MBE Advance Access published May 20, 2009

A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*

Josefa González, J. Michael Macpherson, and Dmitri A. Petrov

Department of Biology, Stanford University, Stanford, CA 94305

Submitted to Molecular Biology and Evolution as a Research Article
Corresponding author:

Josefa González
Department of Biology
Stanford University
Stanford, CA 94305
Phone: 650-7362249
Fax: 650-7236132
jgonzalp@stanford.edu

Keywords: adaptation, transposable elements, Drosophila, purifying selection, life-history traits
Running head: adaptation near constrained loci
Abbreviations: AF, African; *Dmel, D. melanogaster; Dsim, D. simulans; Dyak, D. yakuba*; DT: developmental time; iHS, integrated haplotype score; JHa, Juvenile Hormone analog; NA, North American; TE, transposable element

ABSTRACT

A recent genome-wide screen identified 13 transposable elements that are likely to have been adaptive during or after the spread of *Drosophila melanogaster* out of Africa. One of these insertions, *Bari-Jheh*, was associated with the selective sweep of its flanking neutral variation and with reduction of expression of one of its neighboring genes: *Jheh3*. Here, we provide further evidence that *Bari-Jheh* insertion is adaptive. We delimit the extent of the selective sweep and show that *Bari-Jheh* is the only mutation linked to the sweep. *Bari-Jheh* also lowers the expression of its other flanking gene, *Jheh2*. Subtle consequences of *Bari-Jheh* insertion on life history traits are consistent with the effects of reduced expression of the *Jheh* genes. Finally, we analyze molecular evolution of *Jheh* genes both in the long and the short term and conclude that *Bari-Jheh* appears to be a very rare adaptive event in the history of these genes. We discuss the implications of these findings for the detection and understanding of adaptation.

INTRODUCTION

Transposable elements (TEs) were once considered to be intragenomic parasites leading to almost exclusively detrimental effects to the host genome (Doolittle and Sapienza 1980; Orgel and Crick 1980; Charlesworth, Sniegowski, and Stephan 1994). However, there is growing evidence that TEs sometimes contribute positively to the function and evolution of genes and genomes (Kidwell and Lisch 2001; Daborn et al. 2002; Kazazian 2004; Aminetzach, Macpherson, and Petrov 2005; Biemont and Vieira 2006; Jurka et al. 2007). For example, TEs have contributed to the regulatory and/ or coding sequences of a large number of genes (van de Lagemaat et al. 2003; Marino-Ramirez et al. 2005; Piriyapongsa et al. 2007; Feschotte 2008). Recently, the first comprehensive genome-wide screen for recent adaptive TE insertions in the *Drosophila melanogaster* genome revealed that TEs are a considerable source of adaptive mutations in this species (Gonzalez et al. 2008).

González et al (2008) identified a set of 13 TEs that are likely to have contributed to the adaptation of *D. melanogaster* during its expansion out of Africa (David and Capy 1988; Lachaise 1988). Two lines of evidence pointed to the adaptive roles of these 13 TEs: (1) the flanking regions of all of the investigated TEs (five out of 13) showed signatures of partial selective sweeps (Smith and Haigh 1974; Kaplan, Darden, and Hudson 1988; Kaplan, Hudson, and Langley 1989) and (2) eight of the 13 TEs showed higher frequency in a more temperate compared to a more tropical Australian subpopulation consistent with these TEs playing a role in adaptation to temperate climates. These 13 TEs represent a rich collection for follow-up investigations of adaptive processes in *D. melanogaster*.

The high rate of TE-induced adaptive changes reported by González et al (2008), appeared to be incompatible with the low number of fixed TEs present in the *D. melanogaster* euchromatic genome. The authors suggested that most of these TEs represented local and ephemeral adaptations that were destined to be lost over long periods of time. It is possible then that the loci that are underlying much of the local adaptation over short periods of time would appear conserved when compared across species. If true, this would severely complicate the study of adaptation given that statistical inference of positive selection is often based on the assumption that adaptation is recurring at the same loci or even at the same sites (McDonald and Kreitman 1991; Jensen, Wong, and Aquadro 2007; Macpherson et al. 2007; Yang 2007). The adaptive TEs identified by González et al. (2008) constitute a good starting point to test whether recent adaptation takes place at loci that have shown historically high rate of adaptive divergence.

In this work, we analyzed one of these 13 TEs: FBti0018880, a full-length (1.7 kb) copy of a *Tc1-like* transposon that belongs to the *Bari1* family (Caizzi, Caggese, and Pimpinelli 1993). FBti0018880 is inserted in the 0.7 kb intergenic region between *Juvenile hormone epoxy hydrolase 2* (*Jheh2*) and *Jheh3* genes. Accordingly, we will refer to this insertion as *Bari-Jheh* for the remainder of this paper. Since these two genes have known functions, we can construct plausible hypothesis about the possible phenotypic consequences of the insertion. Both genes code for enzymes involved in the degradation of Juvenile Hormone (JH). JH is a regulator of development, life history and fitness trade-offs in insects (Flatt, Tu, and Tatar 2005; Riddiford 2008). The multiplicity of biological effects of JH requires specific titers of the hormone during different times of

the Drosophila development. The regulation of the JH titer is achieved by a balance between biosynthesis and degradation (de Kort 1996). Changes in the expression of *Jheh* genes are likely to affect JH titer and consequently any of the processes in which this hormone is involved. In Drosophila, these processes include metamorphosis, behavior, reproduction, diapause, stress resistance and aging (Flatt, Tu, and Tatar 2005).

We previously showed that the polymorphism pattern in the 2 kb region flanking *Bari-Jheh* is consistent with the expectations of a partial selective sweep and that *Bari-Jheh* affects the expression of *Jheh3* (Gonzalez et al. 2008). In this work, we expanded the analysis of the flanking region to include the whole coding sequence of the neighboring genes and performed additional allele-specific expression and phenotypic analyses. Altogether, we demonstrate that *Bari-Jheh* insertion is very likely to be an adaptive mutation. We also provide evidence suggesting that the adaptive insertion of *Bari-Jheh* is an extremely unusual event in the history of *Jheh* loci. We discuss the implications of these findings for the understanding of the adaptive process in Drosophila and the challenges that remain to associate *Bari-Jheh* insertion with the adaptively significant phenotype(s).

Material and Methods

Drosophila strains

Table S1 (Supplementary Material online) describes the *D. melanogaster*, *D. simulans* and *D.yakuba* stocks used in this work.

Sequencing and sequence analysis

Genomic DNA was extracted using DNeasy Tissue kit (Qiagen, Valencia, CA). Based on the genome sequence of *D. melanogaster*, *D. simulans* and *D. yakuba* (http://flybase.org) we designed primers in an overlapping fashion to amplify and sequence *Jheh1*, *Jheh2* and *Jheh3* genes. In *D. simulans*, the intergenic regions between these genes were also sequenced. The specific primers used for each species are given in supplementary Table S2 (Supplementary Material online). Only *D. melanogaster* populations from Davis and Raleigh proved isogenic. For the rest of the strains DNA was amplified using a proofreading DNA polymerase (Platinum Pfx; INVITROGEN, Carlsbad, CA) and cloned into Zero Blunt TOPO PCR cloning kit (INVITROGEN, Carlsbad, CA) before sequencing. All the sequences have been deposited in GenBank under accession numbers XXXXXXXX-XXXXXXXX.

D. melanogaster, *D.simulans* and *D. yakuba* sequences available at http://flybase.org were also included in the analysis. For *D.simulans* the genome sequence of 6 additional strains is available (Begun et al. 2007). However, most of the sequences had a poor quality and only North American (NA) strain w501 could be included in our analysis of the coding regions of *Jheh1* and *Jheh2* genes. Sequences were assembled using Sequencher 4.7 software (Gene Codes Corporation, Ann Arbor, MI), aligned with ClustalW (Thompson, Higgins, and Gibson 1994) and edited in MacClade (Maddison and Maddison 1989). *D. simulans* intergenic regions were aligned using DIALIGN (Morgenstern 2004). The repetitive content of the intergenic regions was analyzed using RepeatMasker, available at http://www.repeatmasker.org).

We analyzed the polymorphism pattern of the 5 kb region flanking *Bari-Jheh* insertion in *D. melanogaster* by comparing several summary statistics calculated over the

datasets to the distributions of these statistics obtained by neutral coalescent simulations as described in González et al (2008). The demographic model specified in Thorton and Andolfatto (2006) was incorporated into the simulations. The population was partitioned into two subpopulations, the NA and the African (AF), and the sample was partitioned into two subsamples defined by the presence/ absence of *Bari-Jheh*. We computed the pairwise nucleotide diversity (π) (Tajima 1983), the integrated haplotype score (iHS) (Voight et al. 2006) and the proportion of nucleotide diversity within the haplotypes linked to the TE to the total nucleotide diversity in the sample $f_{TE} = \pi_{TE}/(\pi_{TE} + \pi_{non-TE})$ (Macpherson et al. 2007). Polymorphism data for *D. simulans* was analyzed using DnaSP 4.0 (Rozas et al. 2003).

Detecting presence or absence of *Bari-Jheh*

The presence/absence of *Bari-Jheh* in The Netherlands population was determined by PCR as described in González et al (2008). The following primers were used: L: 5'-AGGGAGCCATCATTGTAATAGCG-3'; R: 5'-TTGTTGGCTTGTGGATTTCAAGT-3'and FL: 5'-CCTACACGGCGAGAAGAGAAAAT-3').

Analysis of the molecular evolution of *Jheh* genes using PAML software

Sequences of the three *Jheh* genes in the 12 Drosophila species were provided by S. Chatterji (S. Chatterji, personal communication). For each gene, sequences were aligned using ClustalW (Thompson, Higgins, and Gibson 1994) and manually edited

when necessary. We checked for duplications of these genes in each of the 12 Drosophila species using tblastx (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

We used PAML to estimate the degree of selective constraint (model M0) and to search for evidence of positive selection in the evolution of *Jheh* genes (models M7 and M8; Yang 2007). We first checked the congruence of the topologies of the gene and species tree. Phylogenetic trees for each of the three genes were built using MEGA 4 (Tamura et al. 2007). While the topology for the species of the melanogaster and obscura group was the same, differences in the topology for the other branches of the tree were found between each of the three *Jheh* genes trees and the species tree. This result could be due to the saturation of substitutions at synonymous sites (Bergman et al. 2002). Consequently, we ran PAML using both the 12 species tree and the tree that included only the 6 species in the melanogaster group.

Reverse Transcriptase-PCR analysis

A total of 50 four day-old female adult flies, 50 third-instar larvae, and 0-18 hour old embryos were collected from one strain with the insertion (Wi3) and one strain without the insertion (Wi1). Total RNA from the three stages was isolated using the TRIzol protocol (INVITROGEN, Carlsbad, CA). RNA was then treated with DNase and purified using RNeasy mini kit (Qiagen, Valencia, CA). First-Strand cDNA was synthesized using SuperScriptIII First-Strand synthesis system for reverse transcriptase PCR (RT-PCR) (INVITROGEN, Carlsbad, CA). To check for genomic contamination RT-PCR reactions without retrotranscriptase were performed.

Specific primers for each of the three *Jheh* genes were designed and are given in supplementary Table S3 (Supplementary Material online). For *Jheh2* gene three different sets of primers were designed. Primers Jheh2F and Jheh2R were designed in different exons to check both for genomic contamination and for the presence of the two alternative transcripts described for this gene. Primers Jheh2-PA_F and Jheh2-PA_R specifically amplified transcript *Jheh2-PA* and primers Jheh2-PB_F and Jheh2-PB_R specifically amplified transcript *Jheh2-PB*. PCRs were run using Pfx polymerase (INVITROGEN, Carlsbad, CA) and the following conditions: 94ºC for 4 min, 30 cycles of 94ºC 1 min, 55ºC 0.5 min, 68ºC 1 min and one last extension step of 10 min at 68ºC.

Allele specific expression analysis

We looked for differences in expression between a *Jheh2* allele carrying *Bari-Jheh* and a *Jheh2* allele lacking *Bari-Jheh* in F1 heterozygous hybrids. We established two different crosses between one strain homozygous for the presence of *Bari-Jheh* (Wi3) and one strain homozygous for its absence (Wi1). In cross 1 the mother was homozygous for the presence of *Bari-Jheh* and in cross 2 the father was homozygous for the presence of *Bari-Jheh*. These two reciprocal crosses allowed us to check for parental effects on the allele expression.

To check for differences in the level of expression between the alleles with and without *Bari-Jheh* we identified a SNP in the coding region of *Jheh2* (position 1756; Figure 1) that is perfectly linked with *Bari-Jheh* (the allele carrying *Bari-Jheh* insertion has a C and the allele lacking *Bari-Jheh* insertion has a T). We used this SNP as the marker for allele specific expression. Differences in expression between the two alleles

were assayed in 3-5 day old adults since the activity of JHEH increases during the first days after eclosion (Khlebodarova et al. 1996). For each cross, we collected males and females separately that were snap frozen in liquid nitrogen and stored at −80 °C until use. We have therefore a total of four samples: males and females progeny from cross 1 and males and females progeny from cross 2. We extracted RNA and synthesized cDNA from each of the four samples as explained above. We then used the cDNA as a template to amplify the diagnostic SNP using primers Jheh2F: 5'-TCGATAAGTTTCTGGTGCAGG -3' and Jheh2R: 5'-CCGGAAAAAGTGAGGCTACAT-3'. A universal sequence was appended to Jheh2R primer for the subsequent pyrosequencing reaction. PCR was done in the presence of 2.5 μM tailed primer, 10 μM non-tailed primer, and 10 μM universal biotin-labelled primer. We analyzed each sample in triplicate. The PCR product was then pyrosequenced to quantify the relative amount of C vs T in the cDNA (EpigenDx; Worcester, MA) using primers Jheh2R and Jheh2FS (TGGCGATTGGGGTTC).

To correct for unequal amplification of the SNP not related to unequal transcription we used the same primers to amplify genomic DNA of the F_1 adults where the ratio was 50:50 (Wang and Elbein 2007). Before testing for statistically significant differences in the expression of the alleles with and without *Bari-Jheh* data was transformed to fit a normal distribution using the arcsin transformation. Significance was then tested by an unpaired *t*-test since genomic DNA and cDNA come from different individuals

Phenotypic assays

Introgressed strains. We introgressed *Bari-Jheh* from two different NA strains, Wi3 and We33, into Wi1 strain. These three strains are isofemale strains that had been further put through 30-60 generations of brother-sister matings. In the first generation we mated virgin females from Wi3 or We33 with Wi1 males. In the second generation we backcrossed virgin F1 females with Wi1 males. In the subsequent generations we individually mated 10 females with Wi1 males (one male and one female per vial). We identified the crosses that involved introgressed strains carrying *Bari-Jheh* by PCR using the primers L/R and FL/R described previously. After 8 generations for Wi3 stock and eleven generations for We33 stock, we carried out brother-sister matings and identified one strain that was homozygous for the presence of *Bari-Jheh* and one strain homozygous for the absence. We named the different lines as follows: Wi3/Bari+ and We33/Bari+ are both lines homozygous for the presence of the element and Wi3/Bari- and We33/Bari- are homozygous for the absence of the element.

We tested the isogenicity of the four introgressed strains through the analysis of TEs known to differ in their presence/absence pattern between the two parental strains. We tested 13 TEs for Wi3 introgression and 14 TEs for We33 introgression. The four introgressed stocks have the same presence/absence pattern for these TEs as the parental Wi1 strain suggesting that the genetic backgrounds of the four strains are very similar with the exception of the presence/absence of *Bari-Jheh*.

Egg-to-adult viability and developmental time assays. We measured egg-to-adult viability and developmental time (DT) on normal food and on food containing a JH syntethic analog (JHa): methoprene (Sigma-Aldrich, St. Louis, MO; 1µg/µl in 95%

ethanol). This JHa is widely used in insect physiology because it mimics JH action (Wilson and Fabian 1986; Riddiford and Ashburner 1991; Flatt and Kawecki 2007). First, we established LD_{50} using 6 different concentrations of JHa: 0 µg/µl, 1.5 µg/µl, 2 µg/µl, 2.5 µg/µl, 3 µg/µl and 3.5 µg/µl. JHa dissolved in ethanol was added to the still liquid, warm food medium to the desired final concentration. An equivalent volume of ethanol was added to the food without JHa. To set up the assay we placed 100 adult flies (50 males and 50 females) of Wi1 strain into egg lying chambers overnight. The next day, eggs were allocated into 6 vials with normal food and 6 vials with food containing the different JHa concentrations, each vial with 50 eggs on 10 ml food (1 line x 6 conditions x 6 replicas = 36 vials). Vials were checked every 12 hours for eclosing adults until all flies had emerged. The JHa concentration that gave approximately 50% mortality of the parental strain Wi1 was 2.5 µg/µl. For the subsequent assays we used the parental strain Wi1 and the four introgressed strains. The experimental design followed was the same as explained before using normal food and food with three different concentrations of JHa: 0 µg/µl, 2.5 µg/µl and 5 µg/µl (5 lines x 3 conditions x 6 replicates = 90 vials).

To estimate the egg-to-adult viability (proportion surviving), vials were checked every 12 h for eclosing adults until all flies had emerged. Average DT was estimated over the mid-point of each successive interval. Analysis of variance was performed using nested ANOVA. We considered the identity of the introgressed strain as a nested factor and investigated the effect of the presence/ absence of *Bari-Jheh* and the JHa concentration plus the interaction between these two factors. We tested whether strains with *Bari-Jheh* were significantly different from strains without *Bari-Jheh* by a Mann-

Whitney test in the case of the viability experiments since the results were expressed in proportions and by a t-test for the DT results.

RESULTS

Bari-Jheh insertion is the mutation causing the selective sweep

We previously analyzed the haplotype configuration of the 2 kb region flanking *Bari-Jheh* insertion and found signatures of a partial selective sweep (González et al 2008). *Bari-Jheh* was located in the center of the sweep and the analysis of 500-bp regions located approximately 10 kb away from *Bari-Jheh* showed that the haplotype structure was decaying on both sides of the insertion suggesting that *Bari-Jheh* was likely to be the causative mutation (González et al 2008). However, it is theoretically possible that *Bari-Jheh* is in perfect linkage with a causative polymorphism located farther away from the 2 kb sequenced region. Furthermore, this region only included partial coding regions of *Jheh2* and *Jheh3* such that mutations in these genes could not be completely discounted as being the cause of the selective sweep. To test this possibility, we further sequenced the flanking region around *Bari-Jheh* to include the complete coding sequence of these two genes. As can be seen in Figure 1, the TE appears to be completely linked to the partial sweep and the sweep decays on both sides of the TE further suggesting that the sweep has its focal point in or close to the element insertion. We estimated several statistical measures of polymorphism and compared them with the distributions obtained by coalescent simulations under the null model specified in González et al. (2008). This null model incorporates the demographic scenario based on the analysis of a European population described in Thornton and Andolfatto (2006). There is uncertainty about the

appropriate demographic model both for European (Li and Stephan 2006; Thornton and Andolfatto 2006) and for NA populations (David and Capy 1988; Caracristi and Schlotterer 2003; Baudry, Viginier, and Veuille 2004). However, our aim in this work was to delimit the extent of the sweep, and to do this we compared the significance of the statistics in the 5kb region with the results previously obtained for the 2kb region. Therefore, we used the same null model for both simulations (González et al. 2008). Results are shown in Table 1. The proportion of nucleotide diversity within the haplotypes linked to the TE to the total nucleotide diversity in the sample, f_{TE}, is not significant for the 2kb region and as expected is not significant for the 5kb region. On the other hand, the integrated haplotype score (iHS) statistic, which is expected to be the most powerful indicator of a partial selective sweep (Voight et al. 2006), is significant when we consider the 2 kb but not the 5 kb region immediately adjacent to *Bari-Jheh*. This result suggested that the mutation causing the sweep was included in this 5 kb region.

We searched for mutations other than *Bari-Jheh* that could have been the target of selection. Besides *Jheh2* and *Jheh3* genes we analyzed *Jheh1*. This gene belongs to the same gene family and is located only 0.6 kb downstream of gene *Jheh2* and therefore approximately 3 kb away from *Bari-Jheh* (Figure 2). As can be seen in Figure 1 and 2, the haplotype of one of the strains withouth *Bari-Jheh* (strain Wi1) is similar to that of the strains with *Bari-Jheh* and could represent the ancestral haplotype in which *Bari-Jheh* inserted. We conclude that *Bari-Jheh* is likely to be the causative mutation generating the partial selective sweep haplotype structure.

Bari-Jheh is present at high frequencies both in NA (93%) and Australian (55%) populations while it is absent in the sampled sub-Saharan AF strains (González et al 2008). Here, we further show that *Bari-Jheh* is present at high frequencies in Europe as well -- we found that it is present in 11 out of 12 strains from one population in The Netherlands. This result confirms that *Bari-Jheh* is present at high frequencies in geographically distant non-AF populations and is consistent with its adaptive role outside of Africa.

Bari-Jheh **affects the expression of its neighboring genes**

We tested whether the presence of *Bari-Jheh* insertion is associated with the loss of expression of any of the three *Jheh* genes. We analyzed one strain with the insertion (Wi3) and one strain without the insertion (Wi1) in embryo, larvae and adult. RT-PCR experiments revealed that the three genes are expressed in the three developmental stages both in the strains with and without the insertion. We couldn't detect one of the two predicted transcripts encoded by *Jheh2* gene (*Jheh2-PB*) in any of the strains. However, the last release of Flybase (r5.13; www.flybase.org), eliminates this transcript and annotates a new one *Jheh2-PC* which revealed that primers were designed in a non coding region.

We further investigated whether *Bari-Jheh* affects expression of *Jheh* genes more qualitatively. We focused on the genes that are more closely linked to *Bari-Jheh*: *Jheh2* and *Jheh3*. Previously we used allele-specific expression analysis in F_1 heterozygous hybrids adults (Wittkopp, Haerum, and Clark 2004) to demonstrate that *Bari-Jheh* leads to reduced expression of the linked *Jheh3* alleles (González et al 2008). Here we further

analyzed the allele-specific expression of *Jheh2* as a function of the presence or absence of *Bari-Jheh* (see Material and Methods). Differences in the expression level between the two alleles under the same cellular conditions, as it is the case for F1 hybrids, indicate a difference in cis-regulatory activity (Wittkopp, Haerum, and Clark 2004). Similarly to *Jheh3* the expression of the *Jheh2* allele linked to *Bari-Jheh* is down-regulated (Figure 3). There is no evidence either for a parental effect or for a sex-specific effect on the expression of these alleles. The results were significant for the female progeny of the two crosses (t-test *P-value* = 0.0031 and *P-value* = 0.0002 for cross 1 and 2 respectively) and for the male progeny of cross 2 (t-test *P-value* = 0.024). Although results were not significant for the male progeny of cross 1 (t-test *P-value*: 0.2680), the level of expression is similar to the male progeny of cross 2 (Figure 3).

Phenotypic effect of the insertion

As mention above, biological effects of JH are often sensitive to the level of this hormone. Application of exogenous JH or JH analogs (JHa), such as methoprene, during larval development results in late pupal inviability, increased developmental time (DT) and increased fecundity (Flatt, Tu, and Tatar 2005; Flatt and Kawecki 2007). Since flies carrying *Bari-Jheh* insertion showed reduced levels of expression of *Jheh2* and *Jheh3* genes, both involved in JH degradation, these flies are likely to have elevated JH titers and could show the same effects. The presence of JHa in the food may therefore enhance the expected effects of *Bari-Jheh* ((Wilson and Fabian 1986; Riddiford and Ashburner 1991; Flatt and Kawecki 2007); see Material and Methods). In this work, we focused on the analysis of viability and DT, results are shown in Figures 4 and 5 respectively.

Results for the parental strain Wi1 are shown for comparison since the genetic background of this strain should be similar to the background of the introgressed strains and therefore constitutes the baseline of the experiment (see Material and Methods).

Both for viability and DT assays the data for all the strains analyzed follows a normal distribution (χ^2 *P-value* = 0.769 and *P-value* = 1 for viability and DT respectively). We found a strong negative correlation between viability and JHa concentration (Pearson's correlation: -0.911, *P-value* = 1.17×10^{-6}) and a strong positive correlation between DT and JHa concentration (Pearson's correlation: 0.917, *P-value* = 7.56×10^{-7}) as expected.

Since two different *Bari-Jheh* alleles were introgressed into the same genetic background, we considered the identity of the introgressed strain as the nested factor in an ANOVA analysis that considers the effects of the presence/ absence of *Bari-Jheh*, the concentration of JHa in the food, and the interaction between these two factors. Both the presence of *Bari-Jheh* (ANOVA *P-value* = 0.0158) and the concentration of JHa (ANOVA *P-value* = 0) have an effect on viability in the expected direction (Figure 4; supplementary Table S4, Supplementary Material online). The interaction between these two factors is also significant (ANOVA *P-value* = 0.0137). When no JHa was added to the food, both introgressed strains carrying *Bari-Jheh* showed reduced viability compared to the strains lacking *Bari-Jheh*, as expected if the down-regulation of *Jheh* genes is increasing JH titer (Mann-Whitney test *P-value* = 0.028 and *P-value* = 0.048 for introgressed strains We33/Bari+ and Wi3/Bari+, respectively). When the food was supplemented with 2.5 µg/µl of JHa, Wi3/Bari+ strain showed reduced viability compared to Wi3/Bari- (Mann-Whitney test *P-value* = 0.0022). As can be seen in Figure

4, the viability differences between Wi3/Bari+ and Wi3/Bari- were more significant when JHa was added to the food suggesting that JHa enhances the effect of *Bari-Jheh*. On the other hand, no significant differences were obtained for We33/Bari introgressed strains (Mann-Whitney test *P-value* = 0.27) suggesting that the genetic background differences can mitigate the effects of *Bari-Jheh*. Finally, when the food was supplemented with 5 µg/µl of JHa, there were no significant differences between the stocks with and without the insertion for any of the two introgressed strains (Mann-Whitney test *P-value* = 0.12 and *P-value* = 0.29 for We33/Bari and Wi3/Bari respectively).

The same ANOVA model was used to test the effects of the different factors on DT. Both the presence/ absence of *Bari-Jheh* (ANOVA *P-value* = 0.0005) and the JHa concentration (ANOVA *P-value* = 0) affects significantly the DT while the interaction between these two factors is not significant (ANOVA *P-value* = 0.18). The effect of the presence of *Bari-Jheh* on the DT is only significant when 5 µg/µl of JHa were added to the food and as expected the strains carrying *Bari-Jheh* insertion showed an increased DT compared to strains lacking *Bari-Jheh* (t-test *P-value* = 0.0061 and *P-value* = 0.014 for We33/Bari and Wi3/Bari strains, respectively) (Figure 5; supplementary Table S5, Supplementary Material online).

In summary, although not all comparisons between the strains carrying and lacking *Bari-Jheh* were significant, when they were, the results were consistent with the effects of the reduced expression of *Jheh* genes, that is, reduced viability and extended DT. This result suggested that *Bari-Jheh* is not only affecting transcription of the neighboring genes, as previously shown, but that it may also have an effect on some fitness components.

Bari-Jheh is inserted closed to highly conserved genes

We analyzed the evolution of the *Jheh* gene family in the 12 Drosophila species sequenced (Clark et al. 2007). The three genes are closely linked in all the species suggesting that they originated from ancient tandem duplication events. Although only two orthologous genes have been identified in *Anopheles gambiae* (AGAP008684 and AGAP008686; (Hubbard et al. 2007)) gene AGAP008685 located between them also shows homology with *Jheh* genes suggesting that the tandem duplications took place before the divergence between *Drosophila* and *Anopheles* about 250 million years ago (Zdobnov et al. 2002).

The number of genes in the *Jheh* family has been conserved along the evolution of the genus Drosophila. Only one species, *D. ananassae*, has four *Jheh* genes instead of three: it has a tandem duplication of *Jheh2* gene. According to its phylogenetic distribution and to its sequence divergence (Ks = 1.2843; (Powell 1997)) this duplication took place in the lineage leading to *D. ananassae*. Therefore, the only exception to the conservation of the gene number in the *Jheh* family is confined to the ananassae subgroup. Both paralogs of *Jheh2* in *D. ananassae* are likely to be functional since no premature stop codon or frameshift mutations were identified in their coding sequence. They show a high level of amino acid identity (78%) and half of the amino acid changes are conservative (supplementary Figure S1, Supplementary Material online). Estimate of Ka/Ks is low (Ka/Ks = 0.1056) indicating that both genes are highly constrained and suggesting that both retained their original function.

Jheh1, *Jheh2* and *Jheh3* genes appear to be functional in the 12 Drosophila species (supplementary Figure S2, Supplementary Material online). *Jheh2* is predicted to encode two alternative transcripts: *Jheh2-PA* and *Jheh2-PC*. The length of the corresponding four proteins is highly conserved across the 12 species and the amino acid identity is high (53%-65%). Seven amino acids previously identified as being functional in epoxy hydrolase enzymes (Barth et al. 2004) are conserved in the 12 species consistent with the functionality of these genes (supplementary Figure S2, Supplementary Material online). One of these functional residues is spliced out in transcript *Jheh2-PC*, however, according to Flybase this transcript may or may not produce a functional polypeptide.

We estimated ω (the ratio of non-synonymous to synonymous divergence) for each *Jheh* gene using PAML (Yang 2007). We did the analysis including either the six species of the melanogaster group or the 12 Drosophila species sequenced. Estimates of dS for all the different branches were ≤ 1 suggesting that synonymous sites are not saturated and therefore the alignments including the 12 Drosophila species can be used to estimate the rate of evolution of *Jheh* genes (Heger and Ponting 2007). For each *Jheh* gene, the estimate of ω was < 0.1 suggesting that *Jheh* genes are evolving under strong purifying selection (Table 2). Similar results were obtained when the analysis was restricted to the six species in the melanogaster group (Table 2).

We also tested for evidence of positive selection by comparing models that allowed heterogeneous ω ratios among sites (see Material and Methods). No evidence for positive selection was found for any of the three genes either when the six melanogaster group species or the 12 Drosophila species were analyzed (*P-value* > 0.05).

Finally, we analyzed if TEs were likely to have played a role in the evolution of *Jheh* genes in the 12 Drosophila species. We did not find any TE insertion in the intergenic region between *Jheh1* and *Jheh2* genes. In the intergenic region between *Jheh2* and *Jheh3* genes, besides the *Bari-Jheh* insertion in *D. melanogaster*, we found small fragments (84-137 bp) that showed similarity with *Penelope* TE in *D. yakuba* and *D. erecta* (supplementary Table S6, Supplementary Material online). Overall, there is no evidence for a recurrent role of TEs in the evolution of *Jheh* genes.

Evidence of constant purifying selection in the recent evolution of *Jheh* genes

We analyzed the evolution of *Jheh* genes in the species of the melanogaster subgroup in greater detail. Besides *D. melanogaster* (Figure 1 and 2), we collected polymorphism data for *D. simulans* (Figure 6), a cosmopolitan species that diverged from *D. melanogaster* approximately 5.4 MYA (Tamura, Subramanian, and Kumar 2004). We also collected polymorphism data for the coding regions of the three *Jheh* genes in *D. yakuba* (Figure 7) which is an endemic AF species that shared a common ancestor with the other two species 12.8 MYA (Tamura et al. 2007).

We first looked for evidence of selective constraint in the coding, non-coding (UTR and introns) and intergenic regions of *Jheh* genes in the three species. The ratio of non-synonymous to synonymous polymorphisms in the coding regions of the three genes (Table 3), the ratio of polymorphisms in non-coding regions to synonymous polymorphism in the coding region of the same gene (Table 3) and the ratio of polymorphism in intergenic regions to synonymous polymorphisms in the two flanking genes (Table 4) was smaller than one. These results suggested that coding and non-

coding and intergenic regions have been evolving under purifying selection. We only found one exception; *D. melanogaster Jheh2* noncoding regions had a higher number of polymorphisms compared to the synonymous polymorphisms within the gene (Table 3). This result is likely explained by the selective sweep associated with the insertion of *Bari-Jheh* in this particular region of the genome (Fig. 1).

The ratios of non-neutral to neutral polymorphism for each of the three analyzed regions -- coding, non-coding (UTR and introns) and intergenic -- are not statistically different between species. This suggests that this region of the genome has been subject to fairly constant levels of purifying selection in the three species (Table 3 and 4).

We further analyzed the intergenic region where *Bari-Jheh* is inserted in 34 *D. melanogaster* strains that we sequenced previously (González et al 2008). Other than *Bari-Jheh*, only a single 43 bp (TA) repeat was found. This simple repeat is flanking the insertion and is also present in the strains without *Bari-Jheh* where its length varies between 6 and 61 bp. This repetitive sequence is not characteristic of *Bari1* insertions since it is not present in the flanking regions of the other five *Bari1* insertions described in the genome (FBti0019419, FBti0019499, FBti0019099, FBti0064232 and FBti0019400). In summary, although *Bari-Jheh* is inserted in an intergenic region likely to be evolving under purifying selection, the exact position where *Bari-Jheh* is inserted is not conserved. Furthermore, the VISTA browser alignment between *D. melanogaster* and *D. simulans* shows that sequence conservation drops in the region immediately adjacent to *Bari-Jheh* (http://genome.lbl.gov/vista/index.shtml; supplementary Figure S3, Supplementary Material online). This result suggests that *Bari-Jheh* may be affecting the expression of its neighbouring genes by altering the physical distance between regulatory

elements and the transcriptional start site or by adding regulatory elements itself rather than by disrupting existing regulatory elements.

No evidence of recurrent adaptive evolution in the recent history of *Jheh* genes

We did not find evidence for recurrent adaptive evolution acting on *Jheh* genes across the phylogeny of the 12 Drosophila species. However, it could be that positive selection has been restricted to the recent history of these species. *Bari-Jheh* insertion most likely played a role in the adaptation to the new environments faced by *D.melanogaster* in its expansion out of Africa (González et al. 2008). Since *D. simulans* has independently undergone a similar migration out of sub-Saharan Africa (Hamblin and Veuille 1999; Baudry et al. 2006), we explored the possibility of a parallel adaptive event in this region of the genome in *D. simulans*. As can be seen in Figure 6, sequence of each *D. simulans* strain represents a different haplotype. Tajima's D and Fu and Li's D and F are not significantly different from the neutral expectations (Table 5). These neutrality tests assume that the population is at equilibrium and as mentioned before, the out-of-Africa *D. simulans* populations are likely to be out of equilibrium (Hamblin and Veuille 1999; Baudry et al. 2006). Not taking into account the demographic history of the species may result in spurious inference of positive selection (Orengo and Aguade 2004; Ometto et al. 2005; Teshima, Coop, and Przeworski 2006; Thornton et al. 2007; Macpherson et al. 2008). However, an expansion of D. simulans out of Africa is unlikely to mask a true selective sweep if it was in fact there.

We performed McDonald Kreitman test to further look for evidence of positive selection in the recent evolution of this genomic region (McDonald and Kreitman 1991;

Andolfatto 2005; Egea, Casillas, and Barbadilla 2008). First, we searched for evidence of positive selection in the *D. melanogaster* and *D. simulans* lineages. We performed the analysis both considering all the positions and excluding variants that are present in only one of the strains analyzed (singletons). We did not find any evidence of positive selection in coding, non-coding or intergenic regions (Table 6 and 7). For coding regions, we also considered the polymorphism data collected for *D. yakuba* and looked for evidence of positive selection during the evolution of these three species. Marginally significant results were obtained for *Jheh2* gene when singletons were excluded from the analysis (Table 6); however, this result is not significant after correcting for multiple tests. Altogether, these results suggest that *Jheh* genes have not been subject to recurrent and pervasive adaptive evolution in the recent past.

DISCUSSION

Bari-Jheh insertion is adaptive

Bari-Jheh insertion was recently identified as being putatively adaptive in a genome-wide screen for recent TE-induced adaptations (González et al. 2008). Here, we provided additional evidence that this insertion was indeed adaptive. By further sequencing the region flanking the insertion, we delimited the extent of the selective sweep and showed that *Bari-Jheh* is the only mutation linked to the sweep (Maynard-(Smith and Haigh 1974; Kaplan, Darden, and Hudson 1988; Kaplan, Hudson, and Langley 1989). Consequently, *Bari-Jheh* appears to be the causative mutation of the sweep (Figure 1). Furthermore, *Bari-Jheh* is associated with changes in the transcription of its flanking genes: it down-regulates the expression of both *Jheh2* and *Jheh3* (Figure

3). Since both genes are involved in the degradation of JH, a plausible consequence of the reduced expression of *Jheh* genes is an increased JH titer. Increased JH titer is expected to lead to reduced viability and extended DT among many other phenotypic effects (Flatt, Tu, and Tatar 2005; Flatt and Kawecki 2007). Although we did not always find significant differences between the strains carrying and lacking *Bari-Jheh*, when we did, the results were consistent with the expectations, suggesting that *Bari-Jheh* has subtle phenotypic consequences (Figure 4 and 5). These two phenotypic effects imply a reduced fitness for the flies carrying the insertion. Interestingly, *Bari-Jheh* is present at high frequency in all the derived non-AF populations analyzed, NA, Australian and European, however, it is not fixed in any of them (González et al. 2008). A plausible explanation for these results is that the reduced viability and increased DT could represent the associated cost of selection for *Bari-Jheh* insertion which would explain why *Bari-Jheh* is not fixed in the derived non-AF populations.

What is the adaptive effect of *Bari-Jheh*? Previous results showed that the frequency of *Bari-Jheh* did not vary between a temperate and a more tropical out-of-Africa population suggesting that the adaptive effect of this insertion was not related to climatic adaptation (González et al 2008). However, JH is a regulator of development, life history and fitness trade-offs (Flatt, Tu, and Tatar 2005; Riddiford 2008). Any of the large number of traits and processes in Drosophila development and life history affected by JH could have been affected by *Bari-Jheh* insertion. In order to understand the adaptive consequences of this insertion a thorough phenotypic analysis will be required. The challenge will be to determine which phenotype or phenotypes to study and under what ecological conditions they should be examined (Jensen, Wong, and Aquadro 2007).

The availability of 192 wild-derived inbred lines that are currently being phenotyped and sequenced will facilitate the understanding of the functional impact of this and other putatively adaptive TEs (Ayroles et al. 2009).

A unique adaptive event near highly constrained loci

Bari-Jheh is inserted near highly constrained genes. The number of genes in the *Jheh* gene family has been conserved for the last 80-124 myr (Tamura, Subramanian, and Kumar 2004). These genes appear to be functional in the 12 Drosophila species sequenced and encode proteins of similar length (Clark et al. 2007). Furthermore, coding, non-coding and intergenic regions seem to have been evolving under purifying selection both in the long term, when the 12 Drosophila species sequenced were analyzed, and short term, when only the species of the melanogaster subgroup were analyzed. In addition, the strength of purifying selection appeared to have been constant at least for the last 12.8 myr (Table 3 and 4). Overall, we can conclude that *Jheh* genes have been evolving under purifying selection for long periods of time and that the strength of purifying selection acting on these genes has not changed in the recent past.

We looked for evidence of parallel adaptive events during the evolution of this gene family. We explored different possibilities; in the long term evolution (1) we looked for evidence of parallel adaptive TE insertions in the intergenic regions and (2) we tested whether a subset of codons in these genes showed evidence for replacement mutations fixing more frequently than silent mutations (Yang 2007). In the short term evolution, (3) we looked for evidence of parallel selective sweeps in the orthologous sequence of *D. simulans* and (4) we tested whether the ratio of non-synonymous to synonymous

divergence was higher than the ratio of non-synonymous to synonymous polymorphism in coding, non-coding and intergenic regions of *D. melanogaster*, *D. simulans* and *D. yakuba* (McDonald and Kreitman 1991; Andolfatto 2005). Overall, we did not find evidence for recurrent and pervasive adaptive evolution acting on *Jheh* genes in the long term or short term evolution of this gene family. In conclusion, *Bari-Jheh* appears to be either unique or at least a very rare adaptive event in the history of *Jheh* genes. No current analysis would suggest that these highly constrained and conserved genes are likely targets of adaptation.

Implications for the study of adaptation

Here, we showed that adaptive variation within species might be found in genes that do not undergo frequent adaptation. These genes would be overlooked by the most widely used approaches to look for positive selection, such as McDonald and Kreitmant test or codon-based tests such as those implemented in PAML, since these approaches are based on the assumption that adaptation is recurring at the same loci or even at the same sites (McDonald and Kreitman 1991; Hughes 2007; Jensen, Wong, and Aquadro 2007; Macpherson et al. 2007; Yang 2007). It is not clear how frequently selection favors repeated amino acid changes at a limited set of sites within a given gene and therefore these type of studies may only give a partial view of the genetics underlying adaptation (Fay, Wyckoff, and Wu 2002; Smith and Eyre-Walker 2002; Andolfatto 2005; Bustamante et al. 2005; Macpherson et al. 2007; Sawyer et al. 2007; Shapiro et al. 2007). In addition, adaptations might be local and ephemeral and therefore destined to be lost over long periods of time (González et al. 2008). This suggests that functional genetic

variation within species might at times be due to different mutations than mutations leading to functional divergence between species. This functional variation will only be identified by approaches that identified mutations that have recently swept through the population such as genome-wide scans for positive selection (see Pavlidis et al. 2008 for a review) or the approach described in González et al. (2008). In conclusion, population genetics methods that are capable of detecting selection on a single recent adaptive mutation and divergence based methods that rely on the repeated selective fixation of amino acid changes followed by appropriate functional studies should be combined in order to get a fuller picture of adaptation.

SUPPLEMENTARY MATERIAL

Table S1. Strains used in this study.

Table S2. Primers used to amplify *Jheh* genes in *D. melanogaster*, *D. simulans* and *D. yakuba*.

Table S3. Primers used for the RT-PCR experiments.

Table S4. Viability assays results.

Table S5. Developmental time assays results.

Table S6. Repetitive content of *Jheh* genes intergenic regions in the 12 Drosophila sequenced species.

Figure S1. Alignment of *Jheh2* paralog genes in *D. ananassae*.

Figure S2. Alignment of the three *Jheh* genes in the 12 Drosophila species sequenced.

Figure S3. Vista browser plot of the intergenic region where *Bari-Jheh* is inserted.

ACKNOWLEDGMENTS

We thank S. Chatterji for providing the *Jheh* genes sequences for the 12 Drosophila species; D. Chung and S. Tran for technical assistance with the generation of the introgressed strains; T. Flatt for helpful advice on the life-history phenotypic assays; R. Hershberg and P. Markova-Raina for help with the PAML analysis and N. Petit and all members of Petrov lab for comments on the manuscript. JG was a Fulbright/Secretaria de Estado de Universidades e Investigacion, MEC postdoctoral fellow. JMM was an HHMI Predoctoral Fellow. This research was supported by grants from the National Institutes of Health (GM077368) and the National Science Foundation (0317171) to DAP.

LITERATURE CITED

Aminetzach, Y. T., J. M. Macpherson, and D. A. Petrov. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science **309**:764-767.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature **437**:1149-1152.

Ayroles, J. F.M. A. CarboneE. A. Stone et al. 2009. Systems genetics of complex traits in Drosophila melanogaster. Nat Genet **41**:299-307.

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

Baudry, E., N. Derome, M. Huet, and M. Veuille. 2006. Contrasted polymorphism patterns in a large sample of populations from the evolutionary genetics model Drosophila simulans. Genetics **173**:759-767.

Baudry, E., B. Viginier, and M. Veuille. 2004. Non-African populations of Drosophila melanogaster have a unique origin. Mol Biol Evol **21**:1482-1491.

Bergman, C. M.B. D. PfeifferD. E. Rincon-Limas et al. 2002. Assessing the impact of comparative genomic sequence data on the functional annotation of the Drosophila genome. Genome Biol **3**:RESEARCH0086.

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

Bustamante, C. D.A. Fledel-AlonS. Williamson et al. 2005. Natural selection on protein-coding genes in the human genome. Nature **437**:1153-1157.

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

Caracristi, G., and C. Schlotterer. 2003. Genetic differentiation between American and European Drosophila melanogaster populations could be attributed to admixture of African alleles. Mol Biol Evol **20**:792-799.

Charlesworth, B., P. Sniegowski, and W. Stephan. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371**:215-220.

Clark, A. G.M. B. EisenD. R. Smith et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature **450**:203-218.

Daborn, P. J.J. L. YenM. R. Bogwitz et al. 2002. A single p450 allele associated with insecticide resistance in Drosophila. Science **297**:2253-2256.

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

de Kort, C. A. D., Granger N. A. 1996. Regulation of JH titers: the relevance of degradative enzymes and binding proteins. Arch Insect Biochem Physiol **33**:1-26.

Doolittle, W. F., and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature **284**:601-603.

Egea, R., S. Casillas, and A. Barbadilla. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res **36**:W157-162.

Fay, J. C., G. J. Wyckoff, and C. I. Wu. 2002. Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature **415**:1024-1026.

Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet **9**:397-405.

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

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

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

Hamblin, M. T., and M. Veuille. 1999. Population structure among African and derived populations of Drosophila simulans: evidence for ancient subdivision and recent admixture. Genetics **153**:305-317.

Heger, A., and C. P. Ponting. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res **17**:1837-1849.

Hubbard, T. J.B. L. AkenK. Beal et al. 2007. Ensembl 2007. Nucleic Acids Res **35**:D610-617.

Hughes, A. L. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity **99**:364-373.

Jensen, J. D., A. Wong, and C. F. Aquadro. 2007. Approaches for identifying targets of positive selection. Trends Genet **23**:568-577.

Jurka, J., V. V. Kapitonov, O. Kohany, and M. V. Jurka. 2007. Repetitive sequences in complex genomes: structure and evolution. Annu Rev Genomics Hum Genet **8**:241-259.

Kaplan, N. L., T. Darden, and R. R. Hudson. 1988. The coalescent process in models with selection. Genetics **120**:819-829.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The "hitchhiking effect" revisited. Genetics **123**:887-899.

Kazazian, H. H., Jr. 2004. Mobile elements: drivers of genome evolution. Science **303**:1626-1632.

Khlebodarova, T. M., N. E. Gruntenko, L. G. Grenback, M. Z. Sukhanova, M. M. Mazurov, I. Y. Rauschenbach, B. A. Tomas, and B. D. Hammock. 1996. A comparative analysis of juvenile hormone metabolyzing enzymes in two species of Drosophila during development. Insect Biochem Mol Biol **26**:829-835.

Kidwell, M. G., and D. R. Lisch. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution **55**:1-24.

Lachaise, D., Cariou, M-L, David, J R, Lemeunier, F, Tsacas F, et al 1988. Historical biogeography of the Drosophila melanogaster species subgroup. Evol Biol **22**:159-225.

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

Macpherson, J. M., J. Gonzalez, D. M. Witten, J. C. Davis, N. A. Rosenberg, A. E. Hirsh, and D. A. Petrov. 2008. Nonadaptive explanations for signatures of partial selective sweeps in Drosophila. Mol Biol Evol **25**:1025-1042.

Macpherson, J. M., G. Sella, J. C. Davis, and D. A. Petrov. 2007. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila. Genetics **177**:2083-2099.

Maddison, W. P., and D. R. Maddison. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) **53**:190-202.

Marino-Ramirez, L., K. C. Lewis, D. Landsman, and I. K. Jordan. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. Cytogenet Genome Res **110**:333-341.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature **351**:652-654.

Morgenstern, B. 2004. DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. Nucleic Acids Res **32**:W33-36.

Ometto, L., S. Glinka, D. De Lorenzo, and W. Stephan. 2005. Inferring the effects of demography and selection on Drosophila melanogaster populations from a chromosome-wide scan of DNA variation. Mol Biol Evol **22**:2119-2130.

Orengo, D. J., and M. Aguade. 2004. Detecting the footprint of positive selection in a european population of Drosophila melanogaster: multilocus pattern of variation and distance to coding regions. Genetics **167**:1759-1766.

Orgel, L. E., and F. H. Crick. 1980. Selfish DNA: the ultimate parasite. Nature **284**:604-607.

Piriyapongsa, J., M. T. Rutledge, S. Patel, M. Borodovsky, and I. K. Jordan. 2007. Evaluating the protein coding potential of exonized transposable element sequences. Biol Direct **2**:31.

Powell, J. R. 1997. Progress and prospects in evolutionary biology. The Drosophila model. Oxford University Press, New York.

Riddiford, L. M. 2008. Juvenile hormone action: a 2007 perspective. J Insect Physiol **54**:895-901.

Riddiford, L. M., and M. Ashburner. 1991. Effects of juvenile hormone mimics on larval development and metamorphosis of Drosophila melanogaster. Gen Comp Endocrinol **82**:172-183.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Sawyer, S. A., J. Parsch, Z. Zhang, and D. L. Hartl. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. Proc Natl Acad Sci U S A **104**:6504-6510.

Shapiro, J. A.W. HuangC. Zhang et al. 2007. Adaptive genic evolution in the Drosophila genomes. Proc Natl Acad Sci U S A **104**:2271-2276.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet Res **23**:23-35.

Smith, N. G., and A. Eyre-Walker. 2002. Adaptive protein evolution in Drosophila. Nature **415**:1022-1024.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437-460.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**:1596-1599.

Tamura, K., S. Subramanian, and S. Kumar. 2004. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol **21**:36-44.

Teshima, K. M., G. Coop, and M. Przeworski. 2006. How reliable are empirical genomic scans for selective sweeps? Genome Res **16**:702-712.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**:4673-4680.

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

Thornton, K. R., J. D. Jensen, C. Becquet, and P. Andolfatto. 2007. Progress and prospects in mapping recent selection in the genome. Heredity **98**:340-348.

van de Lagemaat, L. N., J. R. Landry, D. L. Mager, and P. Medstrand. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet **19**:530-536.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol **4**:e72.

Wang, H., and S. C. Elbein. 2007. Detection of allelic imbalance in gene expression using pyrosequencing. Methods Mol Biol **373**:157-176.

Wilson, T. G., and J. Fabian. 1986. A Drosophila melanogaster mutant resistant to a chemical analog of juvenile hormone. Dev Biol **118**:190-201.

Wittkopp, P. J., B. K. Haerum, and A. G. Clark. 2004. Evolutionary changes in cis and trans gene regulation. Nature **430**:85-88.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24**:1586-1591.

Zdobnov, E. M.C. von MeringI. Letunic et al. 2002. Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science **298**:149-159.

Table 1. Neutrality tests for the 5 kb region flanking *Bari-Jheh* insertion.

Region analyzed	$f_{TE} = \pi_{TE} / \pi$	iHS	π_{NA}	π_{AF}	$\pi_{NA\text{-}TE}$	$\pi_{NA\text{-non TE}}$
Bari-Jheh (5 kb)	0.47	0.19	42.91	72.27	27.56	52.8
	0.45 (0.22, 0.68)	-0.02 (-0.43, 0.38)	28 (13,43.9)	85.5 (74.2, 96)	25.3 (10.7, 41.1)	26 (9.11, 48.2)
	0.12	0.22	7.87	5.66	8	10.2
Bari-Jheh (2 kb)[a]	0.12	**-1.79**	24.24	103.2	8.7	63.19
	0.31 (0.05, 0.61)	-0.21 (-0.86, 0.41)	29.9 (7.85, 57.9)	100 (84.8, 110)	23.8 (3.62, 52.9)	28.6 (3.16, 69.5)
	0.15	0.33	12.8	7.66	12.8	18

NOTE.--Within a cell, the upper number is the observed value of the statistics. The middle number is the mean and the 2.5% and 97.5% confidence interval limits obtained by coalescent simulations. The lower number is the standard deviation. Data for the 2 kb region flanking the same insertion studied previously is shown for comparison

[a]González et al (2008)

Table 2. Ratio of non-synonymous to synonymous substitutions (ω) considering the 12 Drosophila species sequenced and only the six species in the melanogaster group.

Gene	**ω (12 species)**	**ω (6 species)**
Jheh1	0.06623	0.06413
Jheh2	0.09615	0.10795
Jheh3	0.07432	0.05723

Table 3. Ratio of non-synonymous to synonymous polymorphisms (Pn/Ps) and of non-coding to synonymous polymorphisms (Pnc/Ps) for *Jheh* genes in *D. melanogaster* (*Dmel*), *D. simulans* (*Dsim*) and *D. yakuba* (*Dyak*).

	coding			**non-coding (UTR + introns)**		
Species	**N[a]**	**L[b]**	**Pn/Ps**	**N**	**L**	**Pnc/Ps**
Jheh1_Dmel	16	1425	5/22 (1/11)	16	257	11/22 (6/11)
Jheh1_Dsim	15	1425	9/29 (3/14)	15	200	12/29 (8/14)
Jheh1_Dyak	9	1425	12/37 (6/26)	9	225	9/37 (6/26)
χ^2 *P-value*			0.828 (0.698)			0.354 (0.278)
Jheh2_Dmel	16	1392	3/5 (2/3)	16	706	18/5 (12/3)
Jheh2_Dsim	15	1392	7/35 (3/22)	15	568	35/35 (25/22)
Jheh2_Dyak	9	1377	11/42 (5/23)	9	573	45/48 (23/26)
χ^2 *P-value*			0.405 (0.316)			0.034 (0.079)
Jheh3_Dmel	16	1404	5/30 (4/24)	16	296	10/30 (10/24)
Jheh3_Dsim	14	1404	3/19 (1/11)	14	200	9/22 (4/13)
Jheh3_Dyak	9	1404	4/41 (3/17)	9	203	14/24 (3/11)
χ^2 *P-value*			0.724 (0.848)			0.515 (0.815)

Note.-- Values excluding singletons are given in parenthesis

[a]N: number of strains analyzed

[b]L: lenght of the sequence analyzed

Table 4. Ratio of non-coding to synonymous polymorphisms (Pnc/Ps) in the intergenic regions of *Jheh* genes.

	Intergenic regions		
Species	**N[a]**	**L[b]**	**Pnc/Ps**
Dsim Jheh1-Jheh2	15	635	53/64 (33/36)
Dmel Jheh2-Jheh3	16	607	23/35 (14/27)
Dsim Jheh2-Jheh3	14	835	29/57(16/35)
χ^2 *P-value*			0.467 (0.778)

Note.-- Values excluding singletons are given in parenthesis
[a]N: number of strains analyzed
[b]L: lenght of the sequence analyzed

Table 5. Polymorphism and neutrality tests for the 6.8 kb region of *D. simulans* including the three *Jheh* genes.

Strains	**N**[a]	**SS**[b]	**H**[c]	**h_d**[d]	**π(JC)**[e]	**θ_w**[f]	**Neutrality tests**		
							D[g]	**FL-D**[h]	**FL-F**[i]
all	14	242	14	1	0.01095	0.01120	-0.17373	-0.48561	-0.45946
Non-AF	8	117	8	1	0.00776	0.00664	0.87383	0.50397	0.65968
AF	6	217	6	1	0.01354	0.01399	-0.26671	-0.27782	-0.30303

[a]N: number of strains analyzed.
[b]SS: number of segregating sites.
[c]H: number of haplotypes.
[d]h_d: haplotype diversity.
[e]π(JC): average pairwise diversity with Jukes & Cantor correction.
[f]θ_w: Watterson's estimator of nucleotide diversity.
[g]D: Tajima's D statistic.
[h]FL-D: Fu and Li D statistic.
[i]FL-F: Fu and Li F statistic.

Table 6. McDonald Kreitman test for the coding and non-coding regions of *Jheh* genes.

Species		Coding		
Dmel+Dsim	L[a]	Pn/Ps[b]	Dn/Ds[c]	*P-value*
Jheh1	1416	14/48 (4/23)	10/29 (10/34)	0.691 (0.416)
Jheh2	1386	10/38 (5/24)	3/30 (3/32)	0.162 (0.297)
Jheh3	1398	8/47 (5/33)	9/30 (9/32)	0.276 (0.300)
		Non-coding		
Dmel+Dsim	L	Pnc/Ps[d]	Dnc/Ds[e]	*P-value*
Jheh1	198	21/48 (13/23)	12/29 (14/34)	0.998 (0.454)
Jheh2	502	44/38 (30/25)	49/30 (50/34)	0.268 (0.567)
Jheh3	197	14/27 (9/19)	15/14 (15/16)	0.137 (0.209)
		Coding		
Dmel+Dsim vs Dyak	L	Pn/Ps	Dn/Ds	*P-value*
Jheh1	1401	23/70 (14/50)	18/51 (18/52)	0.816 (0.588)
Jheh2	1344	12/57 (8/47)	22/51 (23/52)	0.079 (0.031)
Jheh3	1380	17/69 (14/59)	13/41 (13/42)	0.542 (0.544)

Note.-- Values excluding singletons are given in parenthesis.

[a]L: length of the sequence analyzed.

[b]ratio of non-synonymous to synonymous polymorphisms

[c]ratio of non-synonymous to synonymous divergence corrected by Jukes and Cantor (1969).

[d]ratio of non-coding to synonymous polymorphisms

[e]ratio of non-coding to synonymous divergence corrected by Jukes and Cantor(1969)

Table 7. McDonald and Kreitman tests for the intergenic regions of *Jheh* genes.

	intergenic regions			
	L[a]	Pnc/Ps[b]	Dnc/Ds[c]	*P-value*
Dsim vs Dmel: Jheh1-Jheh2	600	47/86 (28/47)	17/58 (19/66)	0.061 (0.042)
Dmel +Dsim: Jheh2-Jheh3	329	22/85 (14/57)	17/59 (17/64)	0.842 (0.898)

Note.-- Values excluding singletons are given in parenthesis.
[a]length of the sequence analyzed.
[b]ratio of polymorphisms per site
[c]ratio of non-synonymous to synonymous divergence

FIGURE LEGENDS

Figure 1. Sequence of the 5 kb region flanking *Bari-Jheh* in *D. melanogaster*. The figure shows the segregating sites number (SS), the genes associated with the insertion (Gene) and the distance from the insertion (Position). The SS within coding regions are in bold and are identified as replacement (R), synonymous (S) or non sense (NS) polymorphisms. The TE is shown as a black rectangle, the absence of the TE is shown as an empty rectangle. A horizontal line separates the strains with the insertion from the strains without the insertion.

Figure 2. Sequence of *Jheh1* gene in *D. melanogaster*. See Figure 1 for details.

Figure 3. Normalized allelic ratios (allele carrying *Bari-Jheh*/ allele lacking *Bari-Jheh*) for *Jheh2* gene. Ratios for *Jheh3* gene previously published are shown for comparison (González et al. 2008). The bars represent the mean of the ratios for the three replicas and the standard deviation. For each gene, the first two bars correspond to the F_1 progeny (male and female respectively) of cross 1 and the last two bars correspond to the progeny of cross 2. The horizontal line is the ratio expected if there are no differences in the level of expression of the two alleles. Significant ratios are represented in grey.

Figure 4. Egg to adult viability (proportion surviving) of the parental strain Wi1 lacking *Bari-Jheh* insertion and introgressed strains Wi3/Bari and We33/Bari as a function of the JHa concentration in the food used to raise the flies. Data shown are means and standard

errors of replicate lines within a JHa concentration. Significant comparisons are represented in grey.

Figure 5. Developmental time (in hours) of the parental strain Wi1 lacking *Bari-Jheh* insertion and introgressed strains Wi3/Bari and We33/Bari as a function of the JHa concentration in the food used to raise the flies. Data shown are means and standard errors of replicate lines within a JHa concentration. Significant comparisons are represented in grey.

Figure 6. Sequence of the 6.8 kb region including *Jheh1*, *Jheh2* and *Jheh3* genes in *D. simulans.* The black line separates the non-AF from the AF strains. See Figure 1 for details.

Figure 7. Sequence of *Jheh1*, *Jheh2* and *Jheh3* coding regions in *D. yakuba*. See Figure 1 for details.

FIGURE 1

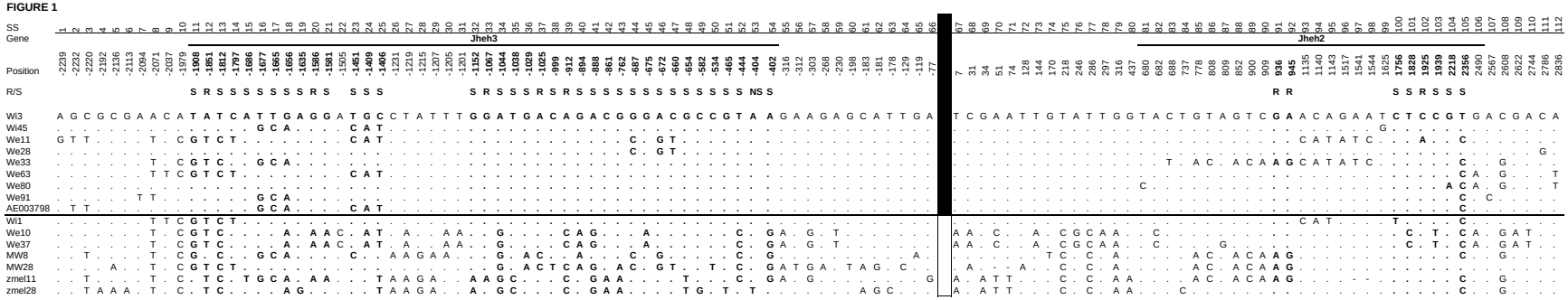

SS
Gene
Jheh3
Jheh2
Position
R/S
Wi3
Wi45
We11
We28
We33
We63
We80
We91
AE003798
Wi1
We10
We37
MW8
MW28
zmel11
zmel28

FIGURE 2

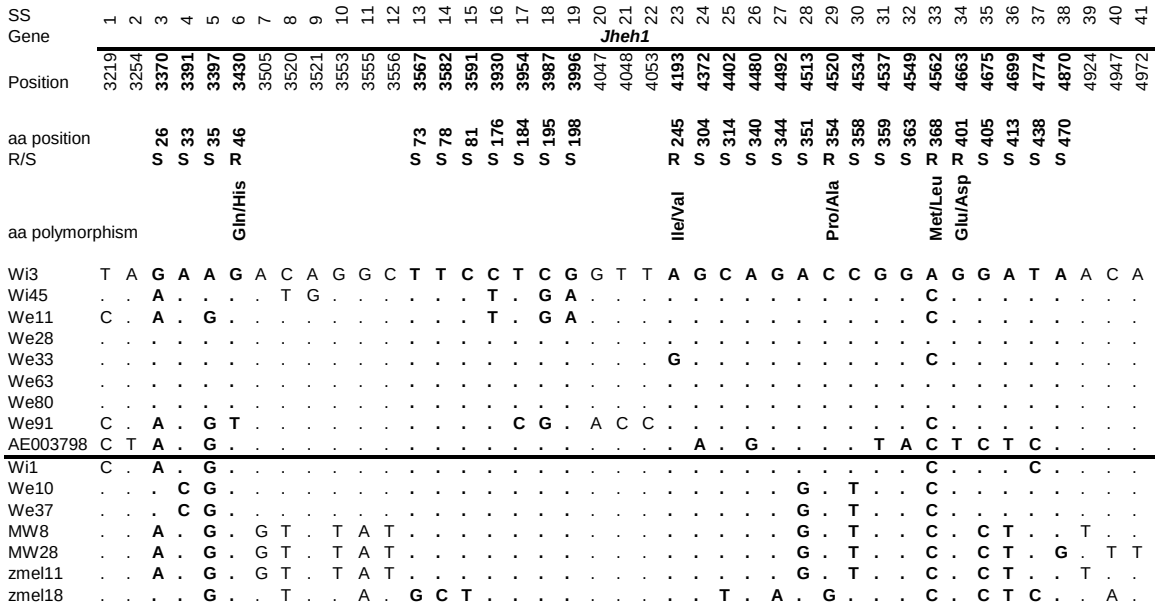

SS
Gene
Jheh1
Position
aa position
R/S
aa polymorphism
Gln/His
Ile/Val
Pro/Ala
Met/Leu
Glu/Asp
Wi3
Wi45
We11
We28
We33
We63
We80
We91
AE003798
Wi1
We10
We37
MW8
MW28
zmel11
zmel18

Figure 3

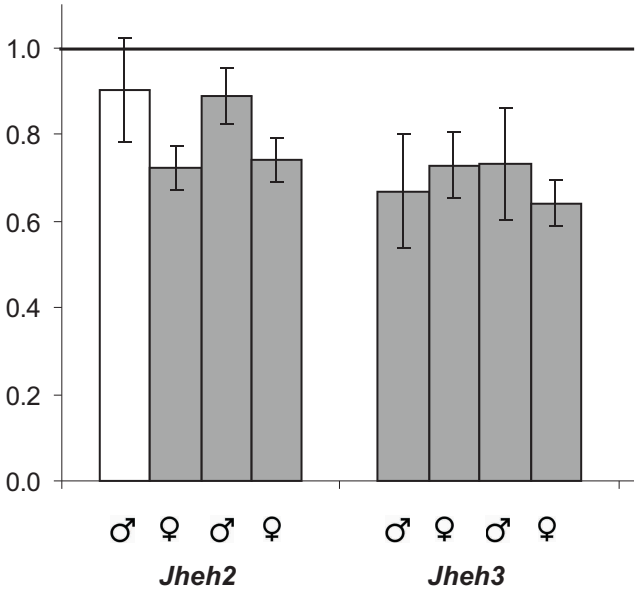


Figure 4

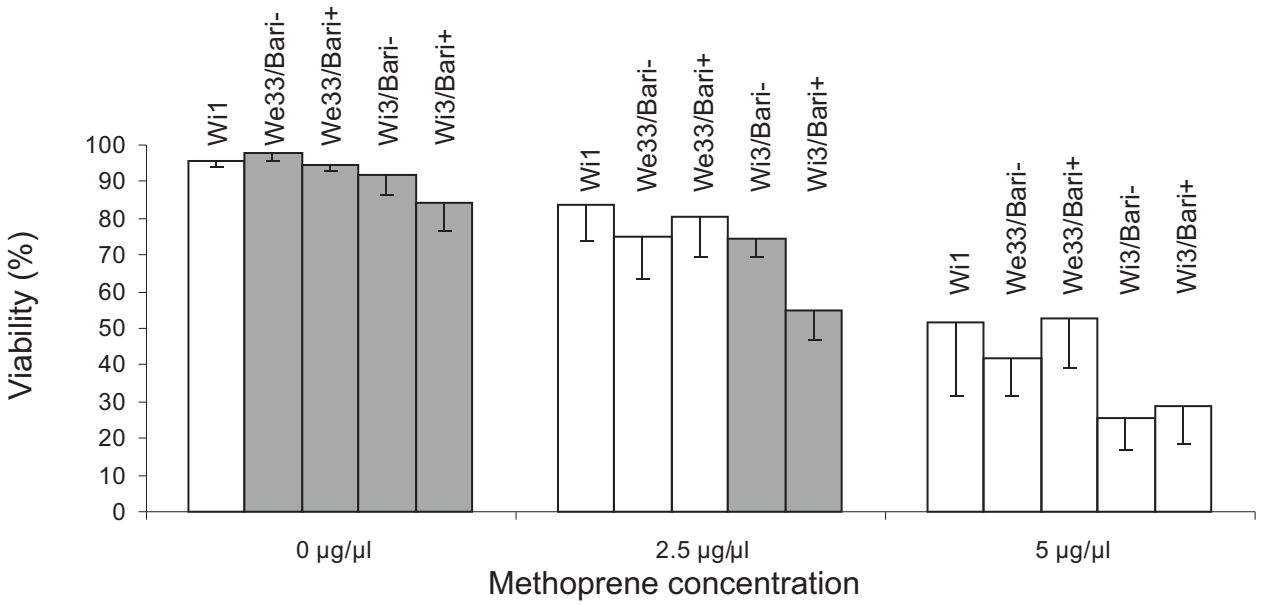

Viability (%)
100
90
80
70
60
50
40
30
20
10
0
Wi1
We33/Bari-
We33/Bari+
Wi3/Bari-
Wi3/Bari+
0 µg/µl
2.5 µg/µl
5 µg/µl
Methoprene concentration

Figure 5

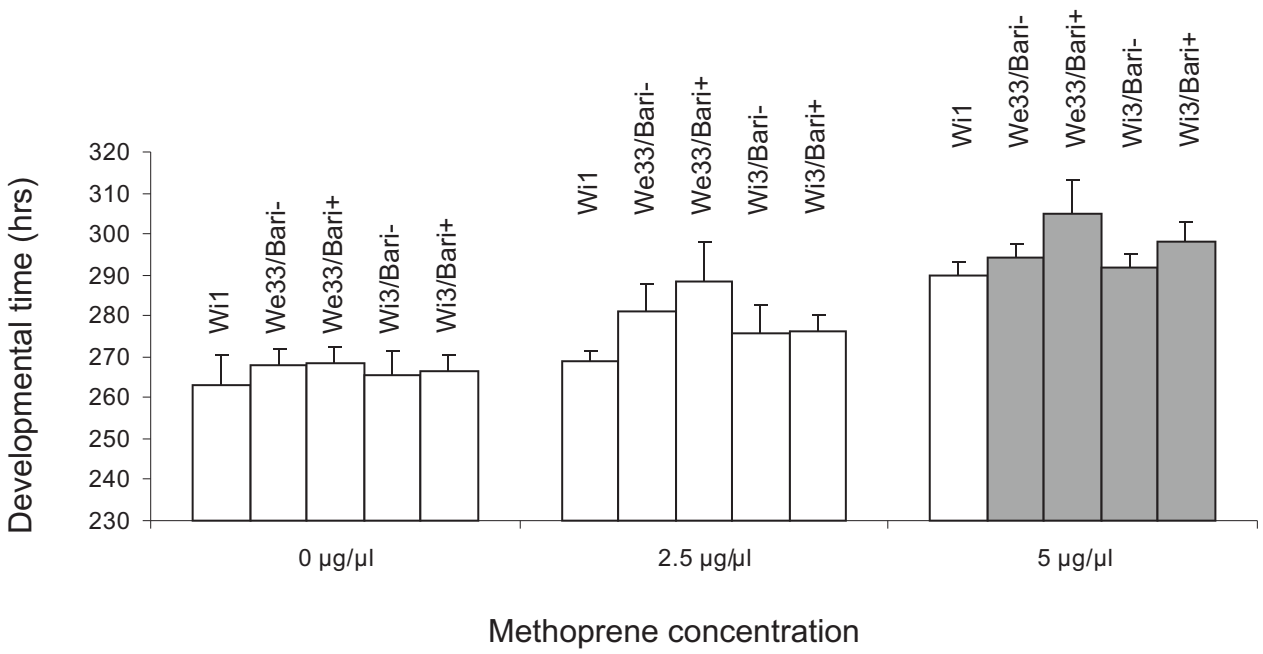

Developmental time (hrs)
320
310
300
290
280
270
260
250
240
230
Wi1
We33/Bari-
We33/Bari+
Wi3/Bari-
Wi3/Bari+
0 µg/µl
2.5 µg/µl
5 µg/µl
Methoprene concentration

FIGURE 6

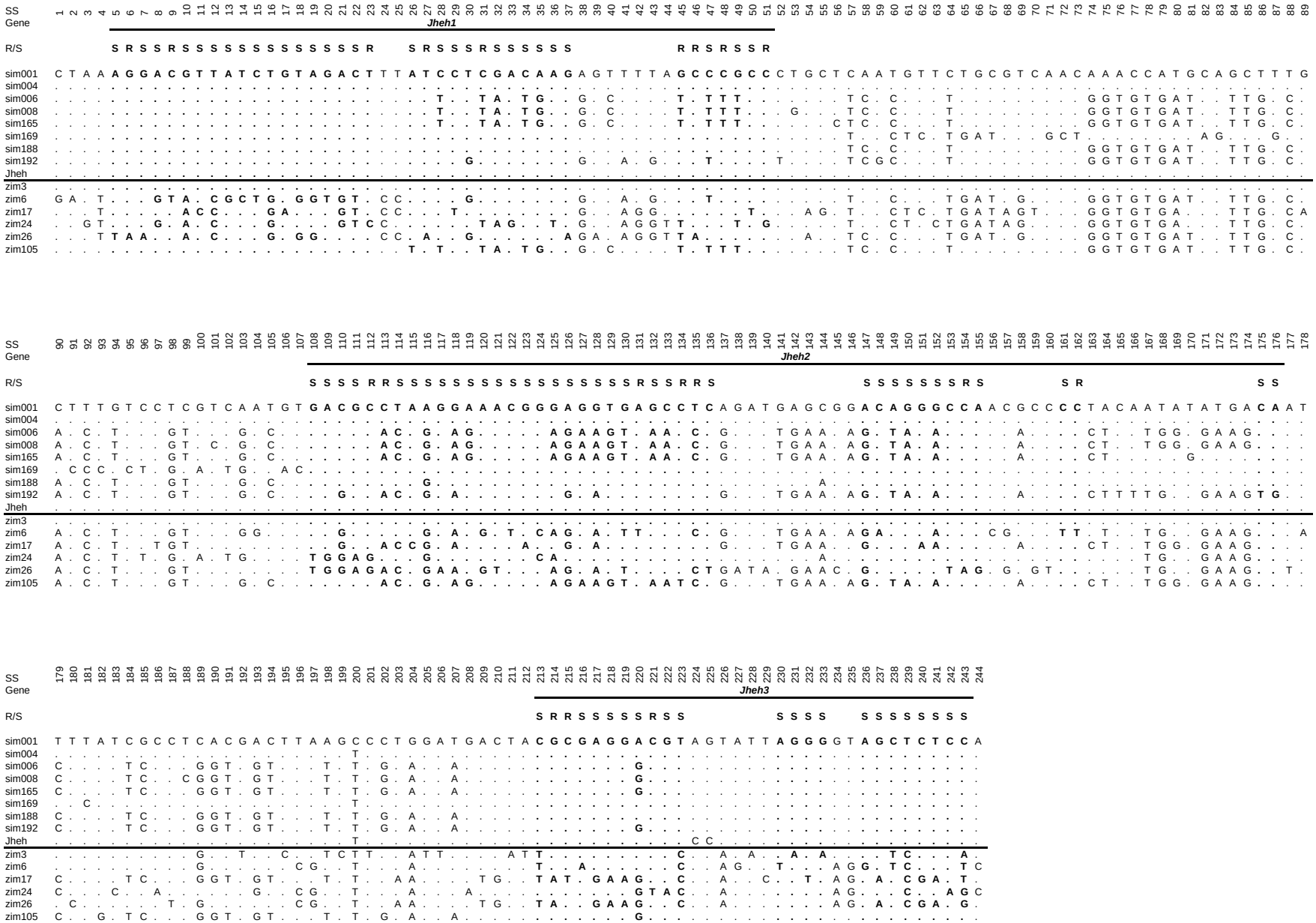
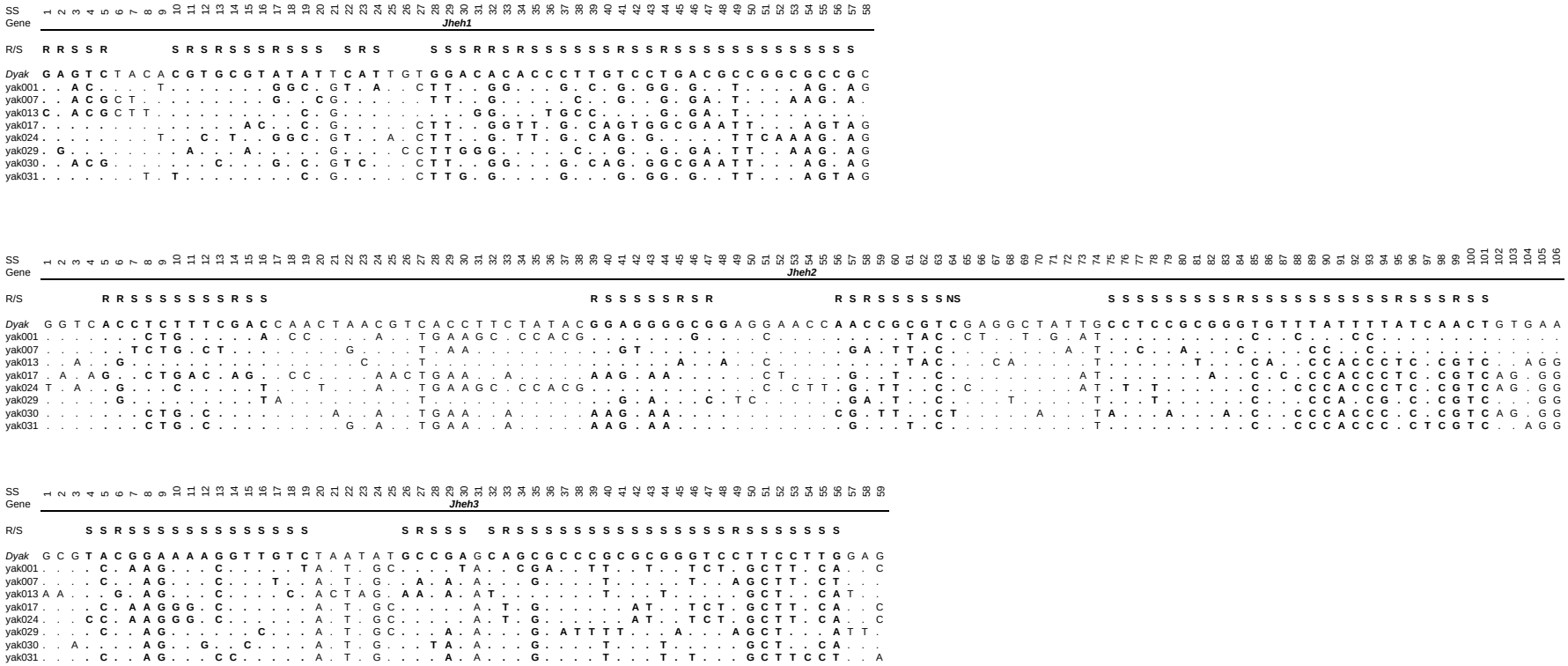

SS
Gene
Jheh1
Jheh2
Jheh3
R/S
sim001
sim004
sim006
sim008
sim165
sim169
sim188
sim192
Jheh
zim3
zim6
zim17
zim24
zim26
zim105

FIGURE 7


SS
Gene
Jheh1
R/S
Dyak
yak001
yak007
yak013
yak017
yak024
yak029
yak030
yak031
Jheh2
Jheh3