



**HAL**  
open science

**Genomic regions harboring insecticide  
resistance-associated Cyp genes are enriched by  
transposable element fragments carrying putative  
transcription factor binding sites in two sibling  
Drosophila species**

Claudia M A Carareto, Eric H Hernandez, Cristina Vieira

► **To cite this version:**

Claudia M A Carareto, Eric H Hernandez, Cristina Vieira. Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling Drosophila species. *Gene*, 2014, 537 (1), pp.93-99. 10.1016/j.gene.2013.11.080 . hal-00922703

**HAL Id: hal-00922703**

**<https://inria.hal.science/hal-00922703>**

Submitted on 4 Jul 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species

Claudia M.A. Carareto <sup>a,\*</sup>, Eric H. Hernandez <sup>a</sup>, Cristina Vieira <sup>b,c</sup>

<sup>a</sup> UNESP—Univ. Estadual Paulista, Departamento de Biologia, Laboratório de Evolução Molecular, 15054-1000 São José do Rio Preto, São Paulo, Brazil

<sup>b</sup> Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

<sup>c</sup> Institut Universitaire de France, France

## ABSTRACT

In the present study, an *in silico* analysis was performed to identify transposable element (TE) fragments inserted in *Cyps* with functions associated with resistance to insecticides and developmental regulation as well as in neighboring genes in two sibling species, *Drosophila melanogaster* and *Drosophila simulans*. The *Cyps* associated with insecticide resistance and their neighboring non-Cyp genes have accumulated a greater number of TE fragments than the other *Cyps* or a random sample of genes, predominantly in the 5'-flanking regions. Most of the insertions were due to DNA transposons, with *DNAREP1* fragments being the most common. These fragments carry putative binding sites for transcription factors, which reinforces the hypothesis that *DNAREP1* may influence gene regulation and play a role in the adaptation of the *Drosophila* species.

## 1. Introduction

Genes encoding hormone biosynthesis enzymes, transcription factors and other factors involved in developmental regulation are essential for an organism's survival. These genes tend to be highly conserved due to their functional nature, which means that the occurrence of deleterious mutations within or close to such genes often has lethal effects (Simons et al., 2006; Wagner et al., 2003). Conversely, genes associated with environmental responses and external stimuli, such as those involved in the metabolism of xenobiotics; with immunological functions; or that are involved in stress responses are less well conserved and contribute to genome plasticity (Chen and Li, 2007; van de Lagemaat et al., 2003). The genome of *Drosophila melanogaster* contains 99 genes belonging to the cytochrome P450 (*Cyps*) family ([http://flybase.org/cgi-bin/quicksearch\\_solr.cgi?caller=quicksearch&tab=basic\\_tab&context=cytochrome%20P450%20&data\\_class=FBgn&species=Dmel&search\\_type=all](http://flybase.org/cgi-bin/quicksearch_solr.cgi?caller=quicksearch&tab=basic_tab&context=cytochrome%20P450%20&data_class=FBgn&species=Dmel&search_type=all)), which encode monooxygenases that function in a wide range of biological processes, including developmental and xenobiotic pathways (Tijet et al., 2001). Among the *Cyps* associated with insecticide resistance in *D. melanogaster*, studies have tended to focus on nine genes:

*Cyp4e2* (Amichot et al., 1994), *Cyp6a2* (Maitra et al., 1996, 2000; Pedra et al., 2004), *Cyp6a8* (Maitra et al., 2000), *Cyp6a9* (Maitra et al., 1996), *Cyp6g1* (Catania et al., 2004; Daborn et al., 2002; Schlenke and Begun, 2004), *Cyp6w1* (Pedra et al., 2004), *Cyp12a4* (Bogwitz et al., 2005), *Cyp12d1* (Brandt et al., 2002) and *Cyp308a1* (Gene Ontology Consortium, 2001). Six of these genes have been associated with high levels of gene transcription in resistant insects (Amichot et al., 2004; Bogwitz et al., 2005; Daborn et al., 2002, 2007; Helvig et al., 2004; Le Goff et al., 2003).

Several studies have related insecticide resistance in *Drosophila* due to the insertion of transposable elements (TEs) in *Cyps*. These sequences can affect the expression of adjacent genes by introducing regulatory binding sites (Conte et al., 2002; Jordan et al., 2003; Kunarso et al., 2010; Molineris et al., 2011; Rebollo et al., 2012; Schmid and Bucher, 2010; Thornburg et al., 2006; Wang et al., 2007, 2009) into gene-flanking regions. Daborn et al. (2002) reported that the insecticide resistance mapping to the DDT-R locus of *D. melanogaster* is due to the over-expression of one gene, *Cyp6g1*, and that a single *Cyp6g1* allele linked to this resistance is distributed worldwide. Unlike other insecticide resistance genes in which overexpression is associated with point mutations, over-expression of *Cyp6g1* is characterized by the insertion of a fragment of the *Accord* retrotransposon upstream of the gene (Catania et al., 2004; Chung et al., 2007; Daborn et al., 2002). Schlenke and Begun (2004) have also reported an association between a TE insertion and insecticide resistance in *Drosophila simulans*. In this species, a *Doc* element inserted in the flanking region of the *Cyp6g1* ortholog promotes its over-expression. In addition, Marsano et al. (2005) and

**Abbreviations:** CDS, coding sequence; Cyp, cytochrome P450 monooxygenases; DINE-1, *Drosophila* interspersed element 1; LTR, long terminal repeat; nt, nucleotide; MITE, miniature inverted-repeat transposable element; SINE, short interspersed nuclear element; TEs, transposable elements; TFBS, transcription factor binding site; TIR, terminal inverted repeat.

\* Corresponding author. Tel.: +55 17 3221 2382.

E-mail address: carareto@ibilce.unesp.br (C.M.A. Carareto).

Bogwitz et al. (2005) have suggested that the presence of the *Bari-1* transposon at the 3' end of *Cyp12a4* in *D. melanogaster* increases gene expression. These reports associate TE insertion with gene over-expression and insecticide resistance.

Chen and Li (2007) analyzed TE insertions in 13 *Cyps* from the *D. melanogaster* genome, eight of which were associated with insecticide resistance (*Cyp4e2*, *Cyp6a2*, *Cyp6a8*, *Cyp6a9*, *Cyp6g1*, *Cyp6w1*, *Cyp12a4* and *Cyp12d1*), while five were involved in ecdysone biosynthesis and developmental regulation (*Cyp302a1*, *Cyp306a1*, *Cyp307a1*, *Cyp314a1* and *Cyp315a1*). Seven of the eight *Cyps* associated with insecticide resistance contained TE insertions, whereas none of those associated with development presented TE insertions in either the 5'- or 3'-flanking regions. The authors hypothesized that TEs might be selectively enriched among environmental response genes but are often excluded from essential housekeeping genes, resulting in the greater genomic plasticity of the former and increased conservation of the latter.

*D. melanogaster* and *D. simulans* are the main species in the *Drosophila* group that are used in research on both TEs and insecticide resistance. Despite being closely related, cosmopolitan species that occupy similar environments, these species differ significantly in terms of the degree of genetic differentiation between populations (Moriyama and Powell, 1996) as well as in their DNA polymorphism (Aquadro et al., 1988; Martín-Campos et al., 1992) and protein sequences (Choudhary and Singh, 1987). These two species also exhibit different responses to insecticides (Windelspecht et al., 1995) and display different proportions of TEs within their genomic sequence, with three times as many TEs being found in *D. melanogaster* than in *D. simulans* (Dowsett and Young, 1982; *Drosophila* 12 genomes consortium, 2007; Hoskins et al., 2002; Kaminker et al., 2002; Vieira et al., 1999). However, the number of TE insertions in *D. simulans* is greater than in *D. melanogaster* because most of the copies in the former are degraded into small fragments (Lerat et al., 2011).

In light of these characteristics of *D. melanogaster* and *D. simulans*, we reappraised the preferential insertion of TEs in genes (and their flanking regions) that belong to different functional categories by analyzing *Cyp* genes associated with resistance to insecticides and developmental regulation. Moreover, we investigated the closest neighboring non-*Cyp* genes with the aim of comparing the dynamics of such insertions in close genomic regions in both species. We showed that the *Cyps* associated with resistance to insecticides generally harbor more TE insertions in both species, as do their neighboring genes, and that the most frequent insertions in these *Cyps* and their neighbors come from two DNAREP1 families, which are enriched with putative transcription factor binding sites (TFBSs).

## 2. Materials and methods

The complete set of analyzed sequences included *Cyp* genes involved in xenobiotic detoxification or insecticide resistance and in developmental processes as well as their 3-kb flanking regions. We also retrieved the closest non-*Cyp* genes neighbors of these *Cyp* genes, together with their flanking regions. Genomic information was obtained from the genomes of *D. melanogaster* (release 5.5) and *D. simulans* (release 1.4) deposited in FlyBase (<ftp://ftp.flybase.net/genomes/>). To define and then extract the flanking regions of the *D. simulans* orthologs of *D. melanogaster* *Cyp* genes, a precomputed file from FlyBase called "orthologs" was used ([http://flybase.org/static\\_pages/docs/datafiles.html#gene\\_orthologs](http://flybase.org/static_pages/docs/datafiles.html#gene_orthologs)). A gene map file ([http://flybase.org/static\\_pages/docs/datafiles.html#gene\\_map\\_table](http://flybase.org/static_pages/docs/datafiles.html#gene_map_table)) of Flybase annotation provided positional coordinates for all genes, and a 3-kb addition at both the 5'- and 3'-ends was edited for the extraction of analyzed sequences.

For each species, *Cyp*-neighboring regions were visually inspected in GBrowse (<http://flybase.org/cgi-bin/gbrowse2/>) to create the neighboring gene set, discarding *Cyp* and tRNA genes as neighbors and selecting only the remaining genes. The gene names, FlyBase identifiers and neighborhood relationships between genes are described in the

Supplementary files (Supplementary File 1, Table S1: *D. melanogaster* and Table S2: *D. simulans*).

A random gene set from each species was created using an in-house-developed shell script based on *awk* and *suf* UNIX tools run on both genomes using its FASTA header identifiers (Table S3: *D. melanogaster* and Table S4: *D. simulans*). At this point, an exception was made in *D. simulans* due to the draft state of its genome relative to that of *D. melanogaster*, and a 100 kb filter was therefore applied for contig size to ensure that the filtered contigs would contain the complete gene and flanking regions of the required size.

All selected sequences were submitted to the RepeatMasker online server (<http://www.repeatmasker.org/>) applying the parameters `-lib (D. melanogaster)`, `-nolow -engine (crossmatch) -gc (query level GC content)`, using the database of *Drosophila* reference TEs stored in Repbase (Jurka et al., 2005). The program was allowed to run the ProcessRepeats script of RepeatMasker to define which fragments were derivative sequences from the same element, and rare assignments were discarded. The number of element insertions was recorded and compared in both the 5'- and 3'-flanking regions for all genes. Transcription factor binding sites were predicted based on DNAREP1\_DM and DNAREP1\_Dyak fragment sequences in the flanking sequences of the *Cyp* and non-*Cyp* neighbor genes of both species using the JASPAR server ([http://asp.iu.uib.no:8090/cgi-bin/jaspar2010/jaspar\\_db.pl](http://asp.iu.uib.no:8090/cgi-bin/jaspar2010/jaspar_db.pl)), JASPAR CORE Insecta and the subset of the *D. melanogaster* database with a 90% cutoff score (Bryne et al., 2008) and passed onto the Consite service <http://asp.iu.uib.no:8090/cgi-bin/CONSITE/consite>.

A generalized linear model, specifically, a negative binomial regression (NBR) with a log-link function, was used to model the number of TE fragment insertions considering the factors gene class (resistance, neighbor and random) and each of species as dummy variables. NBR can be considered a generalization of a Poisson regression, as it exhibits a similar structure to a Poisson regression but with an extra parameter to model the over-dispersion (the magnitude between the mean and variance) of data. A submodeling selection process was implemented by means of stepwise model selection based on Akaike's information criterion; this analysis was performed using the `stepAIC` function included in the package MASS for the R statistical computing program.

## 3. Results

### 3.1. TE insertions in *Cyps* and their neighboring genes

We identified TE insertions within the *Cyps* and non-*Cyp* neighbor genes and throughout their 3-kb flanking regions in *D. melanogaster* (Table 1, Supplementary file 1, Table S1) and *D. simulans* (Table 1, Supplementary file 1, Table S1). The TE insertions found in these genes were remarkably different. First, the *Cyps* associated with developmental regulation did not harbor TE insertions in their introns or flanking regions, while the *Cyps* associated with insecticide resistance did harbor TEs in both *D. melanogaster* and *D. simulans* but only in their flanking regions. Additionally, the genes neighboring these *Cyps*, located at their 5'- and 3'-ends (Table 1, Supplementary file 1, Tables S1 & S2), and a random sample of 43 genes (Supplementary file 1, Tables S6 & S9) found in both species contained many insertions in their introns, which were generally large, as well as in their flanking regions. Second, the *Cyps* associated with resistance to insecticides and their neighboring genes presented the highest percentage of TE insertions in their flanking regions (Table 1, Fig. 1). In *D. melanogaster* and *D. simulans*, 45.4% of the genes associated with resistance harbored TE insertions, as did an even higher percentage of their neighboring genes (*D. melanogaster*: 50%–66.6% and *D. simulans*: 57.1%–83.3%, respectively, for neighbors located at the 5' and 3' ends of the *Cyps*); these values were much higher than found in the neighboring genes of the *Cyps* related to developmental regulation, which harbored insertions only in *D. melanogaster* (5' neighbors: 16.6%; 3' neighbors: 20%). Insertions were also found in the random sample of genes (*D. melanogaster*: 34.9% and *D. simulans*:

**Table 1**

Total number (N) of analyzed genes (*Cyps* associated with insecticide resistance and developmental regulation, *Cyp*-neighboring genes and a sample of randomly selected genes); numbers and percentages of genes with TEs in their 3 kb 5'- and 3'-flanking regions; total numbers of insertions (5'- and 3'-flanking regions and introns); and proportions of insertions in the 3-kb flanking regions of these genes in the genomes of *D. melanogaster* and *D. simulans*.

| Species<br>Gene categories | Genes |                            | Insertion numbers          |    |    | Insertions <sup>a</sup><br>per gene |         |
|----------------------------|-------|----------------------------|----------------------------|----|----|-------------------------------------|---------|
|                            | N     | with TEs <sup>a</sup><br>n | with TEs <sup>a</sup><br>% | 5' | 3' | Introns                             | 5' + 3' |
| <i>D. melanogaster</i>     |       |                            |                            |    |    |                                     |         |
| Cyps-resistance            | 11    | 5                          | 45.4                       | 6  | 4  | 0                                   | 0.9     |
| Cyps-development           | 6     | 0                          | 0                          | 0  | 0  | 0                                   | 0       |
| Cyp neighbors              |       |                            |                            |    |    |                                     |         |
| Resistance 5'              | 6     | 3                          | 50                         | 6  | 3  | 35                                  | 1.5     |
| Resistance 3'              | 6     | 4                          | 66.6                       | 4  | 1  | 17                                  | 0.8     |
| Development 5'             | 6     | 1                          | 16.6                       | 1  | 0  | 4                                   | 0.2     |
| Development 3'             | 5     | 1                          | 20                         | 0  | 1  | 1                                   | 0.2     |
| Gene-nesting Cyps          | 2     | 1                          | 50                         | 1  | 0  | 2 <sup>1</sup>                      | 0.5     |
| Random sample              | 43    | 15                         | 34.9                       | 9  | 10 | 7                                   | 0.4     |
| <i>D. simulans</i>         |       |                            |                            |    |    |                                     |         |
| Cyps-resistance            | 11    | 5                          | 45.4                       | 6  | 8  | 0                                   | 1.3     |
| Cyps-development           | 7     | 0                          | 0                          | 0  | 0  | 0                                   | 0       |
| Cyp neighbors              |       |                            |                            |    |    |                                     |         |
| Resistance 5'              | 7     | 4                          | 57.1                       | 4  | 1  | 2                                   | 0.7     |
| Resistance 3'              | 6     | 5                          | 83.3                       | 10 | 0  | 16                                  | 1.7     |
| Development 5'             | 5     | 0                          | 0                          | 0  | 0  | 2                                   | 0       |
| Development 3'             | 6     | 0                          | 0                          | 0  | 0  | 0                                   | 0       |
| Random sample              | 43    | 9                          | 20.9                       | 4  | 6  | 15                                  | 0.2     |

<sup>a</sup> Insertions within introns were not used in the analyses.

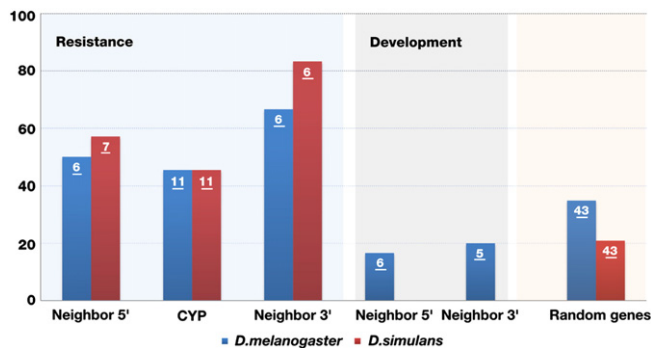
20.9%); these values are also smaller than those in the *Cyps* associated with resistance and their neighbors.

In analyzing our results, we focused on TE fragments inserted in the 3-kb flanking regions of the adjacent genes because the *Cyp* genes, which accounted for the bulk of our analysis, do not harbor TE fragments in their introns. Taking into consideration both the 5'- and 3'-flanking regions (Table 1, Fig. 2 and Supplementary file 1, Tables S1 & S2 and Table S9), the ratio of insertions per gene was observed to be at least twice as high in the genes associated with resistance to insecticides (*D. melanogaster*: 0.9 and *D. simulans*: 1.3) and their neighbors at the 5'- and 3'-ends (*D. melanogaster*: 1.5 and 0.8 and *D. simulans*: 0.7 and 1.7), as in the other gene categories in both *D. melanogaster* (*Cyps* of development: 0 and their neighbors: 0.2; random sample: 0.4) and *D. simulans* (*Cyps* of development and their neighbors: 0 and random sample: 0.2).

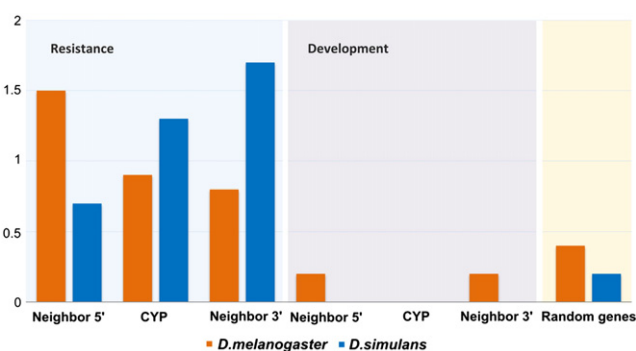
Modeling the number of TE insertions, considering the factors “gene class” (resistance, neighbor and random) and “species” as variables indicated that only the factor “gene class” was responsible for the observed variability. The expected log count of the number of TE insertions showed that the variable “gene random” ( $-1.09447$ ) was significantly different ( $p < 0.001$ ) than in the reference group (intercept: neighboring genes) but was not significantly different than

the variable “resistance genes” ( $p > 0.05$ ). The rate of TE fragment insertions (which is the incidence rate ratio in the model) was three times lower than in the reference group (1.08) and the resistance genes (1.01). Overall, the rate of TE insertions in the flanking regions was lower in the random sample of genes and did not differ between the *Cyps* associated with resistance and their neighboring genes or between the two sibling species, *D. melanogaster* and *D. simulans*.

With a few exceptions, the TE insertions in the 3-kb flanking regions of all of the examined genes consisted of small fragments, generally of less than 200 bp (Table 2). In *D. simulans*, for example, the average sizes varied from 67.1 bp (neighboring genes at the 3'-end of the *Cyps* associated with resistance) to 137.8 bp (*Cyps* associated with resistance). In *D. melanogaster*, the average sizes were larger than in *D. simulans*, varying from 126 bp (neighboring genes at the 5'-end of the *Cyps* associated with resistance) to 688.6 bp (*Cyps* associated with resistance). The latter mean was due to two full-length insertions in the *Cyps* associated with resistance: one Bari\_DM DNA transposon (1740 bp) and one Max retrotransposon (2748 bp of which was within the region under analysis, while the remainder extended to the 3'-intergenic region). Full-length and putatively autonomous sequences (5 TEs), which were defined as such because their size corresponded to the consensus elements deposited in Repbase and the expected domains of an autonomous element in the *Cyp*s and neighboring



**Fig. 1.** Percentage of analyzed genes containing TE sequences in the 3 kb 5'- and 3'-flanking ends in *D. melanogaster* and *D. simulans* (y-axis). The number of genes in each category (*Cyps* associated with resistance, neighbors and random genes) is provided within each column.



**Fig. 2.** Proportion of TE insertions (y-axis) per gene in each category (*Cyps* associated with resistance, neighbors and random genes) in *D. melanogaster* and *D. simulans* in the 5'- and 3'-flanking regions.

**Table 2**  
Total number of TE insertions (N) in the 3-kb flanking regions and means  $\pm$  standard errors of divergence from the consensus element and of the length of TE fragments in the *D. melanogaster* and *D. simulans* genomes.

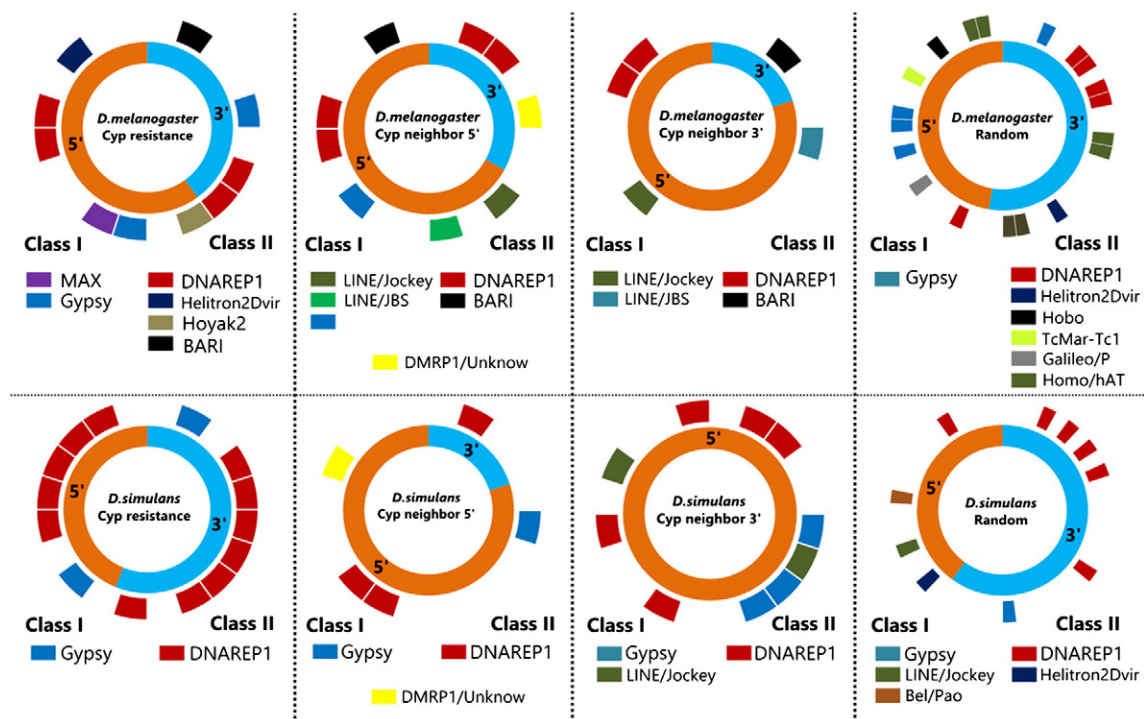
| Gene categories | <i>D. melanogaster</i> |                |                   | <i>D. simulans</i> |                |                  |
|-----------------|------------------------|----------------|-------------------|--------------------|----------------|------------------|
|                 | N                      | Divergence (%) | Length (bp)       | N                  | Divergence (%) | Length (bp)      |
| Resistance      |                        |                |                   |                    |                |                  |
| Cyps            | 10                     | 17.1 $\pm$ 2.5 | 688.6 $\pm$ 292.7 | 14                 | 16.3 $\pm$ 1.4 | 137.8 $\pm$ 40.2 |
| Neighbors 5'    | 9                      | 15.5 $\pm$ 3.3 | 126 $\pm$ 22.0    | 5                  | 15.9 $\pm$ 2.5 | 126.8 $\pm$ 20.1 |
| Neighbors 3'    | 5                      | 13.2 $\pm$ 5.5 | 447 $\pm$ 307.4   | 10                 | 16.9 $\pm$ 2.0 | 67.1 $\pm$ 14.4  |
| Development     |                        |                |                   |                    |                |                  |
| Cyps            | 0                      | -              | -                 | 0                  | -              | -                |
| Neighbors 5'    | 1                      | 26.2           | 87                | 0                  | -              | -                |
| Neighbors 3'    | 1                      | 12.2           | 166               | 0                  | -              | -                |
| Random          | 21                     | 18.3 $\pm$ 1.9 | 137.9 $\pm$ 20.6  | 10                 | 17.5 $\pm$ 1.6 | 80.5 $\pm$ 8.1   |

genes, were found only in the *D. melanogaster* genome. Among the Cyps, a copy of the LTR retrotransposon MAX was found to be inserted in the 5'-end of *Cyp6a8*, and a copy of the DNA transposon BARI\_DM was found in the 3'-end of *Cyp12a4* (this copy is shared with *Ppcs*, the 3'-neighboring gene of *Cyp12a4*). Among the Cyp-neighboring genes, the LTR retrotransposon *Roo* was identified, inserted within an intron of the *luna* gene (3'-neighbor of *Cyp12d1*), as were two NLTR retrotransposons: *BS*, inserted within an intron of the CG6959 gene (5'-neighbor of *Cyp315a1*), and LINEJ1\_DM, inserted within an intron of *luna* (5'-neighbor of *Cyp12d1*). Among these insertions, it is worth noting the retrotransposon MAX, which belongs to the BEL-PAO superfamily and is one of the longest retrotransposons found in *D. melanogaster* (~8550 bp). A full-length MAX is inserted just 254 bp from the 5'-flanking region of *Cyp6a8*, which is located in the opposite orientation (3'-5') of MAX (5'-3'). The LTRs of this MAX element are 99% identical, which means that this insertion is very recent, and it should be emphasized that the 3'-LTR of this sequence contains two copies of each of two TFBSs, potentially affecting the expression of this gene: hunchback, from nucleotides 106–115 (minus strand) and 70–79 (plus strand), and Eip74EF, from nucleotides 16–22 and 191–197 (both in the plus strand).

The number of TE insertions found in the 5'-ends and the 3'-ends was variable (Table 1, Fig. 2 and Supplementary file 1, Tables S1 & S2), but insertions were generally more frequent in the 5'-flanking regions than in those at the 3'-ends. This difference was not observed in the random sample of genes from both *D. melanogaster* and *D. simulans* or the Cyps associated with resistance in *D. simulans*. Although a statistical test is not applicable here because the number of insertions was small and several gene categories included only one (or no) insertions, we can tentatively say that the 5'-flanking region of the Cyps and their neighboring genes are more prone to TE insertions.

### 3.2. TE families in the flanking regions of Cyps and neighboring genes

The identified TE fragments inserted in the flanking regions of all examined gene categories are provided in Supplementary file 1 (Tables S5 to S10). In Fig. 3, information on the insertions is presented including their location in a flanking region (5'- or 3'-region), number and classification. The two species show remarkable differences. The *D. melanogaster* genome harbors higher diversity of TE families in the Cyp flanking regions and their neighboring genes than the *D. simulans* genome. In the latter species, with the exception of one 5'-neighbor



**Fig. 3.** Snapshot overview of the observed TE insertions, by family, in the flanking regions of the Cyps, their neighbors and a random set of genes in *D. melanogaster* and *D. simulans*. The proportions of insertions in both the 5' and 3' regions are shown. The class or order to which each TE belongs is indicated by different colors.

(GD10333) of the GD10398 gene (*Cyp6w1*), which contained a DMRP1 insertion (an unclassified element), and one 3'-neighbor (GD25946) of GD25944 (*Cyp12d1-p*), which presented two TART insertions, all of the other TE fragments belonged to three Gypsy families (*Gypsy19*, *Gypsy8* and *Gypsy3*) and to two Helitrons (DNAREP1\_DM and DNAREP1\_DYak). In addition to the Gypsy families, other LTRs (MAX) and NLTRs (BS) were found in *D. melanogaster*, as were several DNA transposons, such as *Bari*, *Pogo*, *Hoyak* and *Helitrons*. Moreover, the random sample of genes harbored a greater number and diversity of TE insertions in *D. melanogaster* than in *D. simulans* (Table 1, Fig. 3, and Supplementary file 1, Tables S7 and S10, respectively).

Despite the aforementioned differences, the two species share similarities regarding the classes to which the TE fragments belong. DNA transposons (Class II) were the most frequent insertions found in the flanking regions (5' + 3') of all categories of genes, and the Helitrons (DNAREP1\_DM and DNAREP1\_DYak) constituted the majority of this group (Fig. 3). Moreover, DNA transposons occurred in higher percentages in the flanking regions of the resistance *Cyps* (*D. melanogaster*: 70%; *D. simulans*: 85.7%) than in all other categories of genes (*D. melanogaster*: 56.9 to 57.2%; *D. simulans*: 53.3% to 60%). Taking into account only the 5'-flanking regions, where most of the promoter sequences are located and where TE insertions may therefore have a greater effect on gene expression, the enrichment of DNA transposons among the *Cyps* associated with resistance was maintained (*D. melanogaster*: 67.7%; *D. simulans*: 83.3%), while the other gene categories showed no or only slight differences between insertions of retrotransposons and DNA transposons (Supplementary file 1, Table S11). Among the DNA transposons, DNAREP1\_DM and DNAREP1\_DYak, were found to accumulate in the 5'-flanking sequences of the *Cyp* genes and their neighbors in *D. simulans* as well as in the neighbors of the *D. melanogaster* *Cyp* genes (Supplementary file 1, Table S11) but not in the random sample of genes from both species.

DNAREP1 elements were highly abundant in 12 out of 23 sequenced *Drosophila* genomes analyzed by Yang and Barbash (2008). These elements are classified in Repbase as DNA transposons of the *Helitron* family (Kapitonov and Jurka, 2007) but have also been designated DINE-1 or INE-1 and have been proposed to be SINE-like retroelements (Locke et al., 1999), TIR transposons (Bergman et al., 2006) or even MITEs (Yang and Barbash, 2008). Here, we maintain the classification of Repbase, where two consensus sequences of DNAREP1 are deposited: DNAREP1\_DM and DNAREP1\_DYak. The first, DNAREP1\_DM, described in *D. melanogaster* (Kapitonov and Jurka, 1999, 2003), is a non-autonomous 594-bp DNA element that has copies that show an average percentage of identity of approximately 90%; this is the same average identity found in copies in the *D. simulans* and *Drosophila sechellia* genomes (Yang and Barbash, 2008). In contrast, the second consensus sequence, DNAREP1\_DYak (Kapitonov and Jurka, 2003), is a non-autonomous 793-bp element, described in *Drosophila yakuba*, the copies of which display a peak of identity centered approximately 96–100% and a long tail of more divergent sequences. It has been proposed that DNAREP1\_DM and DNAREP1\_DYak are derived from a common ancestor that existed before the divergence of the species of the *melanogaster* subgroup (Yang and Barbash, 2008). It is likely that the sequences identified in this study as DNAREP1\_DYak correspond to relics of those copies that existed in the common ancestor of the *D. melanogaster* and *D. yakuba* complexes of species.

Because of the prevalence of DNAREP1 fragments among the *Cyps* associated with resistance and their non-*Cyp* neighbors compared to the random sample of genes, we performed a more detailed analysis of these insertions. Among the 22 *Cyp* genes associated with resistance (11 from each species), the copies of DNAREP1 were concentrated in the flanking regions of five *Cyps* (*Cyp6a2*, *Cyp6w1*, *Cyp4e1*, *Cyp12d1-p* and *Cyp12d1-d*) in *D. simulans* (12 insertions) and only two genes (*Cyp6a2* and *Cyp6w1*) in *D. melanogaster* (four insertions). Of particular note, from the 14 insertions found in flanking regions of *Cyps* related to

resistance in *D. simulans*, only two were not DNAREP1 sequences. In the neighbors of these *Cyps* (12 in *D. melanogaster* and 13 in *D. simulans*), five insertions were found in four genes in *D. melanogaster* (*BBS4*, *Ecr*, *Lcp4*, CG8343 and *inr-a*), while eight insertions were identified in seven genes in *D. simulans* (GD10333, GD10560, GD10768, GD10301, GD10399, GD25648 and *Mal-A1*).

As a result of these findings, we further investigated these sequences, searching for motifs similar to TFBS in the DNAREP1 copies, assuming that they may be preferentially retained in these genes because they harbor regulatory signals (Jordan et al., 2003). The DNAREP1\_DM and DNAREP1\_DYak consensus elements display putative motifs that are homologous to the transcription factors CF2\_II, in *D. melanogaster*, and CF2-II and hunchback, in *D. simulans* (Supplementary file 2, Table S12). They are preferentially orientated in the sense direction (64.7%).

Using a 90% cutoff score, we found that both the DNAREP1\_DM and DNAREP1\_DYak small fragments maintained in the flanking regions of the *Cyp* genes and their neighbors carried multiple putative motifs similar to *D. melanogaster* TFBSs in both genomes (Supplementary file 2, Table S13 & S14). We checked the orientations of the TFBSs in the DNAREP1 insertions located in the 5'- and 3'-flanking regions of the nearby *Cyps* and their neighboring genes. It was interesting to find that in *D. melanogaster*, the orientation of the TFBS preferentially corresponded to the antisense DNAREP1 strand, independent of the flanking region of the TE insertion (5' insertion: 80%; 3' insertion: 87.5%), whereas the opposite situation was observed in *D. simulans*. Although the differences were not as remarkable as in *D. melanogaster*, the putative TFBSs in the DNAREP1 insertions in the flanking regions of the *D. simulans* genes mostly occurred in the DNAREP1 sense strand, also independent of the gene flanking region in which the TE was inserted (5' insertion: 61.1%; 3' insertion: 57.1%).

#### 4. Discussion

When TEs are inserted in or near essential genes, they are usually deleterious and tend to be eliminated from populations. However, it has been shown that TE insertions in the proximity of *Cyps* can alter their regulation and cause the over-expression of CYP enzymes, resulting in insecticide resistance (Catania et al., 2004; Daborn et al., 2001, 2002; Jouben et al., 2008). TE insertions could therefore also be a source of genetic variability, and the regulatory motifs they provide could be advantageous for organisms in responding to external stimuli. This could be of particular importance in organisms such as insects, which occupy highly variable and selective environments. In this study, we focused on *Cyps* associated with insecticide resistance and developmental regulation as well as their nearest neighboring non-*Cyp* genes in two sibling species that are known to harbor different TE contents.

Most of the identified TE insertions were within non-coding regions (introns and flanking sequences), and such insertions were virtually absent from the CDSs, as previously reported in *Drosophila* (Fontanillas et al., 2007; Kaminker et al., 2002; Miller et al., 2000), reinforcing the view that insertions in non-coding regions are tolerated, whereas TE insertions within or close to coding regions can have deleterious effects and are removed through purifying selection (Lipatov et al., 2005; Sela et al., 2007; Yang and Barbash, 2008). Considering all of the genes in each category studied here, the highest percentage of genes harboring TEs in their flanking regions were found among *Cyps* associated with insecticide resistance and their neighboring non-*Cyp* genes in both *D. melanogaster* and *D. simulans*. Accordingly, these gene categories also included the highest proportions of insertions per gene. One striking observation was that *Cyps* related to development, as well their non-*Cyp* neighbors, showed a low proportion of TE insertions per gene (varying from zero to 0.2 in *D. melanogaster*), not differing from the random sample of genes (0.2 and 0.4 respectively in *D. melanogaster* and *D. simulans*). Such a pattern suggests that the *Cyps* associated with

resistance are more permissive to TE insertions, most likely because these sequences donate transcriptional regulatory signals that could alter host gene expression, resulting in more relaxed purifying selection in the close flanking regions of the genes but also in adjacent regions. Conversely, purifying selection is strong for essential genes related to development, removing TEs that are occasionally inserted in their promoter regions but that also extend to the flanking regions of neighboring genes. This hypothesis is plausible because these non-*Cyp* genes harbor TEs in their introns, where they may also affect gene expression but are more tolerated because their effects are generally not as deleterious as in flanking regions of essential genes.

In addition to the similarities between these two sister species genomes, differences can be observed regarding TE content. As mentioned above, the *D. melanogaster* genome contains three times as many TEs as that of *D. simulans*, essentially because the copies of TEs found in the former species are mainly full-length. With respect to the number of TE fragments present, *D. simulans* displays a greater number than *D. melanogaster*, largely because most of the copies in the former species are degraded and consist of small fragments (Lerat et al., 2011). The *Cyp* flanking regions and the non-*Cyp* neighbors harbor a higher diversity of TEs, retrotransposons and DNA transposons in *D. melanogaster* than in *D. simulans*, which contains few insertions other than *Gypsy* and *DNAREP1*-like (Fig. 3).

The rank order of the abundance of the major TE classes found in the non-pericentromeric regions of the *D. melanogaster* genome was as follows: LTR elements (2.42%, 4.96 copies/Mb) > non-LTR elements (0.62%, 3.24 copies/Mb) > TIR elements (0.15%, 2.06 copies/Mb). *DNAREP1* accounts for only 0.10% of a typical region of the *D. melanogaster* genome but contributes 6.36 copies/Mb because most of its sequences are small and degenerated. The copies of *DNAREP1\_DM* in *D. melanogaster* and *D. simulans* show an average identity of approximately 90% (Yang and Barbash, 2008). The average identity shared with the consensus sequences of *DNAREP1\_DM* and *DNAREP1\_DYak* identified in the flanking regions of the *Cyps* of *D. melanogaster* and *D. simulans* was approximately 83%. These data, together with the small sizes of the insertions, corroborate the hypothesis that the genes analyzed in both species contain old copies of *DNAREP1*. Because long evolutionary periods are necessary for young copies of TEs to acquire new biological functions (Gotea and Makalowski, 2006), the presence of small and divergent copies of the *DNAREP1* sequences, inserted preferentially in the 5'-ends of the *Cyps* and presenting putative TFBSs, suggests that these copies may play an adaptive role in these species by affecting gene expression.

To investigate this possibility, it is important to understand how TE fragments become enriched with *cis*-regulatory elements. One possible explanation is that decaying TE sequences provide materials from which *cis*-regulatory elements can be constructed de novo through the introduction of single or a few point mutations (Feschotte, 2008) because most transcription factor binding sites consist of short degenerate sequences (Wray et al., 2003). Two other possibilities are that *cis*-regulatory elements are already present within the TE at the time of its insertion and are co-opted immediately upon insertion or after modifications in the surrounding environment (Feschotte, 2008) or that its "pre-sites" only require a few mutations to originate functional TFBSs (de Souza et al., 2013). Both scenarios fit with our results. The complete *DNAREP1\_DM* and *DNAREP1\_DYak* sequences harbor motifs similar to *D. melanogaster* TFBSs that play a key role in fly development, such as CF2-II (chorion factor 2, which mediates muscle cell differentiation) and hunchback (involved in salivary gland and embryo development). Moreover, the fragments found in the 5'-flanking ends of the *Cyps* contain motifs similar to TFBSs that are not found in the consensus sequences, such as *usp*, *pan*, *bkr*, *vnd*, *tin*, *br-Z4*, *cad* and *Ubx* in *D. melanogaster*, which could have emerged de novo during the degeneration of these fragments. It was interesting to find that *D. melanogaster* and *D. simulans*, which differ regarding their TE content (in terms of copy number and full-length TEs) in general and in the

diversity of the TEs inserted in the flanking regions of the three gene categories studied here, as cited above, also differ regarding their TFBS orientation bias in the *DNAREP1* strand, which is preferentially anti-sense oriented in *D. melanogaster* but sense oriented (although only slightly) in *D. simulans*. It should be emphasized that due to their orientation in the *DNAREP1* strand, many of these putative TFBSs might be non-functional because transcription factors usually bind in a defined orientation to the DNA double helix, and this orientation depends on the orientation of the DNA sequence they recognize. For example, some TFBSs must have a defined orientation relative to the promoter or the transcription start site, and for TATA boxes, only (+)-strand matches should be considered as true positives, whereas for most other TFBS, both (+)- and (-)-strand matches should be considered equally.

The presence of several *DNAREP1* copies within 3 kb regions of genes, potentially occurring in both untranslated regions and the regulatory regions of genes, led Yang and Barbash (2008) to suggest that this transposon has had a significant impact on the evolution of gene structure throughout the Drosophilidae. Our data show that these insertions in *Cyp* genes carry putative TFBSs, which reinforces the hypothesis that they might play important roles in the adaptation of *Drosophila* species. These particular characteristics may explain why TEs interfere in the transcription of some *Cyp* genes by elevating their expression levels (Catania et al., 2004; Chung et al., 2007; Daborn et al., 2002). Empirical studies involving transfection and expression experiments for these putative regulatory sequences can corroborate the hypothesis that *DNAREP1* sequences have been domesticated to contribute to new, alternative transcription start sites in *Drosophila* genomes. Furthermore, population studies should be conducted to identify insertion polymorphisms for this specific element, followed by a functional analysis, as has been carried out for *D. melanogaster* adaptive insertions (González et al., 2009).

## Acknowledgments

Funding for this project was provided by the Brazilian Funding Agency CNPq (to CMAC) and by the French CNRS, the ANR Genemobile and the Institut Universitaire de France (to CV).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2013.11.080>.

## References

- Amichot, M., et al., 1994. In cytochrome P450. 8th International Conference (Lechner, M.C., Eds), John Libbey Eurotext, Paris, France, pp. 689–692.
- Amichot, M., Tarès, S., Brun-Barale, A., Arthaud, L., Bride, J.M., Bergé, J.B., 2004. Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 *Cyp6a2* enable DDT metabolism. *Eur. J. Biochem.* 271, 1250–1257.
- Aquadro, C.F., Lado, K.M., Noon, W.A., 1988. The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. 1. Contrasting levels of naturally-occurring DNA restriction map variation and divergence. *Genetics* 119, 875–888.
- Bergman, C.M., Quesneville, H., Anxolabéhère, D., Ashburner, M., 2006. Recurrent insertion and duplication generate networks of transposable elements in the *Drosophila melanogaster* genome. *Genome Biol.* 7, R112.
- Bogwitz, M.R., et al., 2005. CYP12A4 confers lufenuron resistance in a natural population of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12807–12812.
- Brandt, A., et al., 2002. Differential expression and induction of two *Drosophila* cytochrome P450 genes near the *Rst(2)DDT* locus. *Insect Mol. Biol.* 11, 337–341.
- Bryne, J.C., et al., 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* 36, D102–D106 (Database issue).
- Catania, F., Kauer, M.O., Daborn, P.J., Yen, J.L., Ffrench-Constant, R.H., Schlotterer, C., 2004. World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. *Mol. Ecol.* 13, 2491–2504.

- Chen, S., Li, X., 2007. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. *BMC Evol. Biol.* 7, 46.
- Choudhary, M., Singh, R.S., 1987. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic-structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans*. *Genetics* 117, 697–710.
- Chung, H., et al., 2007. Cis-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* 175, 1071–1077.
- Conte, C., Dastugue, B., Vaurie, C., 2002. Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. *Mol. Cell. Biol.* 22, 1767–1777.
- Daborn, P.J., Boundy, S., Yen, J.L., Pittendrigh, B., Ffrench-Constant, R., 2001. DDT resistance in *Drosophila* correlates with *Cyp6g1* over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Mol. Genet. Genomics* 266, 556–563.
- Daborn, P.J., et al., 2002. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* 297, 2253–2256.
- Daborn, P.J., Lumba, C., Boey, A., Wong, W., Ffrench-Constant, R.H., Batterham, P., 2007. Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression. *Insect Biochem. Mol. Biol.* 37, 512–519.
- de Souza, F.S., Franchini, L.F., Rubinstein, M., 2013. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol. Biol. Evol.* 30, 1239–1251.
- Dowsett, A.P., Young, M.W., 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 79, 4570–4574.
- Drosophila* 12 genomes consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218.
- Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.
- Fontanillas, P., Hartl, D.L., Reuter, M., 2007. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *PLoS Genet.* 3, 2256–2267.
- Gene Ontology Consortium, 2001. Creating the gene ontology resource: design and implementation. *Genome Res.* 11, 1425–1433.
- González, J., Macpherson, J.M., Petrov, D.A., 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol. Biol. Evol.* 26, 1949–1961.
- Gotea, V., Makalowski, W., 2006. Do transposable elements really contribute to proteomes? *Trends Genet.* 22, 260–267.
- Helvig, C., Tjietb, N., Feyereisen, R., Walker, F.A., Restifo, L.L., 2004. *Drosophila melanogaster* CYP6A8, an insect P450 that catalyzes lauric acid ( $\omega$ -1)-hydroxylation. *Biochem. Biophys. Res. Commun.* 325, 1495–1502.
- Hoskins, R.A., et al., 2002. Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol.* 3 (RESEARCH0085).
- Jordan, I.K., Rogozin, I.B., Glazko, G.V., Koonin, E.V., 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 19, 68–73.
- Jouben, N., Heckel, D.G., Haas, M., Schuphan, I., Schmidt, B., 2008. Metabolism of imidacloprid and DDT by P450 CYP6G1 expressed in cell cultures of *Nicotiana tabacum* suggests detoxification of these insecticides in *Cyp6g1*-overexpressing strains of *Drosophila melanogaster*, leading to resistance. *Pest Manag. Sci.* 64, 65–73.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467.
- Kaminker, J.S., et al., 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomic perspective. *Genome Biol.* 3 (research0084.1-0084.20).
- Kapitonov, V.V., Jurka, J., 1999. DNAREP1\_DM. Repbase Update Release 3.4.
- Kapitonov, V.V., Jurka, J., 2003. Molecular paleontology of transposable elements in *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6569–6574.
- Kapitonov, V.V., Jurka, J., 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* 23, 521–529.
- Kunargo, G., et al., 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634.
- Le Goff, G., et al., 2003. Microarray analysis of cytochrome P450 mediated insecticide resistance in *Drosophila*. *Insect Biochem. Mol. Biol.* 33, 701–708.
- Lerat, E., Burtet, N., Biémont, C., Vieira, C., 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene* 473, 100–109.
- Lipatov, M., Lenkov, K., Petrov, D.A., Bergman, C.M., 2005. Paucity of chimeric gene-transposable elements transcripts in the *Drosophila melanogaster* genome. *BMC Biol.* 3, 24.
- Locke, J., Howard, L.T., Aippersbach, N., Podemski, L., Hodgetts, R.B., 1999. The characterization of DINE-1, a short, interspersed repetitive element present on chromosome and in the centric heterochromatin of *Drosophila melanogaster*. *Chromosoma* 108, 356–366.
- Maitra, S., Dombrowski, S.M., Waters, L.C., Ganguly, R., 1996. Three second chromosome-linked clustered CYP6 genes show differential and barbital-induced expression in DDT-resistant and susceptible strains of *Drosophila melanogaster*. *Gene* 180, 165–171.
- Maitra, S., Dombrowski, S.M., Basu, M., Raustol, O., Waters, L.C., Ganguly, R., 2000. Factors on the third chromosome affect the level of CYP6A2 and CYP6A8 expression in *Drosophila melanogaster*. *Gene* 248, 147–156.
- Marsano, R.M., Caizzi, R., Moschetti, R., Junakovic, N., 2005. Evidence for a functional interaction between the *Bari1* transposable element and the cytochrome P450 *cyp12a4* gene in *Drosophila melanogaster*. *Gene* 357, 122–128.
- Martín-Campos, J.M., Comerón, J.M., Miyashita, N., Aguadé, M., 1992. Intraspecific and interspecific variation at the  $\gamma$ -ac-sc region of *Drosophila simulans* and *Drosophila melanogaster*. *Genetics* 130, 805–816.
- Miller, W.J., Nagel, A., Bachmann, J., Bachmann, L., 2000. Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol. Biol. Evol.* 17, 1597–1609.
- Molineris, I., Grassi, E., Ala, U., Di Cunto, F., Provero, P., 2011. Evolution of promoter affinity for transcription factors in the human lineage. *Mol. Biol. Evol.* 28, 2173–2183.
- Moriyama, E.N., Powell, J.R., 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13, 261–277.
- Pedra, J.H.F., McIntyre, L.M., Scharf, M.E., Pittendrigh, B.R., 2004. Genome-wide transcription profile of field- and laboratory-selected ichlorodiphenyltrichloroethane (DDT)-resistant *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7034–7039.
- Rebollo, R., Romanish, M.T., Mager, D.L., 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet.* 46, 21–42.
- Schlenke, T.A., Begun, D.J., 2004. Strong selective sweep associated with transposon insertion in *D. simulans*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1626–1631.
- Schmid, C.D., Bucher, P., 2010. MER41 repeat sequences contain inducible STAT1 binding sites. *PLoS One* 5, e11425.
- Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A., Ast, G., 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals *Alu*'s unique role in shaping the human transcriptome. *Genome Biol.* 8, R127.
- Simons, C., Pheasant, M., Makunin, I.V., Mattick, J.S., 2006. Transposon-free regions in mammalian genomes. *Genome Res.* 16, 164–172.
- Thornburg, B.G., Gotea, V., Makalowski, W., 2006. Transposable elements as a significant source of transcription regulating signals. *Gene* 365, 104–110.
- Tijet, N., Helvig, C., Feyereisen, R., 2001. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene* 262, 189–198.
- van de Lagemaat, L.M., Landry, J.R., Mager, D.L., Medstrand, P., 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* 19, 530–536.
- Vieira, C., Lepetit, D., Dumont, S., Biémont, C., 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* 16, 1251–1255.
- Wagner, G.P., Amemiya, C., Ruddle, F., 2003. Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14603–14606.
- Wang, T., et al., 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18613–18618.
- Wang, J., Bowen, N.J., Mariño-Ramírez, L., Jordan, I.K., 2009. A c-Myc regulatory subnetwork from human transposable element sequences. *Mol. Biosyst.* 5, 1831–1839.
- Windelspecht, M., Richmond, R.C., Cochrane, B.J., 1995. Malathion resistance levels in sympatric populations of *Drosophila simulans* (Diptera, Drosophilidae) and *Drosophila melanogaster* differ by 2 orders of magnitude. *J. Econ. Entomol.* 88, 1138–1143.
- Wray, G.A., et al., 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419.
- Yang, H.P., Barbash, D.A., 2008. Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. *Genome Biol.* 9, R39.

## Internet resources

Consite service: <http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consiteCyp> genes: [http://flybase.org/cgi-bin/quicksearch\\_soler.cgi?caller=quicksearch&tab=basic\\_tab&context=cytochrome%20P450%20&data\\_class=FBgn&species=Dmel&search\\_type=allGBrowse](http://flybase.org/cgi-bin/quicksearch_soler.cgi?caller=quicksearch&tab=basic_tab&context=cytochrome%20P450%20&data_class=FBgn&species=Dmel&search_type=allGBrowse): <http://flybase.org/cgi-bin/gbrowse2/Gene> localization: ([http://flybase.org/static\\_pages/docs/datafiles.html#gene\\_map\\_table](http://flybase.org/static_pages/docs/datafiles.html#gene_map_table)) *Genomes (FTP)* // *Drosophila melanogaster*: [ftp://ftp.flybase.net/genomes/Drosophila\\_melanogaster/Genomes](ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/Genomes) (*FTP*) // *Drosophila simulans*: [ftp://ftp.flybase.net/genomes/Drosophila\\_simulans/JASPAR](ftp://ftp.flybase.net/genomes/Drosophila_simulans/JASPAR) server: [http://asp.ii.uib.no:8090/cgi-bin/jaspar2010/jaspar\\_db.pl](http://asp.ii.uib.no:8090/cgi-bin/jaspar2010/jaspar_db.pl) ORF Finder: (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) CD-Search Tool (Conserved Domains): <http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi> Orthologs: [http://flybase.org/static\\_pages/docs/datafiles.html#gene\\_orthologs](http://flybase.org/static_pages/docs/datafiles.html#gene_orthologs) Repbase: <http://www.girinst.org/repbase/RepeatMasker>: <http://www.repeatmasker.org/>