28282477	9043	1		9042	Present	28282432	-	-1	ATTACT	
7	CG15394	chr2L	RAL-350	2592009	2592067	8661	9047	386		Present	2592067	-	1	GAGTGT	NO
			RAL-359	2591991	2592111	8663	408		8255	Present	2592067	-	1	GAGTGT	
			RAL-405	2592003	2592100	8664	396		8268	Absent	-	-	-	-	
			RAL-584	2591964	2592134	2	400	398		Present	2592067	-	1	GAGTGT	
			RAL-882	2591992	2592025	9039	409		8630	-	-	-	-	-	
8	plum, CR45226	chr3R	RAL-716	26521022	26521095	8664	401		8263	Present	25621076	-	1	ATAAAC	YES
			RAL-790	26521000	26521117	8662	403		8259	-	-	-	-	-	
			RAL-890	26521019	26521104	8661	411		8250	Present	25621076	427 bp	1	ATAAAC	
			RAL-91	26521503	26521558	404	19		385	Present	26521553	-	-1	AGCAGTT	
9	unc-13	chr4	RAL-189	905371	905527	13	397	384		-	-	-	-	-	NO
			RAL-280	905408	905481	8664	401		8263	Present	905463	-	1	CATTCGCT	
			RAL-336	905376	905479	8744	399		8345	Present	905463	-	1	CATTCGCT	
			RAL-338	905410	905481	8666	401		8265	Present	905463	-	1	CATTCGCT	
10	CG12480	chrX	RAL-852	14169159	14169217	399	8677	8278		Present	14169203	Full lenght	-1	ATTAT	NO
			RAL-801	14169135	14169270	9036	10		9026	Present	14169203	Full lenght	-1	ATTAT	
			RAL-802	14169103	14169273	410	8671	8261		Present	14169203	-	-1	ATTAT	
11	CG15556	chr3R	RAL-41	31130787	31130854	1	396	395		Present	31130840	428 bp	1	CCCAA	NO
			RAL-375	31130785	31130863	8663	405		8258	Present	31130840	428 bp	1	CCCAA	
			RAL-88	31130797	31130863	8675	405		8270	Present	31130840	-	1	CCCAA	
12	CG42637, Gyc76C	chr3L	RAL-861	19790017	19790061	8671	9043	372		Present	19790063	-	1	CCCAC	NO
			RAL-737	19790013	19790090	8668	409		8259	Present	19790063	-	1	CCCAC	
			RAL-908	19789994	19790111	5	410	405		Present	19790063	Full lenght	1	CCCAC	
13	CG11459	chr3R	RAL-91	6027394	6027554	412	11		401	Present	6027532	Full lenght	-1	CAAAATG	YES
			RAL-371	6027440	6027596	405	21		384	Present	6027532	-	-1	CAAAATG	
			RAL-235	6027537	6027646	406	8665	8259		Present	6027608	-	-1	TACTC	
14	CG15152	chr2L	RAL-426	18114246	18114349	8682	409		8273	Absent	-	-	-	-	NO
			RAL-318	18114251	18114324	3	405	402		Present	18114306	-	1	TTGTT	
			RAL-391	18114232	18114348	4	407	403		Present	18114306	-	1	TTGTT	
15	CG43085	chr3L	RAL-555	17646444	17646525	409	1		408	Absent	-	-	-	-	NO
			RAL-812	17646453	17646525	399	1		398	Present	17646499	-	-1	ATACT	
			RAL-810	17646446	17646524	407	8666	8259		Absent	-	-	-	-	
16	CG43329	chr3L	RAL-555	22505402	22505486	8664	412		8252	Present	22505456	-	1	CTAGT	NO
			RAL-812	22505402	22505476	8664	402		8262	Present	22505456	-	1	CTAGT	
			RAL-810	22505415	22505486	13	412	399		Present	22505456	-	1	CTAGT	

17	CG15879	chr3L	RAL-385	2169107	2169159	404	8700	8296		Present	2169152	-	-1	GTTGT	YES
			RAL-441	2169090	2169203	8668	405		8263	Present	2169162	Full lenght	1	TTTTG	
			RAL-395	2169070	2169204	409	8662	8253		Present	2169152	-	-1	GTTGT	
18	CG30033, CG18336, dare	chr2R	RAL-491	11293606	11293690	408	8660	8252		Present	11293659	428 bp	-1	TTAGG	NO
			RAL-820	11293616	11293690	398	8660	8262		Present	11293659	428 bp	-1	TTAGG	
			RAL-853	11293595	11293699	404	7		397	Absent	-	-	-	-	
19	CG10029	chr3R	RAL-385	7317647	7317680	9020	9053	33		Present	7317652	-	1	GGTAT	NO
			RAL-392	7317660	7317692	9033	403		8630	Absent	-	-	-	-	
			RAL-395	7317647	7317698	350	403	53		Present	7317652	-	1	GGTAT	
20	CG33630, CR45205	chr3R	RAL-91	10062357	10062418	11	401	390		Present	10062340	-	1	CATAT	NO
			RAL-359	10062330	10062445	8667	407		8260	Present	10062340	-	1	CATAT	
			RAL-508	10062324	10062445	8661	408		8253	Present	10062340	-	1	CATAT	
21	CG9451, CG9449	chr3L	RAL-385	19497966	19498078	8672	9070	398		Present	19498036	-	1	TTGTT	NO
			RAL-392	19497974	19498077	8680	9069	389		Absent	-	-	-	-	
			RAL-395	19497961	19498075	8673	9061	388		Present	19498036	-	1	TTGTT	
22	CR43280	chr3L	RAL-879	18440617	18440662	9047	11		9036	Present	18440646	-	-1	TATAG	NO
			RAL-774	18440570	18440721	407	8666	8259		Present	18440646	-	-1	TATAG	
			RAL-843	18440571	18440717	406	8670	8264		Present	18440646	-	-1	TATAG	
23	CR44657	chrX	RAL-386	14114650	14114722	8660	404		8256	Present	14114700	-	1	GAGGA	YES
			RAL-765	14115629	14115686	9050	8664		386	Present	14115661	-	-1	GAACA	
			RAL-774	14115560	14115717	9069	21		9048	Present	14115661	Full lenght	-1	GAACA	
24	CR44701	chr2L	RAL-732	9887364	9887438	7	410	403		Present	9887412	-	1	TGCAT	NO
			RAL-727	9887358	9887440	8665	411		8254	Present	9887412	-	1	TGCAT	
			RAL-908	9887336	9887450	8663	401		8262	Present	9887412	Full lenght	1	TGCAT	
25	CR45407, CG42268	chr3L	RAL-861	9760782	9760854	8663	399		8264	Present	9760837	428 bp	1	TGTAC	NO
			RAL-712	9760784	9760853	8664	398		8266	Present	9760837	428 bp	1	TGTAC	
			RAL-352	9760781	9760858	8662	404		8258	Present	9760837	428 bp	1	TGTAC	
26	Rgk3	chr2R	RAL-136	21255575	21255635	411	10		401	Present	21255618	-	-1	CAAAC	NO
			RAL-301	21255520	21255686	403	8673	8270		Present	21255618		-1	CAAAC	
			RAL-379	21256142	21256177	411	9039	8628		-	-	-	-	-	
27	CG14326, CG14324	chr3R	RAL-336	17618978	17619026	391	8		383	Present	17619010	Full lenght	-1	ATGGG	NO
			RAL-338	17618975	17619037	9055	1		9054	Present	17619010	Full lenght	-1	ATGGG	
			RAL-513	17618956	17619053	9054	2		9052	Present	17619010	Full lenght	-1	ATGGG	
28	CG31678, CR45004	chr2L	RAL-853	20667093	20667192	8679	408		8271	Present	20667150	-	1	ACTAT	NO
			RAL-859	20667097	20667170	8664	401		8263	Present	20667150	-	1	ACTAT	
			RAL-890	20667098	20667174	1	405	404		Present	20667150	Full lenght	1	ACTAT	

Additional file 6C. Promoter regions where <3 strains putatively have a *roo* insertion based on TIDAL software predictions.

Genes+A4:I32	Strain	Chr	Chr_coord_5p (TIDAL)	Chr_coord_3p (TIDAL)	TE_start (TIDAL)	TE_end (TIDAL)	TE length	Number of predicted *roo* in the region
AP-1sigma,CG31140	RAL-149	chr3R	23962850	23962919	409	8663	8254	1
Acp54A1	RAL-57	chr2R	17133187	17133290	8669	402	8267	1
Acp65Aa	RAL-85	chr3L	6152914	6152994	8662	406	8256	1
Acp76A,CR44675	RAL-138	chr3L	19052086	19052243	13	403	390	1
Acsl	RAL-306_125b	chr2R	8674680	8674851	3	403	400	1
Argk	RAL-181	chr3L	9066335	9066411	399	8659	8260	1
AstC	RAL-555	chr2L	11081712	11081796	408	8660	8252	1
CCAP-R	RAL-310	chr3R	25956545	25956585	8683	391	8292	1
CG10137,CG33116	RAL-88	chr2L	19533011	19533060	9043	8667	376	1
CG10168	RAL-757	chr3R	23539069	23539140	10	409	399	1
CG10184,CG31145	RAL-884	chr3R	23605608	23605686	8664	406	8258	1
CG10205	RAL-324_100	chr2R	14832627	14832735	8693	411	8282	1
CG10911,Muc55B	RAL-189	chr2R	18052455	18052612	8680	401	8279	1
CG11158	RAL-340_100b	chrX	13682566	13682939	43	8979	8936	1
CG11160	RAL-361	chrX	11044362	11044526	8674	402	8272	1
CG11373	RAL-502	chr3R	5961267	5961377	397	8663	8266	1
CG11379	RAL-426	chrX	1179740	1179827	8686	408	8278	1
CG11458	RAL-822	chr3L	20905401	20905494	9032	8668	364	1
CG11594	RAL-307_125b	chr3L	4026114	4026284	409	11	398	1
CG12078	RAL-45	chr3L	3244368	3244439	8667	402	8265	1
CG12126	RAL-373_125b	chr2R	8263224	8263384	7	395	388	1
CG12209,CG12914	RAL-595	chr2R	10247568	10247637	401	4	397	1
CG12395	RAL-426	chrX	16063181	16063295	8665	9066	401	1
CG12481	RAL-88	chrX	14186828	14186847	8661	17	8644	1
CG12594	RAL-105	chr3R	11722129	11722181	411	31	380	1
CG12605	RAL-884	chr3L	3965116	3965188	408	8670	8262	1
CG12951,CG16749	RAL-897	chr3R	9330565	9330628	4	401	397	1
CG12964	RAL-69	chr2R	15497065	15497127	8663	392	8271	1
CG13183	RAL-703	chr2R	11912692	11912804	409	8	401	1
CG13229,shn	RAL-149	chr2R	11159765	11159840	408	8662	8254	1
CG13408	RAL-359	chr3R	22080059	22080113	22	9027	9005	1
CG1354	RAL-426	chrX	9778390	9778494	7	401	394	1
CG13639,ppk22	RAL-83	chr3R	24859894	24859970	4	408	404	1
CG13675	RAL-38	chr3L	8165984	8166133	8664	407	8257	1
CG13759,Vha36-3	RAL-894	chrX	2580879	2580988	413	14	399	1
CG1387	RAL-38	chrX	8385597	8385713	8664	404	8260	1
CG14082	RAL-228	chr3L	19136715	19136765	8665	9042	377	1
CG14244,CG14246	RAL-738	chr3R	26726879	26726942	8663	397	8266	1
CG14332	RAL-730	chr3R	17422508	17422574	6	406	400	1
CG14446	RAL-843	chrX	6226160	6226338	405	8663	8258	1
CG14626	RAL-367	chrX	1185035	1185107	8669	405	8264	1
CG14627	RAL-897	chrX	1181779	1181860	409	409	0	1
CG14659	RAL-138	chr3R	4831700	4831737	45	8679	8634	1
CG14708	RAL-908	chr3R	11553805	11553909	401	9	392	1
CG14752	RAL-158	chr2R	8590069	8590129	9050	8662	388	1
CG14837	RAL-508	chr3L	7559831	7559937	12	406	394	1
CG14926	RAL-805	chr2L	11343758	11343819	410	8685	8275	1
CG15021,nAChRbeta1	RAL-304_125b	chr3L	4436397	4436517	401	53	348	1
CG15068	RAL-737	chr2R	18390900	18390977	406	8665	8259	1
CG15219,CG31703	RAL-787	chr2L	22104853	22104868	8681	2	8679	1
CG15236	RAL-373_125b	chr2R	6829535	6829685	407	8666	8259	1
CG15278	RAL-356	chr2L	14876281	14876301	8663	25	8638	1
CG15483	RAL-375	chr2L	12761915	12761996	407	8663	8256	1
CG15695	RAL-894	chr3R	20841657	20841776	407	8664	8257	1
CG15711	RAL-280	chr2R	16364865	16364915	10	9054	9044	1
CG15741,CR43908	RAL-42	chrX	11838297	11838406	405	8672	8267	1

CG16772,CR45701,CG10680	RAL-239	chr2L	19963867	19963941	404	2	402	1
CG1678	RAL-176	chrX	21403215	21403298	8661	413	8248	1
CG16826	RAL-336	chr2L	13347334	13347376	389	19	370	1
CG16836	RAL-492	chr2R	18389057	18389089	9022	9053	31	1
CG1688,CR44208	RAL-83	chr2R	9788805	9788875	411	14	397	1
CG16964	RAL-176	chr2L	11947684	11947886	406	8677	8271	1
CG17003	RAL-502	chrX	20002395	20002471	403	8703	8300	1
CG17124	RAL-383	chr2L	10753867	10753980	3	404	401	1
CG17159,CG17162,CR43480	RAL-383	chrX	23091948	23092043	8688	407	8281	1
CG17287	RAL-45	chr2R	17158215	17158281	3	397	394	1
CG17321,CG33120	RAL-884	chr2L	18844894	18844944	401	4	397	1
CG17375	RAL-227	chr2L	6806125	6806193	7	405	398	1
CG17378	RAL-109_125b	chr2L	6774313	6774480	8665	393	8272	1
CG17549,fon	RAL-28	chr2L	19386637	19386686	397	8684	8287	1
CG17570	RAL-391_95b	chr2L	20264747	20264841	401	19	382	1
CG18258	RAL-714	chrX	17211924	17211978	8682	9063	381	1
CG18327,CG18324,CR44368	RAL-821_100b	chr2R	14168426	14168514	8672	9046	374	1
CG18672,CG18673	RAL-818	chr3R	31262111	31262184	405	4	401	1
CG2022	RAL-321	chr3R	4988731	4988849	405	8663	8258	1
CG2120	RAL-761	chrX	8135837	8135905	407	8663	8256	1
CG2574,CG32651	RAL-91	chrX	12408673	12408804	393	8643	8250	1
CG30039	RAL-787	chr2R	12051889	12051964	8669	410	8259	1
CG30187,Nup214	RAL-307_125b	chr2R	22918797	22918902	9034	40	8994	1
CG30369	RAL-28	chr2R	8259191	8259268	1	408	407	1
CG30389	RAL-138	chr2R	21156996	21157096	17	9058	9041	1
CG30416,CG9861	RAL-714	chr2R	23323441	23323526	8661	410	8251	1
CG30438	RAL-405	chr2R	5496474	5496542	404	8712	8308	1
CG30486,CR45468	RAL-373_125b	chr2R	12848796	12848962	404	8669	8265	1
CG31174	RAL-91	chr3R	21869592	21869652	403	7	396	1
CG31313,CG8066	RAL-256	chr3R	14567449	14567509	411	8675	8264	1
CG31998	RAL-748_125b	chr4	206229	206382	393	12	381	1
CG32395	RAL-360_125	chr3L	6668096	6668277	8665	409	8256	1
CG32444,CG32445	RAL-359	chr3L	21639140	21639206	38	392	354	1
CG32633	RAL-93	chrX	13561576	13561644	11	407	396	1
CG32643	RAL-93	chrX	12984427	12984490	8678	405	8273	1
CG32683	RAL-804	chrX	10452444	10452492	9046	7	9039	1
CG32751,CG32750	RAL-309	chrX	6196578	6196662	413	1	412	1
CG33268	RAL-642	chr3L	11591626	11591667	8687	393	8294	1
CG33272	RAL-57	chr3L	11510878	11511061	160	8605	8445	1
CG33286	RAL-804	chr3L	20974733	20974779	2	376	374	1
CG3330	RAL-409_125b	chr3R	27330718	27330879	8675	400	8275	1
CG33483	RAL-819	chr3R	30902450	30902526	356	4	352	1
CG33639	RAL-189	chrX	18225494	18225666	405	8669	8264	1
CG33702	RAL-383	chr3L	9513485	9513611	8660	410	8250	1
CG33970	RAL-714	chr3R	26318807	26318878	406	8665	8259	1
CG34054	RAL-40	chr2R	11410426	11410545	8664	407	8257	1
CG34106,beat-IIIa	RAL-491	chr2L	17130585	17130661	8666	412	8254	1
CG34193,CG4847	RAL-336	chr2R	17511438	17511505	403	8672	8269	1
CG34236	RAL-382	chr2R	13605626	13605802	407	8666	8259	1
CG34238	RAL-381	chr3L	10626135	10626171	411	9038	8627	1
CG34316,CR43471,tefu	RAL-563	chr3R	15231821	15231881	2	402	400	1
CG34320	RAL-820	chrX	1164491	1164521	1	29	28	1
CG34330	RAL-88	chrX	19069369	19069444	8663	403	8260	1
CG34384	RAL-837	chr3R	7814642	7814746	411	19	392	1
CG34402,CG10096,CG10097	RAL-361	chr3R	12393392	12393543	400	8685	8285	1
CG34452	RAL-228	chr3L	15687928	15687983	8663	9045	382	1
CG3604,CG16712	RAL-377	chr2L	3696985	3697086	406	17	389	1
CG42335	RAL-83	chr3R	21758375	21758455	414	8668	8254	1
CG42397	RAL-491	chr3L	11974051	11974114	405	2	403	1
CG42404,CG6499	RAL-373_125b	chr3R	15249623	15249706	9053	8717	336	1
CG42404	RAL-223	chr3R	15249666	15249834	8668	400	8268	1
CG42650	RAL-370	chr3R	7391446	7391554	408	8674	8266	1
CG42694,CG3831	RAL-405	chr2R	22941528	22941622	406	8663	8257	1
CG42758	RAL-370	chr3L	14567568	14567670	8669	396	8273	1
CG42784,CG42810	RAL-589	chr2L	13459746	13459808	397	8668	8271	1
CG42784	RAL-38	chr2L	13423026	13423139	9063	8663	400	1
CG42831	RAL-799	chr3L	10814466	10814535	8664	397	8267	1
CG42852	RAL-26	chr3L	16755188	16755259	409	8674	8265	1
CG43076	RAL-732	chrX	15223880	15223936	8668	396	8272	1
CG43190,CR44304,Sln	RAL-861	chr2R	11845105	11845128	9042	9065	23	1

CG43391,CG33269	RAL-338	chr3L	11581042	11581091	404	27	377	1
CG4374	RAL-21	chr3R	23471812	23471931	1	409	408	1
CG43795,CR43792	RAL-176	chr2R	23476709	23476991	409	37	372	1
CG4382	RAL-721	chr2L	9600915	9600971	8679	400	8279	1
CG43901,v	RAL-320	chrX	10923622	10923733	8665	399	8266	1
CG45060	RAL-712	chrX	9868697	9868713	8663	8679	16	1
CG4520	RAL-321	chr3R	15744506	15744620	405	8667	8262	1
CG45691	RAL-181	chr2L	17321049	17321132	411	8662	8251	1
CG5168,pie	RAL-149	chr2L	10423213	10423270	8673	9062	389	1
CG5397	RAL-362_125b	chr2L	1239726	1239898	3	9068	9065	1
CG5440	RAL-21	chr2L	1334203	1334299	5	9052	9047	1
CG5973	RAL-531	chr2L	7437771	7437873	400	8674	8274	1
CG7140	RAL-208	chr3L	22028378	22028450	8666	402	8264	1
CG7300,CR44181	RAL-707	chr2L	10695616	10695688	3	9066	9063	1
CG7510,scaRNA:MeU4-A65	RAL-818	chr3L	17628201	17628254	8676	393	8283	1
CG7512	RAL-913	chr3L	11513815	11514003	8671	409	8262	1
CG7560	RAL-40	chr3L	11424886	11424989	391	8663	8272	1
CG7582	RAL-894	chr3R	29681292	29681406	8	410	402	1
CG7742	RAL-177	chr2L	5214764	5215060	315	5	310	1
CG7777	RAL-375	chr2R	11439832	11439895	409	8682	8273	1
CG8299	RAL-377	chr2R	15941976	15942073	393	8	385	1
CG8343	RAL-805	chr2R	6179602	6179680	411	5	406	1
CG8483	RAL-362_125b	chr3R	13214224	13214356	19	9042	9023	1
CG8564	RAL-790	chr3L	7388340	7388382	8699	378	8321	1
CG8945	RAL-374	chrX	17088002	17088087	8674	399	8275	1
CG9098,Gal	RAL-437_125	chr2L	6007155	6007327	407	8671	8264	1
CG9313	RAL-757	chr2R	20979321	20979397	8665	405	8260	1
CG9465	RAL-195	chr2L	8772417	8772479	400	8674	8274	1
CG9722,CG42500	RAL-217	chr3R	14152819	14152875	405	8686	8281	1
CG9759	RAL-714	chr3R	13609735	13609763	28	8664	8636	1
CG9967,eys	RAL-820	chr2L	2312008	2312083	412	8670	8258	1
CR42844	RAL-721	chr2L	9305330	9305380	20	401	381	1
CR43334	RAL-646	chr3L	634372	634452	409	1	408	1
CR43426	RAL-894	chr3L	22849867	22849915	8716	8668	48	1
CR43490	RAL-367	chr3R	17288104	17288167	406	8679	8273	1
CR43613	RAL-75	chr3R	29534696	29534834	9063	8667	396	1
CR43625	RAL-426	chr3L	15437068	15437119	9021	409	8612	1
CR43626	RAL-849	chr3L	3252076	3252199	399	8711	8312	1
CR43715	RAL-383	chr2L	5434704	5434747	8663	41	8622	1
CR43751	RAL-765	chr2L	2182916	2182990	403	8665	8262	1
CR43811	RAL-21	chr2R	13792223	13792304	3	9042	9039	1
CR43819	RAL-787	chr2L	17085167	17085220	392	8675	8283	1
CR43883	RAL-315_125b	chr3L	820844	820995	5	9047	9042	1
CR43950	RAL-93	chr3L	16018428	16018477	9052	8676	376	1
CR44077	RAL-352	chr2L	3818522	3818557	18	9053	9035	1
CR44081	RAL-38	chr2L	6813051	6813169	8664	406	8258	1
CR44110	RAL-818	chrX	14519406	14519436	8662	21	8641	1
CR44317	RAL-441	chr3R	4600231	4600346	404	8665	8261	1
CR44367	RAL-101	chr2R	13855850	13855900	8668	9046	378	1
CR44376	RAL-821_100b	chr2R	16964566	16964676	9050	8664	386	1
CR44380	RAL-406	chr2R	16815884	16816007	412	8664	8252	1
CR44423	RAL-440	chrX	15636132	15636234	8666	401	8265	1
CR44458	RAL-409_125b	chr2R	15371727	15371891	11	403	392	1
CR44486	RAL-375	chr2L	18401224	18401292	395	8662	8267	1
CR44498	RAL-317	chrX	6001577	6001638	8675	400	8275	1
CR44503	RAL-409_125b	chr2R	19634502	19634656	399	6	393	1
CR44505	RAL-821_100b	chr2R	19684207	19684327	407	10	397	1
CR44545	RAL-301_125b	chr3L	10789818	10789966	8679	391	8288	1
CR44553	RAL-714	chr3L	12658556	12658620	8662	393	8269	1
CR44621	RAL-377	chrX	13396333	13396429	8689	409	8280	1
CR44685	RAL-40	chr3L	20652481	20652542	394	8700	8306	1
CR44691,Socs16D	RAL-908	chrX	17828020	17828133	7	412	405	1
CR44697	RAL-380	chr2L	4347613	4347683	6	405	399	1
CR44703,Cht9	RAL-406	chr2R	21069676	21069792	411	8670	8259	1
CR44732	RAL-437_125	chr2L	14430939	14431107	405	9	396	1
CR44856	RAL-181	chr2L	14576262	14576332	408	8667	8259	1
CR44967,CG43063	RAL-907	chr3R	13086006	13086053	9023	407	8616	1
CR44999,Gas8	RAL-303_125b	chrX	3504839	3505007	400	8668	8268	1
CR45231	RAL-59	chr2R	19786901	19786976	8664	403	8261	1
CR45239,CR45238	RAL-849	chr3L	6641931	6642092	400	8675	8275	1

CR45278	RAL-426	chr2R	13245092	13245202	401	3	398	1
CR45281,Or45b	RAL-42	chr2R	9559207	9559323	1	9066	9065	1
CR45285	RAL-908	chr2L	16965220	16965335	405	8664	8259	1
CR45319,CR45318	RAL-361	chr2R	11995621	11995784	9061	7	9054	1
CR45349	RAL-508	chr2L	15280325	15280430	399	5	394	1
CR45375,loh	RAL-41	chr2L	10442289	10442310	8690	8669	21	1
CR45377	RAL-774_125	chr2R	11706386	11706539	401	15	386	1
CR45379	RAL-217	chr3L	3864518	3864544	9046	409	8637	1
CR45393	RAL-819	chr3L	17947889	17948019	8667	411	8256	1
CR45403	RAL-732	chr3L	10239959	10240009	397	8683	8286	1
CR45414	RAL-492	chr3L	6937392	6937512	9071	8664	407	1
CR45422,Fie	RAL-395	chr3L	3892242	3892337	40	412	372	1
CR45431	RAL-41	chr3L	22498202	22498278	8666	407	8259	1
CR45434	RAL-818	chr3L	19454673	19454730	408	8687	8279	1
CR45533	RAL-189	chrX	7444953	7445087	8702	9063	361	1
CR45551,CR45550	RAL-136	chr3R	8229381	8229414	6	9052	9046	1
CR45552,CR45553	RAL-897	chr3R	30530415	30530489	1	9054	9053	1
CR45571	RAL-304_125b	chr3R	30612667	30612828	8674	399	8275	1
CR45574	RAL-900_125b	chr3R	8546620	8546784	8676	404	8272	1
CR45583	RAL-461	chr3R	12329209	12329318	406	8672	8266	1
CR45609	RAL-362_125b	chrX	18312118	18312282	403	11	392	1
CR45622	RAL-356	chrX	18228557	18228605	9046	8669	377	1
CR45625	RAL-40	chrX	12912751	12912874	8664	411	8253	1
CR45656,CR45657	RAL-373_125b	chr3R	25847686	25847846	402	8678	8276	1
CR45661	RAL-383	chr3L	22909663	22909776	411	8673	8262	1
CR45662,CG6356	RAL-437_125	chr3R	24283633	24283786	401	12	389	1
CR45673	RAL-492	chr3R	4635233	4635309	408	8708	8300	1
CR45694	RAL-884	chr2L	15841862	15841932	401	8666	8265	1
CR45743	RAL-374	chr3L	5378321	5378431	406	8	398	1
CR45748,CG43894	RAL-217	chr3L	12889846	12889902	16	402	386	1
CanA1	RAL-228	chr3R	31039131	31039212	8664	409	8255	1
Cht9,CR44703	RAL-787	chr2R	21067853	21067909	8675	394	8281	1
Cngl,CG9164	RAL-38	chrX	15385798	15385906	8663	395	8268	1
Con	RAL-320	chr3L	4963164	4963266	8664	400	8264	1
Cpr65Aw	RAL-765	chr3L	6142677	6142750	1	402	401	1
Cpr76Bb	RAL-774_125	chr3L	19519423	19519579	397	8930	8533	1
Cpr76Bd	RAL-440	chr3L	19532555	19532656	407	7	400	1
Csas,fln,CR45161	RAL-820	chr3L	19989245	19989329	411	8663	8252	1
Cul3,yuri,	RAL-584_125b	chr2L	15265044	15265206	6	397	391	1
CycD	RAL-42	chrX	15909743	15909838	401	8679	8278	1
Cyp12a5	RAL-492	chr3R	19132516	19132627	407	8672	8265	1
Cyp6a13	RAL-189	chr2R	8570194	8570351	394	9	385	1
Dfd	RAL-849	chr3R	6791337	6791500	6	394	388	1
Dh44-R1	RAL-83	chr2R	14378282	14378447	407	93	314	1
Dic2	RAL-738	chr3R	20502365	20502438	8664	9064	400	1
Dic61B,p130CAS	RAL-642	chr3L	152656	152708	22	402	380	1
Dscam1,CR45129	RAL-908	chr2R	7333707	7333819	8669	406	8263	1
Eaat2	RAL-301_125b	chr2L	744514	744646	400	8704	8304	1
Elo68beta,Elo68alpha	RAL-91	chr3L	11101780	11101858	411	8669	8258	1
Epac	RAL-732	chr2R	6797170	6797243	8666	403	8263	1
Fie,CR45422	RAL-385	chr3L	3892275	3892316	355	399	44	1
Fkbp14	RAL-338	chr2R	21498347	21498400	17	398	381	1
GABA-B-R3	RAL-324_100	chr2L	749482	749677	9031	8664	367	1
Gr10a,Or10a	RAL-381	chrX	11415813	11415865	23	403	380	1
Gr36c,CG31750	RAL-309	chr2L	17181696	17181744	8694	407	8287	1
Gr89a,decay	RAL-892	chr3R	16498998	16499114	9067	8664	403	1
GstO2,GstO1	RAL-40	chr3L	8523562	8523675	6	9070	9064	1
Gyc-89Da	RAL-42	chr3R	16472127	16472239	408	8	400	1
HP4,CG8042	RAL-325	chr3L	7975345	7975392	392	8665	8273	1
Hs6st,CG4459	RAL-301_125b	chr3R	19929098	19929259	11	400	389	1
Hsromega	RAL-703	chr3R	21310616	21310708	404	8668	8264	1
IM3,CG16836	RAL-26	chr2R	18388751	18388808	407	22	385	1
IM3	RAL-306_125b	chr2R	18388179	18388342	8672	399	8273	1
InR,CR43653	RAL-41	chr3R	21592563	21592616	8664	9046	382	1
Ir21a,CR43609	RAL-301_125b	chr2L	22422	22595	8667	9067	400	1
Jon99Fi	RAL-223	chr3R	30488617	30488779	401	8674	8273	1
Kaz1-ORFB	RAL-142	chr3L	217292	217415	8676	407	8269	1
Keap1	RAL-358	chr3R	17076891	17076922	8661	29	8632	1
Lkr	RAL-517	chr3L	5529648	5529719	8667	403	8264	1
Mipp1,CG43206	RAL-40	chr3L	16566773	16566888	403	8664	8261	1

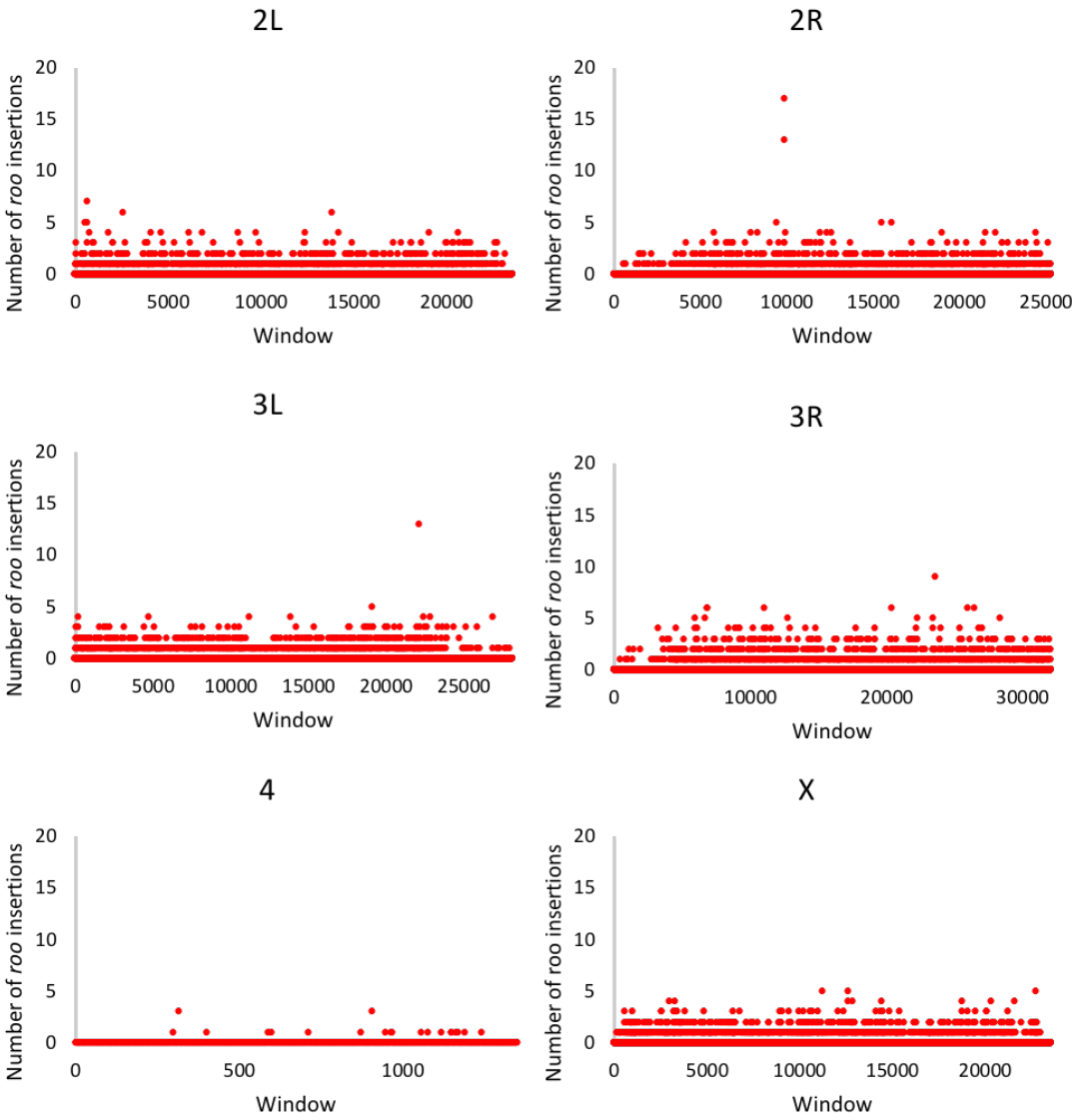
Msr-110	RAL-439	chr3L	5577619	5577742	8663	409	8254	1
Nmt,mthl7	RAL-822	chr3L	8178522	8178667	8670	409	8261	1
Oatp33Ea,CG5421	RAL-75	chr2L	12445932	12445981	9016	402	8614	1
Obp56c	RAL-233	chr2R	19701511	19701538	406	9046	8640	1
Odc1	RAL-382	chr2R	8004646	8004762	9042	8699	343	1
Or33a	RAL-321	chr2L	11934352	11934454	1	9053	9052	1
Or59a	RAL-91	chr2R	23441229	23441312	8664	9062	398	1
Peritrophin-15a,CG13385	RAL-109_125b	chr2L	8396326	8396476	16	394	378	1
Pu,CG4286	RAL-502	chr2R	21178622	21178728	404	7	397	1
RhoGEF64C	RAL-189	chr3L	4693332	4693496	9	401	392	1
SPR	RAL-223	chrX	5446289	5446467	8665	409	8256	1
Set1	RAL-705	chr3L	25107207	25107287	8664	408	8256	1
Sgs8,Sgs7	RAL-908	chr3L	11509447	11509554	8667	397	8270	1
Src64B	RAL-181	chr3L	4598998	4599078	8663	407	8256	1
TSG101	RAL-837	chr3L	16841956	16842072	8664	404	8260	1
Tab2,CG7461	RAL-849	chr2R	19304677	19304868	411	8664	8253	1
TotZ	RAL-761	chr3R	20877302	20877373	5	9067	9062	1
Trpm	RAL-804	chr2R	15307038	15307071	9041	411	8630	1
TyrRII	RAL-730	chr3R	17737009	17737071	6	9059	9053	1
Ugt86Dc,Ugt86Da	RAL-705	chr3R	11156705	11156786	8666	412	8254	1
Ugt86De,Ugt35b	RAL-359	chr3R	11168203	11168321	8662	409	8253	1
beat-Ib	RAL-730	chr2L	15939803	15939886	410	8662	8252	1
beat-Vc	RAL-535	chr3R	12783596	12783668	406	5	401	1
btl,CG8100	RAL-716	chr3L	14083259	14083320	8664	9052	388	1
c-cup	RAL-757	chr2L	1832500	1832573	8667	404	8263	1
cher	RAL-705	chr3R	17117299	17117368	8676	410	8266	1
cpo	RAL-761	chr3R	17937570	17937604	9051	8691	360	1
d4	RAL-189	chr2R	5258521	5258578	9	8730	8721	1
dpr12	RAL-849	chr2R	5804400	5804572	9066	8666	400	1
dpr2	RAL-325	chr2L	10964340	10964433	8662	404	8258	1
dpr3	RAL-307_125b	chr2L	2095149	2095277	400	8686	8286	1
ey	RAL-805	chr4	710934	711009	401	8661	8260	1
frac	RAL-589	chr3L	7562609	7562678	410	13	397	1
glob3	RAL-367	chr3R	5830088	5830151	395	8668	8273	1
grh	RAL-639	chr2R	17803389	17803439	8666	9045	379	1
grsm,CG14384,CG7381	RAL-189	chr3R	12967731	12967886	392	8673	8281	1
hig	RAL-804	chr2R	9247380	9247416	411	9038	8627	1
hoe2,hoe1	RAL-822	chr2L	4932785	4932864	9034	18	9016	1
klar,CR45118	RAL-913	chr3L	527247	527412	403	8674	8271	1
l(2)SH0834,Cdk5alpha	RAL-861	chr2L	10311027	10311105	413	8667	8254	1
l(3)76BDr,CG9279	RAL-513	chr3L	19544303	19544397	7	9052	9045	1
lcs	RAL-405	chrX	21397451	21397570	8663	406	8257	1
lectin-22C,CG42296	RAL-374	chr2L	2036355	2036476	8664	409	8255	1
lectin-29Ca,Acp29AB	RAL-38	chr2L	8393357	8393455	21	409	388	1
ligatin	RAL-382	chr3R	29151737	29151913	7	409	402	1
mbo,CG43630,CG6188	RAL-908	chr3R	12676482	12676513	9051	355	8696	1
mei-P26	RAL-371_125b	chrX	9166367	9166539	404	8669	8265	1
msn	RAL-324_100	chr3L	2586923	2587103	410	8735	8325	1
mthl4	RAL-509	chr2R	17447736	17447815	8663	407	8256	1
mthl6	RAL-217	chr3L	7485589	7485669	8664	408	8256	1
mthl7,Nmt	RAL-832	chr3L	8178606	8178648	9026	405	8621	1
nerfin-1	RAL-913	chr3L	908923	909044	8699	400	8299	1
olf186-F	RAL-639	chr2R	17844370	17844387	21	4	17	1
ord	RAL-441	chr2R	23275055	23275169	3	9070	9067	1
otk	RAL-443	chr2R	12020141	12020236	389	6	383	1
oys,CR44295	RAL-181	chr2R	10044570	10044650	8664	409	8255	1
pcm,CR43260	RAL-176	chrX	19485698	19485778	8666	410	8256	1
pdm3,cola,swif	RAL-646	chr2R	8364770	8364843	409	6	403	1
phl,CR33218	RAL-409_125b	chrX	2330188	2330335	9045	8671	374	1
psh,Hayan	RAL-555	chrX	18484234	18484304	9062	8664	398	1
psq,CG11883	RAL-757	chr2R	10617320	10617362	399	276	123	1
rib	RAL-426	chr2R	19270419	19270511	4	9050	9046	1
santa-maria	RAL-443	chr2L	7450155	7450262	8667	404	8263	1
sick,CR43827	RAL-367	chr2L	19922997	19923024	9045	409	8636	1
skpD	RAL-441	chrX	19812409	19812522	404	2	402	1
slif	RAL-309	chr3L	22881311	22881388	8664	404	8260	1
spri	RAL-307_125b	chrX	10585644	10585806	399	9	390	1
sty	RAL-890	chr3L	3424614	3424674	8665	9053	388	1
thetaTry	RAL-737	chr2R	11342973	11343037	400	8672	8272	1
unc80	RAL-26	chr3R	27662156	27662228	403	8667	8264	1

wb,CR43639	RAL-303_125b	chr2L	14291625	14291786	401	8675	8274	1
yellow-g	RAL-461	chr3L	2273120	2273242	8661	408	8253	1
yin,CG2930	RAL-325	chrX	3869031	3869090	3	9054	9051	1
Btd,Efr	RAL-317	chrX	6219616	6219679	7	398	391	2
Btd,Efr	RAL-318	chrX	6219616	6219689	8671	408	8263	2
CG10748	RAL-799	chr3L	12972348	12972428	8664	408	8256	2
CG10748	RAL-737	chr3L	12972349	12972424	1	404	403	2
CG12990,chas	RAL-804	chrX	17672081	17672153	409	7	402	2
CG12990,chas	RAL-887	chrX	17672076	17672177	394	5	389	2
CG12998	RAL-804	chrX	17218345	17218422	8664	405	8259	2
CG12998	RAL-887	chrX	17218327	17218437	2	401	399	2
CG13544	RAL-229_125b	chr2R	23140948	23141107	404	8681	8277	2
CG13544	RAL-373_125b	chr2R	23140988	23141107	410	50	360	2
CG14662	RAL-280	chr3R	4984152	4984211	8663	387	8276	2
CG14662	RAL-409_125b	chr3R	4984114	4984264	9	390	381	2
CG15661,CG4302	RAL-338	chr2R	21215307	21215324	9047	401	8646	2
CG15661,CG4302	RAL-821_100b	chr2R	21215270	21215323	43	9038	8995	2
CG15767,mir-4963	RAL-801_125b	chrX	5790831	5790965	8704	408	8296	2
CG15767,mir-4963	RAL-802_125b	chrX	5790799	5790965	8	401	393	2
CG15905	RAL-441	chr2R	19648141	19648206	9026	5	9021	2
CG15905	RAL-361	chr2R	19648067	19648231	402	8674	8272	2
CG17234	RAL-338	chr2L	2253010	2253033	9039	400	8639	2
CG17234	RAL-233	chr2L	2253010	2253043	9039	409	8630	2
CG1986	RAL-804	chrX	9911378	9911451	8672	409	8263	2
CG1986	RAL-395	chrX	9911347	9911477	3	409	406	2
CG2291	RAL-913	chr2R	8261443	8261601	8	394	386	2
CG2291	RAL-365	chr2R	8261485	8261548	8664	391	8273	2
CG31183	RAL-908	chr3R	15870201	15870318	8664	407	8257	2
CG31183	RAL-757	chr3R	15870226	15870296	5	404	399	2
CG31345	RAL-716	chr3R	12817426	12817485	19	406	387	2
CG31345	RAL-859	chr3R	12817410	12817487	3	408	405	2
CG3290,CG3292	RAL-280	chr2R	22205994	22206063	403	8670	8267	2
CG3290,CG3292	RAL-897	chr2R	22205988	22206067	409	2	407	2
CG33120,CG17321	RAL-239	chr2L	18844868	18844947	407	8665	8258	2
CG33120,CG17321	RAL-38	chr2L	18844850	18844951	405	8681	8276	2
CG33143	RAL-808_125b	chr2R	22613943	22614091	399	8670	8271	2
CG33143	RAL-727	chr2R	22613977	22614056	408	8663	8255	2
CG34462	RAL-832	chr3L	8328646	8328765	407	8663	8256	2
CG34462	RAL-822	chr3L	8328642	8328780	406	8662	8256	2
CG3777,CG13375	RAL-765	chrX	294819	294897	8661	408	8253	2
CG3777,CG13375	RAL-894	chrX	294806	294914	7	9068	9061	2
CG4213	RAL-350	chr2L	402294	402379	8667	410	8257	2
CG4213	RAL-358	chr2L	402297	402366	8670	397	8273	2
CG43403	RAL-374	chr2L	2620360	2620405	9021	403	8618	2
CG43403	RAL-91	chr2L	2620361	2620390	9042	409	8633	2
CG43750	RAL-303_125b	chr2L	2390676	2390852	9060	8661	399	2
CG43750	RAL-304_125b	chr2L	2390674	2390852	9062	8661	401	2
CG4409	RAL-716	chr2R	16442832	16442912	406	8662	8256	2
CG4409	RAL-732	chr2R	16442832	16442898	407	8676	8269	2
CG7298,obst-F	RAL-306_125b	chr3L	20201867	20202037	8670	404	8266	2
CG7298,obst-F	RAL-303_125b	chr3L	20201862	20202036	1	403	402	2
CG8046	RAL-40	chr2R	9164473	9164593	409	1	408	2
CG8046	RAL-821_100b	chr2R	9164888	9165017	8666	409	8257	2
CG8750	RAL-730	chr3L	13843006	13843061	18	401	383	2
CG8750	RAL-805	chr3L	13842986	13843066	8662	406	8256	2
CG9422,Tdc1	RAL-890	chr2R	6680209	6680290	8664	409	8255	2
CG9422,Tdc1	RAL-563	chr2R	6680219	6680287	8674	406	8268	2
CG9766	RAL-195	chr3R	4277814	4277890	407	8663	8256	2
CG9766	RAL-83	chr3R	4277857	4277921	406	8678	8272	2
CR41443	RAL-57	chr2R	5254667	5254775	391	8668	8277	2
CR41443	RAL-313	chr2R	5254676	5254738	406	8680	8274	2
CR43718	RAL-359	chr2L	396474	396589	8665	404	8261	2
CR43718	RAL-595	chr2L	396493	396574	8664	409	8255	2
CR43987,CG6893	RAL-712	chr3L	18694012	18694095	397	8669	8272	2
CR43987,CG6893	RAL-352	chr3L	18694067	18694086	8695	13	8682	2
CR44135,CG18547	RAL-45	chr3R	11978356	11978421	6	399	393	2
CR44135,CG18547	RAL-136	chr3R	11978374	11978421	8671	412	8259	2
CR44166	RAL-761	chr2R	6841459	6841535	8666	406	8260	2
CR44166	RAL-373_125b	chr2R	6841259	6841413	397	8664	8267	2
CR44575	RAL-367	chr2L	5809625	5809701	3	9070	9067	2

CR44575	RAL-109_125b	chr2L	5810188	5810333	8672	9045	373	2
CR44626	RAL-359	chr2R	19926333	19926447	403	1	402	2
CR44626	RAL-321	chr2R	19926327	19926449	409	8663	8254	2
CR44799,CR44800	RAL-313	chr2L	7636649	7636681	264	9043	8779	2
CR44799,CR44800	RAL-426	chr2L	7636610	7636697	8910	9041	131	2
CR45266	RAL-358	chr2R	12778901	12778928	8669	8696	27	2
CR45266	RAL-350	chr2R	12778896	12778975	8664	409	8255	2
CR45310	RAL-391_95b	chr2R	18226544	18226669	8664	413	8251	2
CR45310	RAL-324_100	chr2R	18226539	18226673	8664	412	8252	2
CR45371	RAL-832	chr2L	8345476	8345583	12	406	394	2
CR45371	RAL-822	chr2L	8345468	8345595	16	403	387	2
CR45447	RAL-365	chr2L	13096746	13096814	399	8667	8268	2
CR45447	RAL-913	chr2L	13096697	13096854	398	8677	8279	2
CR45557	RAL-832	chr3R	32026402	32026466	8682	9055	373	2
CR45557	RAL-822	chr3R	32026393	32026463	33	9025	8992	2
CheB38a	RAL-790	chr2L	20819013	20819140	413	8665	8252	2
CheB38a	RAL-804	chr2L	20818664	20818722	9	398	389	2
DAAM	RAL-374	chrX	1316129	1316228	21	408	387	2
DAAM	RAL-837	chrX	1316104	1316221	8660	401	8259	2
FMRFaR	RAL-358	chr3L	3001285	3001350	8670	400	8270	2
FMRFaR	RAL-350	chr3L	3001300	3001359	22	409	387	2
Gbs-76A	RAL-716	chr3L	19292940	19293013	8666	403	8263	2
Gbs-76A	RAL-382	chr3L	19292893	19293063	5	405	400	2
GstE11	RAL-765	chr2R	18503371	18503461	8671	412	8259	2
GstE11	RAL-849	chr2R	18503352	18503500	8702	401	8301	2
PRL-1	RAL-802_125b	chr2L	16250087	16250251	8665	394	8271	2
PRL-1	RAL-882	chr2L	16250134	16250191	8662	9046	384	2
beat-IIIb,CR44408	RAL-383	chr2L	16983000	16983113	5	406	401	2
beat-IIIb,CR44408	RAL-492	chr2L	16982802	16983110	8680	403	8277	2
futsch	RAL-375	chrX	1409966	1410040	410	4	406	2
futsch	RAL-223	chrX	1409921	1410087	401	8671	8270	2
heph,CR45556	RAL-911	chr3R	32013308	32013387	11	401	390	2
heph,CR45556	RAL-907	chr3R	32013301	32013388	4	9042	9038	2
l(3)mbn	RAL-911	chr3L	6128988	6129102	404	8666	8262	2
l(3)mbn	RAL-461	chr3L	6128983	6129099	409	5	404	2
let-7-C	RAL-362_125b	chr2L	18472460	18472631	403	4	399	2
let-7-C	RAL-310	chr2L	18471850	18471923	402	8664	8262	2
lush,CG9372	RAL-730	chr3L	19606125	19606206	411	8663	8252	2
lush,CG9372	RAL-805	chr3L	19606123	19606204	411	2	409	2
sqa	RAL-513	chr2R	9746663	9746771	408	8674	8266	2
sqa	RAL-377	chr2R	9745659	9745769	7	407	400	2

Additional file 7

Additional file 7. Genome-wide distribution of *de novo roo* elements found in 177 DGRP strains. Number of predicted *de novo roo* elements found in 177 DGRP strains inserted in 1kb windows in chromosomes 2, 3, 4, and X.



Additional file 8

Additional file 8A. PCR results of the five 1kb regions with more *roo* insertions predicted by TIDAL software.

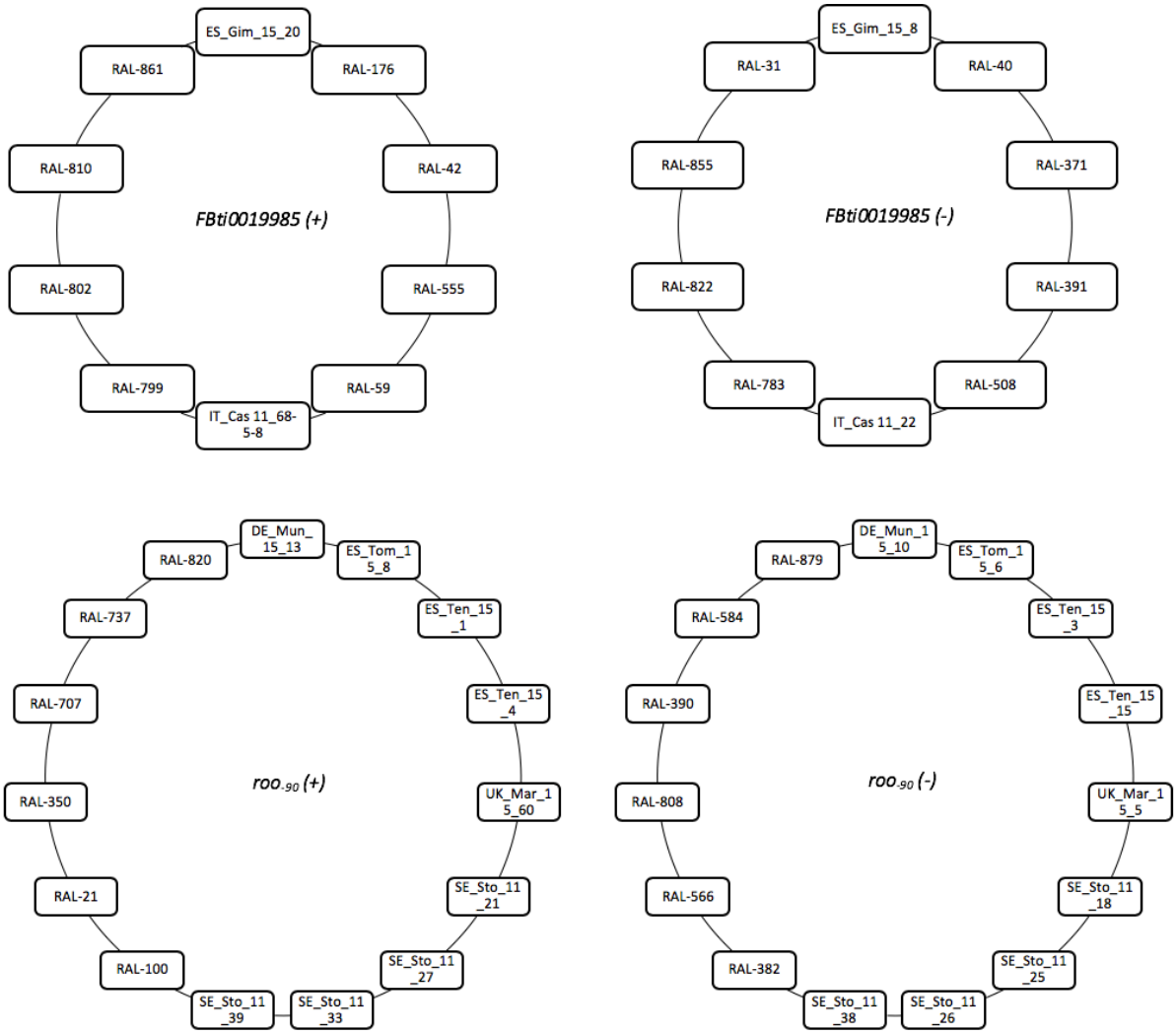
Region	Coordinates	Number of insertions	Location	Gene	Molecular function	Biological process	Comments	Strains used for PCR	Strains homozygous for the presence	Strains heterozygous for the presence	*roo* insertion site/s
1	2R:9,871,001..9,872,000	17	Fist intron, promoter, 5UTR	CG46338 (cbx) and CG18446	Nucleic acid binding	Unknown	Overlapping with FBti0019985	RAL-75, RAL-142, RAL-195, RAL-321, RAL-383, RAL-370, RAL-441, RAL-502, RAL-716	RAL-75, RAL-142, RAL-195, RAL-321, RAL-383, RAL-370, RAL-441, RAL-502, RAL-716		9,871,028 (*roo* $_{-68}$); 9,871,103 (*roo* $_{+7}$); 9,871,096 (*FBti0019985*); 9,871,052 (*roo* $_{-44}$) and 9,871,374 (*roo* $_{+278}$)
2	2R:9,870,001..9,871,000	13	Fist intron	CG46338 (cbx)	Unknown	Unknown	Overlapping with FBti0019984 and FBti0061742	RAL-21, RAL-88, RAL-177, RAL-332, RAL-336, RAL-358, RAL-437, RAL-707, RAL-712	RAL-21, RAL-88, RAL-177, RAL-332, RAL-336, RAL-358, RAL-437, RAL-707, RAL-712		9,871,006 (*roo* $_{-90}$)
3	3L:22,158,001..22,159,000	13	Intron	olf413	copper ion binding; dopamine beta-monooxygenase activity	octopamine biosynthetic process; dopamine catabolic process; oxidation-reduction process; norepinephrine biosynthetic process		RAL-59, RAL-91, RAL-358, RAL-437, RAL-502, RAL-508, RAL-737, RAL-776, RAL-853, RAL-890	RAL-59, RAL-91, RAL-358, RAL-437, RAL-502, RAL-508, RAL-737, RAL-776, RAL-853, RAL-890		22,158,407 and 22,158,736
4	3R:23,617,001..23,618,000	9	Intron	CG31145	phosphotransferase activity, alcohol group as acceptor; protein kinase activity; protein serine/threonine kinase activity	Protein phosphorylation		RAL-75, RAL-584, RAL-646, RAL-776, RAL-802, RAL-820, RAL-908	RAL-75, RAL-646, RAL-776, RAL-802, RAL-908		23,617,889 and 23,617,186
5	2L:634,001..635,000	7	Intergenic					RAL-405, RAL-502, RAL-555, RAL-732, RAL-737, RAL-802, RAL-907	RAL-555, RAL-732, RAL-737, RAL-907	RAL-405, RAL-502, RAL-802	634,692
6	3R:10,990,001..10,991,000	6	Intron	side-VI	Unknown	Unknown	Protein features are: CD80-like, immunoglobulin C2-set; Fibronectin type III; Fibronectin type III superfamily; Immunoglobulin subtype; Immunoglobulin subtype 2; Immunoglobulin-like domain; Immunoglobulin-like domain superfamily; Immunoglobulin-like fold.	RAL-239, RAL-321, RAL-461, RAL-502, RAL-584, RAL-737, RAL-799, RAL-801, RAL-802, RAL-907	RAL-239, RAL-502, RAL-584, RAL-737, RAL-799, RAL-801, RAL-907	RAL-321, RAL-802	10,990,797
7	3R:11,181,001..11,182,000	4	Intergenic					RAL-239, RAL-321, RAL-461, RAL-502, RAL-584, RAL-737, RAL-799, RAL-801, RAL-802, RAL-907	RAL-239, RAL-584, RAL-737, RAL-799, RAL-801, RAL-802	RAL-321, RAL-461, RAL-502, RAL-907	11,181,660

Additional file 8B. 1kb regions with at least 1 *roo* insertion predicted by TIDAL software. Regions checked by PCR are highlighted in yellow.

https://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-019-0152-9#Sec18

Additional file 9

Additional file 9. Schematic representation of the round-robin cross-design for outbred *FBti0019985* (+), *FBti0019985* (-), *roo*$_{-90}$ (+), and *roo*$_{-90}$ (-) generation.



Additional file 10

Additional file 10. Phylogenetic tree of the 20 *roo* solo-LTR found in *CG18446* promoter region and 115 other *roo* insertions annotated in the reference genome.

((FBti0020131:0.00000095275797606147,(((((FBti0019461:0.00439837781624463197,FBti0020047
:0.00023307253464348237):0.00000095275797606147,FBti0020145:0.00035271187171464168):0.00
023281084606104267,((FBti0060681:0.00011604025011133969,FBti0020129:0.000232863423625469
32):0.00046676668233772281,(FBti0019173:0.00000095275797606147,roo:0.00011623405602061341
):0.00058305255811374644):0.00011516023188488629):0.00011663279069259779,(((FBti0019278
:0.00023282041403568698,((FBti0019985inv:0.00469785587259081722,roo-
42:0.00000095275797606147):0.00468926943722107210,(((FBti0019018:0.00058148582611210198,
((((FBti0019019:0.00105104639512688491,(((FBti0059767:0.00000095275797606147,FBti0019752
:0.00023537517799854523):0.00034915006104648563,(((roo-
90inv:0.01447210032380415873,FBti0019435:0.00000095275797606147):0.00011347309127965815,
FBti0019016:0.00000095275797606147):0.00000095275797606147,FBti0019985*:0.00000095275797
606147):0.00000095275797606147):0.00011655679801793160,FBti0019544:0.0003820885247096210
7):0.00000095275797606147):0.00023183158614004543,FBti0019031:0.00050914921056243497):0.
00000095275797606147,((((((FBti0019021:0.00063830473722445956,FBti0019285:0.000382406296
57644982):0.00063851375645350048,(FBti0020063:0.00046536803304739644,(((FBti0019597:0.00
023268039795250702,FBti0020187:0.00023263323817910396):0.00105097371202392742,FBti001902
7:0.00034943707858249752):0.00000095275797606147,(((((roo-
56:0.00722146053247711955,FBti0020085:0.00000095275797606147):0.00000095275797606147,roo
-
393:0.00000095275797606147):0.00000095275797606147,FBti0019665:0.00000095275797606147):0
.00024285920304851474,(FBti0019363:0.00000095275797606147,FBti0019406:0.0000009527579760
6147):0.00034854274332509627):0.00000095275797606147,FBti0019402:0.00081733283113703890)
:0.00023198857846678245):0.00011631448462537597):0.00000095275797606147):0.0002325304253
5864636,FBti0019138:0.00023275896574487991):0.00000095275797606147,(FBti0019175:0.000232
64401472075890,FBti0019068:0.00034928531484234839):0.00000095275797606147):0.00000095275
797606147,((FBti0019431:0.00034906679424455840,(FBti0019436:0.00023558386087907835,FBti0
019539:0.00034916742576682934):0.00023278443508787923):0.00000095275797606147,FBti005976
8:0.00000095275797606147):0.00000095275797606147):0.00000095275797606147,((((FBti0019532
:0.00024286563864339176,FBti0020093:0.00034906694821776639):0.00000095275797606147,((FBt
i0019131:0.00060045060697787040,(FBti0062208:0.00023258269644251331,FBti0019196:0.000000
95275797606147):0.00000095275797606147):0.00023209651149350258,(FBti0019572:0.0000009527
5797606147,(FBti0019025:0.00000095275797606147,FBti0019028:0.00011662655296991606):0.000
00095275797606147):0.00058312147866092028):0.00000095275797606147):0.00000095275797606147,((FBti0019140****:0.00014244582780437872,((FBti0019022:0.00012741342342442502,(roo+278
:0.00000095275797606147,FBti0059830:0.00011250564610576127):0.00000095275797606147):0.00
000095275797606147,(roo+192:0.00839409425538830897,roo-
378:0.03865155723836171536):0.00124302960687344157):0.00000095275797606147):0.0000009527
5797606147,((FBti0019553:0.00081648337627686688,((FBti0019026:0.00046705793366260822,(((
FBti0063895:0.00250101677517159532,FBti0060416:0.00716154255858163398):0.007664082613404
13826,FBti0061311:0.01476771041343427973):0.00839667720731813311,FBti0062794:0.000000952
75797606147):0.00014520467810584970):0.00023215970782298898,(FBti0062347:0.0000009527579
7606147,(FBti0020106:0.00470223796401834232,roo+175**:0.00000095275797606147):0.00000095
275797606147):0.00073821227421529015):0.00000095275797606147):0.00011633789257115705,FBt
i0059731:0.00188597334361628993):0.00000095275797606147):0.00000095275797606147):0.00000
095275797606147,FBti0019148:0.00000095275797606147):0.00000095275797606147):0.0000009527
5797606147):0.00000095275797606147,((((((FBti0020150:0.00023307068913638361,(FBti0019067
:0.00102123635349991014,FBti0019083:0.00024640331085898034):0.00012317717218741889):0.00
011619153182216337,(((FBti0020103:0.00000095275797606147,(FBti0019409:0.0000009527579760
6147,FBti0019458:0.00058311219510323131):0.00000095275797606147):0.00000095275797606147,
((((FBti0019420:0.00232555216629887554,FBti0019029:0.00000095275797606147):0.00000095275
797606147,FBti0020100:0.00000095275797606147):0.00095831546389803669,FBti0020055:0.00100
421921590533049):0.00000095275797606147,FBti0019020:0.00224780379172555191):0.0001404036
1107722134):0.00000095275797606147,FBti0019098:0.00047231205365967421):0.000232811722185
76684):0.00011634057771903445,(FBti0019838:0.00068745147065812581,FBti0019421:0.00034905
259427792643):0.00000095275797606147):0.00011626664259961753,(FBti0019238:0.000349195996
57145048,(FBti0020170:0.00035034600222762761,FBti0059726:0.00083045673795600558):0.00011
567673968155889):0.00011537324750333728):0.00000095275797606147,((FBti0019339:0.00023287
340630343380,(FBti0019463:0.00090584538375607885,FBti0020069:0.00069892697773944583):0.0
0000095275797606147):0.00011623531977988501,FBti0019432:0.00000095275797606147):0.000116
24145390819475):0.00000095275797606147,(FBti0019416:0.00023536417093600616,((((FBti00201
66:0.00011630345824439974,(FBti0020076:0.00027512556430151598,(FBti0019547:0.00012465055
945452518,((FBti0019357:0.00064079839717357174,(FBti0019526:0.00055319361277025327,FBti0
019658:0.00262397180544743803):0.00031423637276460463):0.00064082740169193781,FBti002000
7:0.00023279463993793937):0.00000095275797606147):0.00034908577088545552):0.000116014856
45472104):0.00000095275797606147,FBti0019337:0.00034902746539072755):0.00011463958447734
700,FBti0020022:0.00140242721147641552):0.00069982131163746420,FBti0020154:0.00012318314
620785826):0.00000095275797606147):0.00000095275797606147):0.00000095275797606147):0.000
00095275797606147):0.00011636989875644272,(FBti0059769:0.00046513219141652764,FBti001961
9:0.00062634890493862326):0.00000095275797606147):0.00011618788032198710,(FBti0020014:0.

00057361975107086483,FBti0019060:0.00011629102898008995):0.000000952757976061 47):0.00000095657276053312):0.00011428918850263309):0.00034915851866713042,(FBti0019051:0.00046569309174295898,roo-
68***:0.00000095275797606147):0.00000095275797606147):0.00011605461010942130,(FBti0019556:0.00023250152314356498,((FBti0019393:0.00451722966219725644,FBti0019030:0.00000095275797606147):0.00064429177543738657,FBti0019374:0.00011636271612235554):0.00011612457250825544):0.00000095275797606147):0.00000095275797606147):0.00011620890192864236,((FBti0019615:0.00058156832385670531,((FBti0020121:0.00034882501595982546,(FBti0019439:0.00034900818393230660,FBti0019100:0.00069619359581228656):0.00000095275797606147):0.00011624529163398845,FBti0019101:0.00035129440882294478):0.00000095275797606147):0.00023260954339591379,(((FBti0019205:0.00075366762360008874,FBti0019450:0.00190593902250039255):0.00029905465510746371,FBti0019126:0.00023580523836786292):0.00000095275797606147,(FBti0059737:0.00070761662899994349,FBti0020038:0.00023255506368403066):0.00000095275797606147):0.00000095275797606147):0.00011633075237623261):0.00011621396011360528):0.00064081964024719276,FBti0019231:0.00000095275797606147,FBti0019015:0.00032025644648395283):0.0;

Additional file 11

Additional file 11. List of primers used for insertional cluster validation.

	Forward primer (5'-3')	Reverse primer (5'-3')	roo primer (5'-3')
FBti0020129	TGGCGGCACTTAATACACAT	AAAGGCGCGATGTAAAAAGTCTG	AGTCCCTTAGTGGGAGACCACAG
FBti0019432	ACGACGTTGAAGTGCAGTCT	CGCAGACACATGGTGGCAAT	AGTCCCTTAGTGGGAGACCACAG
FBti0019996	AACCGTAACCGTAGTTCGGC	TCGTGGGGTATACCTGTTGG	CACGTTTATAGCGGAGCCCT
FBti0019436	GGCGGAAAGCCGTGATTTTG	ATACATATCGGCGACCAGCG	AGTCCCTTAGTGGGAGACCACAG
FBti0019556	GGTGGCAAAATCCCCTGGTA	GCGCAACGATAGGGACGATA	TTGAGCAATGCACCCATGTC
FBti0019278	TGCCACGAGGTTCGTTGTT	GTAGAGTCTCACAGGGAAGGGA	CTGTGGTCTCCCACTAAGGGACT
FBti0019017	GAAACGTTGCCAAACACCGA	GCGGAGGAATTACACGCTCT	AGTCCCTTAGTGGGAGACCACAG
FBti0019394	ATGGACCACTGCGGATTCTC	TGTTTCAGAAATCGGTGGGCT	CTGTGGTCTCCCACTAAGGGACT
FBti0019439	TTACCCGCCAGGACACAAAA	GCATGGGATTCATCTTCATTGGC	AGTCCCTTAGTGGGAGACCACAG
FBti0020022	GCAAGAGGGCATCCATAGCA	AGCCGTAAAACCGAAAGGCA	AGTCCCTTAGTGGGAGACCACAG
FBti0020093	CTTTCATTGGCGCCTACGCT	AATGGCCAGGGCAATAAGGG	CTGTGGTCTCCCACTAAGGGACT
FBti0019068	ACAACACTCGGTAGGAAGCG	ACTGCGGTGTATTTCGAATTGTG	CTGTGGTCTCCCACTAAGGGACT
FBti0019597	ATAAGCCAAATCCAGGCGCA	CCACCTGGACGGACTTAACTG	AGTCCCTTAGTGGGAGACCACAG
FBti0019051	CGAATGAACTGCGATTTTGGTCT	CCTCTTTACGGCTCGAGTTGG	AGTCCCTTAGTGGGAGACCACAG
FBti0019532	GATGCGACCACCAACTACCA	TCCCAAAGCTATTCCGTGGC	AGTCCCTTAGTGGGAGACCACAG
FBti0020047	CGCCAACCAATTTGCTCCAC	ATGTGTGACGACACGACTGG	AGTCCCTTAGTGGGAGACCACAG
FBti0059710	AGCCAAACAGTTGGCAGTCT	CGCCACATTTTCGCTCAAACA	CTGTGGTCTCCCACTAAGGGACT
FBti0059714	AGCCAAATGCACACGAGTCT	CTGGAATAGATTACGCAGGTCC	TGCACCTTTTCCACCTTTCC
FBti0062208	ACGGACTTGCGATTCCAGAG	TCCCCTAGAAGGGCTTCTGAAA	AGTCCCTTAGTGGGAGACCACAG
FBti0019285	AGGACACATAGGAAAATGAAAACAA	ACTGAACTGTGAAAGGGGCTG	CTGTGGTCTCCCACTAAGGGACT
FBti0019393	GGAATCGGATCGCTGACACT	ACTGTCGTTTTGGGACAAGGT	AGTCCCTTAGTGGGAGACCACAG
FBti0019463	CCAGGAGCAGACAATGAGCA	GCAAGTGCATGGCCCAATTT	AGTCCCTTAGTGGGAGACCACAG
FBti0019665	CTTTTTGGCTCCACACGGTC	TTTGAACTGGTCCTCCATCCG	CTGTGGTCTCCCACTAAGGGACT
FBti0059664	CCTCTTACGGGAAGTGGTGG	AACTCATTGTGGGTTCGGCT	AAAGGCGTTCTAAGGCGACA
FBti0059662	AGAGGCTGCAAACTAAGCAT	ACAATTCTCTTGCGCCAAATTCA	AGTCCCTTAGTGGGAGACCACAG
FBti0019450	CGCCTGTGGCTTCTTACGAT	TTTTCCCGCACCAGACAACT	AGTCCCTTAGTGGGAGACCACAG

Promoter region	Cluster F primer (5'-3')	Cluster R primer (5'-3')
1	GTCGAAAGTCCGAAGCGGTA	TGTGTTTTGATTGGTCTTGGAAA
3	TGCAGATGGAGATGACGTTGG	AGTTTCGGCTCTGTTTGTTGG
4	GTAGTCAACGCCATTCCCCA	CCCGCCTCCGTTTTTATCCT
5	ACATCGGAGCTTGGAGCTTG	CCGACGAACCCTTGACTGAC
6	TTTTCCGGCCACAGTTTTCC	GCGCCACTGCAGGTTTTATT
7	CGTCACTGCGAGAGATACGC	CAGCTCCTCAAGTAGTGCCTT
8	GCACGGAATCCCTGTCAACT	TTTCAGACAAGCTCGTGGGC
9	CCATTTGTGTTGTCACCTTAAACT	ATGTCACATGTCGCATCAGC
10	GCTGAACCGTGTTTAAGCGAA	TTTGGTCTGCCCAACTGCTA
11	GCGAGTGGCGATAAAACGAG	AGCCCCTAGCTACCGATTGA
12	GCGGGAAAATCTGAAGGGGA	CTTGTTCGCTCTTTGCGCTT
13	GGGATCGACGAGCGATAATTG	CAGATGCGCCTCAAGAACAC
14	GCGGCAAGCATTTTACCGTC	GGGTTATTATCGCCACGGCA
15	TCTGTTATCCTTCTCGGATCTCCT	ATTGAACTGGTGTGTGTGCT
16	CCAGAAGAGCAGAGTCGTCC	AGCTGCACTGAGATCGAGGA
17	GGAGATGTGACCCCAAGGTG	ACCACCCTCGCTATGCTTTT
18	GTCCCCAAAATACCCAACGGA	CCTATGGCCGTTCCACCATC
19	AGCTAGTGGAAGTGCGAACG	GCAACCTTACAGAACGGAGGT
20	AGGAGGGGAATGCAACCATC	GAAATCCAACCGAACCGTCG
21	TGGTTGGGGCTTGGCTAAAA	GCAGAGTTGTGCAAAGCGAA
22	GCCGTTACGTTTGAAGCTCG	TCATTTTTGGTACCCCGCCC
23	TTGTCCTGCGAAAGAGGTCC	AGATCGCACCGATCGAGAAC
24	TTCCAGCGTAGTCATCCCTG	CATTTCATTGTGTTGGCGGC
25	AAGCTCGCTCTCTTACCAACC	GGAGAGTGTGGTGAACGGTG
26	CACTTGTCCGTATGCACGTC	GCGGTCTGGAGCCCATTTAT
27	AGTTGTTGTAGACAGGGGCG	TGAACAAGCGCAAAGCGAAA
28	GTCTACACGTTGGTCGGGAA	TGATCTGCCAACCGGATACC

rooL primer (5'-3')	AGTCCCTTAGTGGGAGACCACAG
rooL2 primer (5'-3')	CTGTGGTCTCCCACTAAGGGACT

Region	Cluster F primer (5'-3')	Cluster R primer (5'-3')
1	GCGAGGGCTTCGCCTTTTAT	GAATTTCTGGTTCGAGCGGC
2	GCGAGGGCTTCGCCTTTTAT	GAATTTCTGGTTCGAGCGGC
3	AGTTTCTCCTGGCCAACGAC	GCTTTTCAACTGCGGAGCAC
4	CCTTTATTTGCCGCTGCTGTC	TTGTTTCGGCCCCGAATATG
5	CCGATCGACTCGGGTAGTGA	CTCACATGGATTCGGGCAGA
6	ACTTCCGGTTCCTGCTGTG	CCGAGCACGGGGGAAAAT
7	AGCAAACAAAGAGCCCCAAC	GTGGTCTGGGTGTATTCAAGC

rooL primer (5'-3')	AGTCCCTTAGTGGGAGACCACAG
rooL2 primer (5'-3')	CTGTGGTCTCCCACTAAGGGACT

7.3. Supplementary material: A versatile transposable element affects the expression of a transcription factor depending on the developmental stage and the environmental conditions in *Drosophila melanogaster*

Figure S1. A) Normalized expression of *CG18446* and *cbx* in the *CG18446*-RNAi strain compared with the *ActGAL4* parental strain. **B)** Normalized expression of *CG18446* in cbx^{c00428} mutant strain compared with the WT strain. We did not used *CG18446*-RNAi and cbx^{c00428} strains for further experiments.

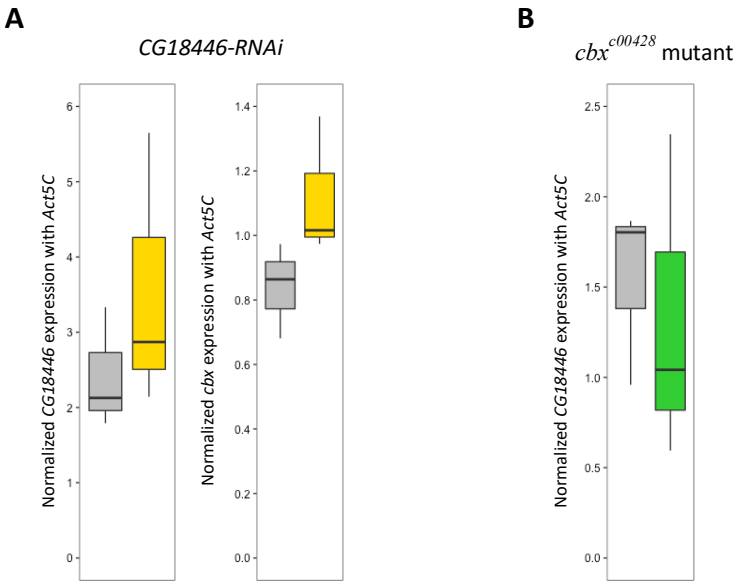


Table S1. Details of the mutant, RNAi and WT strains used.

Name	Gene	Genotype	WT used	Bloomington stock	Expression changes
CG18446^{MI02952}	*CG18446*	y[1] w[*]; Mi{y[+mDint2]=MIC}cbx[MI02952]	y[1] w[67c23] (#6599)	36170	*CG18446*
cbx$^{0741\text{-}G4}$	*cbx*	w[1118]; PBac{w[+mC]=IT.GAL4}CG46338[0741-G4]	w[1118]	63767	*CG18446* and *cbx*
CG18446-RNAi	*CG18446*	y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00618}attP2	w;act GAL4/TM6+tb	33735	-
cbx^{c00428}	*cbx*	w[1118]; PBac{w[+mC]=PB}CG46338[c00428]	w[1118]	10067	-

Table S2. qRT-PCR results of *CG18446* and *cbx* genes in the mutant and RNAi strains used.

	Ct *Actin5C*	Ct *CG18446*	delta Ct (*Actin* - *CG18446*)	2^deltaCt	mean	SD	SEM	t.test
CG18446 ^{MI02952}	23,25	24,05	-0,80	0,574	0,565	0,087	0,050	0,015
CG18446 ^{MI02952}	23,69	24,32	-0,63	0,646				
CG18446 ^{MI02952}	22,89	23,97	-1,08	0,473				
WT #6599	22,76	22,01	0,75	1,682	2,143	0,667	0,385	
WT #6599	23,59	22,71	0,88	1,840				
WT #6599	22,94	21,40	1,54	2,908				
cbx $^{0741\text{-}G4}$	25,24	25,66	-0,42	0,750	0,662	0,089	0,051	0,005
cbx $^{0741\text{-}G4}$	25,68	26,49	-0,81	0,572				
cbx $^{0741\text{-}G4}$	25,43	26,02	-0,59	0,665				
WT w1118	25,09	24,22	0,87	1,834	1,561	0,267	0,154	
WT w1118	25,60	25,22	0,38	1,300				
WT w1118	26,25	25,62	0,63	1,549				
CG18446-RNAi	26,14	24,62	1,52	2,870	3,554	1,850	1,068	0,385
CG18446-RNAi	25,81	23,31	2,50	5,649				
CG18446-RNAi	26,51	25,41	1,10	2,143				
WT ActGAL4	26,66	25,57	1,09	2,127	2,417	0,810	0,468	
WT ActGAL4	26,52	25,68	0,84	1,791				
WT ActGAL4	25,82	24,08	1,74	3,332				
cbx ^{c00428}	24,32	24,26	0,06	1,042	1,328	0,910	0,525	0,739
cbx ^{c00428}	22,95	21,72	1,23	2,346				
cbx ^{c00428}	21,34	22,09	-0,75	0,595				
WT w1118	22,83	22,89	-0,06	0,959	1,543	0,506	0,292	
WT w1118	23,02	22,17	0,85	1,803				
WT w1118	24,22	23,32	0,90	1,866				

	Ct *Actin5C*	Ct *cbx*	delta Ct (*Actin* - *cbx*)	2^deltaCt	mean	SD	SEM	t.test
CG18446 ^{MI02952}	23,25	24,19	-0,94	0,521	0,707	0,217	0,125	0,400
CG18446 ^{MI02952}	23,69	23,77	-0,08	0,946				
CG18446 ^{MI02952}	22,89	23,50	-0,61	0,655				
WT #6599	22,76	23,37	-0,61	0,655	0,911	0,305	0,176	
WT #6599	23,59	23,86	-0,27	0,829				
WT #6599	22,94	22,62	0,32	1,248				
CG18446-RNAi	26,14	26,11	0,02	1,016	1,120	0,217	0,125	0,138
CG18446-RNAi	25,81	25,36	0,45	1,369				
CG18446-RNAi	26,51	26,54	-0,04	0,974				
WT ActGAL4	26,66	26,70	-0,04	0,973	0,840	0,147	0,085	
WT ActGAL4	26,52	27,08	-0,55	0,681				
WT ActGAL4	25,82	26,03	-0,21	0,864				

Table S3. Infection assay results with CG18446^{MI02952} mutant strain.

Males		TIME (hours)										
STRAIN	**TREATMENT**	**13**	**14,5**	**19**	**24,5**	**37**	**62**	**68**	**84,5**	**110**	**132**	**157**
WT #6599	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Nonstress 2**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Nonstress 3**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Infected 1**	0	0	0	0	1	4	4	4	4	4	4
WT #6599	**Infected 2**	1	1	1	2	2	5	5	5	6	6	6
WT #6599	**Infected 3**	0	0	1	1	1	2	2	2	2	2	2
WT #6599	**Infected 4**	0	0	0	0	1	2	2	3	3	3	3
WT #6599	**Infected 5**	0	0	0	1	1	1	1	1	1	1	2
WT #6599	**Infected 6**	0	0	0	0	0	1	1	1	1	1	1
WT #6599	**Infected 7**	0	0	0	0	0	2	2	2	2	3	5
WT #6599	**Infected 8**	1	1	1	1	1	2	2	2	2	2	3
WT #6599	**Infected 9**	1	1	1	1	1	1	1	1	1	1	2
WT #6599	**Infected 10**	0	0	0	0	0	2	2	2	3	3	3
CG18446 ^{MI02952}	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0	0
CG18446 ^{MI02952}	**Nonstress 2**	0	0	0	0	0	0	0	0	0	0	0
CG18446 ^{MI02952}	**Nonstress 3**	0	0	0	0	0	0	0	0	0	0	0
CG18446 ^{MI02952}	**Infected 1**	0	0	0	0	4	8	8	8	8	8	9
CG18446 ^{MI02952}	**Infected 2**	1	1	3	6	8	9	9	9	9	9	10
CG18446 ^{MI02952}	**Infected 3**	0	0	0	3	6	10	10	10	10	10	10
CG18446 ^{MI02952}	**Infected 4**	1	2	2	5	6	7	7	7	7	7	8
CG18446 ^{MI02952}	**Infected 5**	0	1	2	4	6	7	7	7	9	9	9
CG18446 ^{MI02952}	**Infected 6**	0	0	2	4	9	9	9	9	9	9	10
CG18446 ^{MI02952}	**Infected 7**	0	0	1	3	5	7	7	7	7	7	7
CG18446 ^{MI02952}	**Infected 8**	0	0	1	3	4	7	7	8	8	8	9
CG18446 ^{MI02952}	**Infected 9**	0	1	2	3	7	7	7	9	9	9	9
CG18446 ^{MI02952}	**Infected 10**	0	0	1	4	4	8	8	9	9	9	10

Log Rank (Mantel-Cox)
0.000

Females		TIME (min)										
STRAIN	**TREATMENT**	**13**	**14,5**	**19**	**24,5**	**37**	**62**	**68**	**84,5**	**110**	**132**	**157**
WT #6599	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Nonstress 2**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Nonstress 3**	0	0	0	0	0	0	0	0	0	0	0
WT #6599	**Infected 1**	0	0	0	2	2	3	3	3	4	5	5
WT #6599	**Infected 2**	1	1	1	3	7	9	9	9	9	9	9
WT #6599	**Infected 3**	0	0	0	1	9	10	10	10	10	10	10
WT #6599	**Infected 4**	1	1	3	4	6	8	8	8	8	9	9
WT #6599	**Infected 5**	0	0	1	1	1	3	3	3	3	3	3
WT #6599	**Infected 6**	3	3	3	2	4	6	6	6	6	6	6
WT #6599	**Infected 7**	0	0	0	0	0	2	2	2	2	3	3
WT #6599	**Infected 8**	3	3	4	5	8	10	10	10	10	10	10
WT #6599	**Infected 9**	0	0	2	3	3	4	4	5	5	5	5
WT #6599	**Infected 10**	2	2	2	3	5	7	7	8	8	8	8
CG18446 ^{MI02952}	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0	0
CG18446 ^{MI02952}	**Nonstress 2**	0	0	0	0	0	0	0	1	1	1	1
CG18446 ^{MI02952}	**Nonstress 3**	0	0	0	0	0	0	0	0	0	0	0
CG18446 ^{MI02952}	**Infected 1**	4	7	8	9	10	10	10	10	10	10	10
CG18446 ^{MI02952}	**Infected 2**	1	2	3	7	9	10	10	10	10	10	10
CG18446 ^{MI02952}	**Infected 3**	2	2	6	6	7	9	9	9	9	9	9
CG18446 ^{MI02952}	**Infected 4**	2	5	7	7	8	10	10	10	10	10	10
CG18446 ^{MI02952}	**Infected 5**	3	3	7	9	9	10	10	10	10	10	10
CG18446 ^{MI02952}	**Infected 6**	2	3	6	9	9	9	9	9	10	10	10
CG18446 ^{MI02952}	**Infected 7**	5	5	7	8	8	9	9	10	10	10	10
CG18446 ^{MI02952}	**Infected 8**	0	2	4	5	7	9	9	9	9	9	9
CG18446 ^{MI02952}	**Infected 9**	3	3	5	6	7	8	8	9	9	9	9
CG18446 ^{MI02952}	**Infected 10**	7	6	10	10	10	10	10	10	10	10	10

Log Rank (Mantel-Cox)
0.000

Table S4. Cold-stress assay results with *CG18446* and *cbx* mutant strains.

		Total emerged flies	% Survival
WT #6599	**Nonstress 1**	26	86,7
	Nonstress 2	21	70,0
	Nonstress 3	24	80,0
	Nonstress 4	21	70,0
	Nonstress 5	19	63,3
	Cold 1	15	50,0
	Cold2	15	50,0
	Cold 3	15	50,0
	Cold 4	16	53,3
	Cold 5	16	53,3
CG18446 ^{MI02952}	**Nonstress 1**	8	26,7
	Nonstress 2	16	53,3
	Nonstress 3	18	60,0
	Nonstress 4	14	46,7
	Nonstress 5	16	53,3
	Cold 1	4	13,3
	Cold2	4	13,3
	Cold 3	7	23,3
	Cold 4	4	13,3
	Cold 5	2	6,7

Two-way ANOVA

Effect	Sig.
Genotype	0.000
Treatment	0.000
Genotype*Treatment	0.105

		Total emerged flies	% Survival
WT w1118	**Nonstress 1**	13	86,7
	Nonstress 2	14	93,3
	Nonstress 3	14	93,3
	Cold 1	16	53,3
	Cold2	19	63,3
	Cold 3	27	90,0
	Cold 4	20	66,7
	Cold 5	15	50,0
cbx $^{0741\text{-}G4}$	**Nonstress 1**	14	46,7
	Nonstress 2	22	73,3
	Nonstress 3	21	70,0
	Cold 1	12	40,0
	Cold2	16	53,3
	Cold 3	14	46,7
	Cold 4	14	46,7
	Cold 5	11	36,7

Two-way ANOVA

Effect	Sig.
Genotype	0.002
Treatment	0.003
Genotype*Treatment	0.538

Table S5. qRT-PCR results for *cbx* gene in laboratory outbred flies with and without *FBti0019985* under different stress conditions.

Immune-stress	Ct *Actin5C*	Ct *cbx*	delta Ct (*Actin* - *cbx*)	2^deltaCt	mean	SD	SEM
outbred FBti0019985 + C1	28.76	27.67	1.10	2.140	2.336	0.582	0.336
outbred FBti0019985 + C2	26.98	26.08	0.91	1.877			
outbred FBti0019985 + C3	27.52	25.94	1.58	2.990			
outbred FBti0019985 + T1	24.80	24.09	0.71	1.639	2.163	0.687	0.397
outbred FBti0019985 + T2	25.97	25.03	0.93	1.910			
outbred FBti0019985 + T3	26.88	25.32	1.56	2.941			
outbred FBti0019985 - C1	27.96	27.53	0.43	1.348	2.838	1.648	0.952
outbred FBti0019985 - C2	28.02	25.82	2.20	4.608			
outbred FBti0019985 - C3	27.83	26.48	1.35	2.557			
outbred FBti0019985 - T1	27.18	26.62	0.57	1.483	1.321	0.259	0.150
outbred FBti0019985 - T2	26.85	26.89	-0.04	0.971			
outbred FBti0019985 - T3	26.84	26.63	0.21	1.158			

Cold-stress	Ct *Actin5C*	Ct *cbx*	delta Ct (*Actin* - *cbx*)	2^deltaCt	mean	SD	SEM
outbred FBti0019985 + C1	23,23	23,84	-0,61	0,654	0,563	0,127	0,073
outbred FBti0019985 + C2	22,66	23,36	-0,70	0,617			
outbred FBti0019985 + C3	22,07	23,33	-1,26	0,419			
outbred FBti0019985 + T1	21,58	22,72	-1,13	0,456	0,687	0,335	0,193
outbred FBti0019985 + T2	22,68	23,58	-0,90	0,534			
outbred FBti0019985 + T3	22,76	22,66	0,10	1,071			
outbred FBti0019985 - C1	22,74	22,93	-0,19	0,875	0,814	0,114	0,066
outbred FBti0019985 - C2	23,05	23,23	-0,18	0,884			
outbred FBti0019985 - C3	23,91	24,46	-0,55	0,682			
outbred FBti0019985 - T1	20,66	21,35	-0,69	0,621	0,708	0,148	0,086
outbred FBti0019985 - T2	23,44	24,12	-0,68	0,624			
outbred FBti0019985 - T3	22,90	23,08	-0,18	0,880			

Table S6. qRT-PCR results for *CG18446* gene in laboratory outbred and CRISPR-mutant stains under immune-, cold-, and ethanol-stress conditions.

Female guts Rep.1 Immune-stress Total transcript *CG18446*	**Ct *Actin5C***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred *FBti0019985* + C1	25.38	30.22	-4.84	0,035	0.065	0.028	0.016
Outbred *FBti0019985* + C2	23.82	27.63	-3.81	0,072			
Outbred *FBti0019985* + C3	24.28	27.76	-3.48	0,09			
Outbred *FBti0019985* + T1	22.04	25.62	-3.58	0,084	0.185	0.089	0.051
Outbred *FBti0019985* + T2	23.57	25.58	-2.01	0,249			
Outbred *FBti0019985* + T3	24.10	26.26	-2.16	0,223			
Outbred *FBti0019985* - C1	25.19	27.76	-2.57	0,168	0.769	0.132	0.093
Outbred *FBti0019985* - C2	24.89	25.10	-0.21	0,862			
Outbred *FBti0019985* - C3	24.84	25.41	-0.57	0,676			
Outbred *FBti0019985* - T1	24.64	27.50	-2.86	0,138	0.069	0.060	0.035
Outbred *FBti0019985* - T2	24.16	28.85	-4.69	0,039			
Outbred *FBti0019985* - T3	24.32	29.38	-5.06	0,03			

Female guts Rep.1 Immune-stress Long transcript *CG18446*	**Ct *Actin5C***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred *FBti0019985* + C1	22,49	38,89	-16,40	1,15E-05	3,16E-05	1,84E-05	1,06E-05
Outbred *FBti0019985* + C2	20,95	35,74	-14,79	3,53E-05			
Outbred *FBti0019985* + C3	22,19	36,54	-14,35	4,78E-05			
Outbred *FBti0019985* + T1	19,21	32,06	-12,86	1,35E-04	1,15E-04	4,69E-05	2,71E-05
Outbred *FBti0019985* + T2	20,37	34,37	-14,00	6,09E-05			
Outbred *FBti0019985* + T3	21,36	34,09	-12,72	1,48E-04			
Outbred *FBti0019985* - C1	23,00	0,00					
Outbred *FBti0019985* - C2	21,46	0,00					
Outbred *FBti0019985* - C3	21,62	45,75					
Outbred *FBti0019985* - T1	21,35	0,00					
Outbred *FBti0019985* - T2	21,05	0,00					
Outbred *FBti0019985* - T3	21,35	45,25					

Female guts Rep.2 Immune-stress Total transcript *CG18446*	**Ct *Actin***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446***	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred FBti0019985 + C1	18,56	25,22	-6,65	0,010	0,011	0,0012	0,001
Outbred FBti0019985 + C2	19,72	26,08	-6,36	0,012			
Outbred FBti0019985 + C3	18,69	25,27	-6,57	0,010			
Outbred FBti0019985 + T2	18,83	23,35	-4,52	0,044			
Outbred FBti0019985 + T3	18,35	23,12	-4,77	0,037			
Outbred FBti0019985 - C2	20,99	23,77	-2,78	0,146			
Outbred FBti0019985 - C3	21,59	24,66	-3,08	0,119			
Outbred FBti0019985 - T1	18,45	25,25	-6,80	0,009	0,010	0,0018	0,001
Outbred FBti0019985 - T3	19,27	25,70	-6,43	0,012			
FBti0019985^{CRISPR1} C1	32,60	36,13	-3,53	0,086	0,070	0,0231	0,013
FBti0019985^{CRISPR1} C3	32,24	36,46	-4,22	0,054			
FBti0019985^{CRISPR1} T1	18,50	28,56	-10,06	0,001	0,001	0,0009	0,001
FBti0019985^{CRISPR1} T2	19,63	28,31	-8,68	0,002			
FBti0019985^{CRISPR1} T3	18,67	29,04	-10,36	0,001			
FBti0019985^{CRISPR2} C1	18,49	27,85	-9,36	0,002	0,002	0,0005	0,000
FBti0019985^{CRISPR2} C2	19,12	27,78	-8,66	0,002			
FBti0019985^{CRISPR2} C3	18,54	27,28	-8,74	0,002			
FBti0019985^{CRISPR2} T1	32,62	37,61	-4,99	0,031	0,048	0,0229	0,013
FBti0019985^{CRISPR2} T2	32,13	35,88	-3,75	0,074			
FBti0019985^{CRISPR2} T3	32,09	36,77	-4,68	0,039			

Male guts Immune-stress Total transcript *CG18446*	**Ct *Actin***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446***	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred *FBti0019985* + C1	25,34	34,36	-9,02	0,00192	0,00234	0,00155	0,00089
Outbred *FBti0019985* + C2	23,96	31,91	-7,94	0,00406			
Outbred *FBti0019985* + C3	23,82	33,72	-9,90	0,00105			
Outbred *FBti0019985* + T1	23,57	27,27	-3,70	0,07716	0,06997	0,02715	0,01567
Outbred *FBti0019985* + T2	22,99	27,63	-4,65	0,03995			
Outbred *FBti0019985* + T3	21,48	24,91	-3,43	0,09279			
Outbred *FBti0019985* - C1	23,31	34,55	-11,25	0,00041	0,00067	0,00033	0,00019
Outbred *FBti0019985* - C2	23,78	34,60	-10,82	0,00055			
Outbred *FBti0019985* - C3	24,97	34,88	-9,92	0,00103			
Outbred *FBti0019985* - T1	23,52	31,44	-7,93	0,00410	0,00815	0,00410	0,00236
Outbred *FBti0019985* - T2	23,24	29,58	-6,35	0,01229			
Outbred *FBti0019985* - T3	24,87	31,83	-6,96	0,00806			
FBti0019985^{CRISPR1} C1	24,60	40,55	-15,95	0,00002	0,00010	0,00008	0,00004
FBti0019985^{CRISPR1} C2	25,33	38,25	-12,92	0,00013			
FBti0019985^{CRISPR1} C3	24,67	37,28	-12,61	0,00016			
FBti0019985^{CRISPR1} T1	24,44	37,73	-13,30	0,00010	0,00017	0,00013	0,00008
FBti0019985^{CRISPR1} T2	22,67	34,28	-11,61	0,00032			
FBti0019985^{CRISPR1} T3	23,58	37,05	-13,47	0,00009			
FBti0019985^{CRISPR2} C1	25,19	38,64	-13,45	0,00009	0,00014	0,00013	0,00008
FBti0019985^{CRISPR2} C2	25,92	37,69	-11,77	0,00029			
FBti0019985^{CRISPR2} C3	23,61	38,33	-14,72	0,00004			
FBti0019985^{CRISPR2} T1	24,74	37,82	-13,08	0,00012	0,00027	0,00014	0,00008
FBti0019985^{CRISPR2} T2	24,17	35,78	-11,62	0,00032			
FBti0019985^{CRISPR2} T3	24,74	36,10	-11,37	0,00038			

Embryos Cold-stress Total transcript *CG18446*	**Ct *Actin5C***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred *FBti0019985* + C1	23,14	23,53	-0,39	0,764	0,964	0,256	0,148
Outbred *FBti0019985* + C2	23,12	22,80	0,32	1,252			
Outbred *FBti0019985* + C3	22,31	22,51	-0,19	0,875			
Outbred *FBti0019985* + T1	20,76	25,26	-4,50	0,044	0,062	0,015	0,009
Outbred *FBti0019985* + T2	22,50	26,31	-3,81	0,071			
Outbred *FBti0019985* + T3	22,22	26,05	-3,83	0,071			
Outbred *FBti0019985* - C1	21,98	23,09	-1,10	0,465	0,452	0,267	0,154
Outbred *FBti0019985* - C2	22,86	23,35	-0,49	0,712			
Outbred *FBti0019985* - C3	23,19	25,68	-2,48	0,179			
Outbred *FBti0019985* - T1	19,99	22,52	-2,53	0,173	0,273	0,086	0,050
Outbred *FBti0019985* - T2	23,12	24,72	-1,60	0,330			
Outbred *FBti0019985* - T3	22,88	24,55	-1,67	0,314			

Embryos Cold-stress Long transcript *CG18446*	**Ct *Actin5C***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Outbred *FBti0019985* + C1	23,14	26,43	-3,29	0,103	0,129	0,024	0,014
Outbred *FBti0019985* + C2	23,12	25,86	-2,74	0,150			
Outbred *FBti0019985* + C3	22,31	25,19	-2,88	0,136			
Outbred *FBti0019985* + T1	20,76	28,42	-7,66	0,005	0,006	0,001	0,001
Outbred *FBti0019985* + T2	22,50	29,71	-7,20	0,007			
Outbred *FBti0019985* + T3	22,22	29,38	-7,16	0,007			
Outbred *FBti0019985* - C1	21,98	40,69	-18,71	0,000	0,000	0,000	0,000
Outbred *FBti0019985* - C2	22,86	35,88	-13,03	0,000			
Outbred *FBti0019985* - C3	23,19	42,18	-18,98	0,000			
Outbred *FBti0019985* - T1	19,99	38,24	-18,25	0,000	0,000	0,000	0,000
Outbred *FBti0019985* - T2	23,12	40,53	-17,41	0,000			
Outbred *FBti0019985* - T3	22,88	44,54	-21,66	0,000			

Females Ethanol-stress Total transcript *CG18446*	**Ct *Actin5C***	**Ct *CG18446***	**delta Ct (*Actin* - *CG18446*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
outbred FBti0019985 + C1	25,70	27,79	-2,09	0,235	0,224	0,011	0,006
outbred FBti0019985 + C2	24,42	26,65	-2,22	0,214			
outbred FBti0019985 + C3	25,23	27,41	-2,17	0,222			
outbred FBti0019985 + T1	24,44	26,92	-2,48	0,179	0,207	0,025	0,014
outbred FBti0019985 + T2	25,03	27,23	-2,20	0,218			
outbred FBti0019985 + T3	24,91	27,06	-2,15	0,225			
outbred FBti0019985 - C1	24,44	23,14	1,29	2,453	2,260	0,273	0,193
outbred FBti0019985 - C2	24,34	23,29	1,05	2,067			
outbred FBti0019985 - T1	23,67	23,12	0,55	1,468	2,047	0,665	0,384
outbred FBti0019985 - T2	23,69	22,77	0,93	1,900			
outbred FBti0019985 - T3	24,33	22,86	1,47	2,773			

Table S7. Infection assay results with laboratory outbred and CRISPR-mutant strains.

Females Rep. 1		Time (hours)															
STRAIN	TREATMENT	2	14	16	18,5	21	23	25	40	43,5	49	64	74	94	113	119,5	137,5
Outbred *FBti0019985 +*	Nonstress 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Infected 1	2	2	2	3	3	3	3	5	5	5	5	5	7	7	7	8
	Infected 2	2	2	2	2	2	2	5	5	5	5	5	5	7	8	9	9
	Infected 3	1	3	4	5	5	5	5	6	6	6	6	6	7	8	8	8
	Infected 4	0	3	4	4	4	5	5	5	5	6	6	6	6	7	7	8
	Infected 5	1	4	5	5	7	8	9	9	9	9	10	10	10	10	10	10
	Infected 6	1	3	3	4	5	5	5	6	6	6	7	7	9	9	9	10
	Infected 7	1	4	6	7	7	7	7	8	8	8	8	8	8	9	9	9
	Infected 8	1	1	1	3	3	3	3	4	5	5	5	5	5	5	6	6
	Infected 9	0	1	1	2	3	3	3	3	3	4	4	4	6	7	8	8
	Infected 10	2	4	5	5	6	6	6	6	6	6	6	6	6	7	7	7
Outbred *FBti0019985 -*	Nonstress 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Infected 1	0	3	6	6	8	9	9	10	10	10	10	10	10	10	10	10
	Infected 2	0	1	1	2	3	4	4	7	7	7	7	7	8	8	8	9
	Infected 3	0	4	6	9	9	9	9	9	9	9	9	10	10	10	10	10
	Infected 4	0	1	2	2	2	2	3	4	4	7	7	7	8	8	8	9
	Infected 5	0	2	3	3	5	6	6	6	6	6	9	9	9	10	10	10
	Infected 6	0	2	2	4	5	5	6	8	8	9	9	9	9	10	10	10
	Infected 7	1	2	2	2	3	3	3	4	4	4	5	6	8	8	8	9
	Infected 8	1	2	2	3	4	4	5	5	5	5	5	5	6	7	7	8
	Infected 9	1	7	7	8	8	8	8	9	9	9	9	10	10	10	10	10
	Infected 10	2	4	6	8	8	8	8	8	8	8	8	9	9	10	10	10

Log Rank (Mantel-Cox)
0.010

LT50 value	95% CI
29.822	25.619 - 34.324

LT50 value	95% CI
20.89	17.67 - 24.077

Males Rep. 1		Time (hours)															
STRAIN	TREATMENT	2	14	16	18,5	21	23	25	40	43,5	49	64	74	94	113	119,5	137,5
Outbred *FBti0019985 +*	Nonstress 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Infected 1	0	1	1	2	2	2	2	3	3	3	4	5	5	6	6	7
	Infected 2	0	4	6	6	7	7	7	7	7	7	8	8	9	10	10	10
	Infected 3	0	0	0	0	1	1	1	1	1	2	4	6	6	8	8	8
	Infected 4	0	0	1	2	2	2	2	2	2	2	4	4	5	7	7	7
	Infected 5	0	2	2	2	2	2	2	2	2	2	3	3	5	8	8	8
	Infected 6	0	1	5	5	5	6	6	7	7	7	7	8	9	9	9	9
	Infected 7	0	1	1	2	2	4	4	6	6	6	6	6	6	8	9	9
	Infected 8	0	2	4	4	5	6	6	6	6	6	6	6	6	9	9	9
	Infected 9	0	2	2	3	3	3	3	4	4	4	4	4	5	7	7	7
	Infected 10	0	1	2	2	2	2	2	2	2	2	2	2	3	6	6	8
Outbred *FBti0019985 -*	Nonstress 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Infected 1	0	3	3	3	3	4	4	6	6	7	7	7	8	8	8	9
	Infected 2	0	0	0	0	0	0	1	4	6	6	7	7	7	7	7	8
	Infected 3	0	1	1	2	2	2	2	5	5	6	7	8	8	9	9	9
	Infected 4	0	2	2	2	2	2	3	4	4	4	6	8	8	8	8	9
	Infected 5	0	2	2	2	3	3	3	6	6	6	7	7	7	7	7	7
	Infected 6	0	2	2	2	2	2	2	6	7	8	8	8	9	9	9	9
	Infected 7	0	4	4	4	4	4	4	5	5	6	7	7	7	7	9	9
	Infected 8	0	3	3	3	3	3	4	7	8	8	8	8	8	8	8	10
	Infected 9	0	1	2	2	2	2	4	6	7	7	7	8	9	10	10	10
	Infected 10	0	2	2	3	5	5	5	8	8	8	9	9	9	9	9	9

Log Rank (Mantel-Cox)
0.026

LT50 value	95% CI
51.477	45.551 - 58.666

LT50 value	95% CI
37.184	34.536 - 39.968

Females Rep. 2		Time (hours)									
STRAIN	**TREATMENT**	**10**	**13**	**15**	**17**	**21**	**35**	**45**	**60**	**83**	**107**
Outbred *FBti0019985 +*	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0
	Infected 1	0	3	7	7	7	9	9	9	9	10
	Infected 2	0	2	2	5	5	8	8	8	8	9
	Infected 3	0	2	4	7	8	9	9	9	9	9
	Infected 4	0	3	4	7	7	8	9	10	10	10
	Infected 5	0	1	4	7	9	9	10	10	10	10
	Infected 6	0	1	3	3	3	5	5	5	5	5
	Infected 7	0	1	3	3	5	9	9	10	10	10
	Infected 8	0	0	1	3	5	7	8	9	10	10
	Infected 9	0	1	3	4	4	8	8	8	8	8
	Infected 10	2	3	4	5	6	9	9	9	10	10
Outbred *FBti0019985 -*	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0
	Infected 1	0	5	6	7	8	9	9	10	10	10
	Infected 2	1	4	4	6	6	6	6	7	7	7
	Infected 3	2	2	4	5	6	7	7	9	10	10
	Infected 4	0	2	3	5	5	7	7	7	7	7
	Infected 5	3	7	8	8	9	9	9	9	9	10
	Infected 6	0	4	6	6	6	7	7	7	8	9
	Infected 7	3	6	6	6	7	8	8	10	10	10
	Infected 8	3	6	7	7	7	7	7	7	7	7
	Infected 9	2	4	6	7	9	10	10	10	10	10
	Infected 10	0	3	4	7	8	9	9	10	10	10
FBti0019985CRISPR1	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0
	Infected 1	3	3	4	4	6	7	7	7	8	8
	Infected 2	1	1	4	4	7	8	8	8	8	9
	Infected 3	1	2	4	5	7	7	7	9	9	9
	Infected 4	2	4	6	6	7	9	9	9	9	10
	Infected 5	1	3	6	6	8	9	9	10	10	10
	Infected 6	2	2	4	4	5	7	7	7	7	7
	Infected 7	3	6	7	8	9	9	9	10	10	10
	Infected 8	4	6	8	8	8	8	8	9	9	9
	Infected 9	1	7	8	8	8	8	8	9	9	9
	Infected 10	1	4	6	6	6	7	7	8	9	10
FBti0019985CRISPR2	**Nonstress 1**	0	0	0	0	0	0	0	0	0	0
	Nonstress 2	0	0	0	0	0	0	0	0	0	0
	Nonstress 3	0	0	0	0	0	0	0	0	0	0
	Infected 1	3	3	3	3	4	4	4	4	4	6
	Infected 2	5	6	6	6	6	7	7	8	8	9
	Infected 3	4	6	6	6	6	7	7	7	8	9
	Infected 4	1	3	3	4	4	5	5	6	6	6
	Infected 5	1	4	6	6	7	7	7	7	7	8
	Infected 6	1	3	5	6	7	7	8	8	9	9
	Infected 7	2	4	4	4	5	6	6	7	8	10
	Infected 8	5	6	6	6	6	7	8	8	9	9
	Infected 9	1	1	2	4	5	6	6	7	8	9

LT50 value	95% CI
22.09	19.651-24.683

LT50 value	95% CI
15.681	12.902-18.328

LT50 value	95% CI
15.77	13.505-17.949

LT50 value	95% CI
20.939	17.42-24.522

Table S8. qRT-PCR results for *lacZ* reporter gene in laboratory transgenic flies with and without *FBti0019985* sequence under cold- and ethanol-stress.

Embryos Cold-stress	**Ct *Actin5C***	**Ct *lacZ***	**delta Ct (*Actin* - *lacZ*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Empty vector C1	24,17	28,97	-4,80	0,036	0,028	0,010	0,006
Empty vector C2	21,59	27,48	-5,89	0,017			
Empty vector C3	23,08	28,02	-4,94	0,033			
Empty vector T1	23,97	30,82	-6,85	0,009	0,007	0,002	0,001
Empty vector T2	21,90	29,53	-7,64	0,005			
Empty vector T3	22,02	29,14	-7,12	0,007			
***FBti0019985* C1**	22,04	24,61	-2,57	0,168	0,138	0,043	0,025
***FBti0019985* C2**	21,48	24,97	-3,50	0,089			
***FBti0019985* C3**	21,73	24,41	-2,67	0,157			
***FBti0019985* T1**	22,05	26,44	-4,39	0,048	0,026	0,019	0,011
***FBti0019985* T2**	21,34	27,47	-6,13	0,014			
***FBti0019985* T3**	22,09	28,04	-5,95	0,016			

Two-way ANOVA

Effect	**Sig.**
Genotype	0.002
Treatment	0.001
Genotype*Treatment	0.012

Adult Ethanol-stress	**Ct *Actin5C***	**Ct *lacZ***	**delta Ct (*Actin* - *lacZ*)**	**2^deltaCt**	**mean**	**SD**	**SEM**
Empty vector C1	26,48	35,81	-9,33	0,002	0,001	0,001	0,000
Empty vector C2	25,10	36,09	-10,99	0,000			
Empty vector C3	25,49	36,06	-10,57	0,001			
Empty vector T1	25,51	36,02	-10,51	0,001	0,001	0,000	0,000
Empty vector T2	25,59	35,42	-9,83	0,001			
Empty vector T3	25,49	36,12	-10,63	0,001			
***FBti0019985* C1**	25,29	34,14	-8,85	0,002	0,001	0,001	0,001
***FBti0019985* C2**	23,57	34,18	-10,61	0,001			
***FBti0019985* C3**	23,93	35,28	-11,34	0,000			
***FBti0019985* T1**	25,33	34,44	-9,10	0,002	0,001	0,000	0,000
***FBti0019985* T2**	25,43	34,74	-9,31	0,002			
***FBti0019985* T3**	24,91	34,84	-9,94	0,001			

Two-way ANOVA

Effect	**Sig.**
Genotype	0.272
Treatment	0.669
Genotype*Treatment	0.494

Table S9. Cold-stress assay results with laboratory outbred strains with and without *FBti0019985*.

Strain	Treatment	Total emerged flies	% Survival
Outbred *FBti0019985 -*	Nonstress 1	25	83,33
	Nonstress 2	20	66,67
	Nonstress 3	26	86,67
	Nonstress 4	21	70,00
	Nonstress 5	20	66,67
	Cold 1	16	53,33
	Cold 2	26	86,67
	Cold 3	23	76,67
	Cold 4	18	60,00
	Cold 5	18	60,00
	Cold 6	23	76,67
	Cold 7	19	63,33
	Cold 8	19	63,33
Outbred *FBti0019985 +*	Nonstress 1	23	76,67
	Nonstress 2	24	80,00
	Nonstress 3	25	83,33
	Nonstress 4	27	90,00
	Nonstress 5	25	83,33
	Cold 1	23	76,67
	Cold2	26	86,67
	Cold 3	21	70,00
	Cold 4	24	80,00
	Cold 5	17	56,67
	Cold 6	7	23,33
	Cold 7	25	83,33
	Cold 8	23	76,67
	Cold 9	13	43,33
	Cold 10	19	63,33

Two-way ANOVA

Effect	Sig.
Genotype	0.572
Treatment	0.046
Genotype*Treatment	0.411