

***Helitrons* shaping the genomic architecture of *Drosophila*: enrichment of *DINE-TR1* in α - and β -heterochromatin, satellite DNA emergence, and piRNA expression**

Guilherme B. Dias · Pedro Heringer ·
Marta Svartman · Gustavo C. S. Kuhn

Published online: 26 September 2015
© Springer Science+Business Media Dordrecht 2015

Abstract *Drosophila* INterspersed Elements (*DINEs*) constitute an abundant but poorly understood group of *Helitrons* present in several *Drosophila* species. The general structure of *DINEs* includes two conserved blocks that may or not contain a region with tandem repeats in between. These central tandem repeats (CTRs) are similar within species but highly divergent between species. It has been assumed that CTRs have independent origins. Herein, we identify a subset of *DINEs*, termed *DINE-TR1*, which contain homologous CTRs of approximately 150 bp. We found *DINE-TR1* in the sequenced genomes of several *Drosophila* species and in *Bactrocera tryoni* (Acalyptera, Diptera). However, interspecific high sequence identity (~88 %) is limited to the first ~30 bp of each tandem repeat, implying that evolutionary constraints operate differently over the monomer length. *DINE-TR1* is unevenly distributed across the *Drosophila* phylogeny. Nevertheless, sequence analysis suggests vertical transmission. We found that CTRs within *DINE-*

TR1 have independently expanded into satellite DNA-like arrays at least twice within *Drosophila*. By analyzing the genome of *Drosophila virilis* and *Drosophila americana*, we show that *DINE-TR1* is highly abundant in pericentromeric heterochromatin boundaries, some telomeric regions and in the Y chromosome. It is also present in the centromeric region of one autosome from *D. virilis* and dispersed throughout several euchromatic sites in both species. We further found that *DINE-TR1* is abundant at piRNA clusters, and small *DINE-TR1*-derived RNA transcripts (~25 nt) are predominantly expressed in the testes and the ovaries, suggesting active targeting by the piRNA machinery. These features suggest potential piRNA-mediated regulatory roles for *DINEs* at local and genome-wide scales in *Drosophila*.

Keywords *Drosophila virilis* · *Drosophila americana* · Transposable element · Genome evolution · *Helitrons*

Abbreviations

AIC	Akaike Information Criterion
CTRs	Central tandem repeats
DINEs	<i>Drosophila</i> INterspersed Elements
FISH	Fluorescence in situ hybridization
IR	Inverted repeat
miRNA	microRNA
MSL	Male-specific lethal complex
piRNA	Piwi-interacting RNA
satDNA	Satellite DNA
siRNA	Small interfering RNA

Responsible editor: Maria Assunta Biscotti, Pat Heslop-Harrison and Ettore Olmo

Electronic supplementary material The online version of this article (doi:10.1007/s10577-015-9480-x) contains supplementary material, which is available to authorized users.

G. B. Dias · P. Heringer · M. Svartman · G. C. S. Kuhn (✉)
Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
e-mail: gcskuhn@ufmg.br

SPR	Subtree pruning and regrafting algorithm
subTIRs	Subterminal inverted repeats
TE	Transposable element
TR	Tandem repeat

Introduction

Satellite DNAs (satDNA) and transposable elements (TEs) are the main components of heterochromatin and important contributors to the genome architecture and evolution of eukaryotes (Heslop-Harrison and Schwarzacher 2011; Wallrath et al. 2014). These repetitive sequences participate in several biological processes, such as the formation and propagation of heterochromatin (Volpe et al. 2002; de Wit et al. 2005; Vermaak and Malik 2009), centromere maintenance and function (Malik and Henikoff 2009; Ugarkovic 2009; Vermaak and Malik 2009; Plohl et al. 2014; Rošić et al. 2014), and gene expression regulation (Volpe et al. 2002; Menon et al. 2014). They also contribute to the evolution of adaptive traits and the establishment of reproductive barriers between species (Gregory and Johnston 2008; Ferree and Barbash 2009; Brown and O'Neill 2010; Feliciello et al. 2015). There is growing evidence revealing evolutionary connections between TEs and satDNAs in many organisms. In this context, TEs appear to be an important source for satDNA origin, either by creating tandem repeats by ectopic recombination or by the amplification of preexisting internal repeat motifs (e.g., Gaffney et al. 2003; Macas et al. 2009; Brajkovic et al. 2012; Satovic and Plohl 2013; Dias et al. 2014).

Regardless of their functional roles, the indiscriminate spread of repetitive sequences is usually selected against because of the potential deleterious consequences caused by ectopic recombination. The strength of the purifying selection varies according to recombination rates, copy numbers, and element lengths (Petrov et al. 2011). At the molecular level, one of the main mechanisms of defense against transposition events in germ line cells is the expression of piRNAs, which interact with proteins from the Piwi clade to specifically target and silence TEs and other repeats (Kalmykova et al. 2005; Saito et al. 2006; Brennecke et al. 2007; Aravin et al. 2007; Grimson et al. 2008). The majority of piRNAs in *Drosophila melanogaster* derives from variable-sized clusters that lack protein-coding genes and are replete with eroded remnants of ancient TE

insertions and other repeats (Aravin et al. 2007; Brennecke et al. 2007; Jordan and Miller 2008).

Several examples show that TE and satDNA abundance relates with genome size in a wide range of organisms (Kidwell 2002; Gregory and Johnston 2008; Bosco et al. 2007). In this respect, *Drosophila virilis* is of particular interest because it has an atypical large genome with 360–440 Mb (almost twice as big as the genome of many other *Drosophila* species, including *D. melanogaster*) and one of the highest contents of heterochromatin within the genus (50 % in contrast to ~30 % in *D. melanogaster*) (Gatti et al. 1976; Bosco et al. 2007). Moreover, the availability of the *D. virilis* sequenced genome (*Drosophila* 12 Genomes Consortium 2007) provides a great opportunity to study the impact of repetitive DNAs in genome architecture and evolution.

TEs account for ~14 % of the *D. virilis* genome (*Drosophila* 12 Genomes Consortium 2007) and *Drosophila* Interspersed Elements (*DINES*; Locke et al. 1999), with over 3000 copies, are among the most abundant ones (Yang and Barbash 2008). *DINES* are elusive dipteran transposons from the *Helitron* group thought to mobilize via a rolling circle mechanism (Kapitonov and Jurka 2001; Kapitonov and Jurka 2007b; Yang and Barbash 2008). Their general structure includes subterminal inverted repeats (subTIRs), a short inverted repeat (IR), a core region conserved between elements from distantly related species, a microsatellite region of variable size, a central region with tandem repeats (CTRs), and a stem loop at the 3' end (Fig. 1a; Yang and Barbash 2008; Thomas et al. 2014).

DINES are abundant at heterochromatic and euchromatic regions, including several insertions in or around genes (Yang and Barbash 2008). In *D. melanogaster* and *Drosophila simulans*, *DINES* are frequently found inserted near cytochrome P450 genes related to insecticide resistance, which could indicate a possible regulatory role (Carareto et al. 2014). In *Drosophila miranda*, dosage compensation of the neo-X chromosome was achieved through the co-optation of *DINE*-related *Helitrons* that recruit the male-specific lethal (MSL) complex (Ellison and Bachtrog 2013). A large survey of *DINE* elements in 12 sequenced *Drosophila* genomes revealed that the evolution of these elements is highly dynamic and include recent transpositional bursts in several species. The *DINE*-CTRs were also found to be similar within species but very divergent between species and it has been assumed that they have

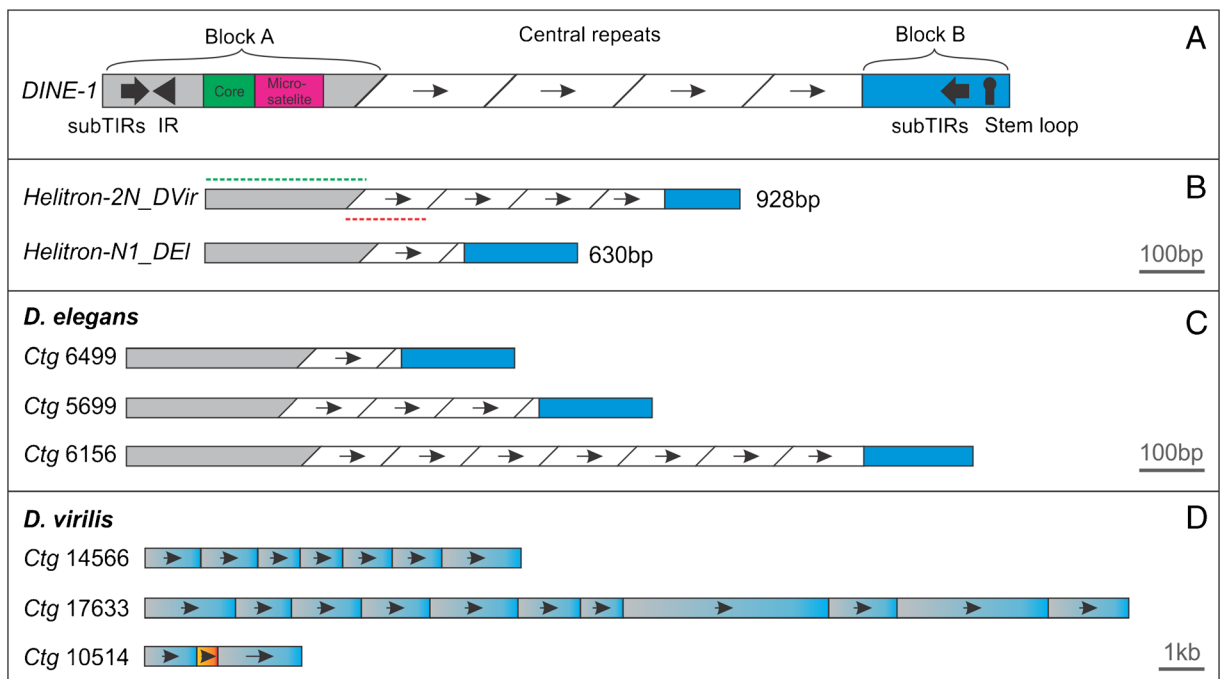


Fig. 1 **a** General organization of *DINE-1* elements. Redrawn from Thomas et al. (2014). **b** *Helitron* consensus identified as *DINE-TR1* from *D. virilis* and *D. elegans*. Dashed lines above block A and below the central repeats indicate segments used as probes in the

FISH experiments. **c** Representation of typical *DINE-TR1* elements found in *D. elegans* contigs. **d** Representatives of *D. virilis* *DINE-TR1* tandem insertions including a *DINE* from a distinct group (*Helitron-1*, orange box) in between two *DINE-TR1*s (*Helitron-2*)

independent origins in the 12 analyzed *Drosophila* species (Yang and Barbash 2008). Despite their perceivable importance in shaping the *Drosophila* genome, our knowledge about *DINEs* is still very limited.

SatDNAs account for ~45 % of the *D. virilis* genome (Bosco et al. 2007) and most of the heterochromatin present in this species consists of three abundant homologous satDNAs (satellites I, II, and III) displaying heptanucleotide repeat units. Despite the fact that these three satellites account for ~40 % of the *D. virilis* genome (Gall et al. 1971), a recent bioinformatic survey in several eukaryotic sequenced genomes identified a 150-bp sequence as the most abundant tandem repeat (TR) of *D. virilis*, potentially corresponding to the DNA underlying the centromeres (Melters et al. 2013). Abdurashitov et al. (2013), using in silico and in vitro DNA digestion, independently identified the same 150-bp repeat as part of a *Helitron* called *Helitron-2_DVir*.

In the present work, we aimed to investigate the association between the 150-bp TR and *Helitron-2_DVir* and to determine their distribution, organization, and impact in the *Drosophila* genome. We found that the *Helitron-2_DVir* containing 150-bp repeats is part of a subgroup of *DINEs* that we called *DINE-TR1*. In

contrast to what was previously assumed for *DINEs*, the CTRs from *DINE-TR1* share homology among several species. Our study revealed that *DINE-TR1* is restricted to Acalypratae (Diptera) but display a patchy distribution within the *Drosophila* genus. After analyzing the chromosomes of *D. virilis* and its closely related species *Drosophila americana*, we found that *DINE-TR1* is highly abundant at several genomic regions. Analysis of the *D. virilis* small RNA profile pointed to the involvement of *DINE-TR1* in piRNA expression. We discuss how our findings shed light on the role played by *DINEs* in several aspects of *Drosophila* genome architecture and evolution.

Material and methods

Identification and phylogenetic distribution of *DINE-TR1*

DINE-TR1 was initially identified in *D. virilis* and *Drosophila elegans* after sequence comparisons between the most abundant TRs identified in 21 *Drosophila* species with sequenced genomes reported

by Melters et al. (2013). In order to identify *DINE-TR1* related elements in other *Drosophila* species, we performed a series of searches. First, we used the Tandem Repeats Finder software (Benson 1999) to look for CTRs in all *Helitrons* from *Drosophila* available in Repbase (Jurka et al. 2005). In addition, dot plots were used for visualization of the organization and size of the arrays (Junier and Pagni 2000). Next, we compared these elements with *Helitron-2_DVir* (a *DINE-TR1* from *D. virilis*) through dot plots in order to identify similarity between the CTRs from *DINE-TR1*.

The remaining *Drosophila* species with available sequenced genomes but without any *Helitron* consensus available at Repbase were queried using the *DINE-TR1* consensus from the available closest species. Besides the sequenced genomes available at Flybase (<http://flybase.org>), we searched for *DINE-TR1* in other species with recently sequenced genomes, including *D. americana* (<http://craacs.fc.up.pt/~nf/dame/>), *Drosophila suzukii* (available in NCBI), and *Drosophila buzzatii* (<http://dbuz.uab.cat/>) (Fonseca et al. 2013; Ometto et al. 2013; Guillén et al. 2014).

To assess the phylogenetic distribution of *DINE-TR1*, we also performed searches in the sequenced genomes of other Diptera, including *Bactrocera tryoni*, *Lucilia cuprina*, *Musca domestica*, and *Glossina morsitans*, all available at NCBI (Gilchrist et al. 2014; Scott et al. 2014; International *Glossina* Genome Initiative 2014).

Multiple sequence alignments and phylogenetic reconstruction

Multiple sequence alignments (MSAs) were performed using the M-Coffee web-server with the default options (Tommaso et al. 2011) and visualized and edited in Jalview (Waterhouse et al. 2009). Maximum likelihood phylogenies were estimated using PhyML (Guindon and Gascuel 2003) with the best substitution model and parameters according to the Akaike Information Criterion (AIC) as determined by JModelTest version 2.1.4 (Darriba et al. 2012). Trees were reconstructed using the subtree pruning and regrafting (SPR) algorithm, and statistical support was calculated after 1000 bootstrap replicates.

Fluorescence in situ hybridizations

Mitotic metaphases were obtained from the neuroblasts of wandering third instar larvae from *D. virilis* (strain

15010–1051.51) and *D. americana* (strain W11) according to the method described in Baimai (1977). Polytene chromosomes were prepared following the acetic acid squash protocol (Ashburner 1989). DNA fibers were isolated from adult flies as described in Kuhn et al. (2008). Specific probes were obtained from *D. virilis* by PCR with primers for *DINE-TR1* block A (forward 5' TTATACCCTTGCAGAGGG 3', reverse 5' GCTGGTTTTTCACATATGTGC 3') and for its CTRs (forward 5' CCATAGGAACGATCGGTCCG 3', reverse 5' CAGCTATATGATATAGTGGTCCG 3'). We cloned PCR products of 240 bp for the block A and of 450 and 600 bp for the CTRs representing 3 and 4 monomers, respectively. These fragments were cloned in the pGEM-T vector (Promega) and sequenced to confirm insert specificity. Recombinant plasmids were labeled with digoxigenin 11-dUTP or biotin 11-dUTP by nick translation (Roche Applied Science). Fluorescence in situ hybridizations (FISH) on chromosomes and extended DNA fibers were performed as described in Kuhn et al. (2008). Briefly, denaturation of metaphase and polytene chromosomes was carried out in 0.07 M NaOH for 3 min and 100–200 ng of each probe were hybridized to the chromosomes for 16–20 h at 37 °C in a moist chamber. Slides were washed twice in 2× SSC at 37 °C for 5 min. DNA fibers were denatured in 70 % formamide/2× SSC at 80 °C. The slides were analyzed under an Axio Imager A2 epifluorescence microscope equipped with the AxioCam MRm camera (Zeiss). Images were captured with AxioVision (Zeiss) software and edited in Adobe Photoshop.

RNA-Seq analysis and identification of *DINE-TR1* at piRNA clusters

We analyzed publicly available small RNA datasets from the Short Read Archive (SRA) under BioProject GSE22067 produced by Rozhkov et al. (2010). Quality checks and filtering were performed using the FASTX toolkit (Gordon and Hannon 2010) implemented in a local Galaxy instance in BioLinux (Field et al. 2006; Goecks et al. 2010). Short-read mapping against the *Helitron-2_DVir* consensus was performed with Lastz (Harris 2007) also implemented on a local Galaxy instance and excluded reads that mapped with less than 85 % identity. Read counts were normalized to one million reads after filtering for abundant degradation transcripts, such as those from tRNAs and rRNAs.

We repeat masked the 20 genomic regions defined as piRNA clusters by Rozhkov et al. (2010) using the CENSOR tool and the *Drosophila* repeat library in Repbase (Kohany et al. 2006). Then, we calculated the contribution of *DINE-TR1* and *Helitrons* in general to each cluster size.

Results

Identification and characterization of *DINE-TR1* in *Drosophila* and other Diptera

After aligning the single most abundant TRs identified in Melters et al. (2013) from the sequenced genomes of 21 *Drosophila* species, we realized that in *D. virilis* and *D. elegans*, the most abundant TR had a similar size of about 150 bp with high sequence similarity (89 %) over a segment of 46 bp. The same 150-bp TR has been found as making short arrays within the *Helitron-2_DVir* of *D. virilis* (Fig. 1b; Abdurashitov et al. 2013). Accordingly, after repeat-masking the TRs in Repbase, we found that the most abundant TR from *D. elegans* also exists as part of a *Helitron* (*Helitron-NI_DEI*; Fig. 1b; Kapitonov and Jurka 2007a).

These two *Helitrons* from *D. virilis* and *D. elegans* belong to the *DINE* group of TEs (Fig. 1a; Kapitonov and Jurka 2007a; Yang and Barbash 2008). We then used the software Tandem Repeats Finder (Benson 1999) to search the Repbase *Helitron* library for elements from other *Drosophila* species that also contained internal TRs similar to those of *D. virilis*. We found that *DINEs* may or may not harbor CTRs and that different and probably unrelated groups of CTRs could be defined based on sequence similarity (data not shown). We named the specific elements with homologous ~150 bp CTRs discussed in this work *DINE-TR1*.

We next surveyed the sequenced genomes of 25 *Drosophila* species with their own *DINE-TR1* consensus from Repbase or with the consensus from the closest related species. The data on the presence or absence of *DINE-TR1* were plotted onto a *Drosophila* plus outgroup phylogeny and showed a discontinuous distribution of *DINE-TR1* across the sampled species (Fig. 2). Entire species subgroups such as *D. melanogaster* and *Drosophila pseudoobscura* appeared to be devoid of *DINE-TR1*, whereas in other lineages (e.g., the *virilis-*

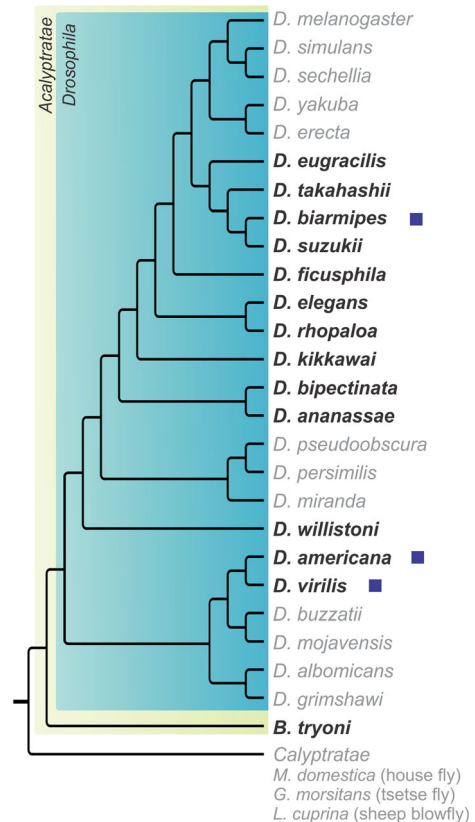


Fig. 2 Phylogeny of Schizophora (Diptera) with representative sequenced species (*Drosophila* phylogenetic relationships were based on Markow 2015 whereas other Diptera species were placed in the tree according to the NCBI Taxonomy Browser classification). Species names in **bold** indicate the presence of *DINE-TR1* based on genomic analyses. **Blue squares** indicate species where *DINE-TR1* CTRs have expanded into abundant satDNA-like arrays

repleta radiation), *DINE-TR1* presence was patchy (Fig. 2).

A BLAST search in the nr/nt (nonredundant nucleotide collection) with the *Helitron-2N_DVir* as query and excluding *Drosophila* revealed the presence of *DINE*-like elements in species from several genera within Schizophora (Diptera), among them *Bactrocera*, *Musca*, and *Stomoxys* (Suppl. Table 1). However, in most cases, similarity was restricted to the core region from block A (Fig. 1a). Because there is a limited and uneven availability of sequences from different species in the nr/nt database (which is strongly focused in euchromatic/coding sequences), we advanced our search by using only species with complete sequenced genomes (see **Methods**). This approach

allowed a more comprehensive search and manual check of the retrieved sequences.

BLAST searches using the *D. virilis* *DINE-TR1* consensus as a query retrieved several *hits* in all species surveyed. Nevertheless, manual verification of the contigs revealed that only in *Drosophila* and *Bactrocera* the similarity extends over the core region of block A and includes part of the CTRs, indicating that *DINE-TR1* might be restricted to Acalypratae (Fig. 2). All the three Calypratae genomes analyzed seem to lack *DINE-TR1* although still possessing other *DINE*-related sequences (Fig. 2; Suppl. Table 1).

To gain insight regarding the discontinuous distribution of *DINE-TR1* inside *Drosophila*, we reconstructed a maximum likelihood phylogeny using the entire block A followed by the first CTR from all surveyed species. Although some nodes showed low bootstrap support, the resulting tree topology showed no major incongruence relative to the species tree (Suppl. Fig. 1). When only the CTRs are used for tree reconstruction, the resulting groupings have significantly lower bootstrap support and do not reflect the species relationships (data not shown). This is probably because most of the CTR sequence is not conserved, resulting in spurious alignment.

Although the CTRs from *DINE-TR1* have preserved an approximate length of 150 bp in all surveyed species, including in the distantly related *B. tryoni*, high sequence identity among all species was found only in the first ~30 bp of each CTR monomer (88 % on average; Fig. 3).

DINE-TR1 with expanded CTRs

Sequence analysis of several contigs of *D. virilis* presenting 150 bp TRs revealed that aside from their organization as short- to medium-sized arrays inside *DINE-TR1*, these repeats also form very large arrays that cover several Kb (Suppl. Table 3). For example, contigs 0 and 6695 are covered by 150-bp repeats forming arrays of 4376 and 5614 bp, respectively (Suppl. Table 3). Fluorescence in situ hybridization (FISH) to extended DNA fibers (fiber FISH) from *D. virilis* using *DINE-TR1* probes specific for the 150-bp TRs and for block A showed intense clustering of *DINE* insertions in some fibers and a marked overabundance of 150-bp TRs in many others (Fig. 4). Altogether, sequence analysis and fiber FISH results suggest that *DINE-TR1* internal CTRs

have undergone amplification generating satDNA-like arrays in *D. virilis*.

The analysis of *DINE-TR1* CTRs in the contigs of *D. elegans* revealed only short arrays (up to six full copies) confined within the *DINE-TR1* structure (Fig. 1c). This result indicates that 150-bp TRs are abundant in the genomes of *D. elegans* due to the high copy number of *DINE-TR1*. Accordingly, we verified that in *D. elegans*, all hits from the block A of *DINE* are highly similar to the Repbase consensus (*Helitron-N1_DEL*) for this species (average similarity 99 %), possibly indicating a recent transpositional burst.

In order to investigate the amplification status of 150-bp CTRs from *DINE-TR1* in other *Drosophila* species, we isolated the typical CTRs from each species and constructed an artificial array with ten monomers. We then used this array as a query in BLAST searches against the sequenced genome of each species. We found that *D. americana* and *Drosophila biarmipes* also displayed the same satDNA-like arrays of *DINE-TR1* CTRs (Fig. 2; Suppl. Table 3). We also checked the expansion of 150-bp CTRs in *D. americana* using fiber FISH and found a very similar pattern as that described for *D. virilis* (Fig. 4). While *D. americana* is a closely related species to *D. virilis* (belonging to the *virilis* subgroup), *D. biarmipes* is a more distantly related species (*melanogaster* group) (Fig. 2). We further noticed, through in silico analysis, that expanded CTR arrays also exist in other *Drosophila* species, albeit much less abundantly. This was the case for *D. sukii*, *Drosophila bipectinata*, and *Bactrocera tryoni* (Fig. 2).

Besides the amplification suffered by the CTRs of *DINE-TR1*, we also found entire *DINE-TR1* elements arranged in tandem (Fig. 1d). From 80 analyzed cases of nearby insertions, more than half (48) were found separated by 60 bp or less, and 37 were less than 10 bp apart. We found up to 11 *DINE-TR1* tandem repeats (ctg17633; Fig. 1d); to our knowledge, the largest number of *Helitron* tandem insertions detected to date. Additionally, we found distinct *Helitrons* intermingled in tandem (ctg10514, Fig. 1d). This type of insertion was previously reported in maize, and, to our knowledge, this is the first reported case in animals (Du et al. 2008).

DINE-TR1 distribution in metaphase and polytene chromosomes of *D. virilis* and *D. americana*

The *D. virilis* karyotype is composed of six acrocentric chromosome pairs ($2n=12$), with large heterochromatic blocks extending from the centromeres of chromosomes

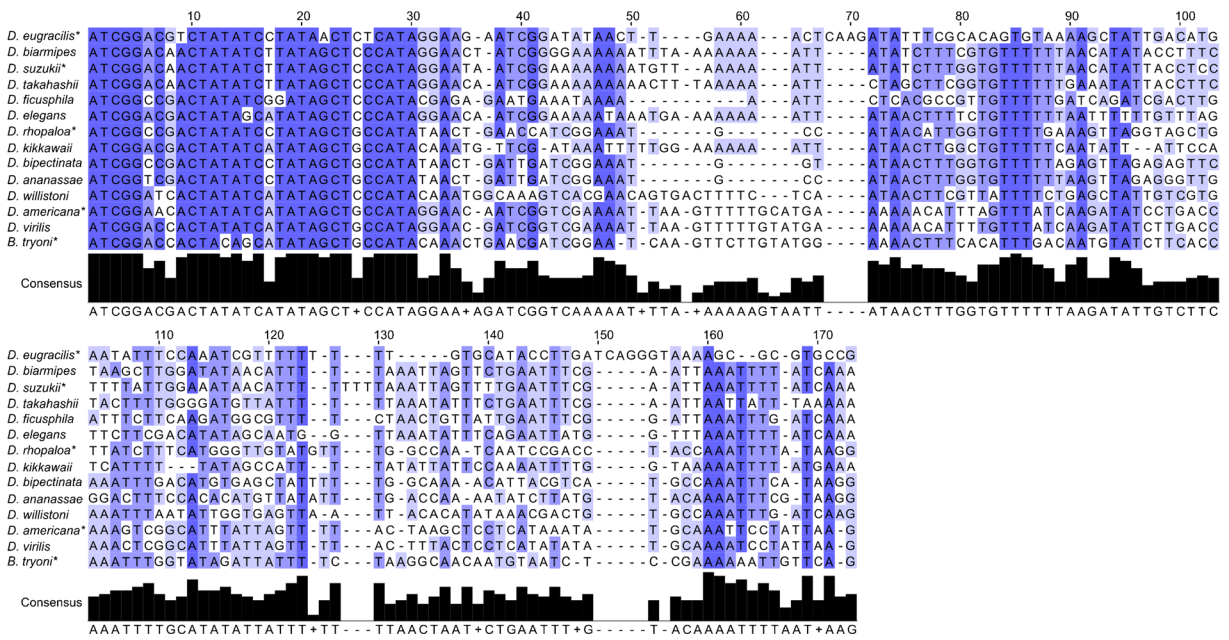


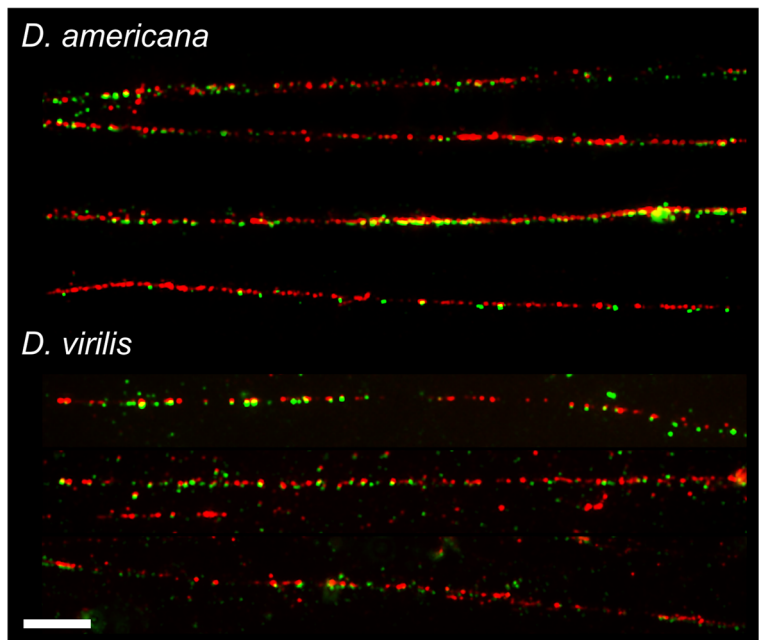
Fig. 3 Multiple sequence alignment (MSA) of the 150-bp CTRs from *DINE-TR1* in several *Drosophila* species and *B. tryoni*. These sequences represent the Repbase *Helitron* consensus for each species. The species with no consensus available are marked with an asterisk and feature the best BLAST hit with the closest species

consensus as query. The conservation histogram and consensus sequence for the entire alignment are included below the MSA. Dashes represent gaps and the plus symbol represents positions where no majority consensus was reached

2, 3, 4, 5, and X and occupying about half of each chromosome. The Y chromosome is entirely heterochromatic, and the microchromosome is predominantly euchromatic (Mahan and Beck 1986). *D. americana*

shows a derived karyotype by two centromeric fusions (2;3 and X;4) and a similarly distributed but less abundant heterochromatin compared to *D. virilis* (Mahan and Beck 1986; Caletka and McAllister 2004). In order to

Fig. 4 Fluorescence in situ hybridization of *DINE-TR1* block A (green) and the CTRs (red) onto extended DNA fibers from *D. americana* and *D. virilis*. Bar represents 10 kb assuming 10 μm=29 kb (Schwarzacher and Heslop-Harrison 2000)



assess the chromosome distribution of both *DINE-TR1* block A and its CTRs, we performed dual-color FISH with specific probes onto the *D. virilis* and *D. americana* metaphase and polytene chromosomes.

In *D. virilis*, the hybridizations revealed a marked enrichment of *DINE-TR1* in the boundaries between the pericentromeric heterochromatin and the euchromatin (i.e., β -heterochromatin) of chromosomes 2, 3, 4, 5, and X (Fig. 5a–b). In *D. americana*, we detected signals in similar regions for chromosomes 2, 3, 5, and X, with chromosome 4 only displaying hybridization of the block A probe (Fig. 5c–d). We confirmed that those regions corresponded to the β -heterochromatin through hybridization of the same probes onto polytene chromosomes, which displayed intense co-localizing signals over the entire chromocenter region (Fig. 6). Apart from β -heterochromatin, *DINE-TR1* is abundant in the centromeric region of chromosome 5 and present at a discrete site in the α -heterochromatin of the X chromosome in *D. virilis*, but not in *D. americana*. *DINE-TR1* also covers much of the Y chromosome length in both species (Fig. 5). The microchromosomes showed

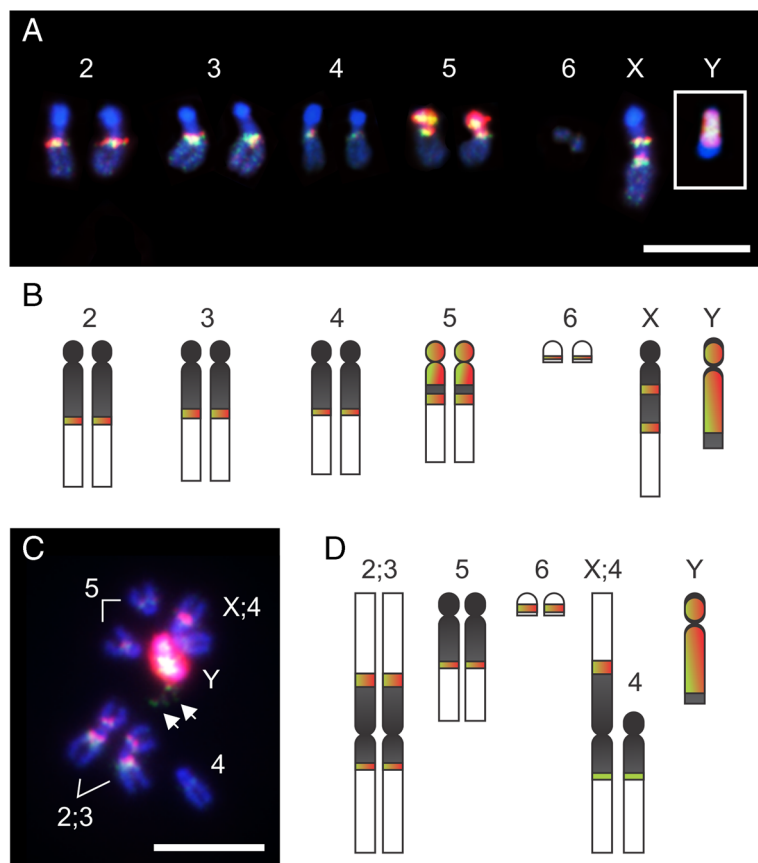
hybridization signals in both metaphases and polytene chromosomes (Figs. 5 and 6).

The hybridization of the probes to the polytene chromosomes evidenced the dispersion of *DINE-TR1* at numerous euchromatic loci in all polytene arms, including some telomeric regions (Fig. 6). BLAST searches revealed that *DINE-TR1* is located near or within several genes in *D. virilis* (Suppl. Table 2). It is noteworthy that we found *DINE-TR1* associated with many development-related genes, including several Homeobox genes (Suppl. Table 2).

DINE-TR1-derived small RNAs are abundant in the gonadal tissues of *D. virilis*

The production of piRNAs is thought to occur from clusters of repetitive DNA located at chromatin boundaries (Brennecke et al. 2007). The overall abundance and enrichment of *DINE-TR1* in the β -heterochromatin suggests its possible participation in the piRNA biogenesis. We addressed this issue by mapping public available short-read RNA sequencing data from *D. virilis*

Fig. 5 Fluorescence in situ hybridization (FISH) of *DINE-TR1* block-A (green) and 150-bp CTRs (red) onto the metaphase chromosomes of **a** *D. virilis* and **c** *D. americana*. Idiograms of the metaphases and FISH signals are depicted in **b** and **d** with colocalization of the probes represented as a red/green mixed pattern. Black-colored regions represent the constitutive heterochromatin visualized by C-banding (Mahan & Beck 1986). Bars represent 5 μ m



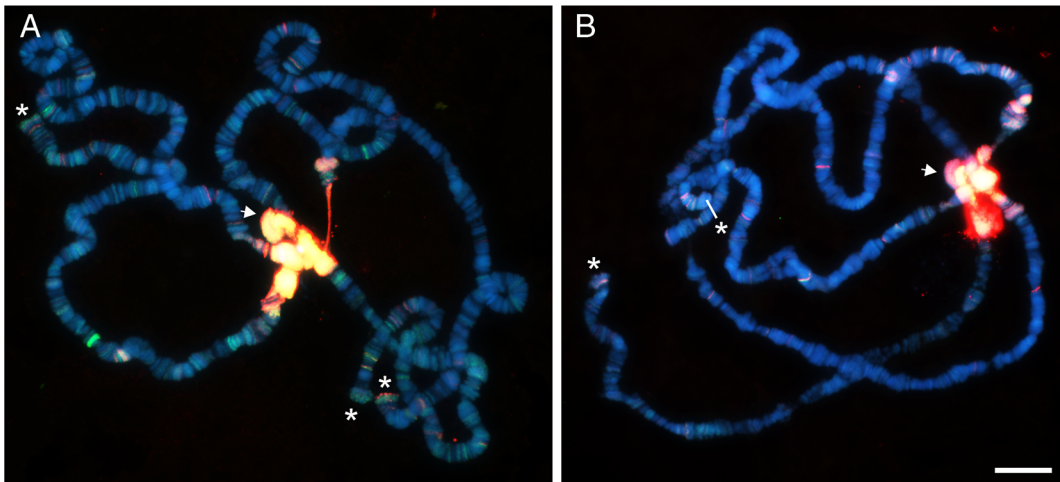


Fig. 6 FISH onto the polytene chromosomes of **a** *D. virilis* and **b** *D. americana* using *DINE-TR1* probes for block A (green) and 150-bp CTRs (red). The chromocenter shows intense hybridization

and the small dot chromosome arm is indicated with an arrowhead. Telomeres with hybridization signals are indicated with an asterisk. The bar represents 10 μm

(strain 160; Rozhkov et al. 2010) to the *Helitron-2_DVir* consensus sequence. The *Helitron-2_DVir* has been specifically chosen because it represents the full-length *Helitron* from the *DINE-TR1* group.

Read counts were calculated for 0- to 2-h embryos and for gonads and carcasses of adult males and females. The results revealed that *Helitron-2_DVir* is almost entirely transcribed, including the CTR region and 5' end (block A). *Helitron-2* small RNA transcripts are relatively abundant in the gonadal tissues from both males and females and almost absent in adult carcasses (Fig. 7a). Interestingly, *Helitron-2_DVir* displays an intermediate transcription level in early embryos (0–2 h) relative to adult gonads (Fig. 7a). The mapped reads from embryos and gonads exhibit a medium size of 25,4 nucleotides (nt) while the few mapped reads from adult carcasses display a medium size of 21,5 nt (Fig. 7b). The mapped reads from embryos and ovaries did not show strand bias, whereas reads from male carcass and gonads, and, to a lesser degree, female carcass, showed abundance of sense over anti-sense transcripts (Fig. 7c).

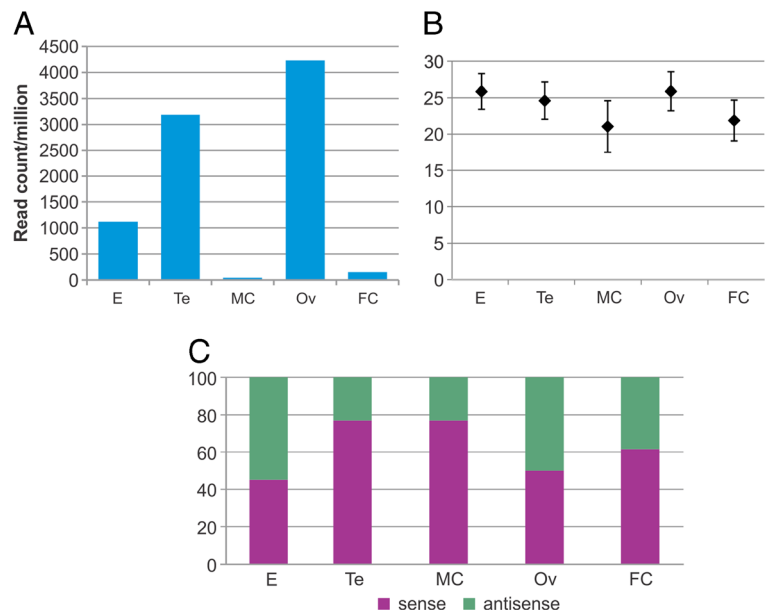
As a second line of investigation, we repeat-masked the genomic regions defined as piRNA clusters by Rozhkov et al. (2010) using the CENSOR tool (Kohany et al. 2006). We found that *DINE-TR1* can be detected in 17 clusters (out of 20) where its abundance range from 0.4 to 9.2 % of the total cluster sizes (Suppl. Table 4). Overall, *Helitrons* are the most abundant DNA transposons in these clusters, spanning from 0.5 to 12.1 % of the total cluster length (Suppl. Table 4).

Discussion

DINE-TR1 is an ancient group of *Helitrons* from Acalyptatae, Diptera

Helitrons are a poorly understood group of DNA transposons that do not possess typical terminal inverted repeats (TIRs) and are thought to transpose via a rolling circle mechanism (Kapitonov and Jurka 2007b). An interesting and widespread group of *Helitrons* from *Drosophila* is the *DINE-1* elements (Locke et al. 1999). Yang and Barbash (2008) observed that the CTRs of *DINEs* from the sequenced genomes of 12 *Drosophila* species are very variable and do not share interspecies homology except for very closely related species, such as in the ones from the *D. melanogaster* subgroup. Nevertheless, we observed sequence similarity between the CTRs from *DINEs* present in *D. virilis* and *D. elegans*, which indicated that even distant *DINE* elements may have homologous CTRs. This new finding prompted us to analyze this *DINE* group and its CTRs in more detail. Our analyses evidenced the existence of a subset of *DINE* elements (*DINE-TR1*) in Acalyptatae (Diptera). We found *DINE-TR1* in the sequenced genomes of 13 *Drosophila* species (out of 25) and in the Queensland fruit fly *B. tryoni* but not in outgroup species from Calyptatae. This result suggests that *DINE-TR1* was already present in the common ancestor of Acalyptatae, some 72 mya (Gaunt et al. 2002). Interestingly, *DINE-TR1* distribution inside *Drosophila* is patchy (Fig. 2).

Fig. 7 Characteristics of small RNAs derived from the *DINE-TRI* elements in *D. virilis*. **a** Read counts for small RNAs derived from *DINE-TRI* in tissues of *D. virilis* strain 160. Counts were normalized to one million reads. **b** Medium size and standard deviation of the small RNAs mapped to *DINE-TRI*. **c** Strand bias of *DINE-TRI* transcription. *E* 0–2 h embryos, *Te* testes, *MC* male carcass, *Ov* ovaries, and *FC* female carcass



The discontinuous distribution of TEs through phylogenies is often explained by means of horizontal transfer (Loreto et al. 2008). However, the phylogenetic reconstruction using *DINE-TRI* sequences did not indicate any major incongruence when compared to the established phylogenetic relationships between these species. In the light of this finding, the patchy distribution of *DINE-TRI* in *Drosophila* could be due to the repeated lineage-specific loss of this element. In fact, Petrov and Hartl (1998) verified that DNA loss is very frequent in *Drosophila*, occurring at 60 times higher estimated rate than in mammals. This could account for *DINE-TRI* loss especially at an evolutionary time point where its genomic abundance was still low. Alternatively, rapid CTR sequence divergence could also prevent the identification of *DINE-TRI* in some species. In fact, only a small segment of CTRs (~30bp) is conserved between distant species. This indicates an ancient origin for the CTRs of *DINE-TRI* but different evolutionary constraints operating over the monomer. In this context, it is interesting to mention that conserved noncoding blocks (intergenic and intronic) in *Drosophila* are usually small (19 bp on average) and some of them may act as *cis*-regulatory elements (Bergman and Kreitman 2001).

DINE-TRI as a potential source for satDNA emergence

Our results showed an expansion of the *DINE-TRI* CTRs in *D. virilis*, *D. americana*, and *D. biarmipes*,

generating satDNA-like arrays. Because *D. virilis* and *D. americana* are both members of the *virilis* species subgroup and share a recent common ancestor dated at ~4 mya (Morales-Hojas et al. 2011), CTR amplification most probably started before the cladogenesis event that separated these two species. On the other hand, the ancient divergence time between *D. biarmipes* and *D. virilis*, at ~62 mya (Tamura et al. 2004), and the lack of a similar pattern of amplification in the other *Drosophila* species (Fig. 2) suggest two independent events of satDNA emergence within *DINE-TRI* in these two lineages.

There is growing evidence of the participation of TEs in the formation of satDNAs, including two reported cases in *D. virilis*. Heikkinen et al. (1995) showed that the pvB370 satDNA share sequence similarity with a TE called pDv and more recently, Dias et al. (2014) showed that a foldback DNA transposon called *Tetris* was involved in the generation of satDNA-like arrays (TIR-220). Herein, we report the participation of a *Helitron* (*DINE-TRI*) in the origin of satDNAs in three *Drosophila* species. To our knowledge, this is the first account on the emergence of satDNAs from preexisting CTRs inside *Helitrons* and also the first report showing the independent emergence of satDNA from the same TE in eukaryotes.

DINEs may be involved in the generation of satDNAs through different mechanisms. For example, the *DINE*-related *SGM* sequences generated a major

satDNA in *Drosophila guanche* (Miller et al. 2000). However, in this case, most of the element's length became tandemly repeated, similar to what happened to the LINE-1 derived centromeric satDNA of cetaceans (Kapitonov et al. 1998).

Both the microsatellite and CTR regions of *DINEs* (Fig. 1a) display copy number variation between different insertions as well as between species (Yang and Barbash 2008). In the microsatellite region, slippage replication may be an important mechanism that promotes array size variation (Charlesworth et al. 1994). For the CTRs, copy number variation is possibly related to nonreciprocal DNA exchanges, such as those promoted by unequal crossing over (Charlesworth et al. 1994). Recent studies discuss copy number variation of TE-associated tandem repeats toward satDNA emergence (Satovic and Plohl 2013). Scalvenzi and Pollet (2014) proposed a model for the evolution of TE-derived satDNAs as the interplay between high recombination rates and the replicative capabilities of the TEs themselves, resulting in genome expansion. Although these copy number variation mechanisms are ubiquitous, only in *D. virilis*/*D. americana* and *D. biarmipes* CTRs expanded leading to the formation satDNA-like arrays. In other species (e.g., *D. sukuzii*, *D. bipectinata*, and *B. tryoni*), the amplification of CTRs seems to be at an earlier stage or other factors may have been operating to prevent their expansion into large arrays. In any case, the current status of CTR array length must result from the balance between size expansion and reduction mechanisms.

It is important to account for possible assemble errors. In the case of tandem repeats, those errors are generally related to the collapsing of similar reads in the same contig thus shrinking the true array size. In this sense, TE-mediated satDNA emergence could be more frequent than what is currently detectable from NGS genome assemblies.

DINE-TRI is abundant at chromatin transition zones

In *D. virilis* and *D. americana*, *DINE-TRI* is particularly enriched at transitional β -heterochromatin regions (Figs. 5 and 6). Wasserlauf et al. (2015) recently microdissected the *D. virilis* chromocenter region and generated a DNA library for FISH in the polytene chromosomes of both *D. virilis* and *Drosophila kanekoi* (*virilis* group). Interestingly, *D. kanekoi* also showed intense hybridization signals in its β -heterochromatin, suggesting that *DINEs* already could have colonized this region

in the common ancestor of the *virilis* group about 8.9 mya (Morales-Hojas et al. 2011; Wasserlauf et al. 2015). A very similar chromosome distribution was reported for another *DINE*-related element (called *PERI*) present in the *Drosophila buzzatii* species cluster (*repleta* group), that diverged from the *virilis* group more than 20 mya (Kuhn and Heslop-Harrison 2011). These examples may reflect a general feature for *DINEs*.

The β -heterochromatin features both euchromatic and heterochromatic characteristics. It is replicated during polytenization but does not develop into a precise banding pattern, appearing as a loose mass of DNA around the chromocenter (Miklos and Cotsell 1990). This region has been regarded as a “transposon graveyard” because it harbors abundant remnants of ancient TE insertions (Vaury et al. 1989). The clustering of *DINE* insertions at the β -heterochromatin could indicate an insertional preference of these TEs for open chromatin regions and/or a reduced effectiveness of natural selection against the deleterious effects of ectopic recombination upon these sequences (Petrov et al. 2011; see also topic on piRNA clusters below). Additionally, *DINE-TRI* abundance in β -heterochromatin may contribute to define the borders between pericentromeric heterochromatin and euchromatin. In this context, it remains to be investigated whether *DINE-TRI* also act as barrier insulators.

In both *D. americana* and *D. virilis*, we found *DINE-TRI* elements located in the vicinity of the telomeres in some chromosomes (Fig. 6). In *D. melanogaster*, three telomeric-specific non-LTR retroelements, *HeT-A*, *TART* and *TAHRE*, are involved in telomere maintenance (Villasante et al. 2008). Previous studies showed that the telomeres of *D. virilis* contain the pVB370 satDNA, and the *TART* and *HeT-A* retroelements (Biessmann et al. 2000; Casacuberta and Pardue 2003; Villasante et al. 2007). However, the pVB370 satDNA is more likely a telomere-associated sequence (TAS) (Casacuberta and Padue 2003; Villasante et al. 2007). *DINE-TRI* could be another TAS in at least some *D. virilis* and *D. americana* chromosomes, defining the borders between euchromatin and telomeric regions.

Locke et al. (1999) used a probe containing the entire *D. melanogaster DINE-1* sequence to assess its distribution in the polytene chromosomes of *D. melanogaster*, *D. simulans*, and *D. virilis*. His results suggest that although being abundant in the dot chromosomes from both *D. melanogaster* and *D. simulans*, *DINE-1* is virtually absent from the *D. virilis* dot. Nevertheless, we did observe

hybridization of *DINE-TR1* over the dot in *D. virilis* and *D. americana* (Figs. 5 and 6). This difference could have been caused by the sequence divergence of the probe derived from *D. melanogaster*. In fact, *DINE* elements are abundant in the dots of several *Drosophila* species including *Drosophila erecta*, *Drosophila mojavensis*, and *Drosophila yakuba* (Leung et al. 2015).

DINE-TR1 in centromeric DNA

Melters et al. (2013) identified 150-bp TRs as the most abundant TR of the *D. virilis* and *D. elegans* genomes and also likely their major centromeric component. Herein, we show that centromeric localization of *DINE-TR1* in *D. virilis* is restricted to chromosomes 5 and Y. This result shows the importance of validating bioinformatic data with cytogenetic tools. In the closely related species *D. americana*, we similarly found *DINE-TR1* covering the Y centromeric region, but not the centromere of chromosome 5. This suggests that either *DINE-TR1* fully colonized the centromere of chromosome 5 only in *D. virilis*, or, less probably, it was completely removed from the homologous region in *D. americana* in the last ~4 my. In any case, this illustrates a high rate of evolutionary change of *DINE-TR1* even between closely related species.

It is also worth mentioning that in *D. melanogaster*, β -heterochromatin has been shown to be a hotspot for neocentromere formation under experimental overexpression of centromeric-specific histone H3 (also known as CID) (Olszak et al. 2011). Therefore, one might speculate that the expansion of tandem repeats and TEs from β -heterochromatin to centromeres may contribute for the high rate of centromeric satDNAs turnover observed in *Drosophila* and in many eukaryotes (reviewed by Plohl et al. 2014).

DINE-TR1 is enriched on the Y chromosome

Initial investigations on the *D. virilis* satDNA content revealed the presence of three abundant simple satellites located in the heterochromatin of all autosomes and the X chromosome, but almost absent in the highly heterochromatic Y (Gall et al. 1971). Our results show that a large segment of the Y chromosome of *D. virilis* and of its sister species *D. americana* is covered with *DINE-TR1* copies (Fig. 5). A similar abundance of the *DINE*-related element *PER1* was also found in the Y chromosome of species from the *D. buzzatii* cluster

(Kuhn and Heslop-Harrison 2011), which may indicate another general feature of *DINEs*. Some studies indicate a clear correlation between sex chromosomes differentiation and repetitive DNA accumulation, a process favored by the absence or low frequency of recombination that is typical of these chromosomes (reviewed in Charlesworth et al. 2005). Accordingly, the colonization and expansion of *DINEs* may have been an important event during the process of Y chromosome differentiation in *Drosophila* species. Furthermore, it may also have affected sex-specific gene expression. For example, differences in heterochromatic blocks harboring TEs and other repetitive sequences seem to have been involved in Y-linked regulatory divergence among *D. melanogaster* populations (Lemos et al. 2008). In that sense, heterochromatic blocks could serve as “chromatin sinks” for the binding of transcription factors or chromatin regulators, depleting or redistributing them throughout the genome (Dimitri and Pisano 1989; reviewed in Francisco and Lemos 2014). Such process is thought to be independent of any specific sequence, being a quantitative phenomenon derived from the amount of heterochromatin (Dimitri and Pisano 1989). Interestingly, Brown and Bachtrog (2014) showed that *Drosophila* males have less repressive chromatin modifications in the assembled portions of the genome, which are mostly euchromatic, probably as a result of the Y-derived genome-wide chromatin regulation.

DINE-TR1-derived piRNAs in *D. virilis*

RNA interference (RNAi), or RNA silencing, is a major genomic regulatory mechanism of eukaryotes that recognizes targets by complementarity with small RNAs from three different classes: siRNAs, miRNAs, and piRNAs. The interaction of piRNAs with proteins from the Piwi clade (PIWI, AUB, and AGO3 in *Drosophila*) is the main genome defense mechanism against transposition events in germ line cells of animals, ensuring stable gametogenesis (Aravin et al. 2007). Nevertheless, this class of small RNAs is the least investigated when compared with siRNAs and miRNAs (reviewed in Ghildiyal and Zamore 2009; Siomi et al. 2011). About 90 % of *Drosophila* piRNAs can be assigned to TEs, satDNAs, and other repetitive sequences (Brennecke et al. 2007; Yin and Lin 2007; Huang et al. 2013).

DINE-1 is the most abundant transposable element in *Drosophila* (Bergman et al. 2006; Yang et al. 2006; Thomas et al. 2014). Despite some investigations on

the piRNA biogenesis in *D. virilis* (Rozhkov et al. 2010, Le Thomas et al. 2014), the involvement of *DINE-1* elements has not been addressed so far.

DINE copies are heavily accumulated at the β -heterochromatin of *D. virilis* and *D. americana* (Figs. 5 and 6). In *D. melanogaster*, the β -heterochromatin is enriched with fragmented and nested TEs (Vaury et al. 1989; Hoskins et al. 2002). In addition, piRNA clusters have also been shown to map to these regions representing chromatin boundaries (Brennecke et al. 2007; Yamanaka et al. 2014).

The piRNA pathway in *D. melanogaster* is mostly active in gonadal tissues (Brennecke et al. 2007; Brower-Toland et al. 2007; Rozhkov et al. 2010; Le Thomas et al. 2014), and piRNA clusters are generally transcribed from both strands, with no pronounced bias (Brennecke et al. 2007; Rozhkov et al. 2010). The piRNAs have a typical size distribution between 23 and 29 nt; with an average of 25.7, 24.7, and 24.1 nt for Piwi, Aub, and Ago3, respectively (Brennecke et al. 2007). We found that small RNA transcripts from *DINE-TR1*, with an average size of 25 nt, are predominantly expressed in *D. virilis* testes and ovaries (Fig. 7). This result strongly points to an active targeting of *DINE-TR1* by the piRNA machinery of *D. virilis*. Interestingly, transcripts from male gonads and carcasses and female carcasses showed strand bias (Fig. 7c). In the case of males, clusters present on the Y chromosome could be skewing piRNA production toward sense strand transcription. Additionally, other classes of small RNAs could be transcribed from *DINE-TR1* in a few loci at low abundances.

When analyzing the genomic regions defined as piRNA clusters by Rozhkov et al. (2010), we found that *DINE-TR1* is present in most of the clusters (Suppl. Table 4). Because there is more than 80 kb of *DINE-TR1* sequences distributed among these clusters, it is plausible to expect that at least some of them are actively transcribed into piRNAs. Altogether, our results strongly suggest the targeting of *DINEs* in *D. virilis* (and probably in other *Drosophila* species) by the piRNA machinery.

DINEs and chromatin modulation

It has become clear in the recent years that RNAi pathways are not only essential for germline stability but can also be considered as key factors influencing heterochromatin dynamics (reviewed in Slotkin and

Martienssen 2007; Biscotti et al. 2015). For example, Huang et al. (2013) demonstrated that Piwi-piRNA complexes interact with chromatin factors such as heterochromatin protein 1a (HP1a) and histone methyltransferases (HMTs) guiding them to specific locations and promoting chromatin changes in a genome-wide scale.

We found that the proportion of *DINE*-derived RNAs in 0–2 h *D. virilis* embryos is intermediate between the values found in gonads and carcasses (Fig. 7a). Because these embryos have no fully onset transcription (Vlassova et al. 1991; Pritchard and Schubiger 1996), their small RNAs and Piwi proteins are essentially the same of the maternal germ cells (Harris and Macdonald 2001; Megosh et al. 2006; Brennecke et al. 2008; Le Thomas et al. 2014). A similar scenario was found for *D. melanogaster* (Brennecke et al. 2008). At early embryogenesis, the maternally inherited piRNAs and piwi proteins could be leading elements in the process of defining heterochromatic domains (Sentmanat et al. 2013). Heterochromatin formation is triggered in embryo cells around 2 h old, coinciding with the first signs of transcription in the embryo cells themselves (Vlassova et al. 1991). The smaller proportion of *DINE*-derived RNAs in 0–2 h embryos could be the result of differences between somatic and germ cell transcripts from the ovaries (with a larger proportion of *DINE-TR1* transcripts in somatic cells), the normal depletion of maternally inherited piRNAs during early development or both.

Our description of *DINE-TR1* association with several *D. virilis* genes agrees with the previous finding that *DINEs* are frequently located within introns and flanking coding regions in several *Drosophila* species (Yang and Barbash 2008). Interestingly, the same study reports that *D. virilis* has the highest number of intronic insertions (1104) among the 12 *Drosophila* species analyzed. It is possible that small RNAs interact with some of these *DINE* elements, establishing local chromatin modifications and affecting the regulation of genes. For example, in the *Arabidopsis* accession *Landsberg erecta* (Ler), the FLC gene (a key factor in flowering pathways), has a mutator-like transposon insertion in the first intron, which is responsible for its low expression (Gazzani et al. 2003; Michaels et al. 2003). Furthermore, it has been demonstrated that this TE insertion is involved in siRNA-mediated silencing by forming a heterochromatin “island” restricted to the element and its vicinity (Liu et al. 2004). More recently, intronic insertions of *Helitrons* on *Arabidopsis* and rice

genes have been shown to be the main targets for heterochromatin establishment (Saze et al. 2013).

In *D. melanogaster*, TE insertions next to genes have been associated with changes in the chromatin state such as the di- and trimethylations of lysine 9 from histone H3 (H3K9me2, H3K9me3) and HP1a assembly; which are typical heterochromatic marks. These chromatin changes are likely piRNA-mediated and result in lower expression of the nearby genes (Sentmanat and Elgin 2012; Lee 2015). Those recent findings indicate that piRNA-mediated TE silencing is not restricted to heterochromatic TE insertions and that abundant TEs such as *DINE-TR1* in *D. virilis* could have a huge impact in both chromatin modulation and gene regulation.

Acknowledgments We are grateful for the anonymous reviewers' comments on the manuscript. This work was supported by grants from "Fundação de Amparo à Pesquisa do Estado de Minas Gerais" (FAPEMIG) (Proc: APQ-01563-14), "Programa Institucional de Auxílio à Pesquisa de Doutores Recém-Contratados da Universidade Federal de Minas Gerais," "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq), and a doctoral fellowship from "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) to GD.

References

- Abdurashitov MA, Gonchar DA, Chernukhin VA, Tomilov VN, Tomilova JE, Schostak NG, Zatssepina OG, Zelentsova ES, Evgen'ev MB, Degtyarev SK (2013) Medium-sized tandem repeats represent an abundant component of the *Drosophila virilis* genome. *BMC Genomics* 14(1):771
- Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318(5851):761–764
- Ashburner M (1989) *Drosophila*. A laboratory handbook. Cold Spring Harbor Laboratory Press
- Baimai V (1977) Chromosomal polymorphisms of constitutive heterochromatin and inversions in *Drosophila*. *Genetics* 85(1):85–93
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580
- Bergman CM, Kreitman M (2001) Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res* 1(8):1335–1345
- Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol* 7(11):R112
- Biessmann H, Zurovcova M, Yao JG, Lozovskaya E, Walter MF (2000) A telomeric satellite in *Drosophila virilis* and its sibling species. *Chromosoma* 109(6):372–380
- Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015) Transcription of tandemly repetitive DNA: functional roles. *Chromosome Research*. doi:10.1007/s10577-015-9494-4
- Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007) Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177(3):1277–1290
- Brajkovic J, Feliciello I, Bruvo-MadWaric B, Ugarkovic D (2012) Satellite DNA-Like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*. *G3 (Bethesda)* 2:931–941
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128(6):1089–1103
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322(5906):1387–1392
- Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SCR, Lin H (2007) *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* 21(18):2300–2311
- Brown EJ, Bachtrog D (2014) The chromatin landscape of *Drosophila*: comparisons between species, sexes, and chromosomes. *Genome Res* 24:1125–1137
- Brown JD, O'Neill RJ (2010) Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet* 11:291–316
- Caletka BC, McAllister BF (2004) A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. *Mol Phylogenet Evol* 33:664–670
- Carareto CM, Hernandez EH, Vieira C (2014) Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species. *Gene* 537(1):93–99
- Casacuberta E, Pardue ML (2003) Transposon telomeres are widely distributed in the *Drosophila* genus: *TART* elements in the *virilis* group. *Proc Natl Acad Sci U S A* 100(6):3363–3368
- Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220
- Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9(8):772
- de Wit E, Greil F, van Steensel B (2005) Genome-wide HP1 binding in *Drosophila*: developmental plasticity and genomic targeting signals. *Genome Res* 15:1265–1273
- Di Tommaso P, Moretti S, Xenarios I, Oróbitg M, Montanyola A, Chang JM, Tally JF, Notredame C (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* gkr 245:W13–W17
- Dias GB, Svartman M, Delprat A, Ruiz A, Kuhn GCS (2014) Tetris is a foldback transposon that provided the building blocks for an emerging satellite DNA of *Drosophila virilis*. *Genome Biol Evol* 6(6):1302–1313

- Dimitri P, Pisano C (1989) Position effect variegation in *Drosophila melanogaster*: relationship between suppression effect and the amount of Y chromosome. *Genetics* 122(4):793–800
- Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218
- Du C, Caronna J, He L, Dooner HK (2008) Computational prediction and molecular confirmation of Helitron transposons in the maize genome. *BMC Genomics* 9(1):51
- Ellison CE, Bachtrog D (2013) Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* 342(6160):846–850
- Feliciello I, Akrap I, Brajković J, Zlatar I, Ugarković Đ (2015) Satellite DNA as a driver of population divergence in the red flour beetle *Tribolium castaneum*. *Genome Biol Evol* 7(1):228–239
- Ferre PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol* 7(10):e1000234. doi:10.1371/journal.pbio.1000234
- Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, Thurston M (2006) Open software for biologists: from famine to feast. *Nat Biotechnol* 24(7):801–804
- Fonseca NA, Morales-Hojas R, Reis M, Rocha H, Vieira CP, Nolte V, Schlötterer C, Vieira J (2013) *Drosophila americana* as a model species for comparative studies on the molecular basis of phenotypic variation. *Genome Biol Evol* 5(4):661–679
- Francisco FO, Lemos B (2014) How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *J Genomics* 2:94–103
- Gaffney PM, Pierce JC, Mackinley AG, Titchen DA, Glenn WK (2003) Pearl, a novel family of putative transposable elements in bivalve mollusks. *J Mol Evol* 56:308–316
- Gall JG, Cohen EH, Polan ML (1971) Repetitive DNA sequences in *Drosophila*. *Chromosoma* 33(3):319–344
- Gatti M, Pimpinelli S, Santini G (1976) Characterization of *Drosophila* heterochromatin. *Chromosoma (Berl)* 57:351–375
- Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19(5):748–761
- Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol* 132:1107–1114
- Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nature Rev Genet* 10(2):94–108
- Gilchrist AS, Shearman DC, Frommer M, Raphael KA, Deshpande NP, Wilkins MR, Sherwin WB, Sved JA (2014) The draft genome of the pest tephritid fruit fly *Bactrocera tryoni*: resources for the genomic analysis of hybridising species. *BMC Genomics* 15(1):1153
- Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86
- Gordon A, Hannon GJ (2010) Fastx-toolkit. FASTQ/A short-reads pre-processing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit
- Gregory TR, Johnston JS (2008) Genome size diversity in the family Drosophilidae. *Heredity* 101:228–238
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455:1193–1197
- Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, Casillas S, Ràmia M, Egea R, Negre B, Mir G, Camps J, Moncuill V, Ruiz-Ruano FJ, Cabrero J, de Lima LG, Dias GB, Ruiz JC, Kapusta A, Garcia-Mas J, Gut M, Gut IG, Torrents D, Camacho JP, Kuhn GC, Feschotte C, Clark AG, Betrán E, Barbadilla A, Ruiz A (2014) Genomics of ecological adaptation in cactophilic *Drosophila*. *Genome Biol Evol* 7(1):349–366
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704
- Harris RS (2007) Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University
- Harris AN, Macdonald PM (2001) Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* 128(14):2823–2832
- Heikkinen E, Launonen V, Müller E, Bachmann L (1995) The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. *J Mol Evol* 41:604–614
- Heslop-Harrison JS, Schwarzacher T (2011) Organization of the plant genome in chromosomes. *Plant J* 66:18–33
- Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, Yasuhara JC, Wakimoto BT, Myers EW, Celniker SE, Rubin GM, Karpen GH (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol* 3(12):research0085
- Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H (2013) A major epigenetic programming mechanism guided by piRNAs. *Dev Cell* 24(5):502–516
- International Glossina Genome Initiative (2014) Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* 344(6182):380–386
- Jordan IK, Miller WJ (2008) Genome defense against transposable elements and the origins of regulatory RNA. In: Lankenau DH, Volff JN, eds. *Transposons and the dynamic genome*. Springer-Verlag Berlin Heidelberg, pp 77–94
- Junier T, Pagni M (2000) Dotlet: diagonal plots in a web browser. *Bioinformatics* 16(2):178–179
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1–4):462–467
- Kalmykova AI, Klenov MS, Gvozdev VA (2005) Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res* 33(6):2052–2059
- Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* 98:8714–8719
- Kapitonov VV, Jurka J (2007a) Helitrons in fruit flies. *Rebase Reports* 7(3):130
- Kapitonov VV, Jurka J (2007b) Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet* 23:521–529
- Kapitonov VV, Holmquist GP, Jurka J (1998) L1 repeat is a basic unit of heterochromatin satellites in cetaceans. *Mol Biol Evol* 15(5):611–612

- Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63
- Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7(1): 474
- Kuhn GCS, Heslop-Harrison JS (2011) Characterization and genomic organization of PERI, a repetitive DNA in the *Drosophila buzzatii* cluster related to DINE-1 transposable elements and highly abundant in the sex chromosomes. *Cytogenet Genome Res* 132:79–88
- Kuhn GCS, Sene FM, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS (2008) Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res* 16(2):307–324
- Le Thomas A, Marinov GK, Aravin AA (2014) A transgenerational process defines piRNA biogenesis in *Drosophila virilis*. *Cell Reports* 8:1617–1623
- Lee YCG (2015) The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. *PLoS Genet* 11(6):e1005269
- Lemos B, Araripe LO, Hartl DL (2008) Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* 319:91–93
- Leung W, Shaffer CD, Reed LK, et al. (2015) *Drosophila* Muller F Elements Maintain a Distinct Set of Genomic Properties Over 40 Million Years of Evolution. *G3 (Bethesda)* 5(5): 719–740
- Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev* 18(23): 2873–2878
- Locke J, Howard LT, Aippersbach N, Podemski L, Hodgetts RB (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
- Loreto ELS, Carareto CMA, Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 100(6):545–554
- Macas J, Koblížková A, Navrátilová A, Neumann P (2009) Hypervariable 3' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene* 448:198–206
- Mahan JT, Beck ML (1986) Heterochromatin in mitotic chromosomes of the *Virilis* species group of *Drosophila*. *Genetica* 68:113–118
- Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138(6):1067–1082
- Markow T (2015) The secret lives of *Drosophila* flies. *eLife* 4: e06793
- Megosh HB, Cox DN, Campbell C, Lin H (2006) The role of PIWI and the miRNA machinery in *Drosophila* germline determination. *Curr Biol* 16:1884–1894
- Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, DeRisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan SW (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* 14:R10
- Menon DU, Coarfa C, Xiao W, Gunaratne PH, Meller VH (2014) siRNAs from an X-linked satellite repeat promote X-chromosome recognition in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 111(46):16460–16465
- Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc Natl Acad Sci U S A* 100(17):10102–10107
- Miklos GLG, Cotsell JN (1990) Chromosome structure at interfaces between major chromatin types: alpha-and beta-heterochromatin. *Bioessays* 12(1):1–6
- Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol Biol Evol* 17(11):1597–1609
- Morales-Hojas R, Reis M, Vieira CP, Vieira J (2011) Resolving the phylogenetic relationships and evolutionary history of the *Drosophila virilis* group using multilocus data. *Mol Phylogenet Evol* 60(2):249–258
- Olszak AM, van Essen D, Pereira AJ, Diehl S, Manke T, Maiato H, Sacconi S, Heun P (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nature Cell Biol* 13(7):799–808
- Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P, Varotto C, Pisani D, Dekker T, Wrobel N, Viola R, Pertot I, Cavalieri D, Blaxter M, Anfora G, Rota-Stabelli O (2013) Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biol Evol* 5(4):745–757
- Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol Biol Evol* 15(3):293–302
- Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, González J (2011) Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol* 28(5):1633–1644
- Plohl M, Meštrović N, Mravinac B (2014) Centromere identity from the DNA point of view. *Chromosoma* 123:313–325
- Pritchard DK, Schubiger G (1996) Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes Dev* 10(9):1131–1142
- Rošić S, Köhler F, Erhardt S (2014) Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell Biol* 207(3):335–349
- Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, McCombie WR, Hannon GJ, Evgen'ev MB (2010) Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA* 16(8):1634–1645
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20:2214–2222
- Satovic E, Plohl M (2013) Tandem repeat-containing MITEs in the clam *Donax trunculus*. *Genome Biol Evol* 5(12):2549–2559
- Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, Ito T, Kakutani T (2013) Mechanism for full-length RNA processing of *Arabidopsis* genes containing intragenic heterochromatin. *Nat Commun* 4:2301
- Scalvenzi T, Pollet N (2014) Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA. *Mol Phylogenet Evol* 81:1–9

- Schwarzacher T, Heslop-Harrison JS (2000) Practical in situ hybridization. BIOS Scientific Publishers Limited, Oxford
- Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, Li M, Meisel RP, Mínx P, Murphy TD, Nelson DR, Reid WR, Rinkevich FD, Robertson HM, Sackton TB, Sattelle DB, Thibaud-Nissen F, Tomlinson C, van de Zande L, Walden KKO, Wilson RK, Liu N (2014) Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol* 15(10):466
- Sentmanat MF, Elgin SC (2012) Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc Natl Acad Sci U S A* 109(35):14104–14109
- Sentmanat M, Wang SH, Elgin SCR (2013) Targeting heterochromatin formation to transposable elements in *Drosophila*: potential roles of the piRNA system. *Biochem Mosc* 78(6):562–571
- Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 12:246–258
- Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8(4):272–285
- Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21(1):36–44
- Thomas J, Vadnagara K, Pritham E (2014) DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helitrons). *Mob DNA* 5:18
- Ugarkovic D (2009) Centromere-competent DNA: structure and evolution. In: Ugarkovic D, ed. Centromere: structure and evolution. Springer-Verlag Berlin Heidelberg, pp. 53–76
- Vaury C, Bucheton A, Pelisson A (1989) The b heterochromatic sequences flanking the I elements are themselves defective transposable elements. *Chromosoma* 98:215–224
- Vermaak D, Malik HS (2009) Multiple roles for heterochromatin protein 1 genes in *Drosophila*. *Annu Rev Genet* 43:467–492
- Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE, de Pablos B (2007) *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res* 17(12):1909–1918
- Villasante A, de Pablos B, Méndez-Lago M, Abad JP (2008). Telomere maintenance in *Drosophila*: rapid transposon evolution at chromosome ends. *Cell Cycle* 15;7(14):2134–2138
- Vlassova IE, Graphodatsky AS, Belyaeva ES, Zhimulev IF (1991) Constitutive heterochromatin in early embryogenesis of *Drosophila melanogaster*. *Mol Gen Genet* 229(2):316–318
- Volpe T, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen R (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297(5588):1833–1837
- Wallrath LL, Vitalini MW, Elgin SCR (2014) Heterochromatin: a critical part of the genome. In: Abmayr SM (ed) Workman JL. Fundamentals of Chromatin, Springer New York, pp 529–552
- Wasserlauf I, Usov K, Artemov G, Anan'ina T, Stegny V (2015) Specific features in linear and spatial organizations of pericentromeric heterochromatin regions in polytene chromosomes of the closely related species *Drosophila virilis* and *D. kanekoi* (Diptera: Drosophilidae). *Genetica* 1–12.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191
- Yamanaka S, Siomi MC, Siomi H (2014) piRNA clusters and open chromatin structure. *Mobile. DNA* 5(1):22
- Yang HP, Barbash DA (2008) Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. *Genome Biol* 9:R39
- Yang HP, Hung TL, You TL, Yang TH (2006) Genome wide comparative analysis of the highly abundant transposable element DINE-1 suggests a recent transpositional burst in *Drosophila yakuba*. *Genetics* 173:189–196
- Yin H, Lin H (2007) An epigenetic activation role of Piwi and a Piwi associated piRNA in *Drosophila melanogaster*. *Nature* 450:304–308