Natural transposable element insertions drive expression changes in genes underlying Drosophila immune response

Anna Ullastres, Miriam Merenciano, and Josefa González*
Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra)
Passeig Marítim de la Barceloneta 37-49. 08003, Barcelona, Spain

*Corresponding author

J. González e-mail: josefa.gonzalez@ibe.upf-csic.es
A. Ullastres e-mail: anna.ullastres@ibe.upf-csic.es
M. Merenciano e-mail: miriam.merenciano@ibe.upf-csic.es

ABSTRACT

Variation in gene expression underlies inter-individual variability in immune response. However, the mutations responsible for gene expression changes remain largely unknown. In this work, we searched for transposable element insertions present at high population frequencies and located nearby immune-related genes in *Drosophila melanogaster*. We identified 12 insertions associated with allele-specific expression changes in immune-related genes. We showed that transgenically induced expression changes in most of these genes are associated with differences in survival to infection with the gram-negative bacteria *Pseudomonas entomophila*. We provide experimental evidence suggesting a causal role for five insertions in the allele-specific expression changes observed. Furthermore, for two insertions we found a significant association with increased tolerance to bacterial infection. Our results showed for the first time that polymorphic transposable element insertions from different families drive expression changes in genes that are relevant for inter-individual differences in immune response.

Keywords: allele-specific expression, gut immunity, adaptation

BACKGROUND

Innate immunity is the first barrier against infections, and many species rely solely in this response to cope with pathogens (1, 2). Mechanisms of pathogen recognition and activation of the innate immune response are conserved across animals (3, 4). In *Drosophila melanogaster*, several signaling pathways participate in the innate immune response (5-8). The Toll and the Imd are the main signaling pathways involved in recognizing and fighting pathogens (9, 10), while the JAK/STAT pathway is involved in cell proliferation (11, 12), and the JNK pathway is required for proper wound healing (13, 14). Cellular processes such as phagocytosis or melanotic encapsulation also play a critical role in the innate immune response, and studies in *D. melanogaster* are also highly relevant to understand them (5, 6).

One of the most likely infection routes happening in nature is oral infection, and the gut epithelium is the first barrier that bacteria encounter in the organism (15, 16). However, the gut immune response is still not completely understood, and it is likely more complex than the systemic immune response. First, both in insects and in vertebrates, the intestinal tract is a single tubule anatomically and functionally compartmentalized (17, 18). Second, the gut is constantly in contact with bacteria composing the microbiota. As such, the host has to differentiate between pathogenic bacteria and gut microbiota (15, 19-21). Thus, there must be a complex transcriptional regulatory toolkit in order to control the expression of immune responsive genes in the gut (22). Indeed, the analysis of gut immunocompetence variation in 140 *D. melanogaster* strains found that small but systematic differences in gene expression exists between resistant and susceptible strains to *Pseudomonas entomophila*, a natural pathogen of this species (23, 24). Variation in gene expression has been shown to underlay inter-individual variability in immune responses also in humans (25-27). However, the causal mutations responsible for these expression changes remain largely unknown (8, 28). Identifying the causal mutations is necessary to establish functional links between the expression phenotypes and the susceptibility/tolerance to infection.

Among the types of mutations that could be responsible for gene expression changes, transposable elements (TEs) are particularly likely to be important contributors. TEs can be a source of *cis*-regulatory elements that can influence gene regulation (29-32). For example, TEs have been shown to add transcription factor binding sites, and transcription start sites, leading to changes in expression of nearby genes (33-36). TEs can also influence gene expression by inducing changes in the chromatin structure (37-39). These changes in expression induced by TE insertions have been associated with several organismal phenotypes such as stress resistance and fertility (40-42). So far, only a few studies have linked individual TE insertions

with pathogen-induced expression changes (40, 43). A recent study conducted in human lymphoblastoid cell lines established from European and African individuals showed that the regulatory effects of polymorphic TEs are associated with immune-related functions (44). These results suggest that TEs may contribute to regulatory variation between individuals, although no functional evidence was provided for the causal role of the insertions in the expression changes identified (44).

In this work, we aimed at assessing the role of polymorphic TE insertions in *D. melanogaster* gut immune response. We first identified polymorphic TE insertions present at high population frequencies and located nearby immune-related genes. We found that 12 of the 14 insertions analyzed were associated with allele-specific expression changes of their nearby immune-related genes. Transgenically induced expression changes in most of these genes were associated with differences in survival after infection, suggesting that expression changes in these genes are phenotypically relevant. Through a combination of experimental approaches including 3'RACE, qRT-PCR, ChIP-qPCR, and *in vivo* enhancer assays, we provided further evidence for the role of five of these insertions in the expression differences observed. Finally, we showed that two of these insertions are associated with increased survival to bacterial infection.

RESULTS

Nineteen TE natural insertions present at high population frequencies are located nearby genes with immune-related functions

To identify polymorphic TEs likely to affect gut immune response, we first looked for insertions present at high population frequencies, and located in genomic regions with a high recombination rate (see Methods). We analyzed 808 TEs annotated in the *D. melanogaster* reference genome and 23 non-reference TEs in four natural populations (Supplementary File 1A and 1B, see Methods) (45-47). We identified 128 insertions present at ≥ 10% frequency in at least one of the four populations analyzed: 109 reference TEs and 19 non-reference TEs (Supplementary File 1C, see Methods). We then surveyed the literature for the functional information available for the genes located nearby each one of these 128 TEs (Supplementary File 1D). We found that 19 of these 128 TEs were associated with 21 immune-related genes (Table 1). Note that for seven of these 19 TEs there is previous evidence suggesting that they have increased in frequency due to positive selection (Table 1) (47).

The functional evidence for the majority of the 21 genes nearby the 19 candidate immune-related TEs comes from transcriptional response to infection experiments (11 genes), infection survival experiments (five genes), or both (three genes) (Table 1). The other two genes, *TM4SF* and *ken*, are members of the JAK-STAT signaling pathway involved in immune response (11). Before investigating whether the identified TEs could be affecting the expression of nearby immune-related genes, we first tested whether transgenically induced changes in the expression of these genes affect survival to bacterial infection. We focused on nine genes: six genes for which survival experiments were not previously available, and three genes for which survival experiments were performed using a different pathogen (Table 2). When available, two different backgrounds were tested (Supplementary File 2A). For three of the 14 strains analyzed, we did not detect differences in expression of the target gene (Supplementary File 2A and 2B). Thus, we did not perform survival experiments for these three strains. For the other 11 strains, survival experiments were performed with the gram-negative bacteria *P. entomophila* (24). As a natural *D. melanogaster* pathogen, experiments with *P. entomophila* have the potential to identify specialized immune responses derived from antagonistic co-evolution (48). We found that mutant, RNAi knockdowns, and overexpression strains of seven of these genes showed differences in survival after oral infection with *P. entomophila*: *NUCB1, CG2233, Bin1*, and *cbx* showed higher survival, *ken*, *CG8008*, and *TM4SF* mutants showed lower survival (Table 2 and Supplementary File 2C). For *CG10943* results were significant, but the mutation effect size was not significant. Finally, *CG15829* RNAi flies did not show differences in survival (Table 2 and Supplementary File 2C).

Overall, we provide additional evidence linking changes in expression with survival differences for four of the six genes for which no phenotypic evidence was previously available, and for the three genes that were previously tested with a different pathogen (Table 2, Supplementary File 2B and 2C). Thus changes in the expression of the genes located nearby TEs present at high population frequencies affect survival to infection. These results suggest that if TEs also affect the change in expression of these genes, TEs could be associated with differences in survival to bacterial infections.

Table 1. TEs present at high population frequencies in four natural *D. melanogaster* populations and located nearby immune-related genes.

The 16 reference TE insertions are listed first, followed by the three non-reference insertions.

TE	TE family (class)	TE length (bp)	Evidences of selection	TE dist. to nearby gene(s)	TE position in the nearby gene	Gene immune-related evidences
FBti0019386	invader4 (LTR)	347	CL test, TajimaD, Phenotypic (Ullastres et al 2015)	0	5'UTR *Bin1*	**Survival.** Mutant larvaes are more sensitive to fungal *A. fumigatus* (fungi) infection (49).
FBti0019457	pogo (DNA)	1,146	F_{ST}, nSL (Rech et al 2019)	4,434	5' *kay*	**Expression.** Component of the JNK pathway, essential for antimicrobial peptide release (50, 51). Up-regulated in *imd* and bsk mutant LPS-induced S2 cells, and down-regulated in *Rel* mutants (52). Up-regulated in larvaes infected with *P. entomophila* (gram-negative) (24), and after 4h of infection with *P. entomophila* (23).
FBti0019985	roo (LTR)	434	TajimaD, iHS, H12, Phenotypic (Merenciano et al 2016)	0	First intron *cbx*	**Survival.** Mutant flies are more sensitive to *S. aureus* (gram-positive) septic infection, but not to *S. typhimurium* (gram-negative) infection (53).
FBti0020046	Doc (non-LTR)	2,305	Allele age (Blumenstiel et al 2014)	281	3' *Jon65Aiv*	**Expression.** Up-regulated after septic injury with mixed bacteria: *M. luteus* (gram-positive) and *E. coli* (gram-negative) (10). Down-regulated after 4h of infection with *P. entomophila (23)*.
FBti0020057	BS (non-LTR)	126	H12, nSL (Rech et al 2019)	338	3' *CG15829*	**Expression.** Up-regulated after infection by septic injury with mixed bacteria (gram-positive and gram-negative), and regulated by *Rel (10)*. Up-regulated after 4h of infection with *P. entomophila* (23).
				739	5' *CG8628*	**Expression.** Up-regulated in microbiota associated flies vs germ free flies (54), after infection with several pathogens (gram-positive and gram-negative, fungi, protozoa) (55), and down-regulated after 4h of infection with *P. entomophila* (23).
FBti0019564	mdg1 (LTR)	189	TajimaD (Kofler et al 2012)	0	Intron *tlk*	**Survival.** Involved in antimicrobial humoral response to gram-negative (51). *tlk* knockdown, together with other five genes knocked-down, reduces phagocytosis of *E. coli* (gram-negative) and *S. aureus* (gram-positive) in S2 cells (56).
FBti0061506	1360 (DNA)	48	iHS (Rech et al 2019)	0	First intron *Dif*	**Survival and expression.** Transcription factor involved in defense response to fungus and gram-positive bacteria and mediates Toll pathway activation (57-61). Up-regulated in guts from *P. entomophila* infected flies (23).
FBti0018877	BS (non-LTR)	131	-	0	First intron *Mef2*	**Survival and expression.** Adult *Mef2* mutant males are more sensitive to *E. cloacae* (gram-negative) and *M. marinum* (gram-positive) septic infection (62). Up-regulated after 4h of infection with *P. entomophila* (gram-negative) (23).
FBti0018883	Burdock (LTR)	6,413	-	136	3' *CG8008*	**Expression.** Induced by LPS (gram-negative) in an IKK-dependent manner in S2 cell cultures (63). Up-regulated after *E.coli* (gram-negative) infection in S2 cells (64).
FBti0019381	Juan (non-LTR)	2,995	-	180	5' *CG42788*	**Expression.** Down-regulated in response to *P. rettgeri* (gram-negative) infection in females (65).
FBti0019602	Juan (non-LTR)	4,249	-	12	3' *CG2233*	**Expression.** Down-regulated in *PEBP1* mutant L3 larvaes, which are more resistant to *M. luteus* (gram-positive) and *E. coli* (gram-negative) infection (66). Latitudinal expression differentiation after infection with *E.coli* and *M. luteus* mix in temperate vs tropical populations (67).
FBti0020119	S (DNA)	1,732	-	0	First intron *AGO2*	**Survival.** Involved in defense response to virus infections (68), and interacts with Imd pathway proteins during gram-negative infection (69).
FBti0020137	S (DNA)	1,732	-	0	First intron *NUCB1*	**Survival.** Mutants are more resistant to *V. cholerae* (gram-negative) oral infection (70).
FBti0018868	297 (LTR)	414	-	1	5' *TM4SF*	**JAK-STAT.** A tetraspanin, which modulate immune-signaling in *Drosophila (71)*.
				340	3' *ken*	**JAK-STAT.** Member of JAK-STAT pathway (72). JAK-STAT pathway plays a role in immune response in *D. melanogaster* (11).
FBti0061105	G5 (non-LTR)	51	-	46	3' *Dscam1*	**Survival and expression.** Required in hemocytes for efficient phagocytosis and binds to *E.coli* (gram-negative) (73).
FBti0062242	BS (non-LTR)	102	-	0	3' UTR *pnr*	**Expression.** *pnr* is a modifier of the Toll pathway and RNAi mutants show Imd pathway hyperactivation when infected with *E. cloacae* (gram-negative) and *M. luteus* and *E. faecalis* (gram-positive) (74).

tdn4	BS (non-LTR)	800	-	479	3' *CG15096*	**Expression.** Down-regulated in Oregon R and *Rel*-mutant flies with microbiota compared to axenic flies (19), and after *P. entomophila* infection (23).
tdn8	Transpac (LTR)	5,500	-	816	5' *CG10943*	**Expression.** Up-regulated in Oregon R and *Rel*-mutant flies with microbiota compared to axenic flies (19), 24h after infection with *O. muscaedomesticae* (protozoan) (55), and after *P. entomophila* infection (23).
tdn17	pogo (DNA)	1,000	-	2,067	5' *lcs*	**Expression.** Involved in virus response, down-regulated in males infected with sigma virus (75). Up-regulated in young flies gut compared to old flies (19).

Table 2. Transgenically induced expression changes in immunity-related genes are associated with differences in survival to bacterial infection.

Survival of mutant strains orally infected with *P. entomophila* compared with flies with a similar background (Supplementary File 2A). *P-values* obtained using log-rank test. OR (CI): odds ratio and confidence interval (95%)

Gene (Stock number)	**Mutant/ RNAi/ Overexpression**	**Previous evidence**	**Survival experiment**	***P-value***	**OR (CI)**
NUCB1 (10581)	PBac{PB} insertion	Survival (different pathogen)	Higher survival	0.006	9.64 (4.41-21.06)
CG2233 (v100849)	RNAi knockdown	Expression	Higher survival	0.001	3.16 (1.64-6.06)
Bin1 (17130)	P{EP} insertion	Survival (different pathogen)	Higher survival	< 0.001	147.43 (46.74-465)
Bin1 (33574)	Gal4/UAS overexpression		Higher survival	0.044	1.36 (0.84-2.18)
cbx (63763)	PBac{IT.GAL4} insertion	Survival (different pathogen)	Higher survival	0.002	3.61 (2.01-6.46)
ken (11244)	P{PZ} insertion	JAK-STAT	Lower survival	0.003	9.68 (2.77-33.79)
CG8008 (25488)	Mi{MIC} insertion	Expression	Lower survival	0.031	2.55 (1.36-4.81)
TM4SF (v8846)	RNAi knockdown	JAK-STAT	Lower survival	< 0.001	1.89 (1.01-3.54)
CG10943 (56051)	Mi{MIC} insertion	Expression	Lower survival	0.045	1.44 (0.83-2.49)
CG15829 (v104642)	RNAi knockdown	Expression	No differences	0.136	0.72 (0.39-1.3)

Immune-related candidate TEs are associated with gene expression changes

In order to explore whether the 19 candidate adaptive TEs were associated with expression changes of their nearby immune-related genes, we measured allele-specific expression (ASE) in flies heterozygous for the presence of each candidate adaptive TE. Because both alleles in the heterozygous shared the same cellular environment, differential expression of the two alleles indicates functional cis-regulatory differences (76). We performed the analysis in flies with two different genetic backgrounds in order to detect possible background-dependent effects in allele-specific expression changes. We were able to analyze with this technique a total of 16 genes located nearby 14 TEs. In non-infected conditions, 10 out of the 16 genes showed statistically significant allele-specific expression differences in at least one of the two genetic backgrounds analyzed (Figure 1, Supplementary File 3). For five of these 10 genes, we found that the allele with the TE was more highly expressed compared to the allele without the TE, and for the other five genes, the allele with the TE was less expressed. In infected conditions, eight out of the 16 genes showed statistically significant allele-specific expression differences in at least one of the two genetic backgrounds analyzed (Figure 1, Supplementary File 3). For three genes, we found that the allele with the TE was more highly expressed, and in the other five genes the allele with the TE was less expressed. Considering both non-infected and infected conditions, five genes showed allele-specific expression differences in the two conditions: for *CG10943* the allele with the TE was more highly expressed, and for *CG8628*, *CG8008*, *CG15096* and *cbx* the allele with the TE was less expressed (Figure 1, Supplementary File 3).

We also checked whether the genetic background affected the allele specific expression differences. In 10 analyses, both backgrounds showed changes in expression in the same direction, more highly expressed or less expressed, and in two of them the differences were statistically significant in the two backgrounds analyzed (Figure 1, Supplementary File 3). On the other hand, seven analyses differed in the direction of the change of expression in the two backgrounds. However, results were always statistically significant in only one of the two backgrounds analyzed (Figure 1, Supplementary File 3).

Overall, we found that most of the candidate immune-related TEs, 12 out of 14, are associated with changes in expression of their nearby gene, in at least one of the two conditions analyzed (Figure 1). While some expression changes are significant only in infected or only in non-infected conditions, a significant proportion of genes (38%)

showed consistent changes in expression in both conditions (Figure 1). Finally, we detected an effect of the genetic background on the allele specific expression differences as has been previously reported (77-79). However, statistically significant results were always consistent between genetic backgrounds (Figure 1).

Candidate TEs affected expression of nearby genes through different molecular mechanisms

We performed several experiments to identify the molecular mechanisms behind the expression changes observed and to further test whether the TEs are the most likely causal mutation behind these changes (Figure 1). We focused on studying the four TEs located in promoter regions and associated with ≥ 1.5-fold higher expression: *FBti0019386*, *FBti0018868*, *tdn8*, and *FBti0061506*. We also studied in detail two other insertions: *FBti0019985* that showed genetic background dependent effects, and *FBti0020057* associated with lower allele-specific expression. In addition, for *FBti0019386* and *FBti0018868*, we also performed survival experiments to bacterial infection.

***FBti0019386* provides a transcription start site to *Bin1* that is only used in infected conditions in the female gut**. *FBti0019386* is an *invader4* element inserted in the 5'UTR region of *Bin1,* a gene required for the expression of immune- and stress-response genes (49), and associated with shorter developmental time (Table 2, Figure 2A) (80). There are two annotated *Bin1* transcripts with the transcription start site (TSS) located in the *FBti0019386* insertion (Figure 2A) (81). We found that homozygous flies with and without *FBti0019386* expressed only the short *Bin1-RA* transcript in non-infected conditions (Figure 2B). However, in infected conditions, flies without *FBti0019386* insertion only expressed *Bin1-RA*, while flies with *FBti0019386* expressed *Bin1-RA*, and three transcripts starting in the TE: *Bin1-RC*, *Bin1-RD* and *Bin1-RE.* We confirmed these results by performing the experiments in a second genetic background (see Methods). Note that *Bin1-RD* and *Bin1-RE* transcripts were not described previously and differ in the 5'UTR length (Figure 2B).

To test whether the transcripts that start in *FBti0019386* insertion are associated with increased expression of *Bin1,* we quantified the expression of the transcripts starting in the TE and the total *Bin1* transcript levels (Figure 2C). In non-infected conditions, flies with and without *FBti0019386* did not differ in *Bin1* expression levels (t-test, p-value >

0.05). In infected conditions, flies with *FBti0019386* overexpressed *Bin1* compared to flies without *FBti0019386* in the two backgrounds (t-test, p-value < 0.001). The contribution of the transcripts starting in *FBti0019386* to the total *Bin1* expression is background dependent: 11.2% in background I, and 66.3% in background II (Figure 2C). To confirm this result, we analyzed a third genetic background homozygous for the presence of *FBti0019386*, and we found that the TE-transcripts contributed 36.2% to *Bin1* total expression (Figure 2C).

Overall, we found that *FBti0019386* adds a TSS for *Bin1* that is only used in infected conditions in the gut (Figure 2B and 2C). We also found that increased expression of *Bin1* in response to infection is only observed in flies with *FBti0019386* insertion, and that the contribution of the transcripts starting in the insertion to the overall level of *Bin1* expression is background dependent (Figure 2C). These results suggest that, besides adding a new TSS for *Bin1*, *FBti0019386*, which is a 347 bp solo-LTR insertion, could also be acting as an enhancer in infected conditions. Moreover, these results are in agreement with the ASE analysis that showed that *FBti0019386* is associated with increased *Bin1* expression only in infected conditions, further suggesting that the TE is the causal mutation (Figure 1). Finally, we found that flies with *FBti0019386* had higher survival to bacterial infection compared with flies without this insertion (Figure 2D).

***FBti0018868* adds a TSS both in infected and non-infected conditions**. *FBti0018868* is a *297* element annotated 1 bp upstream of one of the *TM4SF* transcripts, and 310 bp upstream of the other two transcripts (Figure 3A). *TM4SF* encodes a tetraspanin protein, which plays a role during immune response in *Drosophila* and humans (71). A previous work identified a new TSS for *TM4SF* inside *FBti0018868* (81). We performed RT-PCR to check whether homozygous flies with *FBti0018868* insertion expressed the transcript starting in the TE in non-infected and/or in infected conditions. We detected the presence of the transcript starting in the TE in both conditions (Figure 3A).

To check whether flies with and without *FBti0018868* differ in the expression level of the different *TM4SF* transcripts in infected and non-infected conditions, we performed qRT-PCR. We found that *TM4SF* expression can only be detected in the strains with *FBti0018868* insertion, although at very low levels (Figure 3B). The primers designed to specifically detect the expression of the transcript starting in *FBti0018868* insertion did not detect any expression (Figure 3B).

To test whether *FBti0018868* could be acting as an enhancer, we generated transgenic flies in which *FBti0018868* was cloned in front of the reporter gene *lacZ* (Figure 3C, see Methods). We did not detect *lacZ* expression by qRT-PCR in non-infected or in infected conditions. The *β-GAL* protein expression localization did not differ either from the negative control (Figure 3C).

Overall, we found that *FBti0018868* that was associated with increased expression of *TM4SF* in infected conditions (Figure 1) adds a TSS for its nearby gene *TM4SF*, which is detected both in infected and non-infected conditions (Figure 3A). Only flies with *FBti0018868* insertion showed *TM4SF* expression in both conditions, although the total level of expression was low, and we could not detect the transcripts starting in the insertion using qRT-PCR (Figure 3B). *FBti0018868* is not driving the expression of a reporter gene (Figure 3C) suggesting that the insertion sequence by itself is not enough to increase the expression of a nearby gene, but probably needs additional regulatory sequences. A larger genomic region containing *FBti0018868* insertion should be analyzed before discarding the effect of the insertion on *TM4SF* expression changes. Finally, we found that flies with *FBti0018868* had higher survival to bacterial infection compared with flies without this insertion (Figure 3D).

***tdn8* drives the expression of a reporter gene in non-infected and infected conditions**. *tdn8* is a *Transpac* element located 816 bp upstream of *CG10943*, a gene that is up-regulated in response to immune challenge with different pathogens including *P. entomophila* (Figure 4A and Table 2) (19, 23, 55). We tested whether *tdn8* could be acting as an enhancer (Figure 4B). We found that transgenic strains in which the upstream region of *CG10943* contained the *tdn8* insertion showed higher expression than transgenic strains in which the same region without the insertion was cloned in front of the reporter gene (Figure 4C). Differences in expression were only statistically significant in infected conditions (p-value = 0.046 respectively) (Figure 4C). We found no differences between the two transgenic strains in the localization of the *β-GAL* protein expression in non-infected or infected conditions (Figure 4D).

Overall, we found that *tdn8* is acting as an enhancer. These results are in agreement with our ASE results that showed that *tdn8* is associated with higher expression of *CG10943* in the two genetic backgrounds analyzed (Figure 1).

***FBti0061506* does not drive the expression of a reporter gene**. *FBti0061506* is a *1360* element located in the 5'UTR intron of *Dif-RD* transcript, and 3.8 kb upstream of the other three *Dif* transcripts (Figure 5A). *Dif* is a main transcription factor of the Toll-pathway, and it was found to be up-regulated in gut tissue after *P. entomophila* infection (Table 2) (23).

To test whether *FBti0061506* could act as an enhancer sequence, we generated two reporter gene constructs containing part of the *Dif* intron with and without the insertion (Figure 5B, see Methods). None of the two constructs affected the expression of the reporter gene or the localization of the *β-GAL* protein (Figure 5C).

Overall, our results do not provide evidence for an enhancer role of *FBti0061506*. However, our ASE results showed that *FBti0061506* was associated with *Dif* higher expression in non-infected conditions (Figure 1). Although it is also possible that *Dif* allele-specific expression is due to a *cis*-mutation different from the *FBti0061506*, it could be that the effect of *FBti0061506* is context dependent. Therefore, a larger genomic region with and without the insertion should be analyzed to discard an effect of *FBti0061506* on *Dif* allele-specific expression differences (Figure 1).

***FBti0019985* drives the expression of a reporter gene both in non-infected and infected conditions**. Besides the four TEs located in promoter regions and associated with ≥ 1.5-fold higher allele-specific expression, we also studied in detail *FBti0019985* insertion that showed genetic background dependent effects (Figure 1). *FBti0019985* is a *roo* element inserted in two nested genes: *CG18446* and *cbx*. *FBti0019985* provides a transcript start site for *CG18446* and has been associated with increased cold tolerance (34). *FBti0019985* is also located in the first 5'UTR intron of *cbx-RA* (*CG46338-RA*) transcript, and 700 bp and 5.5 kb upstream of the other two *cbx* transcripts (Figure 6A). *cbx* mutant flies are more tolerant to bacterial infection (Table 2). We first checked whether the TE affects the expression of the different *cbx* transcripts by performing RT-PCR from non-infected guts of homozygous flies with and without the TE. We detected two of the three annotated transcripts, *cbx-RB* and *cbx-RC*, in flies with and without *FBti0019985* (Figure 6A). Thus, we did not find evidence of *FBti0019985* affecting *cbx* transcript choice or transcript structure in non-infected conditions.

The allele containing *FBti0019985* could be acting as an upstream enhancer for *cbx-RB* and *cbx-RC* transcripts. Thus, we performed enhancer reporter assays, and we detected that *FBti0019985* drives the expression of the reporter gene only in infected conditions

(Figure 6B and 6C). *β-GAL* immunostaining showed that the expression was localized in the anterior part of the gut, both in non-infected and in infected conditions (Figure 6D). The localization of the expression only in the anterior part of the gut could explain why we could not detect expression with the qRT-PCR in whole guts in non-infected conditions (Figure 6C).

Overall, we showed that *FBti0019985* does not modify *cbx* transcript structure under non-infected conditions but acts as an enhancer in the anterior part of the gut. These results suggest that the effect of *FBti0019985* could be background dependent as the insertion was associated with lower expression of *cbx* in the second genetic background analyzed (Figure 1).

***FBti0020057* down-regulates the expression of a reporter gene**. We also studied in detail *FBti0020057*, one of the six insertions associated with lower allele-specific expression (Figure 1). *FBti0020057* is a *BS* element annotated in the intergenic region between *CG15829* and *CG8628* (Figure 7A). Both genes are predicted to be involved in Acyl-CoA homeostasis, associated with lipid metabolism (82). Resources redistributions between metabolism and immune response are a key process during infection, and genes involved in lipid metabolism are repressed after infection (5, 83-85). We checked whether *FBti0020057* could be affecting the transcript choice of its upstream gene *CG15829* (Figure 7A). We performed 3'RACE using cDNA of non-infected guts from homozygous flies with and without *FBti0020057*. We only detected the expression of the shorter transcript, *CG15829-RA*, in both strains. Therefore, the TE is not affecting *CG15829* gene transcript structure or transcript choice in the studied conditions.

We then tested whether *FBti0020057* could be down-regulating its downstream gene. To do this, we cloned the whole intergenic region with and without *FBti0020057* in front of the reporter gene *lacZ* (Figure 7B). Consistent with the ASE results for *CG8628*, we found that the transgenic strains with the TE have less expression of the reporter gene both in non-infected and in infected conditions (Figure 7C). We also found that the expression of the reporter gene was localized mostly in the posterior midgut (Figure 7D), a region known to be dedicated to absorption that expresses genes encoding lipid transporters (18). Because TEs can recruit repressive histone marks, such as H3K9me3, that can lead to silencing of nearby genes (38, 86, 87), we then checked whether *FBti0020057* was enriched for H3K9me3. We did not find enrichment for

H3K9me3 when comparing a strain with and without this insertion (Figure 7E). Thus, the TE could be disrupting a regulatory sequence, or it could be adding a binding site for a repressor protein.

Taken together, our results showed that *FBti0020057* is associated with the down-regulation of a reporter gene, consistent with the observed allele-specific expression differences of *CG8628* (Figure 1). On the other hand, we did not find differences in *CG15829* transcript choice associated to the TE. Further experiments are needed to determine whether *FBti0020057* is also responsible for the increased expression of *CG15829* observed in the allele-specific experiments (Figure 1).

DISCUSSION

In this work, we found 19 TE insertions present at high frequencies in *D. melanogaster* natural populations, and located nearby genes enriched for immune-related functions (Table 1). The majority of these insertions, 13 out of 19, have increased in frequency in out-of-Africa populations (Supplementary File 1C). *D. melanogaster* has recently colonized out-of-Africa environments (88, 89). Among the many stressors faced by *D. melanogaster*, our results suggest that response to pathogens has been an important biological process in the colonization of the new environments (Table 2). Immune response has previously been reported to be relevant for local adaptation not only in *D. melanogaster* but also in humans (67, 90-94). Our results are based on the analysis of four natural populations: one population from the ancestral range of the species, and three out of Africa populations: one North American and two European populations (95, 96). Although the three out-of-African populations analyzed come from locations with contrasting climates, analysis of natural populations from other geographical locations is needed to get a more general picture of the biological processes that are relevant for out-of-Africa adaptation. Indeed, a recent analysis of 91 samples from 60 worldwide natural populations suggested that response to stress, behavior, and development are shaped by polymorphic transposable element insertions (47).

We found that our candidate TEs were associated with allele-specific expression differences in 13 out of the 16 immune-related genes analyzed (Figure 1). Recent studies performed in several strains estimated that ~8% to 28% of *D. melanogaster* genes showed allele-specific expression (97, 98). Thus, our results suggest that our candidate TEs are more often associated with genes that show allele specific expression than expected by chance (81%, p-value < 0.0001). We, and others, have shown that

transgenically changes in expression of nine of these 13 genes led to changes in *D. melanogaster* survival rate after infection (Table 1, Table 2, and references therein). Changes in the expression of these genes are thus likely to be relevant for the fly ability to cope with infections. As described previously, we found both gene up-regulation and gene down-regulation in gut immune response (17, 24). Most of the genes showed allele-specific expression changes either in infected conditions, *Bin1*, *TM4SF* and *NUCB1,* or both in non-infected and infected conditions, *CG10943*, *CG8628*, *CG8008*, *CG15906*, and *cbx* (Figure 1). However, we also identified five genes that showed allele-specific expression changes only in non-infected conditions: *CG2233*, *Dif*, *AGO2*, *CG15829*, and *Mef2* (Figure 1). Differences in the basal transcriptomic profile between tolerant and susceptible strains to *P. entomophila* infection have been described previously in *D. melanogaster* (23). Moreover, differences in gene expression pattern before parasitoid attack between control and selected lines for increased resistance to *Asobara tabida,* a *D. melanogaster* endoparasitoid*,* have also been reported (99). Taken together these results suggested that besides the genes that change their expression level in response to the immune challenge, gene expression variability in non-infected conditions also affects the susceptibility of the flies to immune-challenges.

We identified the molecular mechanism underpinning the changes in expression induced by four of the six insertions studied in more detail. We found TEs that add TSS (Figure 2), act as proximal enhancers (Figure 4 and Figure 6), and repressed the expression of nearby genes (Figure 7). These results add to an increasing body of literature showing the multiple ways in which TE insertions affect the expression of nearby genes (30, 40). On the other hand, the other two insertions analyzed did not affect the expression of a reporter gene (Figure 3 and Figure 5). However, it is known that enhancer reporter assays select for compact regulatory elements that can function in an autonomous manner (100). Thus, before discarding the causal role of these insertions in the observed allele-specific expression differences, a larger genomic region including these insertions should be tested for enhancer activity. Indeed, we further showed that *FBti0018868* was associated with increased expression of *TM4SF* (Figure 3C) suggesting that this insertion is likely to be the causal mutation of the allele-specific expression differences previously observed (Figure 1). Moreover, for this insertion we found that it was associated with increased survival to infection (Figure 3E). Still, it could also be that differences in expression are due to polymorphism other than the TE insertions identified. Although we could not identify any other *cis*-variant in the

proximity of the analyzed genes, except for *AGO2*, regulatory regions might not be conserved and there might be other variants also contributing to the differences in expression (Supplementary File 4, see Methods).

The expression changes associated with several of the TE insertions analyzed in this work were consistent with a role of the nearby genes in increased survival after infection (Figure 1 and Table 2). Indeed, for two of the insertions we showed that this is the case (Figure 2 and 3). We showed that *FBti0019386* is associated with increased *Bin1* expression in infected conditions, and increased survival after infection (Figure 1, 2C and 2D) (49). We also found that *FBti0018868* is associated with *TMS4F* higher allele-specific expression and increased survival after infection (Table 3 and Figure 3E). Bou Sleiman et al (23) found that *CG10943* and *Dif* were up-regulated and *CG8628* was down-regulated in strains resistant to *P. entomophila* infection. We indeed found that the candidate adaptive insertion *tdn8* and *FBti0061506* were associated with increased expression of *CG10943* and *Dif* respectively, and *FBti0020057* insertion was associated with *CG8628* lower expression. Finally, we also found that *FBti0019985* located upstream of *cbx* could act as an enhancer (Figure 6). A *cbx* knockout was more sensitive to gram-positive bacterial infection (53), and we found that the same mutant stock was associated with increased survival to *P. entomophila* infection (Table 2). Thus, in this case the change of expression associated with the candidate adaptive TE is also likely to lead to increased survival after infection (Figure 1). Thus, based on results already available in the literature and in our own results, we found that several of the TE insertions analyzed induced changes in nearby genes that are likely to lead to increased *P. entomophila* infection survival.

Overall, we have shown that TEs contribute to immune-related gene expression variation, which could be crucial for a rapid process of adaptation to new environments. For two of the insertions analyzed, we further showed that they are associated with increased survival to bacterial infection. TEs are likely to be key players in immune response in other organisms as well, as has been shown for a particular fixed TE family and a fixed TE insertion in mammals (40, 43). Besides, several polymorphic insertions have been found to be associated with expression changes in immune-related genes in human populations, and our results suggest that polymorphic insertions could also be relevant for immune response variability (44).

METHODS

Fly strains

DGRP strains. 141 DGRP strains (96) were used to estimate the frequencies of TEs annotated in the *D. melanogaster* reference genome using the data in Rech et al (47) (Supplementary File 5). Besides, we used 37 DGRP strains to analyze by PCR a subset of TEs not annotated in the reference genome (46). Finally, DGRP strains were also used to perform allele specific expression analyses (ASE), transcription start site identification (TSS), and enhancer assays (Supplementary File 5). Note that it has previously been shown that differences in the presence/absence of the endosymbiont *Wolbachia*, differences in commensal bacteria and/or feeding behavior has no major effect in the susceptibility of DGRP strains to *P. entomophila* infection (23).

African strains. Frequency estimates for reference TE insertions for a subset of 66 African strains collected in Siavonga (Zambia, (95)) with no evidence of cosmopolitan admixtures were obtained from Rech et al (47) (Supplementary File 5).

European strains. Frequency estimates for reference TE insertions for 73 European strains, 57 from Stockholm (Sweden) and 16 from Bari (Italy), were obtained from Rech et al (47) (Supplementary File 5). Additionally, one strain from Bari (CAS-49) was used for ASE and TSS experiments and one strain from Munich (MUN-8) was used for ASE experiments (Supplementary File 5).

Outbred strains. We generated present and absent outbred strains for *FBti0019386*, and *FBti0018868* insertions. First, we selected all the strains that were present or absent for these TEs based on data generated by *Tlex2* in the DGRP, Zambia, Sweden, and Italy populations (47). Then, in these selected strains, we checked by PCR the presence/ absence of the other nine TEs identified is this work as they are likely to be involved in the immune response as well (*FBti0019985*, *FBti0061506*, *FBti0019602*, *FBti0020119*, *FBti0018883*, *FBti0018877*, *FBti0020137*, *tdn4*, and *tdn8*). For generating both present and absent outbred populations of each TE we chose between seven and nine strains present and absent of a specific TE, respectively (Supplementary File 5). Moreover, present and absent outbred populations have similar frequencies of all the other 10 TEs likely to be involved in immune responses in order to not mask the effect of the studied TE. In every outbred population, we placed 10 males and 10 virgin females of each selected strain in a cage with fresh food. We maintained the population by random mating with a large population size for over four generations before starting the experiments.

Mutant, RNAi knockdown, and overexpression strains. We used three RNAi mutant strains from the VDRC stock center (Supplementary File 2A). To generate the mutants, we crossed the strains carrying the RNAi controlled by an *UAS* promoter with flies carrying a GAL4 driver (a transcription activation system) to silence genes ubiquitously. We performed the experiments with F_1 flies that were obtained from each cross. Based on the phenotypic markers, we separated the RNAi mutant flies from the rest of the F_1 that do not carry the GAL4 driver. The flies without the GAL4 driver were used as the baseline of the experiment. To overcome the lethality of silencing *CG15829* during development, we used an *Act5c*-GAL4 strain regulated by the temperature sensitive repressor GAL80. For this mutant, we transferred flies from 25ºC to 29ºC 24h before performing the experiment.

We also used nine mutant strains generated with different transposable element insertions and two overexpression strains. In this case, we used strains with similar genetic backgrounds as the baseline for the experiments (Supplementary File 2A).

Transposable element datasets

TEs annotated in the reference genome. There are 5,416 TEs annotated in the v6 of the *D. melanogaster* reference genome (101). In this work, we focused on polymorphic TEs present at high population frequencies and located in high recombination regions of the genome. Most TE insertions are expected to be deleterious. Due to its big effective population size, we expect most TE insertions to be present at low frequencies in *D.* melanogaster. Thus, TEs present at high population frequencies are likely to be adaptive. We did not consider the 2,234 INE-1 insertions that are fixed in *D. melanogaster* populations (102-104). We also discarded 1,561 TEs that are flanked by simple repeats, nested TEs, or TEs that are part of segmental duplications because frequencies cannot be accurately estimated for these TEs using T-lex2 (45). Finally, we discarded 813 TEs present in genomic regions with a recombination rate = 0 according to Fiston-Lavier et al (105) or Comeron et al (106). TEs present at low recombination regions are more likely to be linked to an adaptive mutation rather than being the causal mutation (107-110). Moreover, the efficiency of selection is low in these regions and, thus, slightly deleterious TEs could have reached high frequencies (111, 112). Hence, we ended up with a dataset of 808 annotated TEs for which we estimated their population frequencies using T-lex2 (45) (Supplementary File 1A).

231 of the 808 annotated TEs were fixed in the four populations studied. Although some of these fixed TEs might be adaptive, we did not consider them as we cannot perform comparative functional experiments between flies with and without the insertions. We considered high frequent TEs those present at a population frequency ≥ 10%: 109 TEs. Note that varying this threshold does not substantially alter the number of TEs present at high frequencies (e.g. 95 TEs if we consider ≥ 15%).

Non-reference TE insertions. We also analyzed a subset of TEs identified by Rahman et al (46) in DGRP strains that are not annotated in the reference genome (Supplementary File 1B). We analyzed 23 TEs that are present in regions with recombination rate > 0 (105, 106), and were inferred to be present in at least 15 DGRP strains out of the 177 strains analyzed by Rahman et al (46). Then, we obtained from Bloomington Drosophila Stock Center (BDSC) all the strains carrying each of the 23 insertions, and we confirmed by PCR the presence of the insertions in several strains (see below). For each TE, we sequenced at least one of the PCR products to confirm the presence and the family identity of the TE. For those insertions that we could verify, we estimated the frequency of each TE based on TIDAL results in the 177 DGRP strains and considered as high frequent those present at a population frequency ≥ 10%.

Presence/Absence of TEs in the analyzed strains

We performed PCRs to confirm the *in silico* results obtained with *T-lex2* (45) and *TIDAL* (46). We designed specific primers for each analyzed TE using the online software Primer-BLAST (113) (Supplementary File 6). Briefly, we designed a primer pair flanking the TE (FL and R primers), which produces a PCR product with different band sizes when the TE is present and when the TE is absent. For those TEs that are present in the reference genome, we also designed a primer inside the TE sequence (L primer) that, combined with the R primer, only amplifies when the TE is present (114). To perform the PCRs, genomic DNA was extracted from 10 females from each analyzed strain.

Functional annotation of genes nearby candidate adaptive TEs

We looked for functional information of the genes associated to the TEs present at high population frequencies using FlyBase (101). We considered all the genes that were located less than 1kb from the TEs. If the TEs did not have any gene located in the 1kb

flanking regions, we considered only the closest gene. We considered GO annotations based on experimental evidence, and we also obtained functional information based on the publications cited in FlyBase. Several lines of evidence were considered: genome-wide association studies in which SNPs in the analyzed genes were linked to a phenotypic trait, differential expression analyses, and phenotypic evidence based on the analyses of mutant strains (Supplementary File 1D).

qRT-PCR expression analysis of mutant, RNAi, and overexpressing strains

For RNA extraction, three replicates of 20-30 females, males, or guts from each mutant and wild-type strain were flash-frozen in liquid nitrogen and stored at -80ºC until sample processing (Supplementary File 2A). RNA was extracted using the GenElute™ Mammalian Total RNA Miniprep Kit (Merck) following manufacturer´s instructions. 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). Primers used for qPCR experiments are listed in Supplementary File 2D. In all the cases, gene expression was normalized with the housekeeping gene *Act5c*. We performed the qRT-PCR analysis with SYBR Green (BioRad) on an iQ5 Thermal cycler. Results were analyzed using the dCT method following the recommendations of the MIQE guideline (115).

Infection experiments

We infected 5- to 7- day-old female flies with the gram-negative bacteria *Pseudomonas entomophila* (24). Flies were separated into food vials under CO_2 anesthesia two days before the bacteria exposure, and were kept at 25ºC. The experiments were performed as described in Neyen et al (116). Briefly, flies were starved for two hours and then they were flipped to a food vial containing a filter paper soaked with 1.25% of sucrose and bacterial pellet. The bacterial preparation was adjusted to a final OD_{600} = 100, corresponding to 6.5 $x10^{10}$ colony forming units per ml (117). Flies were kept at 29ºC and 70% humidity, which are the optimal infection conditions for *P. entomophila*. In parallel, we exposed non-infected flies to sterile LB with 1.25% sucrose.

Survival experiments

We performed infection survival experiments with mutant strains, RNAi strains, and overexpression strains. We compared the mortality of these strains to the mortality of

strains with similar genetic backgrounds (Supplementary File 2A). We also performed infection survival experiments comparing outbred flies with *FBti0019386* and *FBti0018868* with outbred flies without these insertions, respectively. Female flies were placed in groups of 10 per vial, and we performed the experiments with 5-12 vials (Supplementary File 2C), except for cn^1 considered as a wild-type background for which we used 3 vials. As a control for each experiment, we exposed 3-4 vials containing 10 flies each to sterile LB with 1.25% sucrose.
Fly mortality was monitored at several time points until all the flies were dead. Survival curves were analyzed with log-rank test using SPSS v21 software. If the test was significant, we calculated the odds-ratio and its 95% confidence interval when 50% of the susceptible flies were dead, except for *CG8008* and *cbx* that was estimated when 30% and 96% of the susceptible flies were dead.

RNA extraction and cDNA synthesis from non-infected and infected guts

We dissected 20-30 guts from both non-infected and orally infected 5- to 7-day-old females. Flies were infected with the gram-negative bacteria *P. entomophila* as mentioned above, and they were dissected after 12 hours of bacterial exposure. Samples were frozen in liquid nitrogen and stored at -80ºC until sample processing. RNA from gut tissue was extracted using Trizol reagent and PureLink RNA Mini kit (Ambion). We treated RNA on-column with DNase I (Thermo) during the RNA extraction, and we did an additional treatment after the RNA purification. We synthesized cDNA from a total of 500 ng – 1,000 ng of RNA using the Anchored-oligo (dT) primer and Transcription First Strand cDNA Synthesis kit (Roche).

Allele-specific expression analysis (ASE)

For each TE analyzed, we first identified two strains homozygous for the presence of the TE and two strains homozygous for the absence of the TE according to *T-lex2* or *TIDAL* (45, 46). We then looked for a synonymous SNP linked to the presence of the TE and located in the coding region of the nearby gene. Note that we only selected a SNP when it is present in the coding region of all the alternative transcripts described for that gene. To select the SNP, we downloaded the coding region of the nearby gene from the sequenced DGRP strains available in http://popdrowser.uab.cat/ (118). Once we identified a diagnostic SNP, we re-sequenced the region identified in the used strains to confirm the presence of the SNP, and we performed a PCR to confirm the presence or

the absence of the TE. We selected a synonymous SNP that is not linked to the TE in all the strains analyzed (Supplementary File 7).

We also analyzed the coding region of the gene in order to discard the presence of nonsynonymous SNPs that could be linked to the TE (Supplementary File 4A). Additionally, we analyzed the flanking regions of each TE in order to discard other variants that could be linked to the TE, or that could be potentially modifying the gene regulatory regions (Supplementary File 4B). To do this, we used VISTA to define the conserved regions in the 1 kb TE flanking sequences between *D. melanogaster* and *D. yakuba*, which diverged approximately 11.6 Mya (119). We then checked whether there is any SNP linked to the presence of the TE in the DGRP strains. Only for the *AGO2* gene, we found two SNPs in the coding region that were linked to the TE insertion (Supplementary File 4A). *AGO2* is a gene showing a fast rate of adaptive amino acid substitutions (120, 121), and it is associated to a recent selective sweep (122). However, it is still not clear which is the genetic variant that is under positive selection (122). Thus for 13 out of the 14 TEs analyzed, we could not detect any other *cis*-regulatory change that could be responsible for the observed allele-specific expression differences suggesting that the TE is the most likely mutation.

We were not able to analyze five of the candidate TEs: for three TEs, *FBti0019381*, *FBti0061105* and *FBti0062242,* we could not identify homozygous strains with and without the TE. For *FBti0019564*, we could not identify a diagnostic SNP. Finally, for *tdn17,* we could not design primers to validate the diagnostic SNP due to the presence of repetitive sequences in the nearby gene.

We then crossed a strain with the TE with a strain without the TE differing by the diagnostic SNP to obtain heterozygous flies in which allele-specific expression was measured (Supplementary File 7). Note that for each TE two crosses were performed so that ASE was measured in two different genetic backgrounds.

ASE was measured in non-infected and infected conditions. We obtained cDNA samples from three biological replicates. We also extracted genomic DNA (gDNA) from 15-20 heterozygous females for each cross, which is needed to correct for any bias in PCR-amplification between alleles (123). cDNA and gDNA samples were sent to an external company for primer design and pyrosequencing. We analyzed the pyrosequencing results as described in Wittkopp et al (123). Briefly, we calculated the ratios of the allele with the TE and the allele without the TE of the cDNA samples, and we normalized the values with the gDNA ratio. In order to perform the statistical

analysis, we transformed the ratios with log2, and we applied a two-tailed t-test in order to check whether there were allele expression differences between the alleles. We corrected the p-values for multiple testing using Benjamini-Hochberg's false discovery rate (5% FDR) (124).

Transcript start site detection

To detect whether *FBti0019386* is adding a Transcription Start Site (TSS) to their nearby gene, as suggested by Batut et al (81), we performed RT-PCR in gut tissue of non-infected and infected flies. We used the forward primer 5'-ATCTGAAGCTCGTTGGTGGG-3' and the reverse primer 5' ATGAGACTCCTGTTTCGCCG- 3' to detect *Bin1* transcript starting in the TE, and the same forward primer with the reverse primer 5' AAGAGCAAAGAGAAGCCGGAA-3' to detect *Bin1* short transcript.

3'RACE

We performed 3'RACE to detect whether the *FBti0020057* was affecting the transcript structure or the transcript choice of *CG15829*. We extracted total RNA from gut tissue of non-infected flies, and synthetized the cDNA using SuperScript^{TM} II Reverse Transcriptase (Invitrogen). We amplified the 3' ends with the Universal Amplification Primer (UAP) (5'- CUACUACUACUACUAGGCCACGCGTCGACTAGTAC-3') and nested PCRs specific for *CG15829*: outer primer 5'- CTGCCTAGCAAGGAGGAGTT-3' and the inner primer 5'- GAGAAGAAGGCCCGCTACAA-3'.

Enhancer reporter assays

We generated transgenic flies carrying the TE sequence in front of the reporter gene *LacZ* by using the *placZ.attB* vector (accession number: KC896840) (125). In order to construct a clone with the correct orientation in the promoter region of *lacZ*, two cloning steps were necessary. We first had to introduce specific restriction sites into the flanking regions for each TE sequence. For that, we introduced the restriction sites with the primers used to amplify the region containing the TE sequence (Supplementary File 8). We used a high fidelity Taq DNA polymerase for DNA amplification (Expand High Fidelity PCR system from Sigma). After that, we cloned the PCR product into the vector pCR®4-TOPO® (Invitrogen). Finally, we digested both vectors and ligated the TE sequence into the *placZ.attB*, and we sequenced the cloned insert to ensure that no

polymerase errors were introduced in the PCR step. We purified the vector with the GeneEluteTM Plasmid Miniprep kit (Sigma), and prepared the injection mix at 300 ng/µl vector concentration diluted with injection buffer (5 mM KCl, 0.1 mM sodium phosphate, pH 6.8). The injection mix was sent to an external company to inject embryos from a strain that contain a stable integration site (Bloomington stock #24749). After microinjection, surviving flies were crossed in pairs and the offspring was screened for red eye color, which was diagnostic for stable mutants. We established three transgenic strains for each analyzed TE, which were considered as biological replicates in the expression experiments. As a negative control, we also established transgenic strains with the *placZ.attB* empty vector, in order to control for possible *lacZ* expression driven by the vector sequence.

For two TEs, *FBti0018868* and *FBti0019985*, we designed primers flanking the TE and cloned the PCR product in front of the reporter gene *lacZ* (Supplementary File 8). For the other three TEs, we constructed two different clones to generate two transgenic strains: one strain with the TE and the other strain without the insertion. For the TE *FBti0061506*, which spans only 48 bp, one strain carries the TE and part of the flanking intronic region, and the other strain contains the same genomic region without the TE. For the TE *tdn8,* one strain carries the upstream region of *CG10943*, including the 5'UTR, with *tdn8*, and the other strain carries the same genomic region without *tdn8*. Finally, for the TE *FBti0020057,* we cloned the whole intergenic region, including the UTRs of the flanking genes (Supplementary File 8).

We analyzed the flanking regions of *FBti0020057* to check whether the insertion is disrupting a regulatory region. For that, we analyzed the 150 bp flanking sequence with and without the insertion looking for predicted transcription factor binding sites with JASPAR (126), using a relative profile score threshold of 90% (Supplementary File 9).

qRT-PCR expression analysis

For the transgenic strains generated in the enhancer assays, we checked *lacZ* expression in female guts in non-infected and infected conditions. We used the forward primer 5'-CCTGCTGATGAAGCAGAACAACT-3', and reverse primer 5'-GCTACGGCCTGTATGTGGTG-3' to check *lacZ* expression.

We measured the gut total expression of *TM4SF* and *Bin1* genes in homozygous strains with and without *FBti0018868* and *FBti0019386*, respectively. We used the following primers to detect *Bin1* total expression: forward 5'- TGTCGTCCCGTAGAGCAGAA-

3' and reverse 5'-CAAGCAGATTGACCGCGAGA-3', and *TM4SF* total expression: forward 5'-GCAGCGAGGATAACGGGAAA-3' and reverse 5'-AGTAGACCGAGTGACCCCAG-3'.

We also designed primers to detect specifically those transcripts starting in the TE for each gene. For transcript starting in *FBti0019386,* we used forward 5'-TGCAGCAGATGGCTCATATT-3' and reverse 5'-AGTGCTCAAGACCCTAATGGAA-3', and for transcripts starting in *FBti0018868,* we used forward 5'-CTTGGCGTTGTCCTTAGTCA-3' and reverse 5'-ACTGATTTATATCGTATGGGGTGCT-3'. We analyzed the two pairs of genetic backgrounds used for the ASE experiments, and one extra genetic background for *Bin1*. In all the cases, gene expression was normalized with the housekeeping gene *Act5c*. We performed all RNA extractions, cDNA synthesis and qRT-PCR analysis as mentioned above.

Immunofluorescence staining

We performed immunofluorescence gut staining to localize β-*GAL* expression in the transgenic flies from the enhancer assays, both in non-infected and infected conditions. Flies were dissected and gut tissue was fixed with 4% Formaldehyde. The tissue was then stained by using the primary antibody mouse anti-βGalactosidase (Hybridoma bank 40-1a), and the secondary antibody anti-mouse Alexa Fluor ® 555 (Sigma). Images were analyzed and captured using a Leica SP5 confocal microscope.

Chromatin immunoprecipitation-qPCR

We performed ChIP-qPCR experiments to detect whether *FBti0020057* that was associated with allele-specific lower expression was adding H3K9me3 repressive marks (38, 87). For that, we compared the histone mark levels in homozygous flies with the TE with the levels in homozygous flies without the TE. We used $y^1;cn^1bw^1sp^1$ strain (127), the strain used to obtain the *D. melanogaster* reference genome sequence (128-130), as the homozygous strain with *FBti0020057* insertion, and RAL-908, as the homozygous strain without those insertions (96). We first confirmed by PCR the presence or absence of *FBti0020057* in these strains. To detect H3K9me3 levels associated to the TE, we designed primer pairs in the TE flanking regions ("left" and "right"): one primer inside the TE sequence and one primer outside the TE sequence (Supplementary File 10). To detect H3K9me3 levels in the strains without the TE, we

used the left forward primer and the right reverse primer. Primer efficiencies ranged from 90-110%. We used a total of 45-55 guts per strain and performed three biological replicates for each strain. In order to obtain the chromatin, we followed Magna-Chip™ A/G kit (from Merck) protocol. After dissection, we homogenated the samples in the buffer A1 with a dounce 30 times, and we crosslinked the guts using formaldehyde at a final concentration of 1.8% for 10 minutes at room temperature. We stopped the crosslink by adding glycine at a final concentration of 125 mM, we incubated samples three minutes at room temperature, and kept them on ice. Then, we washed the samples three times with buffer A1, and we incubated the sample for three hours at 4ºC with 0.2 ml of lysis buffer. After lysis, we sonicated the samples using Biorruptor® pico sonication device from Diagenode: 14 cycles of 30 seconds ON, 30 seconds OFF. We kept 20 µl of input chromatin for the analysis (see below), and we immunoprecipitated 80 µl of the sample with antibody against H3K9me3 (#ab8898 from Abcam). As a control for the immunoprecipitation, we checked the H3K9me3 levels in the genes *18S* and *Rpl32* that are expected to be, respectively, enriched and depleted for this histone mark (Supplementary File 10). We quantified the immunoprecipitation by qRT-PCR analysis with SYBR Green (BioRad) on an iQ5 Thermal cycler. We quantified H3K9me3 immunoprecipitation normalizing with the input chromatin for each sample. Results were analyzed using the dCT method and following the recommendations of the MIQE guideline (115).

FIGURE LEGENDS

Figure 1. Allele-specific expression analysis. Results from female guts in non-infected conditions (in green) and in infected conditions (in purple). Each dot represents the average ratio of gene expression levels between the allele with the TE and the allele without the TE for the three replicas analyzed. Each gene has two dots representing each one of the two genetic backgrounds analyzed. Statistically significant differences are depicted with dark color (t-test p-values < 0.05, corrected for FDR 5%). Error bars represent SEM.

Figure 2. *FBti0019386* adds a new TSS to its nearby gene *Bin1* used under infected conditions. (A) Transcripts annotated for *Bin1*. *FBti0019386* overlaps with two of the annotated transcripts. (B) Transcripts detected by RT-PCR in flies with and without *FBti0019386*, both in non-infected (N-Inf) and infected (Inf) conditions. Transcript

regions wave-patterned are inferred from FlyBase transcript annotation and were not sequenced in this work. *Bin1-RD* and *Bin1-RE* transcripts are, respectively, 318 bp and 172 bp shorter compared to *Bin1-RC* transcript. (C) *Bin1* expression levels in the gut in the two backgrounds analyzed in the ASE, and a third homozygous background with *FBti0019386* insertion (CAS-49). Error bars represent SEM. ND: not detected. (D) Survival curves in non-infected (discontinuous lines) and infected conditions (continuous lines) for flies with *FBti0019386* insertion (red) and without this insertion (grey). Error bars represent SEM.

Figure 3. *FBti0018868* adds a new TSS to its nearby gene *TM4SF*. (A) *FBti0018868* is located upstream of *TM4SF* gene and it has three annotated transcripts described. We detected a new *TM4SF* transcript overlapping with *FBti0018868* both in flies in non-infected and infected conditions. Transcript regions wave-patterned are inferred from FlyBase transcript annotation and were not sequenced in this work. (B) *TM4SF* total expression levels in the gut in the two different backgrounds used in the ASE (I and II) with and without *FBti0018868*. Error bars represent SEM. ND: not detected. (C) Vector constructions with the empty vector as a negative control, and a vector carrying *FBti0018868* in front of the reporter gene *lacZ*. Below, β-GAL immunostaining (in green), and DAPI staining (in grey) of guts from transgenic females with the empty vector and with the *FBti0018868* construct. The scale bar represents 500 µm. (D) Survival curves in non-infected (discontinuous lines) and infected conditions (continuous lines) for flies with (red) and without *FBti0018868* (grey). Error bars represent SEM.

Figure 4. *tdn8* acts as an enhancer regulatory sequence. (A) *tdn8* is located upstream of *CG10943* gene. (B) Vector construction without *tdn8* and with *tdn8* in the promoter region of the reporter gene *lacZ*. (C) Expression levels of the reporter gene *lacZ* in transgenic female guts without and with *tdn8*, both in non-infected and in infected conditions. (D) β-GAL immunostaining (in green), and DAPI staining (in grey) from female non-infected and infected guts. The scale bar represents 500 µm.

Figure 5. *FBti0061506* does not drive the expression of a reporter gene. (A) *FBti0061506* is located in the first intron of one of *Dif* transcripts, and upstream of the other transcripts. (B) Vectors construction for the enhancer assays with and without

FBti0061506 in front of the reporter gene *LacZ*. (C) β-GAL immunostaining (in green), and DAPI staining (in grey) from transgenic female non-infected and infected guts. The scale bar represents 500 µm.

Figure 6. *FBti0019985* acts as an enhancer regulatory sequence. (A) Transcripts annotated for the gene *cbx*. (B) Vector constructs for the enhancer assays. (C) Expression levels of the *lacZ* reporter gene in transgenic female guts, both under non-infected and infected conditions. Empty vector showed no detectable expression levels in any of the two conditions. (D) β-GAL immunostaining (in green), and DAPI staining (in grey) from transgenic female guts with *FBti0019985*. Scale bar represents 500 µm.

Figure 7. *FBti0020057* down-regulates the expression of a reporter gene. (A) *FBti0020057* is located in the intergenic region of *CG15829* and *CG8628.* (B) Vectors construction for the enhancer assays with and without *FBti0020057* in front of the reporter gene *LacZ*. (C) Expression levels of the *lacZ* reporter gene under non-infected and infected conditions. Error bars represent SEM. (D) β-GAL immunostaining (in green), and DAPI staining (in grey) from transgenic female non-infected and infected guts. The scale bar represents 500 µm. (E) ChIP qRT-PCR analysis for H3K9me3 in the genomic region were *FBti0020057* is inserted. "Absent": H3K9me3 enrichment in the reference strain that does not contain *FBti0020057* (grey), "Left": H3K9me3 enrichment in the left flanking region of *FBti0020057* in RAL-903 that contains this TE (red), and "Right": H3K9me3 enrichment in the right TE flanking region of *FBti0020057* in in RAL-903 (red).

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the European Commission (H2020-ERC-2014-CoG-647900 and FP7-PEOPLE-2011-CIG-293860), by the MEC/FEDER (BFU2014-57779-P). A.U. was a FPI fellow (BES-2012-052999) and JG was a Ramon y Cajal fellow (RYC-2010-07306). The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

Authors' contributions

AU participated in the conception and design of the experiments, performed the experiments, analyzed the results, and drafted the manuscript. MM performed experiments and analyzed the results. JG participated in the conception and design of the experiment, analyzed the results, and wrote the manuscript.

Acknowledgements

We would like to thank all the members of González Lab for help with the experiments, and for their comments on the manuscript.

REFERENCES

1. Hoffmann JA, Reichhart JM. Drosophila innate immunity: an evolutionary perspective. Nat Immunol. 2002;3(2):121-6.
2. Quintana-Murci L, Clark AG. Population genetic tools for dissecting innate immunity in humans. Nat Rev Immunol. 2013;13(4):280-93.
3. Bergman P, Seyedoleslami Esfahani S, Engström Y. Drosophila as a Model for Human Diseases-Focus on Innate Immunity in Barrier Epithelia. Curr Top Dev Biol. 2017;121:29-81.
4. Buchmann K. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. Front Immunol. 2014;5:459.
5. Buchon N, Silverman N, Cherry S. Immunity in Drosophila melanogaster--from microbial recognition to whole-organism physiology. Nat Rev Immunol. 2014;14(12):796-810.
6. Lemaitre B, Hoffmann J. The host defense of Drosophila melanogaster. Annu Rev Immunol. 2007;25:697-743.
7. Martins NE, Faria VG, Teixeira L, Magalhães S, Sucena É. Host adaptation is contingent upon the infection route taken by pathogens. PLoS Pathog. 2013;9(9):e1003601.
8. Teixeira L. Whole-genome expression profile analysis of Drosophila melanogaster immune responses. Brief Funct Genomics. 2012;11(5):375-86.
9. Boutros M, Agaisse H, Perrimon N. Sequential activation of signaling pathways during innate immune responses in Drosophila. Dev Cell. 2002;3(5):711-22.
10. De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B. The Toll and Imd pathways are the major regulators of the immune response in Drosophila. EMBO J. 2002;21(11):2568-79.

11. Myllymäki H, Rämet M. JAK/STAT pathway in Drosophila immunity. Scand J Immunol. 2014;79(6):377-85.
12. Raftery N, Stevenson NJ. Advances in anti-viral immune defence: revealing the importance of the IFN JAK/STAT pathway. Cell Mol Life Sci. 2017;74(14):2525-35.
13. Arthur JS, Ley SC. Mitogen-activated protein kinases in innate immunity. Nat Rev Immunol. 2013;13(9):679-92.
14. Rämet M, Lanot R, Zachary D, Manfruelli P. JNK signaling pathway is required for efficient wound healing in Drosophila. Dev Biol. 2002;241(1):145-56.
15. Bonfini A, Liu X, Buchon N. From pathogens to microbiota: How Drosophila intestinal stem cells react to gut microbes. Dev Comp Immunol. 2016;64:22-38.
16. Capo F, Charroux B, Royet J. Bacteria sensing mechanisms in Drosophila gut: Local and systemic consequences. Dev Comp Immunol. 2016;64:11-21.
17. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B. Drosophila intestinal response to bacterial infection: activation of host defense and stem cell proliferation. Cell Host Microbe. 2009;5(2):200-11.
18. Buchon N, Osman D, David FP, Fang HY, Boquete JP, Deplancke B, et al. Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep. 2013;3(5):1725-38.
19. Broderick NA, Buchon N, Lemaitre B. Microbiota-induced changes in drosophila melanogaster host gene expression and gut morphology. MBio. 2014;5(3):e01117-14.
20. Kayama H, Takeda K. Functions of innate immune cells and commensal bacteria in gut homeostasis. J Biochem. 2016;159(2):141-9.
21. Lazzaro BP, Rolff J. Immunology. Danger, microbes, and homeostasis. Science. 2011;332(6025):43-4.
22. Buchon N, Broderick NA, Lemaitre B. Gut homeostasis in a microbial world: insights from Drosophila melanogaster. Nat Rev Microbiol. 2013;11(9):615-26.
23. 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.
24. 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.
25. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. Science. 2014;343(6175):1246949.
26. Lee MN, Ye C, Villani A-C, Raj T, Li W, Eisenhaure TM, et al. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science. 2014;343(6175):1246980.
27. Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. Proc Natl Acad Sci U S A. 2018;115(3):E488-E97.
28. Sackton TB, Lazzaro BP, Clark AG. Genotype and gene expression associations with immune function in Drosophila. PLoS Genet. 2010;6(1):e1000797.
29. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18(2):71-86.
30. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016;351(6274):aac7247.
31. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21-42.

32. Sundaram V, Wang T. Transposable Element Mediated Innovation in Gene Regulatory Landscapes of Cells: Re-Visiting the "Gene-Battery" Model. Bioessays. 2018;40(1).
33. Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, et al. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep. 2015;10(4):551-61.
34. 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.
35. Sundaram V, Choudhary MN, Pehrsson E, Xing X, Fiore C, Pandey M, et al. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. Nat Commun. 2017;8:14550.
36. 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.
37. Estécio MR, Gallegos J, Dekmezian M, Lu Y, Liang S, Issa JP. SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. Mol Cancer Res. 2012;10(10):1332-42.
38. Lee YCG, Karpen GH. Pervasive epigenetic effects of. Elife. 2017;6.
39. Guio L, Vieira C, González J. Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster. Sci Rep. 2018;8(1):12197.
40. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083-7.
41. Flemr M, Malik R, Franke V, Nejepinska J, Sedlacek R, Vlahovicek K, et al. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell. 2013;155(4):807-16.
42. Guio L, Barrón MG, González J. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Mol Ecol. 2014;23(8):2020-30.
43. 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.
44. Wang L, Rishishwar L, Jordan IK, Mariño-Ramírez L. Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements. Nucleic Acids Research. 2016;45(5):2318-28.
45. Fiston-Lavier AS, Barrón MG, Petrov DA, González 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.
46. 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.
47. 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.
48. Keebaugh ES, Schlenke TA. Insights from natural host-parasite interactions: the Drosophila model. Dev Comp Immunol. 2014;42(1):111-23.

49. Costa E, Beltran S, Espinàs ML. Drosophila melanogaster SAP18 protein is required for environmental stress responses. FEBS Lett. 2011;585(2):275-80.
50. Kallio J, Leinonen A, Ulvila J, Valanne S, Ezekowitz RA, Rämet M. Functional analysis of immune response genes in Drosophila identifies JNK pathway as a regulator of antimicrobial peptide gene expression in S2 cells. Microbes Infect. 2005;7(5-6):811-9.
51. Kleino A, Valanne S, Ulvila J, Kallio J, Myllymäki H, Enwald H, et al. Inhibitor of apoptosis 2 and TAK1-binding protein are components of the Drosophila Imd pathway. EMBO J. 2005;24(19):3423-34.
52. Kim T, Yoon J, Cho H, Lee W-b, Kim J, Song Y-H, et al. Downregulation of lipopolysaccharide response in drosophila by negative crosstalk between the AP1 and NF-κB signaling modules. Nature Immunology. 2005;6:211.
53. 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.
54. Erkosar B, Erkosar Combe B, Defaye A, Bozonnet N, Puthier D, Royet J, et al. Drosophila microbiota modulates host metabolic gene expression via IMD/NF-κB signaling. PLoS One. 2014;9(4):e94729.
55. Roxström-Lindquist K, Terenius O, Faye I. Parasite-specific immune response in adult Drosophila melanogaster: a genomic study. EMBO Rep. 2004;5(2):207-12.
56. Ulvila J, Vanha-aho L-M, Kleino A, Vähä-Mäkilä M, Vuoksio M, Eskelinen S, et al. Cofilin regulator 14-3-3ζ is an evolutionarily conserved protein required for phagocytosis and microbial resistance. Journal of Leukocyte Biology. 2011;89(5):649-59.
57. Brown AE, Baumbach J, Cook PE, Ligoxygakis P. Short-term starvation of immune deficient Drosophila improves survival to gram-negative bacterial infections. PLoS One. 2009;4(2):e4490.
58. Christofi T, Apidianakis Y. Drosophila immune priming against Pseudomonas aeruginosa is short-lasting and depends on cellular and humoral immunity. F1000Res. 2013;2:76.
59. Cornwell WD, Kirkpatrick RB. Cactus-independent nuclear translocation of Drosophila RELISH. J Cell Biochem. 2001;82(1):22-37.
60. Gobert V, Gottar M, Matskevich AA, Rutschmann S, Royet J, Belvin M, et al. Dual activation of the Drosophila toll pathway by two pattern recognition receptors. Science. 2003;302(5653):2126-30.
61. Rutschmann S, Jung AC, Hetru C, Reichhart JM, Hoffmann JA, Ferrandon D. The Rel protein DIF mediates the antifungal but not the antibacterial host defense in Drosophila. Immunity. 2000;12(5):569-80.
62. Clark RI, Tan SW, Péan CB, Roostalu U, Vivancos V, Bronda K, et al. MEF2 is an in vivo immune-metabolic switch. Cell. 2013;155(2):435-47.
63. Silverman N, Zhou R, Erlich RL, Hunter M, Bernstein E, Schneider D, et al. Immune activation of NF-kappaB and JNK requires Drosophila TAK1. J Biol Chem. 2003;278(49):48928-34.
64. Valanne S, Kleino A, Myllymäki H, Vuoristo J, Rämet M. Iap2 is required for a sustained response in the Drosophila Imd pathway. Dev Comp Immunol. 2007;31(10):991-1001.
65. Short SM, Lazzaro BP. Reproductive status alters transcriptomic response to infection in female Drosophila melanogaster. G3 (Bethesda). 2013;3(5):827-40.
66. Reumer A, Bogaerts A, Van Loy T, Husson SJ, Temmerman L, Choi C, et al. Unraveling the protective effect of a Drosophila phosphatidylethanolamine-binding

protein upon bacterial infection by means of proteomics. Dev Comp Immunol. 2009;33(11):1186-95.
67. Juneja P, Quinn A, Jiggins FM. Latitudinal clines in gene expression and cis-regulatory element variation in Drosophila melanogaster. BMC Genomics. 2016;17(1):981.
68. Zhang Q, Zhang L, Gao X, Qi S, Chang Z, Wu Q. DIP1 plays an antiviral role against DCV infection in Drosophila melanogaster. Biochemical and Biophysical Research Communications. 2015;460(2):222-6.
69. Fukuyama H, Verdier Y, Guan Y, Makino-Okamura C, Shilova V, Liu X, et al. Landscape of protein-protein interactions in Drosophila immune deficiency signaling during bacterial challenge. Proc Natl Acad Sci U S A. 2013;110(26):10717-22.
70. Berkey CD, Blow N, Watnick PI. Genetic analysis of Drosophila melanogaster susceptibility to intestinal Vibrio cholerae infection. Cell Microbiol. 2009;11(3):461-74.
71. Levy S, Shoham T. The tetraspanin web modulates immune-signalling complexes. Nat Rev Immunol. 2005;5(2):136-48.
72. Arbouzova NI, Bach EA, Zeidler MP. Ken & barbie selectively regulates the expression of a subset of Jak/STAT pathway target genes. Curr Biol. 2006;16(1):80-8.
73. Watson FL, Püttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science. 2005;309(5742):1874-8.
74. Valanne S, Myllymäki H, Kallio J, Schmid MR, Kleino A, Murumägi A, et al. Genome-wide RNA interference in Drosophila cells identifies G protein-coupled receptor kinase 2 as a conserved regulator of NF-kappaB signaling. J Immunol. 2010;184(11):6188-98.
75. Carpenter J, Hutter S, Baines JF, Roller J, Saminadin-Peter SS, Parsch J, et al. The transcriptional response of Drosophila melanogaster to infection with the sigma virus (Rhabdoviridae). PLoS One. 2009;4(8):e6838.
76. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430(6995):85-8.
77. Hawkins JS, Delgado V, Feng L, Carlise M, Dooner HK, Bennetzen JL. Variation in allelic expression associated with a recombination hotspot in Zea mays. Plant J. 2014;79(3):375-84.
78. Tung J, Akinyi MY, Mutura S, Altmann J, Wray GA, Alberts SC. Allele-specific gene expression in a wild nonhuman primate population. Mol Ecol. 2011;20(4):725-39.
79. Von Korff M, Radovic S, Choumane W, Stamati K, Udupa SM, Grando S, et al. Asymmetric allele-specific expression in relation to developmental variation and drought stress in barley hybrids. The Plant Journal. 2009;59(1):14-26.
80. 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.
81. 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.
82. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45(D1):D190-D9.
83. De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A. 2001;98(22):12590-5.

84. Dionne MS, Pham LN, Shirasu-Hiza M, Schneider DS. Akt and FOXO dysregulation contribute to infection-induced wasting in Drosophila. Curr Biol. 2006;16(20):1977-85.
85. Felix TM, Hughes KA, Stone EA, Drnevich JM, Leips J. Age-specific variation in immune response in Drosophila melanogaster has a genetic basis. Genetics. 2012;191(3):989-1002.
86. Lee YC. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. PLoS Genet. 2015;11(6):e1005269.
87. Yin H, Sweeney S, Raha D, Snyder M, Lin H. A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster. PLoS Genet. 2011;7(12):e1002380.
88. David JR, Capy P. Genetic variation of Drosophila melanogaster natural populations. Trends Genet. 1988;4(4):106-11.
89. Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2006;2(10):e166.
90. Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, et al. Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. Mol Ecol. 2012;21(19):4748-69.
91. Fan S, Hansen ME, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. Science. 2016;354(6308):54-9.
92. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Ferrer-Admettla A, Pattini L, et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 2011;7(11):e1002355.
93. Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster. Genetics. 2011;187(1):245-60.
94. Quach H, Quintana-Murci L. Living in an adaptive world: Genomic dissection of the genus. J Exp Med. 2017;214(4):877-94.
95. 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.
96. 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.
97. Osada N, Miyagi R, Takahashi A. *Cis*- and *Trans*-regulatory Effects on Gene Expression in a Natural Population of *Drosophila melanogaster* and. Genetics. 2017;206(4):2139-48.
98. Glaser-Schmitt A, Parsch J. Functional characterization of adaptive variation within a cis-regulatory element influencing Drosophila melanogaster growth. PLoS Biol. 2018;16(1):e2004538.
99. Wertheim B, Kraaijeveld AR, Hopkins MG, Walther Boer M, Godfray HC. Functional genomics of the evolution of increased resistance to parasitism in Drosophila. Mol Ecol. 2011;20(5):932-49.
100. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613-26.

101. 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.
102. Kapitonov VV, Jurka J. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A. 2003;100(11):6569-74.
103. Sackton TB, Kulathinal RJ, Bergman CM, Quinlan AR, Dopman EB, Carneiro M, et al. Population genomic inferences from sparse high-throughput sequencing of two populations of Drosophila melanogaster. Genome Biol Evol. 2009;1:449-65.
104. Singh ND, Petrov DA. Rapid sequence turnover at an intergenic locus in Drosophila. Mol Biol Evol. 2004;21(4):670-80.
105. Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. Drosophila melanogaster recombination rate calculator. Gene. 2010;463(1-2):18-20.
106. Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in Drosophila melanogaster. PLoS Genet. 2012;8(10):e1002905.
107. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134(4):1289-303.
108. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8(3):269-94.
109. Hudson RR, Kaplan NL. Deleterious background selection with recombination. Genetics. 1995;141(4):1605-17.
110. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23(1):23-35.
111. Barrón MG, Fiston-Lavier AS, Petrov DA, González J. Population genomics of transposable elements in Drosophila. Annu Rev Genet. 2014;48:561-81.
112. Castellano D, Coronado-Zamora M, Campos JL, Barbadilla A, Eyre-Walker A. Adaptive Evolution Is Substantially Impeded by Hill-Robertson Interference in Drosophila. Mol Biol Evol. 2016;33(2):442-55.
113. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.
114. González 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.
115. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55(4):611-22.
116. Neyen C, Bretscher AJ, Binggeli O, Lemaitre B. Methods to study Drosophila immunity. Methods. 2014;68(1):116-28.
117. Vallet-Gely I, Novikov A, Augusto L, Liehl P, Bolbach G, Péchy-Tarr M, et al. Association of hemolytic activity of Pseudomonas entomophila, a versatile soil bacterium, with cyclic lipopeptide production. Appl Environ Microbiol. 2010;76(3):910-21.
118. Ràmia M, Librado P, Casillas S, Rozas J, Barbadilla A. PopDrowser: the Population Drosophila Browser. Bioinformatics. 2012;28(4):595-6.
119. Junqueira AC, Azeredo-Espin AM, Paulo DF, Marinho MA, Tomsho LP, Drautz-Moses DI, et al. Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. Sci Rep. 2016;6:21762.
120. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol. 2006;16(6):580-5.

121. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying adaptive evolution in the Drosophila immune system. PLoS Genet. 2009;5(10):e1000698.
122. Obbard DJ, Jiggins FM, Bradshaw NJ, Little TJ. Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of Drosophila. Mol Biol Evol. 2011;28(2):1043-56.
123. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2011;13(1):59-69.
124. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
125. Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A. 2007;104(9):3312-7.
126. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 2018;46(D1):D1284.
127. Brizuela BJ, Elfring L, Ballard J, Tamkun JW, Kennison JA. Genetic analysis of the brahma gene of Drosophila melanogaster and polytene chromosome subdivisions 72AB. Genetics. 1994;137(3):803-13.
128. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of Drosophila melanogaster. Science. 2000;287(5461):2185-95.
129. Celniker SE, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, Patel S, et al. Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. Genome Biol. 2002;3(12):RESEARCH0079.
130. Hoskins RA, Nelson CR, Berman BP, Laverty TR, George RA, Ciesiolka L, et al. A BAC-based physical map of the major autosomes of Drosophila melanogaster. Science. 2000;287(5461):2271-4.

FIGURE 1

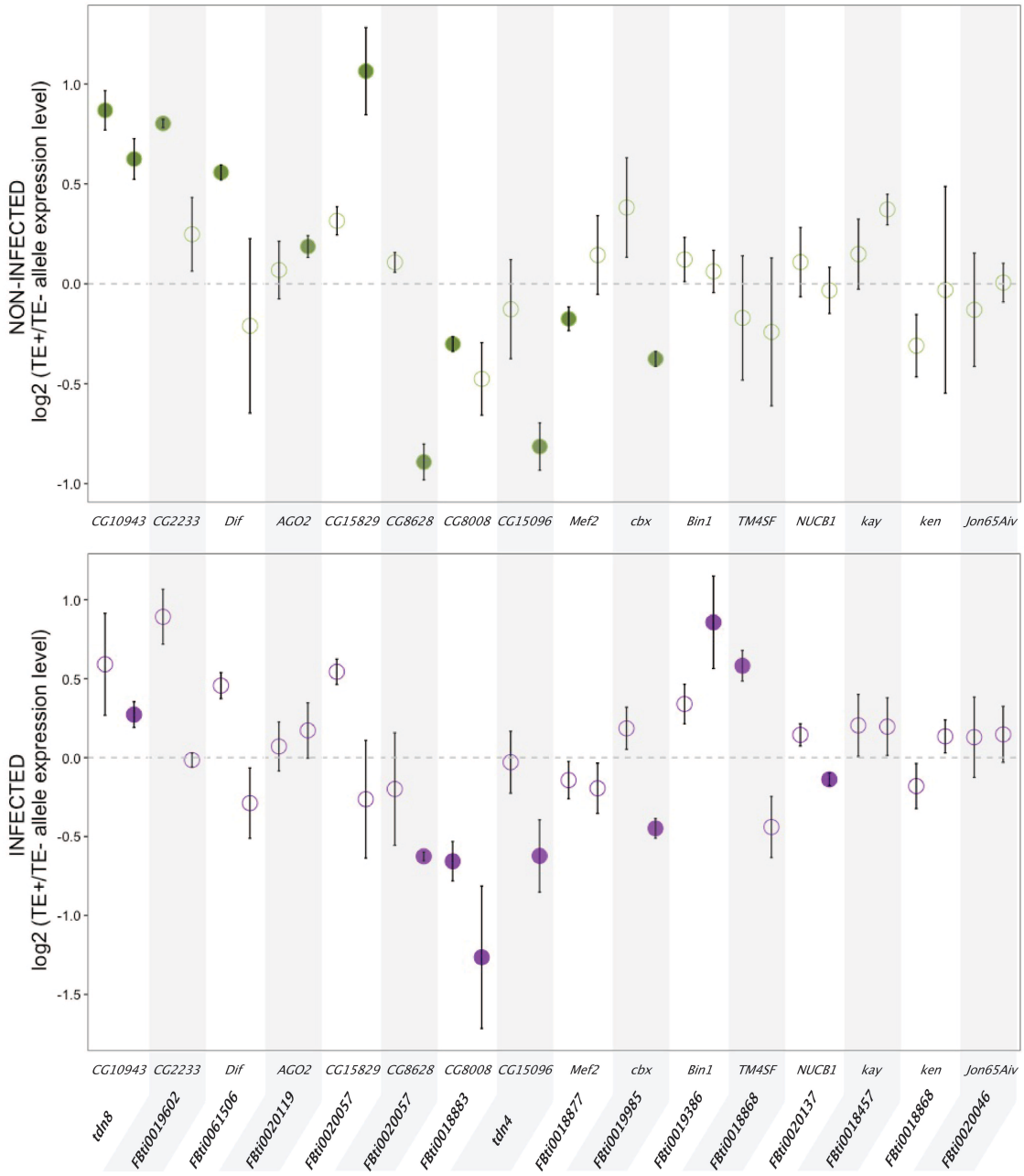

NON-INFECTED
log2 (TE+/TE- allele expression level)
1.0
0.5
0.0
-0.5
-1.0
CG10943
CG2233
Dif
AGO2
CG15829
CG8628
CG8008
CG15096
Mef2
cbx
Bin1
TM4SF
NUCB1
kay
ken
Jon65Aiv
INFECTED
log2 (TE+/TE- allele expression level)
1.0
0.5
0.0
-0.5
-1.0
-1.5
CG10943
CG2233
Dif
AGO2
CG15829
CG8628
CG8008
CG15096
Mef2
cbx
Bin1
TM4SF
NUCB1
kay
ken
Jon65Aiv
tdn8
FBti0019602
FBti0061506
FBti0020119
FBti0020057
FBti0020057
FBti0018883
tdn4
FBti0018877
FBti0019985
FBti0019386
FBti0018868
FBti0020137
FBti0018457
FBti0018868
FBti0020046

FIGURE 2

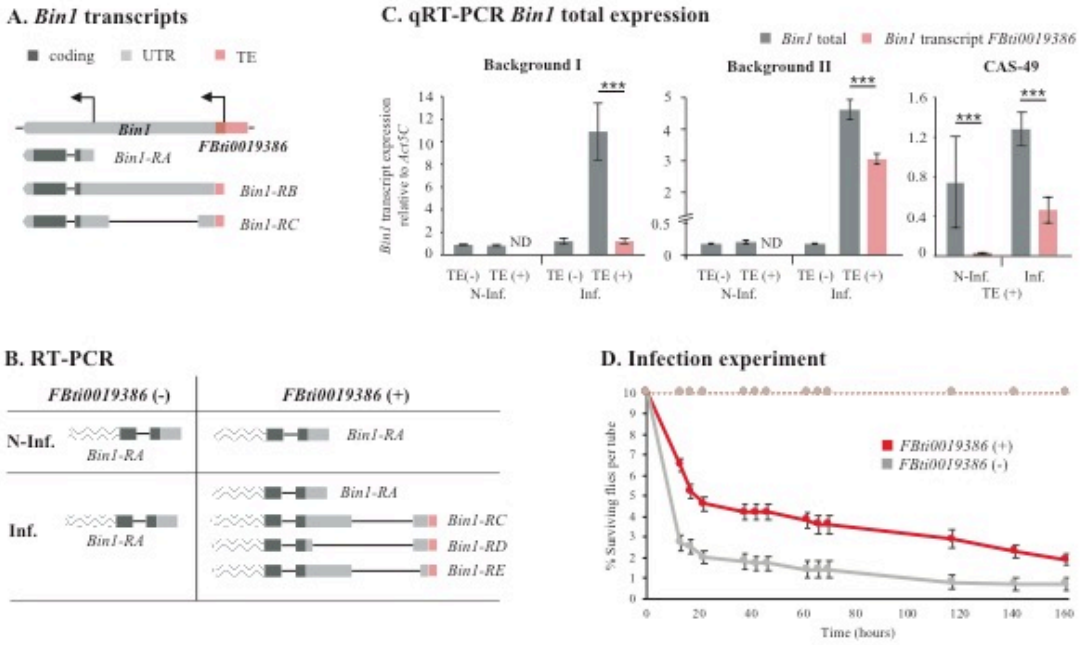

A. Bin1 transcripts
coding
UTR
TE
Bin1
FBti0019386
Bin1-RA
Bin1-RB
Bin1-RC
C. qRT-PCR Bin1 total expression
Bin1 total
Bin1 transcript FBti0019386
Background I
Background II
CAS-49
Bin1 transcript expression relative to Act5C
ND
***
TE(-) TE (+) TE (-) TE (+)
N-Inf.
Inf.
B. RT-PCR
FBti0019386 (-)
FBti0019386 (+)
N-Inf.
Inf.
Bin1-RA
Bin1-RC
Bin1-RD
Bin1-RE
D. Infection experiment
FBti0019386 (+)
FBti0019386 (-)
% Surviving flies per tube
Time (hours)

FIGURE 3

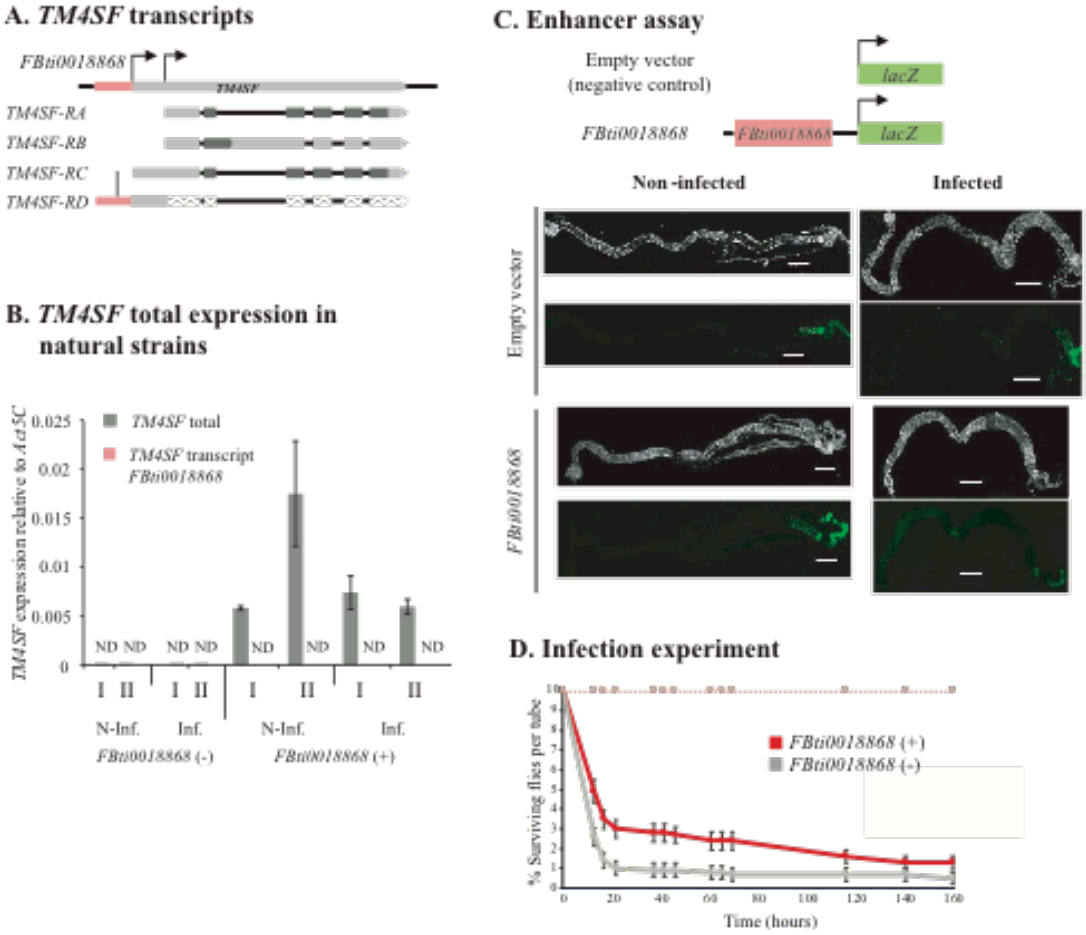

A. TM4SF transcripts
FBti0018868
TM4SF
TM4SF-RA
TM4SF-RB
TM4SF-RC
TM4SF-RD
B. TM4SF total expression in natural strains
TM4SF expression relative to Ad5C
0.025
0.02
0.015
0.01
0.005
0
TM4SF total
TM4SF transcript FBti0018868
ND ND ND ND ND ND ND ND
I II I II I II I II
N-Inf. Inf. N-Inf. Inf.
FBti0018868 (-)
FBti0018868 (+)
C. Enhancer assay
Empty vector (negative control)
lacZ
FBti0018868
FBti0018868
lacZ
Non -infected
Infected
Empty vector
FBti0018868
D. Infection experiment
% Surviving flies per tube
FBti0018868 (+)
FBti0018868 (-)
0 20 40 60 80 100 120 140 160
Time (hours)

FIGURE 4

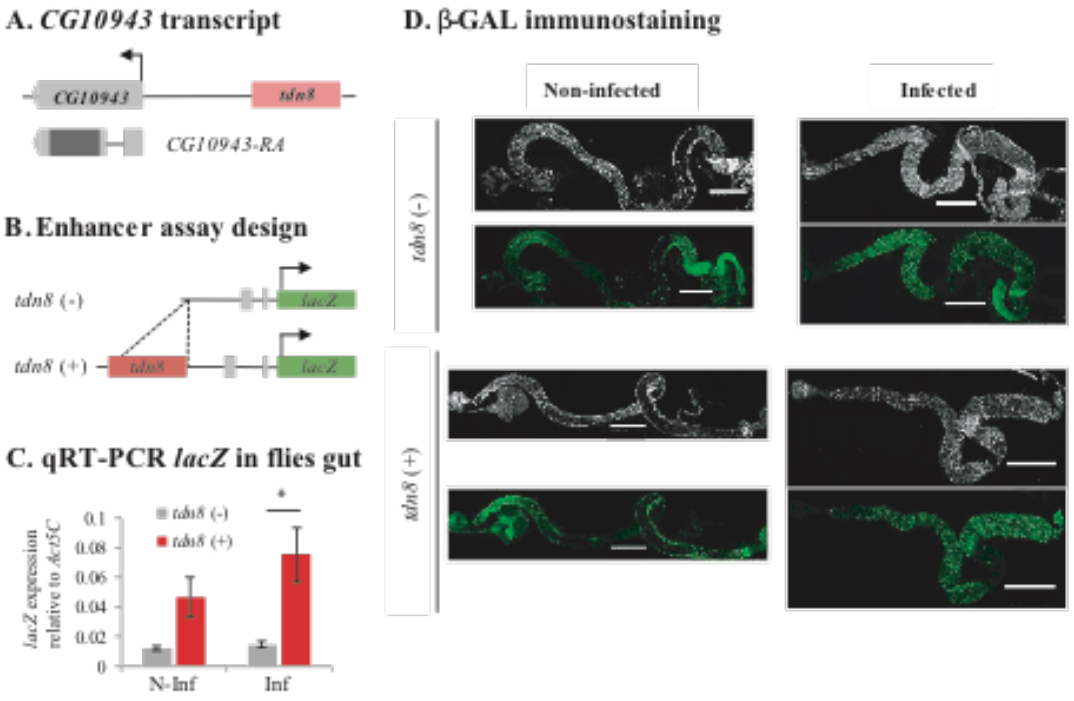

A. CG10943 transcript
CG10943
tdn8
CG10943-RA
B. Enhancer assay design
tdn8 (-)
lacZ
tdn8 (+)
tdn8
lacZ
C. qRT-PCR lacZ in flies gut
lacZ expression relative to Act5C
0.1
0.08
0.06
0.04
0.02
0
tdn8 (-)
tdn8 (+)
*
N-Inf
Inf
D. β-GAL immunostaining
Non-infected
Infected
tdn8 (-)
tdn8 (+)

FIGURE 5

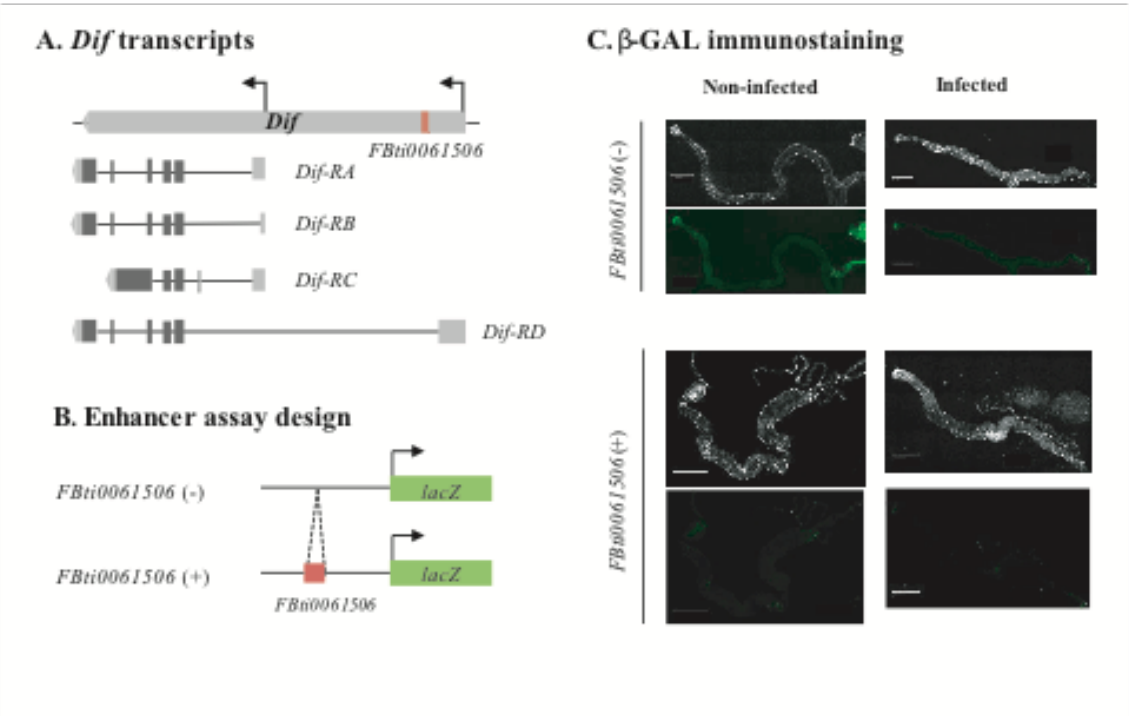

A. Dif transcripts
Dif
FBti0061506
Dif-RA
Dif-RB
Dif-RC
Dif-RD
B. Enhancer assay design
FBti0061506 (-)
lacZ
FBti0061506 (+)
lacZ
FBti0061506
C. β-GAL immunostaining
Non-infected
Infected
FBti0061506 (-)
FBti0061506 (+)

FIGURE 6

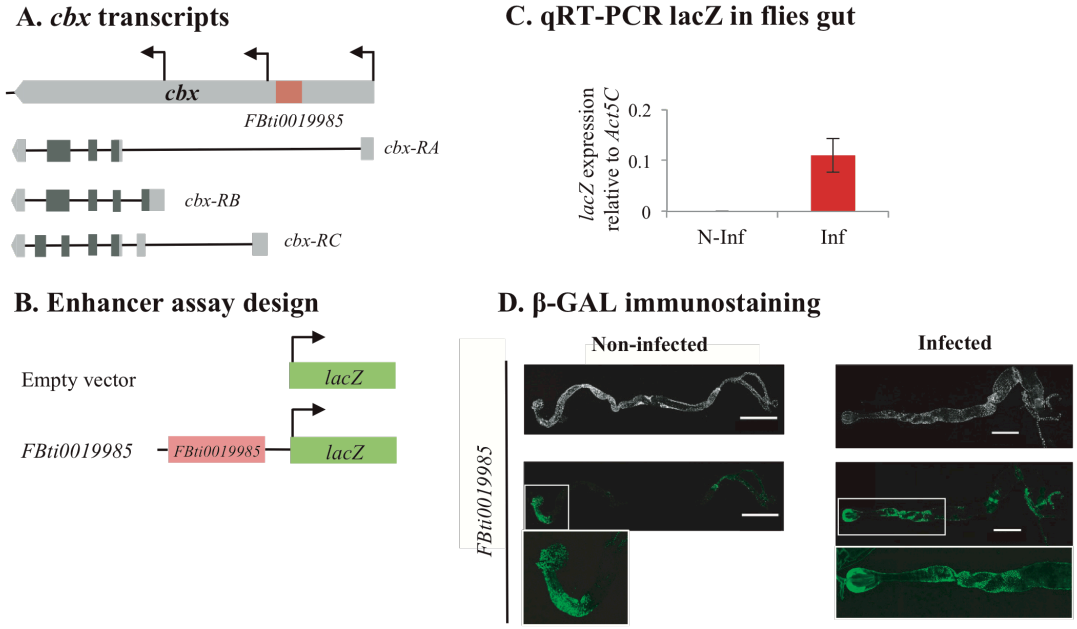
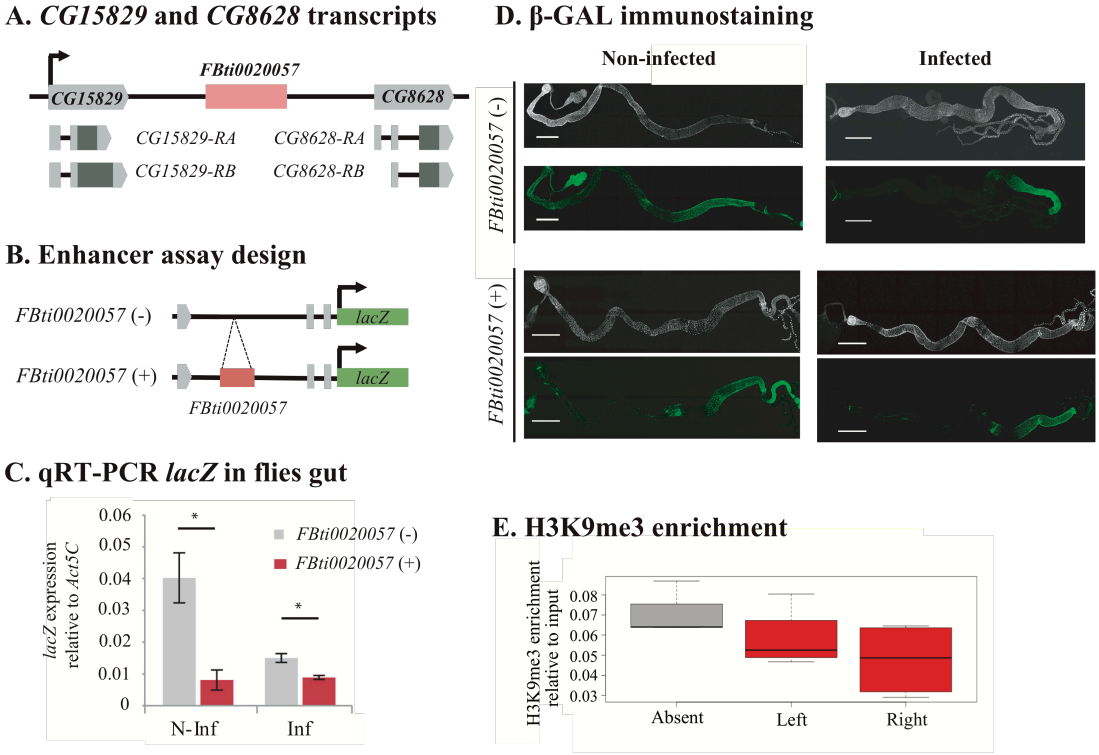

A. cbx transcripts
cbx
FBti0019985
cbx-RA
cbx-RB
cbx-RC
B. Enhancer assay design
Empty vector
lacZ
FBti0019985
FBti0019985
lacZ
C. qRT-PCR lacZ in flies gut
lacZ expression relative to Act5C
0.2
0.1
0
N-Inf
Inf
D. β-GAL immunostaining
Non-infected
Infected
FBti0019985

FIGURE 7


A. CG15829 and CG8628 transcripts
FBti0020057
CG15829
CG8628
CG15829-RA
CG8628-RA
CG15829-RB
CG8628-RB
B. Enhancer assay design
FBti0020057 (-)
FBti0020057 (+)
lacZ
lacZ
FBti0020057
C. qRT-PCR lacZ in flies gut
lacZ expression relative to Act5C
0.06
0.05
0.04
0.03
0.02
0.01
0
*
*
FBti0020057 (-)
FBti0020057 (+)
N-Inf
Inf
D. β-GAL immunostaining
Non-infected
Infected
FBti0020057 (-)
FBti0020057 (+)
E. H3K9me3 enrichment
H3K9me3 enrichment relative to input
0.08
0.07
0.06
0.05
0.04
0.03
Absent
Left
Right