HAL open science

# Transposable elements in Drosophila

Vincent Mérel, Matthieu Boulesteix, Marie Fablet, Cristina Vieira

**To cite this version:**

Vincent Mérel, Matthieu Boulesteix, Marie Fablet, Cristina Vieira. Transposable elements in Drosophila. Mobile DNA, 2020, 11 (1), 10.1186/s13100-020-00213-z . hal-02957906

**HAL Id: hal-02957906**

**https://hal.science/hal-02957906v1**

Submitted on 22 Dec 2023

**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.

Mérel et al. Mobile DNA (2020) 11:23
https://doi.org/10.1186/s13100-020-00213-z

Mobile DNA

REVIEW Open Access

# Transposable elements in *Drosophila*

Vincent Mérel, Matthieu Boulesteix, Marie Fablet and Cristina Vieira*

## Abstract

*Drosophila* has been studied as a biological model for many years and many discoveries in biology rely on this species. Research on transposable elements (TEs) is not an exception. *Drosophila* has contributed significantly to our knowledge on the mechanisms of transposition and their regulation, but above all, it was one of the first organisms on which genetic and genomic studies of populations were done. In this review article, in a very broad way, we will approach the TEs of *Drosophila* with a historical hindsight as well as recent discoveries in the field.

**Keywords:** Population genomics, *Drosophila*, intra and interspecific TE diversity, epigenetics

## Background

### A few words about Transposable Elements

Transposable elements (TEs) are selfish genetic elements that are able to multiply in a genome by copying themselves to other locations. This particular property allows them to persist and multiply in populations without the need of providing any advantage to the host [1–3]. Discovered in maize in the late 1940's by Barbara McClintock, they were understudied for decades [4, 5]. With the advent of molecular biology, notably their use for genetic engineering, an enormous amount of work has been done on TEs. The first sequencing projects stimulated the interest in these sequences, as they underscored their ubiquitous character. Indeed, TEs are found in virtually all eukaryotic species investigated so far [6–9]. They may represent up to 80% of a genome, as in Maize [10]. Additionally, one may expect these large elements, up to 20 kb, possessing coding sequences, regulatory sequences, and a unique epigenetic profile, to produce large-effect mutations [11, 12]. Actually, TEs have been shown to profoundly impact not only genomes, from chromosomal rearrangements to genome size, but also individuals, from deleterious to adaptive effects. Like many other research topics in biology, research on TEs owes much to *Drosophila*.

* Correspondence: cristina.vieira@univ-lyon1.fr
Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, F-69622 Villeurbanne, France

### A few words about *Drosophila*

The *Drosophila* genus is estimated to include several thousand species [13] sharing their most recent common ancestor ~25-40 My ago [14]. So far, ~1500 drosophilid species have been described. The most extensively studied *Drosophila* species is, by far, *Drosophila melanogaster*. Originating from Sub-Saharan Africa, it has colonized all continents, except for Antarctica, as a human commensal [15, 16]. During the last 15,000-20,000 year, it expanded its range to Europe and Asia and was only recently introduced to Australia and the Americas (~200 years ago) [17]. *D. melanogaster* is raised in the lab since the beginning of the XXth century [16, 18]. Easy to maintain and having a short generation time, this species has been extensively studied since then. Nowadays, a search for the terms "*Drosophila*" and "*melanogaster*" on *pubmed* returns approximately 55,000 references, with more than 2000 published in 2018.

A great number of genetic tools, such as genetic transformation vectors using TEs, and the P-element in particular [19], the GAL4/UAS system to study gene expression, or more recently, the CRISPR/Cas9 system for site-specific genome engineering, are available for *Drosophila* species (see [18] for review). In addition to genetic tools, genome sequencing is relatively easy in this genus. Due to their relatively small size, *Drosophila* genomes can be sequenced at relatively low cost [20]. *D. melanogaster* genome was

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

among the first eukaryotic genomes sequenced, and is arguably the best annotated genome so far. A lot of sequencing data are available in the *Drosophila* genus. The genome of at least 46 species were sequenced and assembled [21]. In addition, in *D. melanogaster*, several studies aimed at sequencing either individuals or populations (PoolSeq) [22–29]. This sequencing effort benefited largely from diverse consortia. One of the first, and probably one of the best-known, the *Drosophila melanogaster Genetic Reference Panel (*DGRP) consortium made available the genomic sequence of more than two hundred inbred lines from an American population [22, 24]. At a broader geographical scale, the global diversity lines consortium sequenced a panel of 84 worldwide strains [29]. We also should mention the European *Drosophila* Population Genomics Consortium (*DrosEU*) which recently produced PoolSeq data fom 48 European population samples [28]. Nowadays, more than 1,121 individual Drosophila genomes are available [30], as well as pooled genomes from 30 localities in Europe and 23 in North America. For some individual genomes of the DGRP, data about gene expression and various phenotypic traits are also available [22, 31–34]. DGRP lines and a large variety of mutants and natural strains of *D. melanogaster,* collected from all over the world at different times, are currently maintained and available for researchers [35]. In addition, more than 250 species are accessible [36]. From an ecological/genomics perspective, *Drosophila* species offer a unique opportunity to perform comparative studies. For instance, the pair *D. melanogaster*/*D. simulans*, with a short time of divergence (around 1.5 My), share a common geographical range, as both are cosmopolitan species, but have very different ecologies, the former being close to human habitats and the second being found only in forest environments [14, 37]. Other *Drosophila species*, such as *D. suzukii*, are classified as invasive species, and represent an opportunity to study the genomic determinants of the invasive process. A last example that we can cite is the use of *Drosophila* species as models for speciation studies. This has been done extensively using the species close to *D. melanogaster* (*D. simulans*, *D. sechellia* and *D. mauritiana*) [38–40] and species from the *repleta* group (*D. mojavensis* and *D. arizonae* [41–44]; *D. buzzatii* and *D. koepferae* [45–47].

### A few words about Transposable Elements & *Drosophila*

*Drosophila* has been used as a model to study TEs for more than forty years now. The activity of the then-called "mobile dispersed genes" was already studied at the beginning of the 80's [48, 49]. Even before, they were studied as the uncharacterized inducers of the hybrid dysgenesis phenomenon [50, 51], in which the transmission of some genetic factor by the male but not the female resulted in a sterile progeny. Since then, research on TE in *Drosophila* heavily benefited from the advantages provided by this model, from genetic engineering to sequencing techniques. Not only the molecular mechanisms beyond the hybrid dysgenesis are now much better understood, but the study of this phenomenon also led to major discoveries in TE regulation, such as regulation by small RNAs. In this review, we aimed at giving an overview of the accumulated knowledge on Transposable Elements from molecular aspects to populations genomics in *Drosophila*, comparing the *D. melanogaster* to other *Drosophila* species where relevant.

## TE diversity

### About the classification

The abundance and ubiquity of TEs rapidly brought the necessity of a unified classification system for these sequences. The question of TE classification has been, and continues to be, a subject of debate [11, 52–54], especially the necessity for such system to reflect the phylogeny of TEs. From an evolutionary perspective, a purely phylogenetic classification seems ideal, however this may be hard to achieve. Beyond the polyphyletic nature of TEs, there are several other difficulties. One is that TE phylogeny does not necessarily reflect the organism phylogeny. Another is that the phylogenetic analysis of TE protein sequences may be arduous, because some TEs do not possess any coding sequence, some TEs possess several coding sequences with different phylogenetic signals due to recombination events, and some TEs are present in thousands of copies in the genome. In the sequencing era, when genome annotation is fundamental, Wicker *et al.* (2007) proposed a set of rules to rapidly classify TEs [11]. This widely used classification relies on transposition mechanisms, sequence similarities and structural relationships. In decreasing hierarchical order, we find the following classification levels: class, sometimes subclass, order, superfamily and family (and sometimes subfamily). The highest-level category, *i.e.* class, divides TE sequences into those with or without an RNA transposition intermediate. Next, the order category distinguishes sequences according to the insertion mechanism. Orders are further divided into superfamilies. The superfamily category discriminates sequences on the basis of particular features, for instance protein or non-coding domain structure, presence and length of direct repeats generated on both sides of a TE upon insertion (Target Site Duplication, TSD). The lowest-level category, *i.e.* family, includes sequences with a high rate of identity at the DNA level (at least 80% of identity over at least 80% of their internal or coding domain, or within their terminal repeat regions, or in both). Note that a

distinction also exists between autonomous TEs, *i.e.* TEs able to move by themselves, and non-autonomous TEs, *i.e.* TEs relying on other TEs to move, usually because they lack a certain protein.

### Class I TEs: retrotransposons

Class I TEs are also called retrotransposons. They transpose via an RNA intermediate. The RNA intermediate is transcribed from a genomic copy, then reverse-transcribed into DNA by a TE-encoded reverse transcriptase. Each complete replication cycle produces one new copy. Retrotransposons can be divided into five orders: long terminal repeat (*LTR*) *retrotransposons*, Dictyostelium intermediate repeat sequence (DIRS)-like elements, Penelope-like elements (PLEs), long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). All of them are present in *Drosophila,* but LTR retrotransposons and LINEs are by far the most abundant [20, 55].

In *Drosophila*, LTR retrotransposons usually range from 5 to 7 kb (Fig. 1) [11, 57–59]. They owe their names to the direct Long Terminal Repeats (~300-400 bp) flanking them. They typically display two genes: *gag* and *pol*. *gag* encodes the capsid, and *pol* encodes a protease (Prot), an integrase (Int) and a reverse transcriptase (RT) with an RNase domain. After the transcription step, some transcripts will be translated while the others may end up transposed (Fig. 1) (see [60] for more details on transposition mechanisms). The protease of *pol* cleaves Pol into a protease, an integrase and a reverse transcriptase [61]. The Gag protein assemble into a capsid that makes a particle around untranslated transcripts, the integrase, reverse transcriptase and a tRNA [62]. Because the formed ribonucleoprotein (RNP) does not comprise the transcript from which proteins were translated, we typically refer to a trans-preference mechanism of RNP assembly. Using the tRNA as a primer for synthesis, the reverse transcriptase initiates the production of double stranded DNA from the TE transcript [63]. After reverse transcription, the particle falls apart, the integrase recognizes the two ends of the cDNA and inserts them into the host genome. Upon integration, LTR retrotransposons produce a TSD of 4-6 bp [11]. Note that the LTR order is further divided into five superfamilies: *Copia* (*e.g. Copia* and *1731* families), *Gypsy* (*e.g. HMSBEAGLE* and *412* families), *Bel-Pao* (*e.g. BEL, Roo* and *Max*

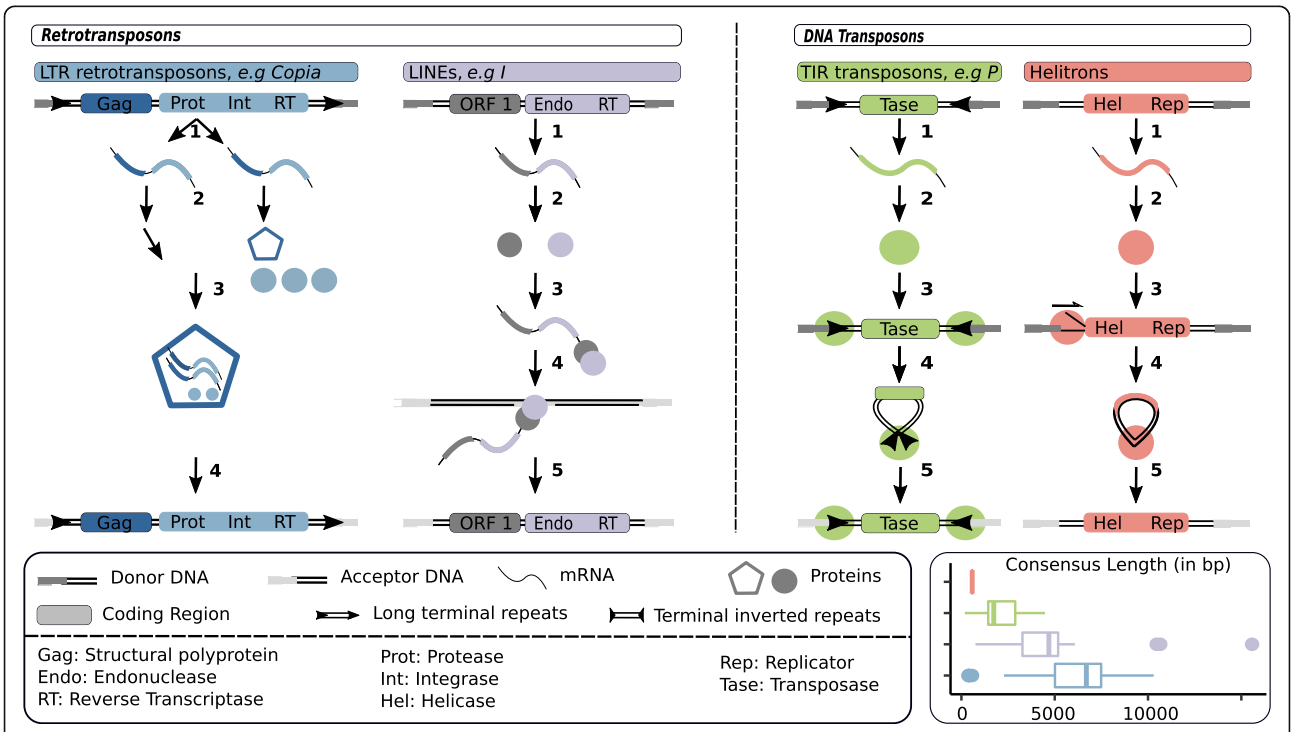



**Fig. 1** TE structure and transposition mechanisms. LTR retrotransposons: 1. Transcription. 2. Translation of one part of the transcripts. The protease (Prot) cleaves *pol* polyprotein. 3. *gag* proteins assemble around untranslated transcripts, the integrase (Int), reverse transcriptase (RT) and a tRNA. 4. Reverse transcription and integration. LINE retrotransposons: 1. Transcription. 2. Translation. 3. Protein(s) bind to the transcript. 4. A strand of donor DNA is cut, target-primed reverse transcription starts at the exposed 3′ extremity. 5. The TE is integrated. TIR DNA transposons: 1. Transcription. 2. Translation. 3. Two transposases bind to the TIRs. 4. Transposases dimerize and cut TIR extremities forming a free complex. 5. The complex binds to donor DNA and is integrated. Helitron DNA transposons: 1. Transcription. 2. Translation. 3. At the donor site, the plus strand is cut. A replication fork is formed. 4. Replication results in a double stranded transposon circle. 5. Integration. The bottom right panel represents the distribution of the lengths of *D. melanogaster* consensus sequences (RepBase [56]), using the same color code as above.

families), *Retrovirus* and *Endogenous RetroViruses (ERV)*. According to Wicker and colleagues classification, *Retroviruses* and *ERVs* also have an envelope gene (*env*). The corresponding protein allows *Retroviruses* to infect other cells. In *Drosophila*, few families have been shown to possess an *env* coding ORF, for example *Idefix*, *Gypsy*, *Tirant* and *ZAM* families [58, 64, 65]. Note that the insect endogenous retroviruses belong to the *Gypsy* superfamily, and that their origin is distinct from that of vertebrate ERVs [66]. Infectious properties have been demonstrated for *Gypsy* and *ZAM* families [67, 68].

LINEs are 3 to 5 kb-long, and generally contain two ORFs (Fig. 1) [11, 59, 69–71]. The first ORF encodes a protein with both RNA binding and nucleic acid chaperone properties [72, 73]. The second ORF encodes a protein that displays two domains: an endonuclease (Endo) and a Reverse Transcriptase [74, 75]. Contrary to LTR retrotransposons, LINEs exhibit a cis-preference mechanism of RNP assembly. After translation, the protein(s) bind to the mRNA molecule from which they originate, and form an RNP in the cytoplasm [76] (see [77] for more details on transposition mechanisms). The ribonucleoprotein particle moves back to the nucleus, and the protein cuts a single strand of the host genome at the point of insertion. The exposed 3' end allows the initiation of reverse transcription (target-primed reverse transcription). Subsequent events remains unclear, however the following has been proposed. During or after reverse transcription, the second strand of the host genome is cleaved. The newly reverse transcribed single-stranded DNA binds to the generated 3' extremity, and this extremity acts as a primer for the synthesis of the second strand of DNA. LINEs generate TSDs of various sizes upon insertion. Note that, probably as a consequence of early termination of reverse transcription, transposition may result in creation of 5' - truncated copies [78].

As mentioned above, besides LTR retrotransposons and LINEs that are abundant in *Drosophila* genomes, Class I comprises three other orders: DIRS, PLEs and SINEs. To our knowledge, DIRS and SINEs have not been found in Drosophila so far [20, 79]. PLEs were initially discovered in *D. virilis* and are involved in the hybrid dysgenesis phenomenon (Table 1). These TEs are present at least in the *virilis* group and in *D. willistoni* [89]. PLEs resemble LINEs, in a sense that they encode an endonuclease and a reverse transcriptase. However, they possess terminal repeats that can be in a direct or an inverse orientation.

### Class II TEs: DNA transposons

Class II TEs are DNA transposons. They do not transpose via an RNA intermediate but via a DNA intermediate. There are four orders: terminal inverted repeat (TIR) transposons, Crypton, Helitron and Maverick. TIRs and Helitrons are the most abundant in *Drosophila*.

TIR Transposable Elements are typically ranging from 1.5 to 3 kb in *D. melanogaster*, and are characterized by their TIRs of variable lengths (Fig. 1) [11, 59, 90, 91]. TIRs encode one unique protein called transposase (Tase). The transposition mechanism begins with two transposases recognizing and binding to the TIRs [92]. Transposases dimerize and cleave the ends of TIRs forming a free complex containing the TE [93]. The formed entity binds to the target DNA locus, where the transposon is integrated. The TSD size and the sequences of TIRs are highly variable across the nine known superfamilies [11]. Although the transposition mechanism in itself is not replicative, such TEs can increase their copy numbers in two ways. First, by transposing during chromosomal replication from a position that has already been replicated to a position ahead of the replication fork [94]. Second, they can exploit gap repair following excision to create an extra copy at the donor site [95].

The Helitron order, which is represented by the unique Helitron superfamily, gave rise to rather small TEs in *D. melanogaster* (< 1 kb, Fig. 1) [11, 96, 97]. Helitrons encode one unique protein with both a DNA helicase (Hel) and a replicator (Rep) domain. Because Helitrons were discovered only in 2001, and the lack of active Helitron examples limits experimental work, Helitron transposition mechanisms remain murky. However, using an artificially reconstructed active Helitron, Grabundzija and colleagues provided new insights and suggested the model synthesized hereafter [98]. First, the plus strand, the original donor strand, is nicked at the 5'-extremity of the TE and a replication fork is created. DNA replication results in a reconstituted double stranded donor site and a double stranded TE circle. This step may be repeated several times, producing several TE circles. Moreover, on the TE circles, a second DNA cleavage may occur on the original donor strand, a new replication fork established, and two double stranded transposon circles obtained from one. Finally, the double stranded TE may be integrated at the acceptor site. Note that the small sizes of Helitrons in *D. melanogaster* are explained by their non-autonomous character.

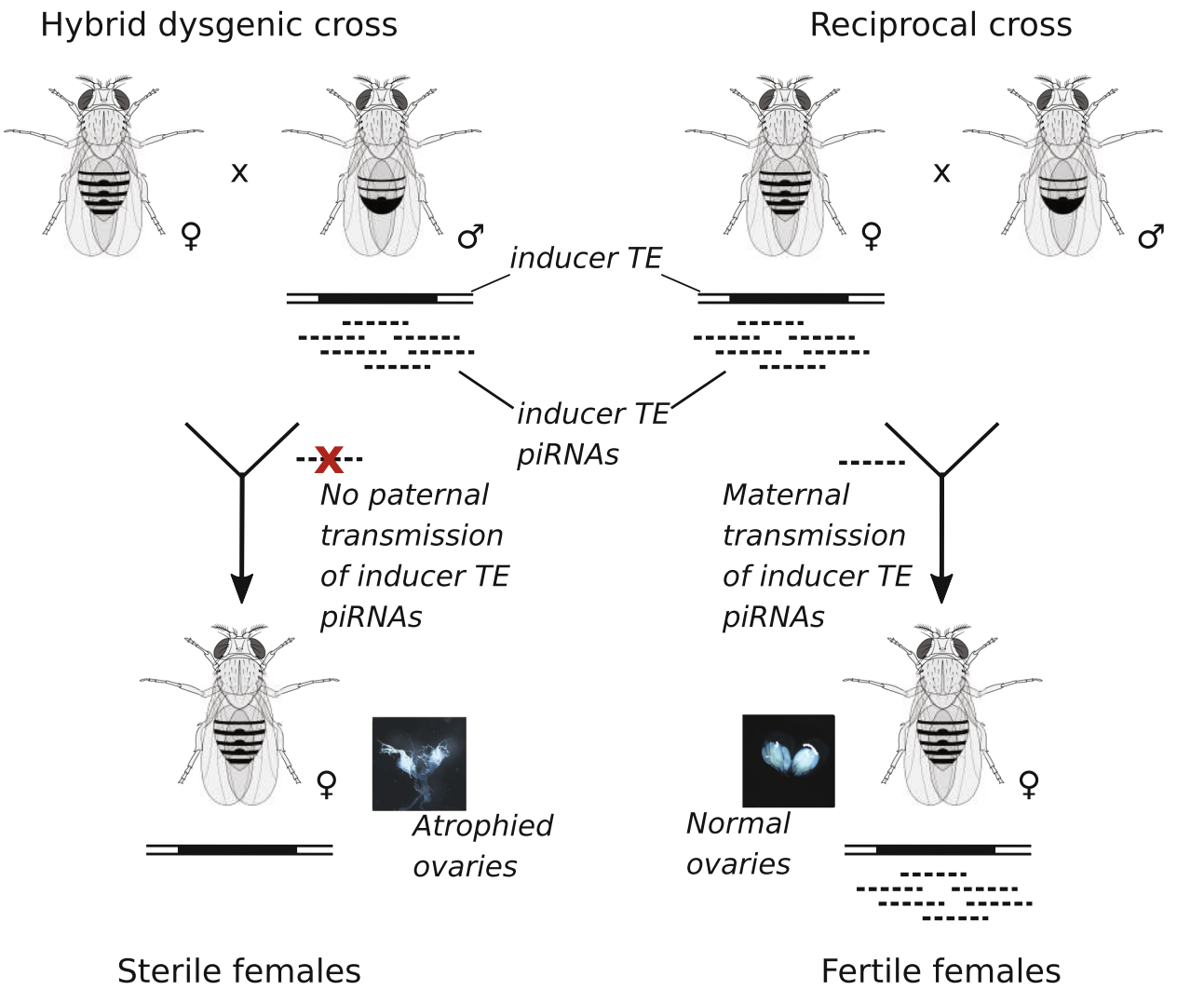
## TE abundance

### The Drosophila melanogaster reference genome

To obtain a picture of TE content in *D. melanogaster* genome, we investigated TE copy numbers and TE sequence occupancy in the last release of the reference genome assembly (Fig. 2). We used a combination of RepeatMasker, to identify genomic fragments

**Table 1** Hybrid dysgenesis

In *Drosophila*, some intraspecific crosses were observed to produce sterile females [50, 51, 80–85]. This phenomenon is called hybrid dysgenesis (Ovaries pictures from ref 81). It happens when males possessing a particular TE, hereafter referred as the inducer TE, are crossed with females whose genome is devoid of this TE. On the contrary, the reciprocal cross leads to viable and fertile individuals. The explanation is related to piwi-interacting RNAs (piRNAs), small RNAs repressing TEs with sequence complementarity (see the piRNA section). Because piRNAs are maternally transmitted, in dysgenic crosses the inducer TE insertion is transmitted to the progeny without the piRNAs directed against it [86]. In the reciprocal cross, both the inducer TE insertion and its piRNAs are transmitted, allowing the control of the TE family in the progeny, and the hybrid to be fertile.
Hybrid dysgenesis was documented in three systems in *D. melanogaster:* associated with *P-element*, *I-element* or *Hobo* [50, 51, 80]. *P-element* also appears to induce hybrid dysgenesis in *D. simulans* [81]. In addition, in *D. virilis,* a hybrid dysgenic cross potentially implying several TEs was reported [82–85]. From a historical perspective, the hybrid dysgenesis phenomenon played an important role not only in the discovery of horizontal transfers of TEs, but also in the study of host defenses against TEs [86–88].




homologous to a library of *Drosophila* TE consensus sequences available in the RepBase database, and the bioinformatic tool OneCodeToFindThemAll to reconstitute TE copies [56, 100, 101]. As previously reported, *D. melanogaster* genome contains ~20% of TEs [55, 102]. Note that a significant variation exists regarding these estimates [103–105]. These differences are likely to be at least partly explained by the genome assembly, or the part of the genome assembly that is analyzed, or both. For example, the *Drosophila 12 genomes* consortium considered only the best-assembled part of the genome, likely representative of the euchromatic portion of the genome, and found the TE content ranging from 2 % to 8 % (see Population Genomics section for details about TE density in different genomic regions). On the contrary, even if far from reporting the entire sequence of heterochromatic regions, the assembly used in Fig. 2 comprises at least 20 Mb of heterochromatic sequences, *i.e.* ~15% of the 140 Mb assembly [106]. Nevertheless,

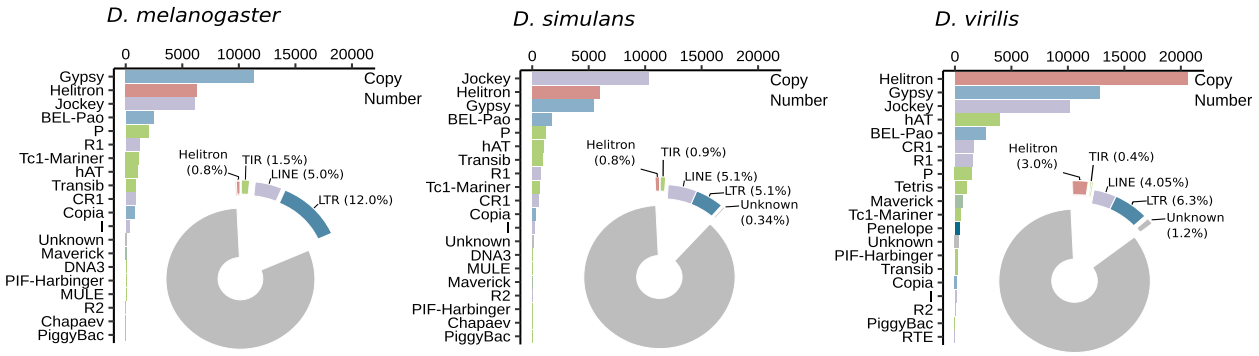



**Fig. 2** TE contents in *D. melanogaster*, *D. simulans* and *D. virilis* (from left to right). Barplots represent TE copy numbers for the top 20 TE superfamilies. Piecharts illustrate genomic sequence occupancy of each TE order (in percentages of the assemblies). These results were obtained using the *D. melanogaster* reference genome assembly (r6.29), and recently produced long-reads assemblies of *D. simulans* and *D. virilis* [99]. RepeatMasker was used to recover TE fragments and TE genomic sequence occupancy (RepeatMasker v1.332, -nolow, -norna, -species drosophila; Repbase-derived RepeatMasker libraries 20181026 [100],). TE fragments were assembled into TE copies using OneCodeToFindThemAll [101].

the relative abundance of the different TE orders is globally conserved across studies and similar to what is represented in Fig. 2 [55, 102, 103, 105]. Retrotransposons, and essentially LTRs and LINEs (respectively 12% and 5% of the genome in our analysis), contribute substantially to *D. melanogaster* TE content. DNA transposons correspond to a smaller proportion of the genome: we found that they represent less than 2%, including 0.9% for Helitrons and 0.7% for TIR elements. This ten-fold difference in terms of genomic sequence occupancy between retrotransposons and DNA transposons is mostly due to the larger size of retrotransposons (Fig. 2). Indeed, in terms of insertion numbers we found 11,657 DNA transposons (6,284 Helitrons and 5,373 TIR elements) and 23,148 retrotransposons (14,540 LTR retrotransposons and 8,608 LINEs) (see also [103] and [101]). For each of the four major orders, one superfamily is often over-represented: *Gypsy* for LTR elements, *Jockey* for LINEs, *P* for TIR elements, *Helitron* for Helitrons. According to our analysis, the different TE orders exhibit different numbers of families: indeed, we found insertions belonging to 721 LTR families, 331 LINE families, 213 TIR families and 63 Helitron families. The mean copy number per family is 26, but large variations exist. The family having the highest number of insertions is DNAREP1_DM, for which we found 1,746 copies. This sequence is annotated as a non-autonomous Helitron [107] (but see [97, 108] concerning classification).

### Interspecific variation

When it comes to TE contents across *Drosophila* species, a direct comparison of studies may be difficult. Indeed, authors are free to choose among a large number of programs and methods dedicated to identifying TEs, which leads to widely different results [105, 109]. For example, using the same TE sequence library but two different tools to annotate the *D. willistoni* genome, the 12 genomes consortium estimated TE content to be either 9 % or 16 %. The library used may also greatly affect results. In the same study, using the same tool, but a *D. melanogaster* TE sequence library or a *de novo* library, the authors found either 12 or 20 % TEs in the *D. ananassae* genome. Overall, in this study seven combinations of library-detection tools were used, leading to a TE content ranging from less than 10 % to up to 30 % in *D. ananassae*. The direct comparison of studies may thus be risky. A further layer of complexity comes from the sequencing technology, which impacts the quality of genome assemblies. Short paired-end read based assemblies lead to underestimation of TE contents compared to Sanger and long read based assemblies [110–112]. For all these reasons, to describe variation of TE contents in the *Drosophila* genus, here we focus on studies directly aiming at comparing TE amounts across species, and we remain cautious when linking them. For illustrative purposes, in addition to the annotation of TE contents in *D. melanogaster*, we estimated TE genomic sequence occupancy and copy numbers in two species: *D. simulans* and *D. virilis* (Fig. 2). We used the exact same methods as for *D. melanogaster*, and we do not expect the TE library to strongly bias the results, as it contains sequences constructed from the three species, which are among the most - studied with regard to TEs [113, 114]. Beyond that, we chose these two species because of their different positions relatively to *D. melanogaster* in the *Drosophila* phylogeny. On one hand, *D. simulans* is a close relative to *D. melanogaster*; they diverged approximately 1.5 Mya. Both species belong to the *melanogaster* subgroup within the *melanogaster* group, itself in the *sophophora* subgenus [14]. On the other hand, *D.*

*melanogaster* and *D. virilis* diverged about 25 Mya and *D. virilis* belongs to a different subgenus, the *drosophila* subgenus.

The first study intending to compare global TE contents across a significant number of *Drosophila* species was performed by the *Drosophila 12 genomes* consortium. This consortium investigated TE genomic sequence occupancy in eight species from the *sophophora* subgenus, mostly from the melanogaster subgroup, and four species from the *drosophila* subgenus. As stated above, the researchers focused on genomic parts likely to be euchromatic, and they used different methods. Using the method giving the lowest estimates, they found a global range of variation going from 1% to 9% of TEs in the genome. The method leading to the highest estimates resulted in genome containing from 3% to 30% of TEs. Invariably, *D. ananassae* was the species with the highest proportion of TEs. The authors chose the most unbiased and conservative method to compare the relative abundance of LTR retrotransposons, LINEs, TIR elements and so-called OTHERs among species. They found that the pattern LTRs>LINEs>TIRs>OTHERs is globally conserved across the phylogeny, with LTR retrotransposons usually constituting more than 50% of the repeatome. The two exceptions are *D. mojavensis* and *D. pseudoobscura.* In *D. mojavensis*, LTR elements represent only 45% of the repeatome, and in *D. pseudoobscura*, LTR retrotransposons and LINEs each contribute to roughly 33% of the repeatome. Our analysis shows a slightly different pattern, with equivalent genomic sequence occupancy for LTR elements and LINEs in *D. simulans*, and more Helitrons than TIR elements in *D. virilis* (Fig. 2). Recently, Hill and colleagues investigated both the proportion of TEs and their number of insertions in the genomes of five species. Four of these species were already in the set analyzed by the *Drosophila 12 genomes* consortium, except for *D. innubila*. The LTRs>LINEs>TIRs>OTHERs pattern for TE genomic proportions was not respected by any of the considered species. The dominant category differed: the most abundant elements are LTR retrotransposons in *D. ananassae,* while they are LINEs in *D. pseudoobscura*, and DNA transposons in *D. innubila*. *D. ananassae* was also the species with the highest TE content, with approximately 35% of TEs in the genome. Considering TE copy numbers, the authors found a total ranging from 2,000 to 14, 000 depending on the species. Once again, the difference with the previous results may probably be explained by data/method differences. Relative abundances of the different TE categories were found to differ across genomes. For example, DNA transposons were the most abundant in *D. willistoni,* whereas in *D. ananassae* they were as numerous as LINEs or LTR elements. The study with the largest dataset of species compared in terms of TE content was published by Sessegolo and collaborators [20]. These authors investigated the TE contents of 26 *Drosophila* species. Once again, the LTRs>LINEs>TIRs>OTHERs pattern did not hold for many species. The genomic content of repeats ranged from 4.65% in *D. busckii* to 30.80% in *D. suzukii*. The authors found a significant effect of phylogenetic inertia on TE content, but because of uneven sampling across the phylogeny, it was difficult to extract a pattern for each subgroup, many being represented by only one species. Overall, the data suggest large variations in the abundance of TEs across the *Drosophila* genus.

### Intraspecific variation

At the intraspecific level, genome size, which is correlated to TE abundance in *Drosophila*, is variable within populations of both *D. simulans* and *D. melanogaster*. This suggests that TE contents may change between populations, at least quantitatively [20, 55, 115]. In addition, the discovery of hybrid dysgenesis, *i.e.* the generation of a sterile hybrid by crossing particular parental strains differing by TE families, has highlighted qualitative differences in TE content at the intraspecific level (Table 1) [50, 51, 69]. TE contents in populations were extensively studied by *in situ* hybridization on polytene chromosomes, restricting the results to a few families. Quantitative differences related to the hybrid dysgenesis phenomenon have been observed for *I-Element*, *P-Element* and *Hobo* in *D. melanogaster* [69, 80, 116]. It has been demonstrated that the P-element has recently been acquired by horizontal transfer, likely from *D. willistoni*, and then spread step by step in worldwide populations between 1950 and 1990 [87, 117–119]. The history seems to repeat itself with the current invasion of *D. simulans* by the *P-element* after a horizontal transfer event from *D. melanogaster* [81, 120]. Horizontal transfers of TEs have now been extensively described in eukaryotes [121] and the study of TEs in the genomes of *D. melanogaster*, *D. simulans* and *D. yakuba* suggests that one-third of TE families has originated by recent horizontal transfers between these species [122]. In addition to hybrid dysgenesis, the study of 34 TE families from various populations of *D. simulans* by Vieira and colleagues showed fairly large qualitative differences between populations. Indeed, they found at least 14 families of TEs that were present only in certain populations [123, 124]. Quantitatively, and as an example, a study of the *412* element in *D. simulans* showed a gradient in copy numbers ranging from 1–10 in South Africa to 23 in Europe [125]. Genome size and TE content variations parallel the worldwide colonization of *D. melanogaster* but not that of *D. simulans* [115]. In *D. subobscura*, *Bilbo* and *Gypsy* families show slightly more copies in colonizing than original populations [126]. Similar

results were obtained when contrasting copy numbers of *Bilbo* and *Osvaldo* between colonizing and original populations of *D. buzzatii* [127]. In both cases, the study of insertion frequencies suggested that genetic drift associated with a founder effect that accompanied the colonization was responsible for the observed variation of copy numbers. Recently, genomic analyses of European *D. melanogaster* populations from *DrosEU* confirmed that intraspecific variation of TE contents may be substantial, and reveals TE proportions ranging from 16% to 21% of genomes [28].

## TE activity

### Spontaneous rate of transposition

A recent study by Adrion and colleagues [128] provided the first genome-wide estimate of TE movement rate in *D. melanogaster*. These authors used NGS data to compare TE contents across laboratory lines before and after ~150 generations of mutation accumulation. They found that the TE movement rate is slightly lower than the point mutation rate: 2.45 × 10(-9) per site per generation against 2.8 × 10(-9) per site per generation, respectively [129]. The rate of insertions is higher than the rate of deletions: 2.11 × 10(-9) per site per generation against 1.37 × 10(-10) per site per generation, respectively. Considering that there are 270 millions sites in the genome assembly, these numbers correspond to approximately 0.57 insertions and 0.037 deletions per generation. Those estimates were obtained across all TE superfamilies and are consistent with previous reports using *in situ* hybridization to determine transposition events for one or a few families [130–132]. Adrion and colleagues found superfamily-specific insertion and deletion rates to range between 0 and 5.13 × 10(-3) per copy per generation, and between 0 and 1.29 × 10(-4) per generation, respectively. They also found a significant effect of the genetic background, as previously reported [133–135].

### Transposition bursts

Beyond the spontaneous rate of transposition, a significant number of studies have shown that transposition bursts could occur in *Drosophila* (see [136] for a review). A burst is characterized by movement of large numbers of TE sequences through the genome during a short evolutionary time [137]. Although these bursts can happen without any apparent reason, they are commonly associated with stressful conditions such as extreme temperatures, irradiation, chemical exposure, or viral infection [138–142]. For example, Vasil'eva and colleagues showed that gamma radiation could increase the *412* transposition rate up to 5.6 events per genome per generation. Note that the attempts to induce TE mobilization with thermal shocks led to contradictory results in *Drosophila*, potentially due to the differences between tested genetic backgrounds, or tested TEs, or both, but also to methodological considerations (see [136]). Furthermore, although to our knowledge it has not been observed in *Drosophila* so far, stress may also lead to repression of TE activity [143]. Another stress widely studied in *Drosophila* for its effect on transposition is the genomic stress occurring when two somehow divergent genomes are united after hybridization (Table 1). In several biological systems it increases TE activity with potentially dramatic consequences on the phenotype, including sterility [144, 145]. It was observed when crossing individuals from different species, but also when crossing particular strains from the same species which corresponds to the hybrid dysgenesis phenomenon mentioned above [47, 50, 51, 80]. The causes of the TE bursts are not completely elucidated yet. Concerning hybridization, it has been shown that a failure of the host defense against TEs could be at stake (see below and Table 1). Regarding TE activation in response to stressful conditions, it has long been suggested that it could be due to TEs displaying binding sites for stress specific transcription activators, such as transcription factors [146]. In agreement with this idea, the temperature responding *Mariner* and *Copia* elements were shown to display sequences homologous to the promoter of heat shock proteins [147, 148]. More recently, a transcriptomic study demonstrated that temperature dependent TE expression is TE family specific and dependent on the genetic background. The authors proposed that TE transcription is indeed regulated by an interaction between TE family-specific regulatory sequences and host *trans*-acting factors [149]. Note, however, that this study was done on a range of temperatures that are not necessarily stressful (13–29°C). It is also important to consider that all the reports mentioned above concern laboratory experiments in conditions that are potentially unlikely *in natura*. The mechanisms at play in natural populations still remain poorly understood. One study demonstrated a burst of transposition for *DINE-1* in *D. yakuba* [150], and its causes are still unknown. In *D. simulans,* the copy numbers of the *412* element increase with latitude following the minimum temperature, and in *D. melanogaster*, significant correlations were found between TE abundance and different geographical and environmental variables for four families [125, 151]. However, in both cases, a possible confounding effect of demographic history cannot be excluded. Only one study established a direct link between TE activity and a geo-climatic variable: in *D. simulans,* the *Mariner* element somatic activity varies along a latitudinal cline between tropical Africa and Europe [152].

### Interspecific variation

So far, few studies tried to compare TE activity across *Drosophila* species. In 2011, Lerat and colleagues compared the TE contents of four *Drosophila species* from the *melanogaster* subgroup: *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba* [153]. They found that *D. simulans, D. sechellia* and *D. yakuba* genomes contained a large fraction of degraded copies compared to *D. melanogaster*. The authors suggested a recent TE activity in *D. melanogaster*, compared to the three other species. This can partially be observed when comparing the so-called TE landscapes of *D. melanogaster* and *D. simulans* (Fig. 3). These landscapes constitute an easy way to visualize TE activity through time. The X axis corresponds to the divergence of the TE sequences from the consensus, and it can be seen as a proxy of the time passed since the last wave of transposition. In Fig. 3, we can see a recent peak of activity of LTR elements, especially in *D. melanogaster*. In *D. simulans*, the peak of activity is also recent but much smaller. Another study was aimed at comparing TE activity between *D. melanogaster* and *D. simulans* using NGS population data [155]. Based on TE insertion frequency data, the authors determined that more than 58 families are probably highly active in both species. Half of the TE families show evidence of variation of activity through time, and are not the same depending on the species. Finally, they found that retrotransposons were the most active TEs in *D. melanogaster,* while DNA transposons were the most active TEs in *D. simulans*. A recent study compared TE frequencies in five distant points of the *Drosophila* phylogeny [55]. These species shared a common ancestor around 30 Mya [14]. The authors found evidence that an excess of low frequency insertions is prevailing in the phylogeny and is observed for most TE families. This suggests that an active repeatome is frequent, at least in the *Drosophila* genus.

## Impacts of TEs

### On the genome

TEs play an important role in the structural evolution of genomes through the generation of various types of mutations: chromosomal rearrangements, gene disruption and changes in gene expression. The simplest mechanism by which TEs can cause chromosomal rearrangements is through participation in an ectopic recombination event [156]. Ectopic recombination corresponds to recombination between more-or-less identical sequences inserted at different locations in the genome, such as TEs [157]. Depending on their relative positions and orientations, their recombination can result in different kinds of chromosomal rearrangements: duplication, deletion, inversion, or translocation. TEs were associated with chromosomal rearrangements *in natura* in various species of *Drosophila*, and mainly with inversions [158–161]. In several cases, ectopic recombination was identified as the cause of these rearrangements [159, 160]. When they insert into genes or their regulatory sequences, TEs can disrupt gene function. A perfect example is the use of the *P-element* in the Berkeley *Drosophila* Genome Project [162–164]. The Berkeley *Drosophila* Genome Project aimed at disrupting each *D. melanogaster* gene using the *P-element* in order to decipher gene functions. More than 5,000 genes were disrupted in that way. TEs can affect gene expression in two principal ways. First, they may bring regulatory sequences (see [165] for a review). For example, *Bari-Jheh* adds extra antioxidant response elements upstream of the *Jheh1* and *Jheh2* genes and is associated with

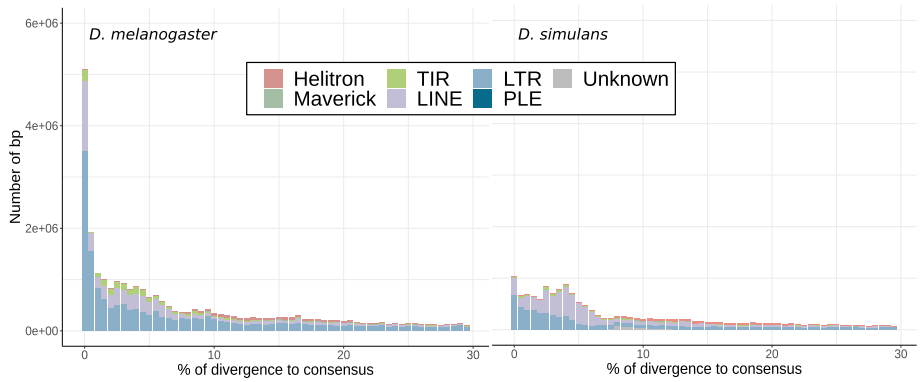



**Fig. 3** TE landscapes in *D. melanogaster* and *D. simulans*. For each TE fragment the divergence to consensus was estimated. For each TE order the total amount of DNA (in bp) is shown as a function of the percentage of divergence. The percentage of divergence to the consensus sequences is a proxy for age: old TEs have accumulated mutations, young TEs are similar to consensus sequences. RepeatMasker was used to recover TE fragments in genomic assemblies (same method as Figure 2 [100],). Percentages of divergence to consensus were evaluated from RepeatMasker output .align file using A. Kapusta script [154]

upregulation of *Jheh1* and *Jheh2* [166]. Second, the spread of repressive epigenetic marks targeting TEs can reduce the expression of nearby genes (see below, host defenses against TEs), as it was also demonstrated in the *Jheh* cluster [167]. Lee and Karpen demonstrated recently that the spread of repressive epigenetic marks to nearby DNA occurs for more than half of euchromatic TEs, and can extend up to 20 kb [12]. This effect is TE dependent, copy number dependent, but also species dependent, with stronger epigenetic effect in *D. simulans* compared to *D. melanogaster*.

### On the individual

While some of the aforementioned genomic changes might remain phenotypicaly silent, others may have dramatic repercussions at the individual level. TEs are responsible for up to 80% of the phenotypic spontaneous mutations observed in *D. melanogaster* [168] and many observations suggest deleterious effects of TEs in *Drosophila*. Five to 10 % insertions of active *P-elements* are estimated to cause recessive lethal mutations in *D. melanogaster* [169]. In *D. simulans,* somatic transposition of *Mariner* decreases lifespan [170]. In 2004, a study used two *D. melanogaster* lines with the same genetic background, but different TE copy numbers, to evaluate the impact of TE number on fitness. The authors found differences in fitness and egg hatchability between the two lines, the line with more TEs performing worse than the other. Both homozygous and heterozygous TE insertions were shown to have deleterious effects on fitness and its components [134]. Overall, TE insertions are expected to be generally neutral or deleterious to the host genome [171]. Considering that adaptive mutations are supposed to quickly reach fixation in populations, the low numbers of fixed insertions in *D. melanogaster* and *D. simulans* support this theory. In 2006, Burt and Trivers calculated the number of insertions since the divergence between the two species and concluded that, given both genome size and number of fixed insertions, the occurrence and fixation of a beneficial insertion is a really rare event [156]. However, they also underscored the difficulty to detect fixed insertions using *in situ* hybridization, and suggested it would have been interesting to estimate the rate of fixation from sequencing data. In 2015, using population sequencing data, Kofler and colleagues estimated the number of fixed insertions in *D. melanogaster* since its divergence from *D. simulans* to be approximately 200 [155]. Considering a 1.4 Mya divergence [14], we computed a fixation rate of 1.4 fixed insertions every 10,000 years, *i.e.* maximum 1.4 beneficial fixed insertions every 10,000 years. If we update the Burt and Trivers calculation and compare the number of fixed insertions to the total number of insertions over this period: Population size × Insertion rate per genome per generation × Divergence time between *D. melanogaster* and *D. simulans* × Number of generations per year = 10 (6) × 0.57 × 1.4 × 10 (6) × 24 = 1.9 × 10 (13) insertions, that is to say 200/(1.9 × 10 (13)) = 1.0 × 10(-11) insertions reaching fixation. Finally, we estimated maximum 1.4 beneficial fixed insertions every 10,000 years, or maximum 1 out of 1e11 insertions, being beneficial and fixed. These numbers are upper bounds because all fixed insertions are unlikely to be beneficial. Indeed, most of the fixed insertions are present in regions where the effect of selection is weak, and are essentially old. Therefore, they are more likely to have reached fixation slowly by drift than quickly by positive selection [172, 173]. So far, 21 fixed insertions have been identified within or near genomic regions showing low Tajima's D values, and 12 fixed insertions are relatively young. Considering the above, one could expect to find very few putatively adaptive insertions among unfixed insertions. Surprisingly, there are at least 57 of such insertions in the reference genome [173], suggesting a high rate of TE mediated adaptation recently or even ongoing. The discrepancy between the number of candidates for recent adaptation and the fixation rate was discussed considering the three following points: 1. The migration of *D. melanogaster* out of Africa may have caused a significant augmentation of the adaptation rate. 2. TE derived adaptations might be ephemeral. 3. Adaptive TE sequences may evolve quicker than neutral insertions, resulting in an underestimation of the number of fixed insertions [174]. One may also add that the TE mutation rate has potentially increased recently [175]. It is worth noting that few insertions were clearly associated with an adaptive phenotype so far [166, 176–178]. Interestingly, candidate adaptive insertions are often close to, or within genes associated with stress response, behavior and development. Moreover, two of the historical examples of adaptation associated with TEs correspond to two different insertions in the same gene implicated in the response to oxidative stress, *cyp6g1,* in two different species: *D. melanogaster* and *D. simulans* [176, 178, 179].

### The case of telomeric elements

A few TEs appear to have evolved a new function in *Drosophila* genomes. Because of the DNA replication mechanism, a *Drosophila* chromosome end loses 70-80 bp each generation [180]. This gradual reduction of

chromosome ends is threatening internal regions containing essential genes and may contribute to ageing [181]. Organisms have evolved different mechanisms that protect their chromosomes. Usually in eukaryotic genomes a ribonucleoprotein enzyme, the telomerase, mediates the RNA dependent synthesis of tandemly repeated simple sequences at chromosome ends [182]. In *D. melanogaster*, the three families, *HeT-A*, *TART* and *TAHRE,* transpose to chromosome extremities, and protect them from shortening [180, 183–186]. Many phylogenetically distinct telomeric retrotransposons have been found in more distant species [187]. All these telomeric elements belong to a single monophyletic clade inside the *Jockey* superfamily. The telomeric element phylogeny and species phylogeny are congruent, suggesting vertical transmission from a common ancestor and a conserved host-element relationship [187]. Furthermore, the clade presents evidence of specialization to transpose at chromosome ends [188]. Because of this, the relationship between TEs and their host in this case was referred to as genomic "symbiosis" [188]. However, Saint-Leandre and colleagues investigated more species of the *melanogaster* group [189]. They suggest that these *Jockey* telomeric elements may have evolved to selfishly over-replicate. In agreement with this hypothesis, they found recurrent gains, losses, and replacements of *Jockey* telomeric elements. Moreover, in *D. biarmipes*, the telomere-specialized elements have disappeared completely.

## Host defenses

Because of the above-mentioned deleterious effect of TE insertions, several mechanisms of TE control have evolved. Among these, epigenetic modifications play an important role [190]. For example, in mammals and plants, TE insertions are usually associated with DNA methylation and histone modifications. Both are related to repressive chromatin states. In *Drosophila*, DNA methylation has been shown to be almost completely absent, and small RNAs are central to TE regulation [191, 192]. They may also trigger histone tail modifications and chromatin conformation modifications. There are two small RNA pathways controlling TEs in *Drosophila*: the piRNA and the siRNA pathways. Our purpose here is to give a brief overview of these pathways and their role in shaping TE dynamics. In particular, we refer the reader to [193, 194] for comprehensive reviews on the mechanistic aspects of the piRNA pathway.

### The piRNA pathway

The piRNA pathway produces small, single stranded RNAs that were first called rasiRNAs (repeat associated small interfering RNAs); however, contrary to regular small interfering RNAs, they are 23-30 nt long, and are associated with the Piwi-subfamily Argonaute proteins, which led to their new designation as piRNAs (piwi-interacting RNAs). These piRNAs silence TEs in germ cells, where maintaining the integrity of the genome is of primary importance, as new mutations are passed on to future generations. This pathway is also active in the ovarian somatic follicle cells, which support oogenesis. It prevents endogenous retroviruses, such as *Gypsy,* from infecting the adjacent oocyte [195]. Research studies in *Drosophila* were seminal in the piRNA field. Much of what we know today was discovered using this model. In fact, piRNAs were identified for the first time in 2001 in fly testis [196]. They were found to silence *Stellate,* a gene involved in male sterility. Some of them were even found to be homologous to TEs and assumed to be involved in transposon regulation. Moreover, a long-term study of the *Gypsy* family activity led to the discovery of *flamenco*, a non protein-coding locus producing piRNAs, which was subsequently shown to be involved in the control of other TE families, essentially LTR retrotransposons [197, 198] (Table 2).

piRNAs originate from discrete genomic loci called piRNAs clusters. These loci contain mainly defective TEs and are transcribed into long piRNA precursors (Fig. 4 [202]). Approximately 150 clusters have been identified in the genome of *D. melanogaster*, representing 3.5% of the assembled genome [208]. The vast majority of them appear to be heterochromatic. The size of piRNA clusters varies substantially, with the largest being 240 kb. Overall, the largest 15 clusters produce a

**Table 2** the *flamenco* story

Much of what we know today on the piRNA pathway was discovered using the *Drosophila* model. Especially, a considerable effort started 40 years ago in *D. melanogaster* led to the early discovery of a gene producing piRNAs and silencing *Gypsy* in a piwi dependent manner (see [199] for a review). This gene named *flamenco* was the first piRNA cluster identified.

In the 1980s, the *ovoD* dominant mutation was identified in *D. melanogaster* and associated with female sterility [200]. Interestingly, crosses between *ovoD* males and females from a particular strain led to the reversion of the phenotype and the recovery of fertility of the daughters, in addition to numerous mutations at other loci. Further work revealed that the particular strain used for the mothers actually displayed high copy numbers of uncontrolled *Gypsy*, whose transposition into *ovo* led to a null allele and reversion of sterility [201]. And then, it was demonstrated that the locus controlling *Gypsy* activity was the *flamenco* locus, located on the X chromosome and containing a lot of TE sequences [202, 203]. In 2004, Sarot and collaborators found out that *Gypsy* transposition was sensitive to a mutation in *piwi*, a gene known to affect RNA-mediated silencing [204]. They also demonstrated that small RNAs homologous to *Gypsy* where present in silenced tissues. For the first time a gene producing small RNAs was associated with TE silencing, and Piwi was implicated in this process.

Despite the subsequent discovery of many other piRNA clusters in *D. melanogaster*, flamenco is still widely studied as a model and produces most of the piRNAs in ovarian somatic cells [197, 205–207].

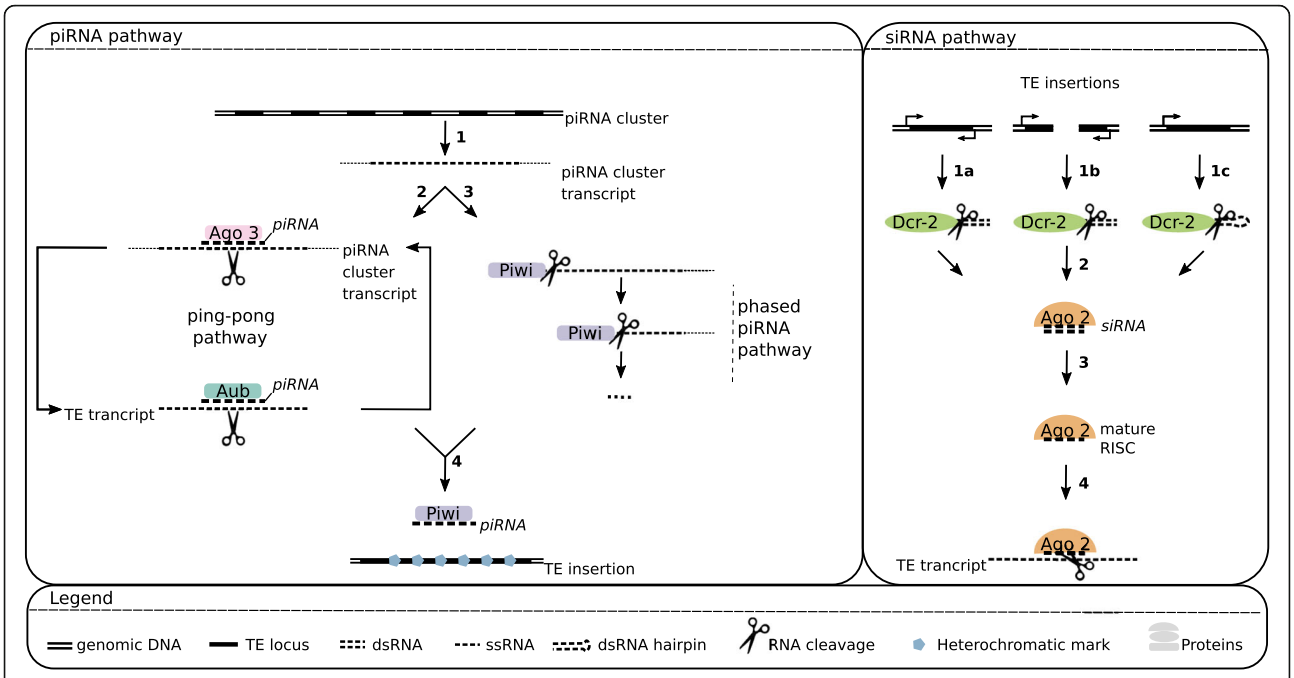



**Fig. 4** small RNA pathways controlling TEs. piRNA pathway: 1. RNA PolII transcribes a genomic piRNA cluster into a long single stranded RNA. 2. The transcript thus formed enters the ping-pong pathway, which is ensured by Aub and Ago3, generates sense and antisense piRNAs, and ensure post-transcriptional silencing by transcript slicing. 3. Piwi directs the cleavage of the piRNA cluster transcript and generates a piRNA. This step may be repeated. 4. Transcriptional silencing: in the nucleus, a piRNA guides Piwi and promotes H3K9 methylation of TE DNA sequences. siRNA pathway: 1. Generation of a long dsRNA by: a. bi-directional transcription of a unique TE locus, b. interaction of two complementary transcripts from distinct TE loci. c. hairpin formation, due for example to inverted repeats binding 2. Dcr-2 processes long dsRNAs into siRNAs, which are loaded on Ago2. 3. The passenger strand of the siRNA is sliced by Ago2, only the guide strand remains. 4. The RISC binds to a TE transcript with sequence complementarity to the guide strand and Ago2 cleaves it.

large proportion of the total amount of piRNAs: 70% of the piRNAs uniquely mapped to the genome originate from these clusters.

The beginning of piRNA biogenesis is similar in germline and somatic cells (see [193, 194] for detailed reviews). PiRNA cluster transcription is ensured by RNA Pol II and leads to a single stranded long RNA (Fig. 4). Then, piRNA cluster transcripts may enter either the ping-pong pathway or the phased piRNA pathway [208–213]. The ping-pong pathway occurs in germline cells. In this case, guided by a sense piRNA, Argonaute3 (Ago3) binds to a complementary piRNA cluster transcript and cleaves it. Then, Aubergine (Aub) attaches to the newly formed 5' extremity, slices the transcript and forms an antisense piRNA. Finally, guided by an antisense piRNA, Aub operates a cut in a TE transcript, Ago3 recognizes the resulting 5' extremity, cleaves the transcript and forms a sense piRNA. This is the ping-pong pathway or ping-pong loop. The phased piRNA pathway is not specific to germline cells and may also occur in ovarian somatic follicle cells. Piwi is loaded at the 5' extremity of the piRNA precursor and Zucchini (Zuc) performs cleavage, generating the piRNA. Piwi is then loaded again at the 5' extremity of the precursor piRNA, and the process is repeated in a step-by-step cleavage generating multiple piRNAs. Note that, for clarity, piRNA maturation steps such as trimming are not mentioned here.

After synthesis, piRNAs mediate silencing both at the transcriptional and post-transcriptional levels [214]. The post-transcriptional silencing occurs in the cytoplasm of germline cells only, and corresponds to the ping-pong pathway (Fig. 4). At the transcriptional level, a piRNA guides the Piwi protein to a TE insertion, probably due to sequence complementarity with nascent TE transcripts, and mediates local heterochromatin formation by addition of the repressive mark H3K9me3 to histone tails [215–221]. Note that, despite the fact that as early as 2001 piRNAs were detected in testes, so far most of the work on TE regulation by piRNAs has been done on ovaries [196]. Regulation in testes seems to be quite similar to what happens in female germline, with both ping-pong and phased piRNA pathways being active [222–224]. However, contrary to ovaries, the data suggest an Ago-3 independent amplification loop in spermatogenesis.

### The siRNA pathway

In addition to piRNAs, sequencing of small RNAs revealed the existence of another class of interfering RNAs targeting TEs: endogenous small interfering RNAs, or endo-siRNAs [225–227]. These small RNAs are present in both somatic and germline cells. endo-siRNA precursors are double strand RNAs (dsRNAs). These precursors may be produced through three distinct mechanisms (Fig. 4) [228]. 1. Transcription of the same genomic region in both sense and antisense directions (convergent transcription), then base pairing of the overlapping region between sense and antisense transcripts. 2. Transcription of complementary sense and antisense transcripts from different genomic regions and base pairing. 3. Base pairing of inverted repetitive elements of one transcript to form a hairpin RNA. The resulting long dsRNA is loaded on Dicer-2 (Dcr-2) and its cofactor Loquacious-PD (Loqs-PD) and then processed into 21 nt small double stranded RNAs. They are then loaded on the *RNA-induced silencing complex (*RISC) including the Ago2 protein. One strand is held and guides the complex to target transcripts that are then cleaved by the RNase domain of Ago2.

### Evolution

Several studies demonstrated rapid evolution of anti-TE RNAi genes in *Drosophila* [47, 229–232]. Indeed, these genes often present signatures of recurrent positive selection. By analogy to the signatures of positive selection observed for genes involved in host-parasite interactions, the rapid evolution of anti-TE RNAi genes is often interpreted as a consequence of an arms race occurring between TEs and TE immunity effectors. Focusing on the piRNA pathway, Blumenstiel and colleagues propose that selection for sensitivity to TE content but also selection for specificity to TE content may drive the rapid evolution of host defense mechanisms [233]. More precisely, concerning the specificity aspect, the authors propose that a too efficient piRNA pathway may induce a too efficient silencing of TE copies that could spread to neighboring genes, which would constitute a cost. They designated this form of off-target gene silencing as "genomic autoimmunity", an analogous to classic forms of autoimmunity which are caused by an immune response that incorrectly targets self. Despite the rapid evolution of anti-TE RNAi genes in *Drosophila*, suggesting that host defense mechanisms may vary a lot across the genus, most of the literature on this subject concerns *D. melanogaster*. A recent study of 20 arthopod species suggests that somatic piRNAs were probably produced in the ancestral arthropod more than 500 Mya and demonstrated that, in contrast to *D. melanogaster*, *D. virilis* presents somatic piRNAs [234]. This suggests a loss of the piRNA pathway in the soma of *D. melanogaster*.

## Population genomics

The *Drosophila* model has been of outstanding importance in the field of population genomics of TEs. The ease to get and maintain wild type strains was obviously a key factor, but so was the development of the *in situ* hybridization method on *Drosophila* polytene chromosomes more than 40 years ago [235, 236]. *In situ* hybridization allows to detect and localize genomic DNA sequences using a labeled sequence (probe) homologous to the targeted sequence. The giant polytene chromosomes are found only in some species and tissues, and offer to the researcher a high degree of resolution [237, 238]. Using TE probes on salivary gland polytene chromosomes of *Drosophila* third instar larvae, researchers were able to detect and localize TE insertions in individuals and thus to accurately estimate TE insertion frequencies in natural populations [239–241].

### About the nature of selection acting on TEs

The first *in situ* hybridization studies evaluating TE insertion frequencies in natural populations of *D. melanogaster* demonstrated a predominance of insertions segregating at low frequencies [239–241]. This result obtained for specific families was later confirmed at a broader scale. Population sequencing data showed that, in *D. melanogaster* and *D. simulans*, more than 80% of TE copies have insertion frequencies lower than 0.2 [155]. This observation is often interpreted as the result of purifying selection acting on TEs. So far, three main hypotheses have been formulated concerning the nature of selection against TEs: 1) the gene-disruption hypothesis [3, 242], 2) the ectopic recombination hypothesis [243, 244], 3) the deleterious TE-product expression hypothesis [245].

The gene disruption hypothesis assumes that insertions inside genes or regulatory regions are under strong purifying selection because of their negative effect on the host fitness [242]. A large amount of work supports this hypothesis, demonstrating a depletion of TE insertions in exons and untranslated regions [172, 246–248]. Moreover, Lee and Karpen demonstrated that repressive histone marks affecting euchromatic TEs can spread up to 20 kb both in *D. melanogaster* and *D. simulans*, and that this phenomenon is associated with selection against TEs [12]. Therefore, we may extend this hypothesis beyond insertions inside genes or regulatory regions to include insertions close to genes.

The ectopic recombination hypothesis states that purifying selection acts against chromosomal rearrangements resulting from recombination events between TE sequences showing sequence identity and located at distinct loci [243, 244]. According to this hypothesis, TE size, TE family copy number, and

meiotic recombination rate, expected to be positively correlated with ectopic recombination rate, should be associated with the strength of purifying selection [137]. First, since long insertions provide longer targets for recombination, one can indeed expect a stronger effect of purifying selection against long TEs in the ectopic recombination hypothesis. The negative correlation between TE size and population frequencies suggests that it is actually the case [172, 249]. Second, because ectopic recombination is more likely to occur when TEs are heterozygous, ectopic recombination should happen more frequently for TE families with a high copy number of polymorphic TEs. Therefore, the negative correlation between TE insertion frequencies and copy numbers also supports the ectopic recombination hypothesis [172, 249]. Finally, because ectopic recombination is intrinsically related to the local recombination rate, the fact that low-recombining regions are highly enriched in TEs, and that a negative correlation exists between insertion frequencies and recombination rate [172, 246, 249, 250], constitute one more argument in favor of the ectopic recombination hypothesis. However, this last point may be explained by the Hill-Robertson effect, or the lower density of genes in low-recombining regions, or both. The Hill-Robertson effect corresponds to a reduction in the efficiency of selection on a locus due to selection on related loci. If slightly deleterious insertions are close to adaptive mutations, they will be less efficiently removed in low-recombining regions than in high-recombining regions. The lower density of genes in low-recombining regions may explain the higher TE density in these regions because one may expect that TE insertions are strongly counter-selected close to genes (gene disruption hypothesis). However, one paradox exists when considering the ectopic recombination hypothesis. Indeed, considering the higher rate of recombination on the X chromosome, and the ectopic recombination hypothesis, TE density should be lower on the X chromosome [251]. However, recent studies of *D. melanogaster* natural populations show different results. TE density was found to be either higher on the X chromosome [246], or similar between the X chromosome and autosomes when taking into account differences in the amount of low recombining regions [172]. A higher transposition rate in the X chromosome relatively to autosomes has been proposed as a plausible explanation to the observed paradox [137]. Mutation accumulation data recently showed such tendency with a 1.86 fold change for insertion rate on the X chromosome relatively to autosomes [128].

One last hypothesis remains concerning the nature of the purifying selection affecting TEs: the deleterious TE-product expression hypothesis [245]. Under this model, transcription and translation of TEs may be resource consuming for the host and TE proteins could disrupt cellular processes. According to this hypothesis, and assuming that full length TEs are more transcribed than nearly complete copies, one may expect complete copies to be under more intense purifying selection than nearly complete copies. However, Petrov and colleagues did not find such effect investigating TE frequencies genome wide [249].

### Models of TE dynamics

So far, two main models have been formulated to conceptualize TE dynamics in *Drosophila* populations. The historical model is the transposition-selection balance model: it assumes that TE abundance is regulated by a balance between transposition and selection against TEs [3, 252]. According to this model, insertions with low frequency in populations are expected to be mainly insertions subjected to strong purifying selection. However, because transposition rates are not constant over time, another model has been proposed: the transposition burst model [175]. This model proposes that TE dynamics in populations is explained by transposition bursts. Under this hypothesis, a large proportion of low frequency insertions may result from recent TE activity rather than strong selection against TEs. Data, especially on TE genomic distribution (see above), suggest a pre-eminent role of purifying selection in TE dynamics, and thus support the transposition-selection balance model. Furthermore, an excess of rare TEs compared to the standard neutral model is found, as expected if selection acts against TEs [246]. However, confronting population data with simulation, Kofler and colleagues showed that both in *D. melanogaster* and *D. simulans*, 50% of families have temporally heterogeneous transposition rates and that a correlation exists between insertion frequencies and their age [155, 172]. So far, it is clear that both purifying selection and variation in transposition rate act on TE population dynamics. Until now, TE regulation has been poorly integrated in the models of TE dynamics. In 2010, Lu and colleagues incorporated piRNAs in a population genetics framework [253]. They used simulations to investigate the dynamics of TEs. They focused on retrotransposons, studying the retrotransposons that are targeted by piRNAs but also the retrotransposons generating piRNAs. The results indicate that: piRNAs may reduce TE fitness cost; TEs generating piRNAs may easily reach fixation because they confer a selective advantage; and TEs targeted by piRNAs may also reach fixation because host defenses reduce their deleterious effect. In 2013, the observation that a TE insertion inside a piRNA cluster was able to silence the corresponding TE family led to the formulation of the trap model

[197]. In this model, after invasion of a host genome, a TE family proliferates until it is trapped, *i.e.* one insertion occurs into a piRNA cluster, then the subsequent production of piRNAs silences the invading family. This model was validated and enriched with populational considerations by Kofler and colleagues [88]. Monitoring the *P-Element* invasion, in connection with the piRNA pathway, in experimentally evolving populations of *D. simulans*, they suggested the following three-step model for a TE invasion: 1) TE copies colonize the genome, 2) the first TE insertions in piRNA clusters occur but are not yet sufficient to stop TE proliferation and 3) the TE family is inactivated by the fixation of an insertion within a piRNA cluster. Using simulated data, they were able to demonstrate that this "trap model" accurately describes TE abundance in *D. melanogaster* germline. They also showed that the suppression of TE activity by segregating cluster insertions is reversible. Importantly, they demonstrated that transposition rates and population sizes affected mostly the duration of the invasion steps but not the amounts of accumulating TEs. In fact, the major factor capable of affecting the number of accumulating TEs was the piRNA cluster size.

## Conclusions

In today's biology research, increasing weight is given to the study of non-model species. This is clearly justified by the diversity of the living world, and even more so for the study of genetic elements as diverse and dynamic as TEs. However, we should not overlook model organisms, because the vast amount of techniques, data collected and knowledge will help us develop and test new hypotheses. Furthermore, the dissection of conserved pathways in these organisms, such as the piRNA pathway, should provide results valid for a broad range of species. Despite the fact that *Drosophila* is an old biological model, it still presents many opportunities for TE research. In general, studies of TEs could benefit from unified approaches to identifying and quantifying TEs. As we demonstrated above, the ultimate model *D. melanogaster* appears slightly different from its sister species regarding TEs —maybe related to the fact that it ended up as the ultimate model species— however, it is clear that the research community greatly benefits from comparative genomics in the *Drosophila* genus, and a great deal of work remains to be done in *Drosophila* and the species in the group in order to do proper comparative genomics. It is clear that the development of long-read technologies will greatly facilitate this work. Another challenge is to understand the activity of TEs and how, *in natura*, this activity is triggered and controlled. Once again, *Drosophila* is a model of excellence with the possibility of doing experimental evolution with a follow-up of TE dynamics. At the same time, this will allow a better understanding of the fine regulation systems of TE activity. Finally, it seems to us that one of the most exciting challenges is to understand the true impact of TEs in adaptive processes, even more so now, with all the gross changes in our environment. Experimental evolution, with different species and different environmental factors, are a real opportunity to move forward in this field.

### Abbreviations

Ago2: Argonaute2; Ago3: Argonaute3; Aub: Aubergine; Dcr-2: Dicer-2; DIRS: Dictyostelium Intermediate Repeat Sequence; Endo: Endonuclease; Env: Envelope gene; ERV: Endogenous RetroViruses; Int: Integrase; Hel: DNA helicase; LINE: Long Interspersed Nuclear Element; Loqs-PD: Loquacious-PD; LTR: Long Terminal Repeat; piRNA: piwi-interacting RNA; PLE: Penelope-Like Element; Prot: Protease; rasiRNA: repeat associated small interfering RNA; Rep: Replicator; RISC: *RNA-induced silencing complex*; *RNP*: *RiboNucleoProtein*; RT: Reverse Transcriptase; Tase: Transposase; TSD: Target Site Duplication; TE: Transposable Element; TIR: Terminal Inverted Repeat; SINE: Short Interspersed Nuclear Element; siRNA: small interfering RNA; Zuc: Zucchini

### Acknowledgments

The authors sincerely thank the anonymous reviewers.

### Authors' contributions

VM has drafted the initial version of the review and designed the figures; MB, MF and CV have contributed to the writing of the manuscript. All authors have approved the final version.

### Funding

This work was supported by the ANR Exhyb and ANR SWING (grant overseen by the French National Research Agency) and the CNRS.

### Availability of data and materials

The datasets analyzed during the current study are available in the following repositories: https://github.com/danrdanny/Drosophila15GenomesProject/raw/master/assembledGenomes/ [99], ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r6.29_FB2019_04/

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests

Received: 16 February 2020 Accepted: 14 April 2020
Published online: 03 July 2020

### References

1. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980;284:601–3.
2. Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. Nature. 1980;284:604–7.
3. Charlesworth B, Charlesworth D. The population dynamics of transposable elements. Genet Res. 1983;42:1–27.
4. Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. Nat Rev Genet. 2002;3:329–41.
5. Ravindran S. Barbara McClintock and the discovery of jumping genes. Proc Natl Acad Sci. 2012;109:20198–9.
6. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000;408:796.
7. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420:520.
8. C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. 1998;282:2012–8.

9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.
10. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5.
11. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
12. Lee YCG, Karpen GH. Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. eLife. 2017;6 https://doi.org/10.7554/eLife.25762.
13. Singh BN. Species and genetic diversity in the genus Drosophila inhabiting the Indian subcontinent. J Genet. 2015;94:351–61.
14. Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. Estimating divergence dates and substitution rates in the Drosophila phylogeny. Mol Biol Evol. 2012;29:3459–73.
15. Keller A. Drosophila melanogaster's history as a human commensal. Curr Biol CB. 2007;17:R77–81.
16. Markow TA. The secret lives of Drosophila flies. eLife. 2015;4 https://doi.org/10.7554/eLife.06793.
17. David JR, Capy P. Genetic variation of Drosophila melanogaster natural populations. Trends Genet TIG. 1988;4:106–11.
18. Hales KG, Korey CA, Larracuente AM, Roberts DM. Genetics on the Fly: A Primer on the Drosophila Model System. Genetics. 2015;201:815–42.
19. Rubin GM, Spradling AC. Genetic transformation of Drosophila with transposable element vectors. Science. 1982;218:348–53.
20. Sessegolo C, Burlet N, Haudry A. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. Biol Lett. 2016;12:20160407.
21. Genome List - Genome - NCBI. https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/drosophila. Accessed 4 Aug 2019.
22. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482:173–8.
23. Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, et al. Genomic Variation in Natural Populations of Drosophila melanogaster. Genetics. 2012;192:533–98.
24. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014;24:1193–208.
25. Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture. PLoS Genet. 2012;8 https://doi.org/10.1371/journal.pgen.1003080.
26. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, et al. The Drosophila Genome Nexus: A Population Genomic Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population. Genetics. 2015;199:1229–41.
27. Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, et al. Broad geographic sampling reveals predictable, pervasive, and strong seasonal adaptation in Drosophila. bioRxiv. 2019:337543. https://doi.org/10.1101/337543.
28. Kapun M, Barrón MG, Staubach F, Vieira J, Obbard DJ, Goubert C, et al. Genomic analysis of European Drosophila melanogaster populations on a dense spatial scale reveals longitudinal population structure and continent-wide selection. bioRxiv. 2019:313759. https://doi.org/10.1101/313759.
29. Grenier JK, Arguello JR, Moreira MC, Gottipati S, Mohammed J, Hackett SR, et al. Global diversity lines - a five-continent reference panel of sequenced Drosophila melanogaster strains. G3 Bethesda Md. 2015;5:593–603.
30. Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. Mol Biol Evol. 2016;33:3308–13.
31. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. Systems genetics of complex traits in Drosophila melanogaster. Nat Genet. 2009;41:299–307.
32. Shorter J, Couch C, Huang W, Carbone MA, Peiffer J, Anholt RRH, et al. Genetic architecture of natural variation in Drosophila melanogaster aggressive behavior. Proc Natl Acad Sci U S A. 2015;112:E3555–63.
33. Durham MF, Magwire MM, Stone EA, Leips J. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. Nat Commun. 2014;5:4338.
34. Weber AL, Khan GF, Magwire MM, Tabor CL, Mackay TFC, Anholt RRH. Genome-wide association analysis of oxidative stress resistance in Drosophila melanogaster. PLoS One. 2012;7:e34745.
35. Bloomington Drosophila Stock Center. Bloomingt. Drosoph. Stock Cent. https://bdsc.indiana.edu/stocks/stockdata.html. Accessed 27 Oct 2019.
36. The National Drosophila Species Stock Center | College of Agriculture and Life Science. http://blogs.cornell.edu/drosophila/. Accessed 4 Aug 2019.
37. Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. Historical biogeography of the Drosophila melanogaster species subgroup. Hist Biogeogr Drosoph Melanogaster Species Subgr. 1988;22:159–225.
38. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 2010;20:816–25.
39. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in Drosophila. Genome Res. 2014;24:797–808.
40. Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. Genome Res. 2014;24:84–95.
41. Bono JM, Markow TA. Post-zygotic isolation in cactophilic Drosophila: larval viability and adult life-history traits of D. mojavensis/D. arizonae hybrids. J Evol Biol. 2009;22:1387–95.
42. Lohse K, Clarke M, Ritchie MG, Etges WJ. Genome-wide tests for introgression between cactophilic Drosophila implicate a role of inversions during speciation. Evol Int J Org Evol. 2015;69:1178–90.
43. Sanchez-Flores A, Peñaloza F, Carpinteyro-Ponce J, Nazario-Yepiz N, Abreu-Goodger C, Machado CA, et al. Genome Evolution in Three Species of Cactophilic Drosophila. G3 GenesGenomesGenetics. 2016;6:3097–105.
44. Lopez-Maestre H, Carnelossi EAG, Lacroix V, Burlet N, Mugat B, Chambeyron S, et al. Identification of misexpressed genetic elements in hybrids between Drosophila-related species. Sci Rep. 2017;7:40618.
45. Vela D, Fontdevila A, Vieira C, García Guerreiro MP. A genome-wide survey of genetic instability by transposition in Drosophila hybrids. PLoS One. 2014;9:e88992.
46. García Guerreiro MP. Changes of Osvaldo expression patterns in germline of male hybrids between the species Drosophila buzzatii and Drosophila koepferae. Mol Genet Genomics MGG. 2015;290:1471–83.
47. Romero-Soriano V, Modolo L, Lopez-Maestre H, Mugat B, Pessia E, Chambeyron S, et al. Transposable Element Misregulation Is Linked to the Divergence between Parental piRNA Pathways in Drosophila Hybrids. Genome Biol Evol. 2017;9:1450–70.
48. Green MM. Transposable Elements in Drosophila and Other Diptera. Annu Rev Genet. 1980;14:109–20.
49. Ananiev EV, Ilyin YV. A comparative study of the location of mobile dispersed genes in salivary gland and midgut polytene chromosomes of Drosophila melanogaster. Chromosoma. 1981;82:429–35.
50. Picard G. Non-Mendelian Female Sterility in DROSOPHILA MELANOGASTER : Hereditary Transmission of I Factor. Genetics. 1976;83:107–23.
51. Kidwell MG, Kidwell JF, Sved JA. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics. 1977;86:813–33.
52. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet. 2008;9:411–2 author reply 414.
53. Seberg O, Petersen G. A unified classification system for eukaryotic transposable elements should reflect their phylogeny. Nat Rev Genet. 2009;10:276.
54. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. Reply: A unified classification system for eukaryotic transposable elements should reflect their phylogeny. Nat Rev Genet. 2009;10:276.
55. Hill T. Transposable element dynamics are consistent across the Drosophila phylogeny, despite drastically differing content. bioRxiv. 2019:651059. https://doi.org/10.1101/651059.
56. GIRI. https://www.girinst.org/repbase/. Accessed 16 Feb 2020.
57. Mount SM, Rubin GM. Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol. 1985;5:1630–8.
58. Marlor RL, Parkhurst SM, Corces VG. The Drosophila melanogaster gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol Cell Biol. 1986;6:1129–34.
59. Lindsley DL, Zimm GG. The Genome of Drosophila Melanogaster. San Diego: Academic Press; 2012.
60. McCullers TJ, Steiniger M. Transposable elements in Drosophila. Mob Genet Elem. 2017;7:1–18.

61. Dunn BM, Goodenow MM, Gustchina A, Wlodawer A. Retroviral proteases. Genome Biol. 2002;3 https://doi.org/10.1186/gb-2002-3-4-reviews3006.
62. Shiba T, Saigo K. Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. Nature. 1983;302: 119–24.
63. Arkhipova IR, Mazo AM, Cherkasova VA, Gorelova TV, Schuppe NG, Llyin YV. The steps of reverse transcription of Drosophila mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. Cell. 1986;44:555–63.
64. Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C. Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of Drosophila melanogaster. Mol Biol Evol. 1999;16:54–66.
65. Akkouche A, Grentzinger T, Fablet M, Armenise C, Burlet N, Braman V, et al. Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. EMBO Rep. 2013;14:458–64.
66. Terzian C, Pélisson A, Bucheton A. Evolution and phylogeny of insect endogenous retroviruses. BMC Evol Biol. 2001;1:3.
67. Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. Proc Natl Acad Sci. 1994; 91:1285–9.
68. Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, et al. Life Cycle of an Endogenous Retrovirus, ZAM, in Drosophila melanogaster. J Virol. 2000;74:10658–69.
69. Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. The molecular basis of I-R hybrid Dysgenesis in Drosophila melanogaster: Identification, cloning, and properties of the I factor. Cell. 1984;38:153–63.
70. Fawcett DH, Lister CK, Kellett E, Finnegan DJ. Transposable elements controlling I-R hybrid dysgenesis in D. melanogaster are similar to mammalian LINEs. Cell. 1986;47:1007–15.
71. Priimägi AF, Mizrokhi LJ, Ilyin YV. The Drosophila mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene. 1988;70:253–62.
72. Dawson A, Hartswood E, Paterson T, Finnegan DJ. A LINE-like transposable element in Drosophila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. EMBO J. 1997;16:4448–55.
73. Martin SL, Bushman FD. Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. Mol Cell Biol. 2001;21:467–75.
74. Finnegan DJ. Transposable elements: How non-LTR retrotransposons do it. Curr Biol. 1997;7:R245–8.
75. Pélisson A, Finnegan DJ, Bucheton A. Evidence for retrotransposition of the I factor, a LINE element of Drosophila melanogaster. Proc Natl Acad Sci U S A. 1991;88:4907–10.
76. del Carmen SM, Disson O, Robin S, Brun C, Teninges D, Bucheton A. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. Nucleic Acids Res. 2005;33:776–85.
77. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob DNA. 2010;1:15.
78. Petrov DA, Hartl DL. High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol Biol Evol. 1998;15:293–302.
79. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity. 2011;107:487–95.
80. Yannopoulos G, Stamatis N, Monastirioti M, Hatzopoulos P, Louis C. hobo is responsible for the induction of hybrid dysgenesis by strains of Drosophila melanogaster bearing the male recombination factor 23.5MRF. Cell. 1987;49: 487–95.
81. Hill T, Schlötterer C, Betancourt AJ. Hybrid Dysgenesis in Drosophila simulans Associated with a Rapid Invasion of the P-Element. PLoS Genet. 2016;12:e1005920.
82. Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER. Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A. 1995;92:8050–4.
83. Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A. 1997;94: 196–201.
84. Vieira J, Vieira CP, Hartl DL, Lozovskaya ER. Factors contributing to the hybrid dysgenesis syndrome in Drosophila virilis. Genet Res. 1998;71:109–17.
85. Blumenstiel JP. Whole genome sequencing in Drosophila virilis identifies Polyphemus, a recently activated Tc1-like transposon with a possible role in hybrid dysgenesis. Mob DNA. 2014;5:6.
86. Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, Pelisson A, et al. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. Proc Natl Acad Sci. 2008;105:14964–9.
87. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. Evidence for Horizontal Transmission of the P Transposable Element between Drosophila Species. Genetics. 1990;124:339–55.
88. Kofler R, Senti K-A, Nolte V, Tobler R, Schlötterer C. Molecular dissection of a natural transposable element invasion. Genome Res. 2018;28:824–35.
89. Evgen'ev MB. What happens when Penelope comes? Mob Genet Elem. 2013;3:e24542.
90. Jacobson JW, Medhora MM, Hartl DL. Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci U S A. 1986;83:8684–8.
91. O'Hare K, Rubin GM. Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell. 1983; 34:25–35.
92. Zhang L, Dawson A, Finnegan DJ. DNA-binding activity and subunit interaction of the mariner transposase. Nucleic Acids Res. 2001;29:3566–75.
93. Tang M, Cecconi C, Kim H, Bustamante C, Rio DC. Guanosine triphosphate acts as a cofactor to promote assembly of initial P-element transposase–DNA synaptic complexes. Genes Dev. 2005;19:1422.
94. Chen J, Greenblatt IM, Dellaporta SL. Molecular Analysis of Ac Transposition and DNA Replication. Genetics. 1992;130:665–76.
95. Nassif N, Penney J, Pal S, Engels WR, Gloor GB. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol. 1994;14:1613–25.
96. Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 2001;98:8714–9.
97. Kapitonov VV, Jurka J. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 2007;23:521–9.
98. Grabundzija I, Hickman AB, Dyda F. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. Nat Commun. 2018;9 https://doi.org/10.1038/s41467-018-03688-w.
99. Miller DE, Staber C, Zeitlinger J, Hawley RS. Highly Contiguous Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. G3 GenesGenomesGenetics. 2018;8:3131–41.
100. Smit A, Hubley R, Green P. RepeatMasker Home Page. 2013. http://www.repeatmasker.org/. Accessed 30 Jan 2020.
101. Bailly-Bechet M, Haudry A, Lerat E. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mob DNA. 2014;5:13.
102. Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). Genome Biol Evol. 2015;7:1192–205.
103. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, et al. The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. 2002;3:research0084.1–2.
104. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. PLoS One. 2011; 6:e16526.
105. Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007;450:203–18.
106. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. 2015;25:445–58.
107. Dfam. https://www.dfam.org/family/DF0001586/summary. Accessed 2 Feb 2020.
108. Thomas J, Vadnagara K, Pritham EJ. DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). Mob DNA. 2014;5:18.
109. Lerat E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity. 2010;104:520–33.
110. Rius N, Guillén Y, Delprat A, Kapusta A, Feschotte C, Ruiz A. Exploration of the Drosophila buzzatii transposable element content suggests underestimation of repeats in Drosophila genomes. BMC Genomics. 2016;17 https://doi.org/10.1186/s12864-016-2648-8.
111. Chiu JC, Jiang X, Zhao L, Hamm CA, Cridland JM, Saelao P, et al. Genome of Drosophila suzukii, the Spotted Wing Drosophila. G3 GenesGenomesGenetics. 2013;3:2257–71.

112. Paris M, Boyer R, Jaenichen R, Wolf J, Karageorgi M, Green J, et al. Near-chromosome level genome assembly of the fruit pest Drosophila suzukii using long-read sequencing. bioRxiv. 2020. https://doi.org/10.1101/2020.01.02.892844.
113. Repbase Reports - 2012, Volume 12, Issue 9. https://www.girinst.org/2012/vol12/issue9/. Accessed 31 Jan 2020.
114. Kidwell MG, Evgen'ev MB. How valuable are model organisms for transposable element studies? Genetica. 1999;107:103.
115. Vieira C, Nardon C, Arpin C, Lepetit D, Biémont C. Evolution of Genome Size in Drosophila. Is the Invader's Genome Being Invaded by Transposable Elements? Mol Biol Evol. 2002;19:1154–61.
116. Bingham PM, Kidwell MG, Rubin GM. The molecular basis of P-M hybrid dysgenesis: The role of the P element, a P-strain-specific transposon family. Cell. 1982;29:995–1004.
117. Kidwell MG. Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1983;80:1655–9.
118. Anxolabéhère D, Kidwell MG, Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol Biol Evol. 1988;5:252–69.
119. Bonnivard H. Stability of European natural populations of Drosophila melanogaster with regard to the P -M system: a buffer zone made up of Q populations. J Evol Biol. 1999;12:633–47.
120. Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. The recent invasion of natural Drosophila simulans populations by the P-element. Proc Natl Acad Sci. 2015;112:6659–63.
121. Panaud O. Horizontal transfers of transposable elements in eukaryotes: The flying genes. C R Biol. 2016;339:296–9.
122. Bartolomé C, Bello X, Maside X. Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. Genome Biol. 2009;10:R22.
123. Vieira C, Lepetit D, Dumont S, Biémont C. Wake up of transposable elements following Drosophila simulans worldwide colonization. Mol Biol Evol. 1999;16:1251–5.
124. Biémont C, Vieira C, Borie N, Lepetit D. Transposable elements and genome evolution: the case of Drosophila simulans. Genetica. 1999;107:113–20.
125. Vieira C, Biémont C. Geographical variation in insertion site number of retrotransposon 412 in Drosophila simulans. J Mol Evol. 1996;42:443.
126. García Guerreiro MP, Chávez-Sandoval BE, Balanyà J, Serra L, Fontdevila A. Distribution of the transposable elements bilbo and gypsy in original and colonizing populations of Drosophila subobscura. BMC Evol Biol. 2008;8:234.
127. García Guerreiro MP, Fontdevila A. Osvaldo and Isis retrotransposons as markers of the Drosophila buzzatii colonisation in Australia. BMC Evol Biol. 2011;11 https://doi.org/10.1186/1471-2148-11-111.
128. Adrion JR, Song MJ, Schrider DR, Hahn MW, Schaack S. Genome-Wide Estimates of Transposable Element Insertion and Deletion Rates in Drosophila Melanogaster. Genome Biol Evol. 2017;9:1329–40.
129. Keightley PD, Ness RW, Halligan DL, Haddrill PR. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics. 2014;196:313–20.
130. Harada K, Yukuhiro K, Mukai T. Transposition rates of movable genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1990;87:3248–52.
131. Nuzhdin SV, Mackay TF. The genomic rate of transposable element movement in Drosophila melanogaster. Mol Biol Evol. 1995;12:180–1.
132. Maside X, Bartolomé C, Assimacopoulos S, Charlesworth B. Rates of movement and distribution of transposable elements in Drosophila melanogaster: in situ hybridization vs Southern blotting data. Genet Res. 2001;78:121–36.
133. Biémont C, Aouar A, Arnault C. Genome reshuffling of the copia element in an inbred line of Drosophila melanogaster. Nature. 1987;329:742–4.
134. Pasyukova EG, Nuzhdin SV. Doc and copia instability in an isogenic Drosophila melanogaster stock. Mol Gen Genet MGG. 1993;240:302–6.
135. Díaz-González J, Vázquez JF, Albornoz J, Domínguez A. Long-term evolution of the roo transposable element copy number in mutation accumulation lines of Drosophila melanogaster. Genet Res. 2011;93:181–7.
136. Guerreiro MPG. What makes transposable elements move in the Drosophila genome? Heredity. 2012;108:461–8.
137. Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. Population Genomics of Transposable Elements in Drosophila. Annu Rev Genet. 2014;48:561–81.
138. Biémont C, Arnault C, Heizmann A, Ronsseray S. Massive changes in genomic locations of P elements in an inbred line of Drosophila melanogaster. Naturwissenschaften. 1990;77:485–8.
139. Vasilyeva LA, Bubenshchikova EV, Ratner VA. Heavy heat shock induced retrotransposon transposition in Drosophila. Genet Res. 1999;74:111–9.
140. Zabanov SA, Vasil'eva LA, Ratner VA. Induction of transposition of MGE Dm412 using gamma-irradiation of an isogenic line of Drosophila melanogaster. Genetika. 1995;31:798–803.
141. Vasil'eva LA, Ratner VA, Antonenko OV, Lopukhova ED, Bubenshchikova EV. Induction of MGE 412 transposition in an isogenic strain of Drosophila melanogaster by different doses of ethanol fumes. Genetika. 2003;39:717–20.
142. Nabirochkin SD, Gabitova L, Ossokina MA, Soldatov AV, Gazaryan TG, Gazaryan KG. Oncoviral DNAs induce transposition of endogenous mobile elements in the genome of Drosophila melanogaster. Mutat Res Mol Mech Mutagen. 1998;403:127–36.
143. Horváth V, Merenciano M, González J. Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. Trends Genet. 2017;33:832–41.
144. Baack EJ, Whitney KD, Rieseberg LH. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. New Phytol. 2005;167:623–30.
145. Metcalfe CJ, Bulazel KV, Ferreri GC, Schroeder-Reiter E, Wanner G, Rens W, et al. Genomic instability within centromeres of interspecific marsupial hybrids. Genetics. 2007;177:2507–17.
146. Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites? Heredity. 2000;85:101–6.
147. Strand DJ, McDonald JF. Copia is transcriptionally responsive to environmental stress. Nucleic Acids Res. 1985;13:4401–10.
148. Chakrani F, Capy P, David J. Developmental temperature and somatic excision rate of mariner transposable element in three natural populations of Drosophila simulans. Genet Sel Evol. 1993;25:121.
149. Jakšić AM, Kofler R, Schlötterer C. Regulation of transposable elements: Interplay between TE-encoded regulatory sequences and host-specific trans-acting factors in Drosophila melanogaster. Mol Ecol. 2017;26:5149–59.
150. Yang H-P, Hung T-L, You T-L, Yang T-H. Genomewide Comparative Analysis of the Highly Abundant Transposable Element DINE-1 Suggests a Recent Transpositional Burst in Drosophila yakuba. Genetics. 2006;173:189–96.
151. Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A-B, Vieira C, et al. Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. Mol Ecol. 2019;28:1506–22.
152. Giraud T, Capy P. Somatic activity of the mariner transposable element in natural populations of Drosophila simulans. Proc Biol Sci. 1996;263:1481–6.
153. Lerat E, Burlet N, Biémont C, Vieira C. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene. 2011;473:100–9.
154. Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. Ann N Y Acad Sci. 2017;1389:164–85.
155. Kofler R, Nolte V, Schlötterer C. Tempo and Mode of Transposable Element Activity in Drosophila. PLoS Genet. 2015;11:e1005406.
156. Burt A, Trivers R. Genes in conflict: the biology of selfish genetic elements. Cambridge: Belknap Press of Harvard University Press; 2006.
157. Lim JK, Simmons MJ. Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. BioEssays. 1994;16:269–75.
158. Lyttle TW, Haymer DS. The role of the transposable element hobo in the origin of endemic inversions in wild populations of Drosophila melanogaster. Genetica. 1992;86:113–26.
159. Cáceres M, Ranz JM, Barbadilla A, Long M, Ruiz A. Generation of a Widespread Drosophila Inversion by a Transposable Element. Science. 1999;285:415–8.
160. Delprat A, Negre B, Puig M, Ruiz A. The Transposon Galileo Generates Natural Chromosomal Inversions in Drosophila by Ectopic Recombination. PLoS ONE. 2009;4 https://doi.org/10.1371/journal.pone.0007883.
161. Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikodvorskaja V, Pyatkov KI, et al. Mobile Elements and Chromosomal Evolution in the Virilis Group of Drosophila. Proc Natl Acad Sci U S A. 2000;97:11337–42.
162. Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM. Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci. 1995;92:10824–30.

163. Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, et al. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. Genetics. 1999; 153:135–77.
164. Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, et al. The BDGP Gene Disruption Project: Single Transposon Insertions Associated With 40% of Drosophila Genes. Genetics. 2004;167:761–81.
165. Rebollo R, Romanish MT, Mager DL. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annu Rev Genet. 2012;46:21–42.
166. Guio L, Barrón MG, González J. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Mol Ecol. 2014;23:2020–30.
167. Guio L, Vieira C, González J. Stress affects the epigenetic marks added by natural transposable element insertions in Drosophila melanogaster. Sci Rep. 2018;8 https://doi.org/10.1038/s41598-018-30491-w.
168. Green MM. Mobile elements and spontaneous gene mutations (1988) Banbury Rep. 30: Eukaryotic transposable elements as mutagenic agents. NY Cold Spring Harb Lab 1988:41 -50.
169. Cooley L, Kelley R, Spradling A. Insertional mutagenesis of the Drosophila genome with single P elements. Science. 1988;239:1121–8.
170. Nikitin AG, Woodruff RC. Somatic movement of the mariner transposable element and lifespan of Drosophila species. Mutat Res. 1995;338:43–9.
171. Charlesworth B, Langley CH, Sniegowski PD. Transposable element distributions in Drosophila. Genetics. 1997;147:1993–5.
172. Kofler R, Betancourt AJ, Schlötterer C. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. PLoS Genet. 2012;8:e1002487.
173. Rech GE, Bogaerts-Márquez M, Barrón MG, Merenciano M, Villanueva-Cañas JL, Horváth V, et al. Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila. PLoS Genet. 2019;15:e1007900.
174. González J, Petrov DA. The Adaptive Role of Transposable Elements in the Drosophila Genome. Gene. 2009;448:124–33.
175. Bergman CM, Bensasson D. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in Drosophila melanogaster. Proc Natl Acad Sci. 2007;104:11340–5.
176. Daborn PJ, Yen JL, Bogwitz MR, Goff GL, Feil E, Jeffers S, et al. A Single P450 Allele Associated with Insecticide Resistance in Drosophila. Science. 2002; 297:2253–6.
177. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, et al. Cis-Regulatory Elements in the Accord Retrotransposon Result in Tissue-Specific Expression of the Drosophila melanogaster Insecticide Resistance Gene Cyp6g1. Genetics. 2007;175:1071–7.
178. Schlenke TA, Begun DJ. Strong selective sweep associated with a transposon insertion in Drosophila simulans. Proc Natl Acad Sci. 2004;101:1626–31.
179. Carareto CMA, Hernandez EH, Vieira C. 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;537:93–9.
180. Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen F. Transposons in place of telomeric repeats at a Drosophila telomere. Cell. 1993;75:1083–93.
181. Whittemore K, Vera E, Martínez-Nevado E, Sanpera C, Blasco MA. Telomere shortening rate predicts species life span. Proc Natl Acad Sci. 2019;116: 15122–7.
182. Blackburn EH. Telomerases. Annu Rev Biochem. 1992;61:113–29.
183. Young BS, Pession A, Traverse KL, French C, Pardue ML. Telomere regions in Drosophila share complex DNA sequences with pericentric heterochromatin. Cell. 1983;34:85–94.
184. Traverse KL, Pardue ML. A spontaneously opened ring chromosome of Drosophila melanogaster has acquired He-T DNA sequences at both new telomeres. Proc Natl Acad Sci U S A. 1988;85:8116–20.
185. Biessmann H, Mason JM, Ferry K, d'Hulst M, Valgeirsdottir K, Traverse KL, et al. Addition of telomere-associated HeT DNA sequences "heals"broken chromosome ends in Drosophila. Cell. 1990;61:663–73.
186. Abad JP, de Pablos B, Osoegawa K, de Jong PJ, Martín-Gallardo A, Villasante A. TAHRE, a Novel Telomeric Retrotransposon from Drosophila melanogaster, Reveals the Origin of Drosophila Telomeres. Mol Biol Evol. 2004;21:1620–4.
187. Villasante A, Abad JP, Planello R, Mendez-Lago M, Celniker SE, de Pablos B. Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res. 2007;17:1909–18.
188. Pardue M-L, DeBaryshe PG. Drosophila telomeres: A variation on the telomerase theme. Fly (Austin). 2008;2:101–10.
189. Saint-Leandre B, Nguyen SC, Levine MT. Diversification and collapse of a telomere elongation mechanism. Genome Res. 2019;29:920–31.
190. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007;8:272–85.
191. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA Methylation across Insects. Mol Biol Evol. 2017;34:654–65.
192. Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola. Genome Biol Evol. 2018;10:1185–97.
193. Ozata DM, Gainetdinov I, Zoch A, OCarroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019;20:89–108.
194. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41:324–37.
195. Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, et al. Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. EMBO J. 1999;18:2659–69.
196. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol. 2001;11: 1017–27.
197. Zanni V, Eymery A, Coiffet M, Zytnicki M, Luyten I, Quesneville H, et al. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci U S A. 2013;110:19842–7.
198. Mevel-Ninio MT, Pelisson A, Kinder J, Campos AR, Bucheton A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics. 2007. https://doi.org/10.1534/genetics.106.068106.
199. Goriaux C, Théron E, Brasset E, Vaury C. History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. Front Genet. 2014;5 https://doi.org/10.3389/fgene.2014.00257.
200. Busson D, Gans M, Komitopoulou K, Masson M. Genetic Analysis of Three Dominant Female-Sterile Mutations Located on the X Chromosome of DROSOPHILA MELANOGASTER. Genetics. 1983;105:309–25.
201. Mével-Ninio M, Mariol M-C, Gans M. Mobilization of the gypsy and copia retrotransposons in Drosophila melanogaster induces reversion of the ovoD dominant female-sterile mutations: molecular analysis of revertant alleles. EMBO J. 1989;8:1549–58.
202. Robert V, Prudhomme N, Kim A, Bucheton A, Pélisson A. Characterization of the flamenco Region of the Drosophila melanogaster Genome. Genetics. 2001;158:701–13.
203. Prudhomme N, Gans M, Masson M, Terzian C, Bucheton A. Flamenco, a Gene Controlling the Gypsy Retrovirus of Drosophila Melanogaster. Genetics. 1995;139:697–711.
204. Sarot E, Payen-Groschêne G, Bucheton A, Pélisson A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics. 2004;166:1313–21.
205. Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep. 2014;15:411–8.
206. Dennis C, Brasset E, Sarkar A, Vaury C. Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in Drosophila. Nat Commun. 2016;7:1–12.
207. Sokolova OA, Ilyin AA, Poltavets AS, Nenasheva VV, Mikhaleva EA, Shevelyov YY, et al. Yb body assembly on the flamenco piRNA precursor transcripts reduces genic piRNA production. Mol Biol Cell. 2019;30:1544–54.
208. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128:1089–103.
209. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348:812–7.
210. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315:1587–90.
211. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137: 509–21.

212. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137:522–35.
213. Han BW, Wang W, Li C, Weng Z, Zamore PD. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348:817–21.
214. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 2006;20: 2214–22.
215. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013;27:390–9.
216. Darricarrère N, Liu N, Watanabe T, Lin H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc Natl Acad Sci U S A. 2013;110:1297–302.
217. Iwasaki YW, Murano K, Ishizu H, Shibuya A, Iyoda Y, Siomi MC, et al. Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. Mol Cell. 2016;63:408–19.
218. Klenov MS, Lavrov SA, Korbut AP, Stolyarenko AD, Yakushev EY, Reuter M, et al. Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res. 2014;42:6208–18.
219. Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev. 2013;27:400–12.
220. Sienski G, Dönertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151:964–80.
221. Wang SH, Elgin SCR. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A. 2011;108:21164–9.
222. Vagin VV, Sigova A, Li C, Seite H, Gvozdev V, Zamore PD. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science. 2006;313:320–4.
223. Nagao A, Mituyama T, Huang H, Chen D, Siomi MC, Siomi H. Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. RNA N Y N. 2010;16:2503–15.
224. Quénerch'du E, Anand A, Kai T. The piRNA pathway is developmentally regulated during spermatogenesis in Drosophila. RNA N Y N. 2016;22:1044–54.
225. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. 2008;320:1077–81.
226. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature. 2008; 453:793–7.
227. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. An endogenous small interfering RNA pathway in Drosophila. Nature. 2008;453: 798–802.
228. Claycomb JM. Ancient endo-siRNA pathways reveal new tricks. Curr Biol CB. 2014;24:R703–15.
229. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc B Biol Sci. 2009;364:99–115.
230. Kolaczkowski B, Hupalo DN, Kern AD. Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol Biol Evol. 2011;28:1033–42.
231. Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD. Recurrent and recent selective sweeps in the piRNA pathway. Evol Int J Org Evol. 2013;67:1081–90.
232. Fablet M, Akkouche A, Braman V, Vieira C. Variable expression levels detected in the Drosophila effectors of piRNA biogenesis. Gene. 2014;537: 149–53.
233. Blumenstiel JP, Erwin AA, Hemmer LW. What Drives Positive Selection in the Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis. Yale J Biol Med. 2016;89:499–512.
234. Lewis SH, Quarles KA, Yang Y, Tanguy M, Frézal L, Smith SA, et al. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol. 2018;2:174–81.
235. Pardue ML, Gerbi SA, Eckhardt RA, Gall JG. Cytological localization of DNA complementary to ribosomal RNA in polytene chromosomes of Diptera. Chromosoma. 1970;29:268–90.
236. Biémont C, Monti-Dedieu L, Lemeunier F. Detection of transposable elements in Drosophila salivary gland polytene chromosomes by in situ hybridization. Methods Mol Biol Clifton NJ. 2004;260:21–8.
237. Saunders R. In Situ Hybridization to Polytene Chromosomes. Methods Mol Biol Clifton NJ. 2000;123:103–13.
238. Stormo BM, Fox DT. Polyteny: still a giant player in chromosome research. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2017;25: 201–14.
239. Brown AJL, Moss JE. Transposition of the I element and copia in a natural population of Drosophila melanogaster. Genet Res. 1987;49:121–8.
240. Charlesworth B, Lapid A, Canada D. The distribution of transposable elements within and between chromosomes in a population of Drosophila melanogaster. I. Element frequencies and distribution. Genet Res. 1992;60: 103–14.
241. Biémont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, Aulard S, et al. Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of Drosophila melanogaster. Genet Res. 1994;63:197–212.
242. Montgomery E, Charlesworth B, Langley CH. A test for the role of natural selection in the stabilization of transposable element copy number in a population of Drosophila melanogaster. Genet Res. 1987;89:435–45.
243. Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. On the role of unequal exchange in the containment of transposable element copy number. Genet Res. 1988;52:223–35.
244. Finnegan DJ. Transposable elements. Curr Opin Genet Dev. 1992;2:861–7.
245. Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica. 1999;107:129–37.
246. Cridland JM, Macdonald SJ, Long AD, Thornton KR. Abundance and Distribution of Transposable Elements in Two Drosophila QTL Mapping Resources. Mol Biol Evol. 2013;30:2311–27.
247. Deloger M, Cavalli FMG, Lerat E, Biémont C, Sagot M-F, Vieira C. Identification of expressed transposable element insertions in the sequenced genome of Drosophila melanogaster. Gene. 2009;439:55–62.
248. Lipatov M, Lenkov K, Petrov DA, Bergman CM. Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. BMC Biol. 2005;3:24.
249. Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. Population genomics of transposable elements in Drosophila melanogaster. Mol Biol Evol. 2011;28:1633–44.
250. Bartolomé C, Maside X, Charlesworth B. On the Abundance and Distribution of Transposable Elements in the Genome of Drosophila melanogaster. Mol Biol Evol. 2002;19:926–37.
251. Comeron JM, Ratnappan R, Bailin S. The Many Landscapes of Recombination in Drosophila melanogaster. PLoS Genet. 2012;8:e1002905.
252. Charlesworth B, Jarne P, Assimacopoulos S. The distribution of transposable elements within and between chromosomes in a population of Drosophila melanogaster. III. Element abundances in heterochromatin. Genet Res. 1994; 64:183–97.
253. Lu J, Clark AG. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res. 2010;20:212–27.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

**Learn more** biomedcentral.com/submissions

BMC