Population Genomics of Transposable Elements in *Drosophila*

Maite G. Barrón,[1] Anna-Sophie Fiston-Lavier,[2] Dmitri A. Petrov,[3,*] and Josefa González[1,*]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain 08003; email: maite.barron@ibe.upf-csic.es, josefa.gonzalez@ibe.upf-csic.es

[2]Institut des Sciences de l'Evolution (ISEM), UMR5554 CNRS-Université Montpellier 2, France 34090; email: asfiston@univ-montp2.fr

[3]Department of Biology, Stanford University, Stanford, California 94305; email: dpetrov@stanford.edu

Annu. Rev. Genet. 2014. 48:561–81

First published online as a Review in Advance on October 1, 2014

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev-genet-120213-092359

*Contributed equally to this review

Keywords

evolutionary models, next-generation sequencing, adaptation

Abstract

Studies of the population dynamics of transposable elements (TEs) in *Drosophila melanogaster* indicate that consistent forces are affecting TEs independently of their modes of transposition and regulation. New sequencing technologies enable biologists to sample genomes at an unprecedented scale in order to quantify genome-wide polymorphism for annotated and novel TE insertions. In this review, we first present new insights gleaned from high-throughput data for population genomics studies of *D. melanogaster*. We then consider the latest population genomics models for TE evolution and present examples of functional evidence revealed by genome-wide studies of TE population dynamics in *D. melanogaster*. Although most of the TE insertions are deleterious or neutral, some TE insertions increase the fitness of the individual that carries them and play a role in genome adaptation.

TRANSPOSABLE ELEMENTS ARE AN ABUNDANT, DIVERSE, AND ACTIVE COMPONENT OF GENOMES

Nonautonomous TE elements: elements that do not encode the machinery to transpose but can still be mobile using the machinery of the autonomous TE copies

Autonomous TE elements: elements that encode the machinery to transpose

Ectopic recombination: exchange of DNA sequences, between two similar sequences located at nonhomologous regions of the genome leading to a chromosomal rearrangement

Polytene chromosomes: giant chromosomes that display a characteristic band-interband morphology; present in several tissue, they are commonly found in the salivary glands of flies; formed as the result of several rounds of DNA replication without cell division

Transposable elements (TEs) are mobile DNA sequences that encode an ability to copy themselves to other sites in the genome and increase their copy number in the process. Certain TEs, known as nonautonomous, rely on the enzymatic machinery of autonomous copies to move around the genome. Owing to mobility and self-replication ability, TEs can be abundant, diverse, and active components of genomes.

TEs are present in virtually all eukaryotic organisms studied to date and in 80% of the sequenced prokaryotes (112). In all of these organisms, TEs represent a sizable portion of the genome that can vary from ~1% (e.g., in the filamentous fungus *Fusarium graminearum*) to ~85% in *Zea mays* and *Zea luxurians* (32, 111).

TEs are diverse components of genomes. They are classified into two different classes on the basis of whether their transposition mechanism is DNA-based or RNA-based; into different orders on the basis of structural relationships; and into families on the basis of sequence similarities (62, 116). Within each TE family, order, and class, the age and the number of copies can also drastically vary.

TEs are also active components of genomes that generate mutations both when they transpose from one genomic location to another and when they induce structural rearrangements, most commonly via ectopic recombination between TE copies. TE-induced mutations vary greatly in size, ranging from small-scale nucleotide changes, e.g., when a few nucleotides are left behind after a TE excises, to large chromosomal rearrangements. TE-induced mutations are also diverse in terms of their molecular effect: TEs can inactivate or duplicate genes, add or remove regulatory regions, induce new patterns of alternative splicing, and cause epigenetic changes, affecting the expression and/or structure of nearby genes (2).

TE abundance, diversity, and activity are highly variable from one species to another. For example, mammalian genomes tend to harbor a small number of high copy-number families comprising few currently active TEs and thus primarily old TEs, whereas organisms such as plants and insects tend to harbor a much larger number of smaller copy-number families composed of very young TEs (114). TE content also varies among individuals from the same species: Whereas the total copy number might be similar between different individuals, the location of particular TE insertions can vary substantially. Understanding the dynamics of TEs in populations, i.e., which factors explain the number of TEs that belong to a particular TE family, order, or class, the diversity of TEs present in a given genome, and the frequency distribution of individual TE insertions, is crucial if we want to understand the complex organization, function, and evolution of genomes.

DROSOPHILA MELANOGASTER AS A MODEL ORGANISM TO STUDY TRANSPOSABLE ELEMENT DYNAMICS

Drosophila melanogaster has been used as a model organism for the study of TE population dynamics for more than 25 years. The possibility of physically mapping, by in situ hybridization into polytene chromosomes, the location of individual TE copies in the genome provided the first insights into TE dynamics in populations (24, 90). The small genome size of *D. melanogaster* and its relatively small TE content (approximately 20% of the genome) made it an obvious choice for obtaining the first genome sequence of a complex animal (1). *D. melanogaster* is still to date one of the best sequenced, assembled and annotated eukaryotic genomes (87). Additionally, the wealth of functional information available for *D. melanogaster* and the multiple genetic manipulation techniques that are available make this organism ideal for the study of the evolutionary forces shaping genetic variation in natural populations, including TE-induced variation. The availability

of the reference genome sequence and of new sequencing techniques has accelerated our understanding of TE population dynamics.

NEW INSIGHTS OFFERED BY NEXT-GENERATION SEQUENCING TECHNOLOGIES

Next-generation sequencing (NGS) technologies represent a quantum leap in our ability to study TE population dynamics. It is now possible to sequence several individuals or pooled fly samples collected in different geographical locations and/or at different time points and analyze TE dynamics across space and time. Several tools have been designed to discover and annotate TEs in an assembled genome or in raw NGS data (84, 94). More recently, new tools were designed specifically to call (i.e. define whether a TE is present and/or absent in individual strains or pooled NGS data) TEs and to estimatheir population frequencies when NGS data sets are available for multiple strains and/or pooled samples (43, 68). The analysis of numerous genomes from the same population also offers the possibility to improve the annotation of TEs in a species and the accuracy of the frequency estimates of TEs in the population (42, 43) (**Figure1*a***).

One of the challenges of analyzing the TE content of whole-genome data sets is the identification of novel TE insertions, i.e., TE copies not present in the reference genome. Such TE copies can belong to known or novel TE families. During the past two years, numerous tools were designed to discover and annotate TE insertions [ngs_te_mapper (77), RelocaTE (103), RetroSeq (63), PoPoolationTE (68), TEA (transposable element analyzer) (73), TE-locate (100), and T-lex2 (42)]. All of these approaches analyze the mapping of the NGS data on reference sequences in order to search for evidence of TE presence and/or absence. In order to specifically identify nonreference (i.e., novel) insertions, the approach consists of searching for discordant mapping of pair-ends (specifically, pairs in which only one read of the pair is mapped) and single reads that are partially mapped (i.e., soft-clipped reads) (**Figure 1*b***). To distinguish among the insertions corresponding to TE insertions, the approach relies on the availability of a library of representative known TE sequences. All the evidence of the presence of an insertion can be then aligned against the TE library. However, because of that, only the variation in the number of individual TE copies from known families can be reported, not the variation in the number of TE families or TE orders. New approaches that allow thorough annotation of all the TE insertions present in a given genome, from known and unknown TE families and/or orders, are being generated by several groups (44, 86).

One of the limitations of the currently available methodologies is the short length of the reads, which limits our ability to reconstruct individual TE sequences (64). TE length is one of the parameters that influence TE population dynamics (see below), and as such it is crucial to get accurate estimates of individual TE sizes and to identify TEs of all lengths with similar accuracy. Being able to reconstruct individual TE sequences allows us to determine the number of full-length active copies, relic copies evolving as pseudogenes, and nonautonomous copies. The relative proportion of these three functional categories is likely to affect the dynamics of TE families, as has been shown in simulation studies (13, 72). New technologies that allow one to obtain much longer reads, such as Illumina TruSeq synthetic long-read (87, 115) and Pacific Biosciences' single molecule real-time (SMRT) technologies (58), hold great promise for the improved understanding of TE dynamics.

However, even with the limitations of the currently available analytical and technological approaches, we have moved from having small data sets focused on a limited number of families and a limited number of population samples to genome-wide data for all TEs in a genome and frequency estimates based on hundreds of genomes. As expected, these new data sets have confirmed how diverse TE content is and have prompted new studies of TE dynamics.

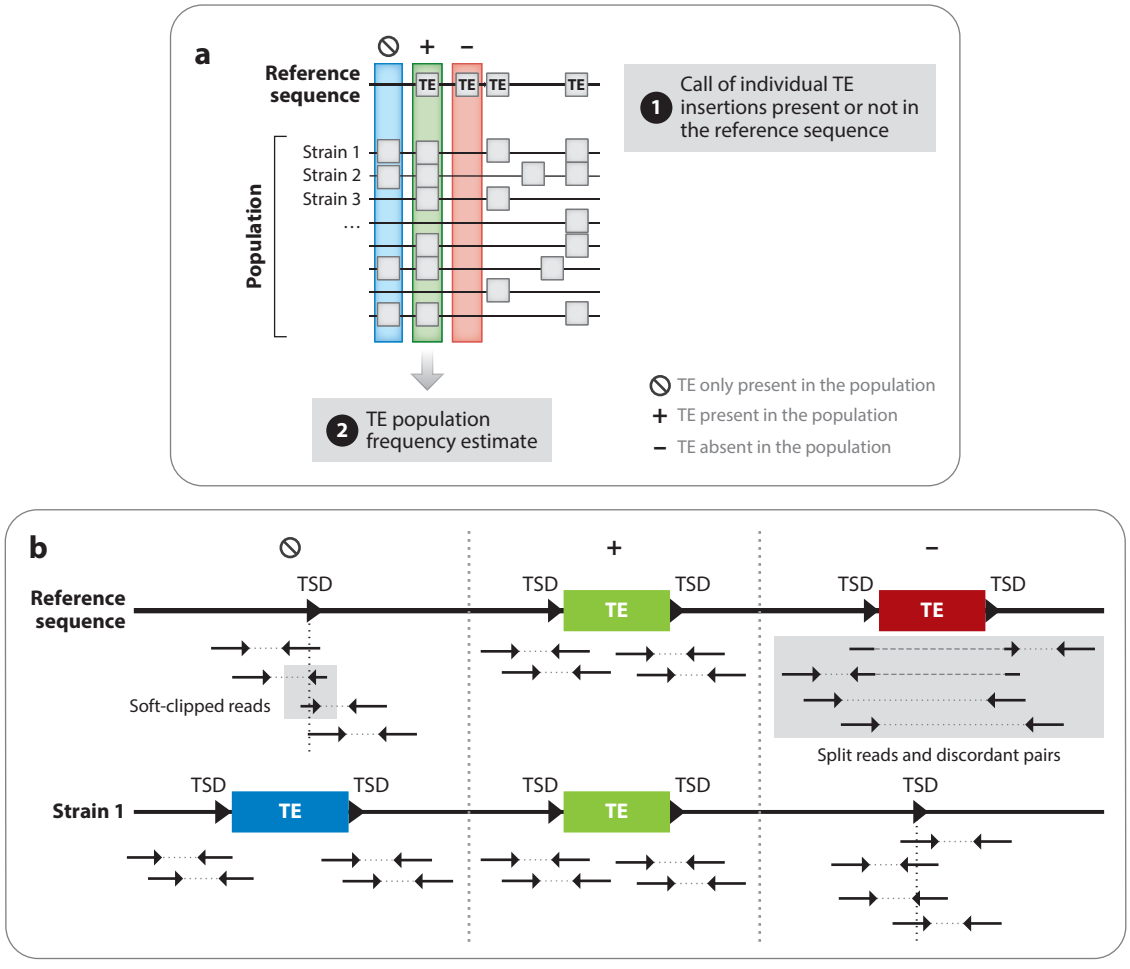

Figure 1

Short-read sequencing technologies offer new insights for transposable element (TE) population genomics study. (*a*) Nonreference TE insertions (i.e., absent from a reference sequence), reference TE insertions (i.e., annotated in a reference sequence) present in the population, and reference TE insertions absent in the population. Short reads in individual strains or those pooled can be used in profiling the TE polymorphism. Nonreference and reference TE insertions could be discovered and then found in other strains or population data. By combining the presence and/or absence detection of TEs from individual strains from the same population (or a pooled sample), the TE population frequency can be estimated. (*b*) Short-read data are represented by pairs of reads pointing in opposite directions and a distance defined during the library preparation (also called insert size). All the pairs represented here come from strain 1 (or a pooled sample). Mapping the reads as single ends or pairs on a reference sequence allows the detection of the TE presence and/or absence. The presence of a TE is supported by the presence of TSDs (target site duplications), short and direct repeats flanking the TE that are usually created after the TE insertion. If a TE is not present in the reference sequence, the reads spanning the TE junction are partially mapped (soft-clipped reads). If a reference TE insertion is also present in the strain, all the reads/pairs from the strain map correctly along the TE sequence, whereas no read maps onto the TE sequence if the reference TE insertion is absent from the strain. The absence is represented by discordant pairs, which refer to a read and its mate, with the insert size greater than the expected insert size distribution of the data set, and by split-reads.

DISTRIBUTION OF TRANSPOSABLE ELEMENTS IN THE *DROSOPHILA MELANOGASTER* GENOME

Several genome population projects provided sequencing data of *D. melanogaster* populations: a European population (68), a North American population [*Drosophila* Genetic Reference Panel (DGRP)] (31, 81), and a laboratory population [*Drosophila* Synthetic Population Resource (DSRP)] (67). The comparison between the TE annotations of these three NGS data sets and the most recent release of the reference genome (Flybase v5.49) yielded a number of insights. The number of TEs identified in each one of these data sets is different. On top of the 5,434 TE insertions annotated in the reference genome, they discovered 10,208 TE insertions in the European population, 17,639 TEs in the North American populations (DGRP data set) and 7,104 TEs in the laboratory population (DSRP data set). The differences in the total number of TEs among the different populations are most likely explained by the different number of strains used for each population: 113 and 131 for European and North American populations, respectively, whereas the analyzed laboratory population contained only 15 strains. However, in this last population (DSRP) the number of TE insertions per strain is larger compared with the other populations, most probably because of the higher sequence coverage and/or the increased amount of time that these strains have remained in laboratory conditions (31, 93). The latter possibility would suggest that TEs continue to transpose in laboratory conditions but are less subject to selection against new copies.

In the reference genome, the TE density per chromosome is similar for all chromosomes (~4%–~10%) except for the fourth chromosome that shows a much higher density (~66%) (**Figure 2**). When NGS data sets are considered, a homogeneous TE density between all the chromosomes is detected but not with the fourth chromosome. The TE density of the fourth chromosome is lower compared with the other chromosomes in the DGRP and DSRP populations, suggesting

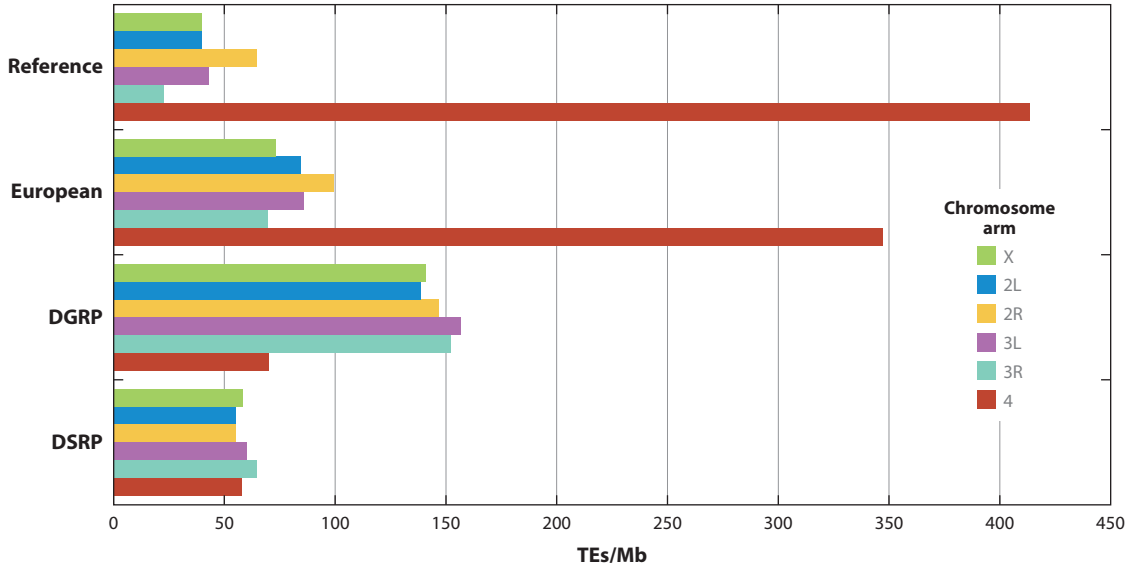

Figure 2

Number of transposable elements (TEs) per megabase (Mb) for each chromosome arm (X, 2L, 2R, 3L, 3R, 4) for the reference genome (Flybase, version 5.49), for a European population (68), for a North American population [*Drosophila* Genetic Resource Panel (DGRP)] (31), and for a laboratory population [*Drosophila* Synthetic Population Resource (DSRP)] (31).

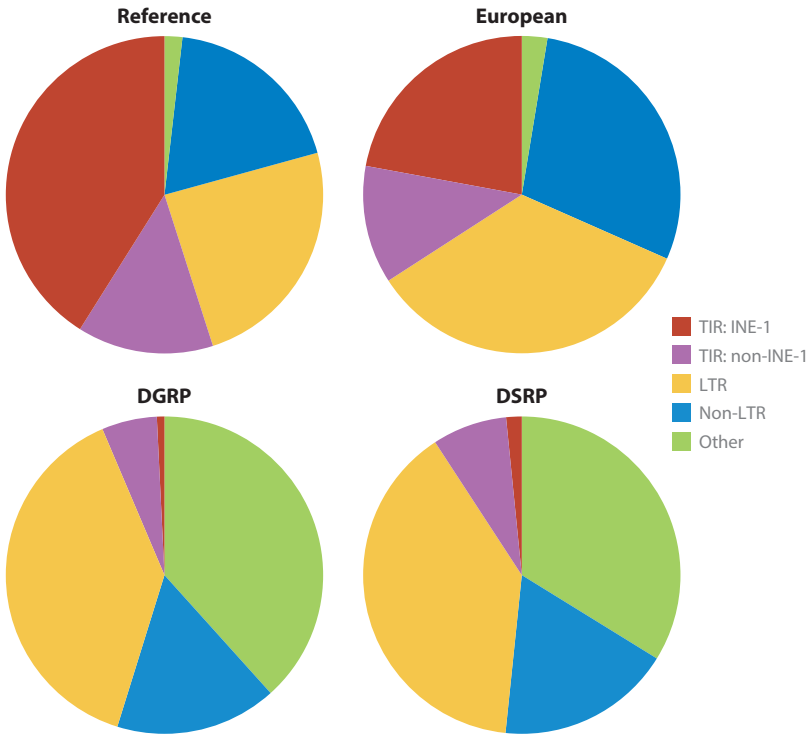

Figure 3

Distribution of transposable element (TE) groups for reference genome and three genome-wide studies that included nonreference insertions: a European population (68), a North American population [*Drosophila* Genetic Resource Panel (DGRP)] (31), and a laboratory population [*Drosophila* Synthetic Population Resource (DSRP)] (31). The five TE groups represent the three main TE orders [i.e., long terminal repeat (LTR), non-LTR, and terminal inverted repeat (TIR)]. The TIR elements are represented in two groups that distinguish the highest copy-number repeats in *D. melanogaster*, which are called INE-1, from the other elements from this order. The last TE group, called "Other," includes TEs that are not part of the main TE orders, except for the North American population, which also includes nonclassified TEs.

that new approaches can underestimate the detection of TEs, specifically the old, fixed, partial, and/or nested TE insertions that are mainly observed in low-recombining regions of the genome such as the fourth chromosome. In contrast to initial findings and theoretical predictions (5, 90), there is no evidence for a reduction in TE density on the X chromosome relative to autosomes (**Figure 2**) (60, 97), and this pattern does not change when nonreference TEs are included (31, 68).

Similarly, most of the TEs in the reference genome belong to the DNA transposon class (**Figure 3**). The INE-1 family, known to have been inactive for approximately the last three million years, contains 2,235 of the 2,986 DNA transposons and is thus composed of old fixed TEs (61, 108). The proportion of DNA transposons in the NGS data sets is smaller. This is probably explained at least in part by the nondetection of old and fixed TEs by the new approaches and the considerable proportion of TEs that cannot be classified (**Figure 3**).

Overall, a substantially increased number of individual TE insertions detected with the approaches that detect nonreference TE insertions highlights the relevance of these techniques to obtain a global view of the TE content in the populations and in the genome. These data sets also

show that current methodologies detect TEs with a skew toward younger, non-nested, euchromatic TEs. Hence, future studies that allow annotation of all individual TE copies are needed to get a complete view of the distribution of TEs along the genome in different populations.

Burst: movement of large numbers of TE sequences through the genome in a short evolutionary time

NEW INSIGHTS ON TRANSPOSABLE ELEMENT DYNAMICS IN *DROSOPHILA MELANOGASTER* POPULATIONS

The general model to address the study of TE dynamics assumes that each TE is transposed at a given rate, and it is subsequently removed by a combined effect of a given excision rate and purifying selection (21). An example of this logic is the well-known transposition-selection balance model, in which the maintenance of TEs in the population is explained by equilibrium between the increase in copy number by a constant transposition rate and elimination from the population by natural selection acting against the deleterious effects of TEs (22). This model allows us to make predictions about the changes in the copy number per genome per generation as well as about TE frequency distribution in the populations under different evolutionary hypotheses (see below). However, the assumption of constant transposition rate has been questioned, given that TE transpositions are known to occur in bursts (34, 35, 66, 69, 80, 104). Hence, the burst transposition model, which relaxes the assumption of transposition-selection balance, has also been proposed to explain TE population dynamics (7, 11, 18).

Analysis of TE dynamics under the assumption of transposition-selection balance starts with the estimate of TE population frequencies (10, 11, 48, 68, 90, 96, 117). If TE insertions are neutral, then their frequency in the population should be indicative of their time of insertion in the genome (i.e., age), with rare TEs expected to be young and frequent and fixed TEs expected to be old. When selection is acting, this logic remains true with some caveats. Deleterious TEs should primarily be rare and should still be young; however, the adaptive TEs might reach high frequencies or even fix quickly and thus might be either young (if we analyzed them soon after their increase in frequency) or old (if we analyzed them much later). Thus, the distribution of TE frequencies should inform us about how selection is acting on the TE families, with families composed primarily of rare TEs likely under purifying selection and/or having just undergone a transposition burst, whereas unusually frequent TEs and TEs that are too young for their high frequency might be suspected of having an adaptive effect.

Other factors, such as horizontal TE transfer and host regulation and/or self-regulation of transposition, do affect TE dynamics as well. However, horizontal transfer of TEs is not likely to be a very common event in *Drosophila* and as such is expected to play a limited role in TE dynamics on a short timescale (79, 89). Host control mechanisms, such as the piRNA pathway (51, 65) and self-regulation of transposition by TEs (23, 54, 59), cannot explain the observed patterns of TE frequencies, i.e., the majority of TEs are present at low population frequencies. In the following section, we summarize the most recent insights on TE dynamics that have been obtained from genome-wide analyses of TEs in *D. melanogaster*.

Transposition-Selection Balance Model

Attempts to accurately estimate TE population frequencies have been carried out for more than 25 years. Early experimental analysis performed in a limited number of TE families showed that most TEs are present at low population frequencies (9, 10, 24, 25, 90, 109). Recent genome-wide TE frequency estimates have confirmed these initial findings: A large proportion of TEs, ranging from 47.9% to 76% in the different studies, are present at low frequencies, suggesting that purifying selection plays a major role in TE population dynamics (**Table 1**) (11, 31, 48, 68).

Table 1 Genome-wide studies of transposable element (TE) population dynamics

Study	Approach[a]	Population	Data set	Number of strains	Number of TEs	Number of filtered TEs[b]	Percentage of TEs detected at low frequency
Petrov et al. 2011 (97)	Pooled-PCR	5 NA, 1 Af	Approximately 50% of all euchromatic reference, non-nested, non-INE-1	75	755	755	75%
Kofler et al. 2012 (68)	Pair-end sequencing[c]	1 EU	Reference and nonreference, non-nested, in NGS pooled DNA	113	10,208	7,843	47.9%
Cridland et al. 2013 (31)	Pair-end sequencing[c]	1 NA, 1 lab strain[d]	Reference and nonreference in NGS individual strains	146	23,087	–	>83.3%
Blumenstiel et al. 2014 (11)	Single-strain PCR	1 NA, 1 Af	Pseudogene-like evolving TEs	24	190	190	70%, 76%

[a]Approach: Approach to estimate TE population frequencies.
[b]Number of TE insertions for which population frequency has been estimated.
[c]Bioinformatical approach based on pair-end sequencing data in which a TE is identified if one read of the pair-end fragment maps to the unique region of a reference genome and the other maps to a TE.
[d]DSPR (*Drosophila* Synthetic Population Resource) strains.
Abbreviations: Af, African; EU, European Union; NA, North American; NGS, next-generation sequencing; PCR, polymerase chain reaction.

Erratum

Three nonmutually exclusive hypotheses have been described to explain the nature of purifying selection acting against TE insertions (reviewed in 92): (*a*) the gene-disruption hypothesis (41, 88); (*b*) the deleterious TE-product expression hypothesis (92); and (*c*) the ectopic recombination hypothesis (90).

Gene-disruption hypothesis. The gene-disruption hypothesis is a widely accepted model in which purifying selection is assumed to be strongly against TE insertions when they are inside a gene or regulatory region (29, 41, 88, 101, 106). In *D. melanogaster*, the analyses of laboratory-induced TE mutations show that TEs do not exclusively transpose outside of coding regions (4, 6). However, in the first in-depth analysis of the euchromatic reference genome (release 3), among the 1,572 TEs identified, none was annotated in coding regions (60). However, if we take into account that 18.3% of the genome corresponds to exons, under the null hypothesis of homogenous insertions genome-wide, we would expect 283 TE insertions to be found in exons. Hence, strong deleterious selection can be invoked to explain this pattern. Follow-up analysis of the reference genome identified only one of these TEs inserted in a protein-coding region (3, 82, 96). Lipatov and colleagues (78) specifically looked for transcripts containing TE and host-gene sequences and found only four that were part of protein-coding regions. Finally, in a genome-wide study in which TEs not present in the reference genome were additionally analyzed, Kofler and colleagues (68) found 249 TEs in coding regions (~2.5% of all TEs), but only 16 of them were fixed (<0.2%). Overall, there are fewer TE insertions in exons and untranslated regions than expected based on the proportion of these sequences in the genome (31, 36, 60, 68, 78, 97), suggesting that selection acting against the deleterious effects of TE insertions inside genes is strong.

Deleterious transposable element–product expression hypothesis. The deleterious TE-product expression model is based on the assumption that the replication of active and inactive copies and the translation of TE-encoded proteins can have a metabolic cost for the cell (16, 92). Additionally TE-encoded proteins could be deleterious because they can disrupt cellular processes (92). Petrov and colleagues (97) made an attempt to test this hypothesis genome-wide by comparing the population frequency of full-length TEs versus truncated TEs (>90% size of the canonical element) in transcriptionally active TE families [according to Deloger and colleagues (36)]. Full-length TEs should be transcribed at higher levels or at least more often than incomplete copies, and, consequently, deleterious selection against them should result in decreased population frequency. However, they were not able to observe this effect and hence, at present there is no direct evidence of selection against the expression of TE-encoded proteins at a genome-wide scale.

Meiotic recombination: process by which a pair of homologous DNA sequences exchanges some portion of the DNA; also known as homologous recombination

Ectopic recombination hypothesis. The previous models alone cannot explain why TEs are also observed at low frequency in nonfunctional regions of the genome. An alternative, nonexclusive hypothesis is the ectopic recombination model. Ectopic recombination between TE copies that belong to the same family, and thus share sequence identity, and are located in different genomic regions can generate chromosomal rearrangements that often lead to inviable gametes (45, 70, 91). Under this model, we expect that (*a*) meiotic recombination rate, (*b*) size of the TE insertion, and (*c*) family copy number should affect the probability of ectopic recombination and therefore the intensity of selection acting against TE insertions. Under the ectopic recombination model, we expect a negative correlation between TE frequency and rate of meiotic recombination. It is assumed that meiotic recombination is correlated with ectopic recombination, and, consequently, meiotic recombination rate should be a good estimator of the ectopic recombination intensity (45, 46, 70, 91). As expected according to this model, low-recombining regions, such as pericentromeric regions and the fourth chromosome, are highly enriched in TEs and harbor most fixed TE insertions (31, 68, 74). After removing heterochromatic regions, a negative correlation is still observed between recombination rate and TE population frequency (5, 68, 97), even among polymorphic TEs (68, 97), suggesting that ectopic recombination is an important factor affecting TE dynamics.

However, there are at least two alternative hypotheses to the ectopic recombination model that could also explain the observed negative correlation between recombination rate and TE population frequency: Hill-Robertson interference (see sidebar, Hill-Robertson Interference)

HILL-ROBERTSON INTERFERENCE

The Hill-Robertson interference is the reduction in the efficiency of selection operating on a locus as a consequence of simultaneous selection operating on linked loci. Two distinct scenarios can be described: (*a*) Two or more adaptive mutations appear in two different haplotypes in the population. If the recombination rate is low, both haplotypes compete against each other until one of them is fixed and the other disappears from the population. With recombination, the two haplotypes could exchange alleles and generate a new haplotype that carries both adaptive alleles. Hence, having a low rate of recombination reduces the rate of adaptive fixation. (*b*) Slightly deleterious and adaptive mutations are found in a haplotype. Some slightly deleterious mutations become fixed owing to the selective sweep of adjacent adaptive mutations. In addition, the elimination of deleterious alleles by selection also eliminates adjacent, weakly adaptive mutations from the population. Overall, the lack of recombination reduces the efficiency of selection, increasing the fixation of slightly deleterious mutations and decreasing the rate of adaptive substitutions.

and gene density. Recombination rate correlates with the efficiency of selection due to Hill-Robertson interference (56). Indeed, genomic regions with reduced recombination rate show an excess of deleterious mutations and a dearth of adaptive substitutions among different *Drosophila* species consistent with reduced efficiency of selection in these regions (8, 17, 27, 53, 57, 107, 118). However, according to computer simulations performed by Dolgin & Charlesworth (38), the observed pattern of TE fixation in low-recombining regions could be explained only by Hill-Robertson interference under a highly unlikely combination of conditions: When recombination rate is extremely low, excision rate is effectively absent and synergism among TEs is weak.

However, low-recombination regions also have low gene density (1). This could decrease negative selection pressures in low-recombination regions, allowing higher TE density or higher TE population frequencies. Under this hypothesis, the observed correlation between recombination rate and TE population frequency beyond these specific regions will need additional evidence, such as a positive correlation between recombination rate and gene density. Therefore, although it is possible that the negative correlation between recombination rate and TE population frequency could be explained by the Hill-Robertson effect and/or gene density, these two alternative hypotheses are both unlikely to explain the global observed correlation.

TE insertion size is also expected to affect the probability of undergoing ectopic recombination: Longer TEs should recombine more often, as they represent longer targets for homologous pairing (39, 96). Among TEs annotated in the reference genome, a negative correlation has been described between TE length and TE population frequency (68, 97), implying that longer TEs are more deleterious and tend to be removed more efficiently from the genome. Although reliable length estimates for nonreference detected TEs cannot be obtained, indirect estimates based on the canonical sequence length also suggest that TE length and population frequency are negatively correlated (68).

Another prediction of the ectopic recombination model is that population frequency should negatively correlate with family copy number. Indeed, ectopic recombination is more likely to happen when TEs are heterozygous and hence, the probability of undergoing ectopic recombination should increase with the copy number of polymorphic TEs (90, 91). As expected, a negative correlation has been described between TE population frequencies and the copy number of polymorphic TEs (68, 97).

Note that a significant statistical interaction between TE length and family TE copy number indicates that families with longer TEs tend to have a larger copy number in *Drosophila* (97). If long TEs exerted a more deleterious effect on its nearby genes, correlation between length and copy number of TEs within a family and population frequency of these TEs could be explained without additional need to invoke the ectopic recombination model. Under this alternative hypothesis, TEs of similar length or similar recombination backgrounds should suffer similar intensity of negative selection independently of the family to which they belong. Petrov and colleagues (97) tested this alternative explanation and found that it does not hold: Two TEs within the same family are more likely to have similar frequencies over and above the frequency predicted by their length and local recombination rate compared with TEs belonging to different families.

Finally, another testable prediction of the ectopic recombination model is that we should observe lower TE density, and TEs at lower population frequencies in the X chromosome compared with autosomes, because of the higher rate of recombination on the X chromosome (28). Higher ectopic recombination rates increase the strength of negative selection in the X chromosome, leading to a faster elimination of TEs and preventing their increase in frequency. This idea is reinforced by data showing higher efficiency of selection in the X chromosome (17, 71, 81), probably due to the faster X hypothesis (see sidebar, Faster X Hypothesis) (20).

FASTER X HYPOTHESIS

Genes on the X chromosomes evolve more rapidly than genes on autosomes. In males, new recessive X-linked mutations are hemizygous and are directly exposed to selection, whereas new recessive autosomal mutations are masked from expression in heterozygotes. This results in increased efficiency of selection for novel X-linked mutations. The X chromosome also has a higher recombination rate (28), which reduces Hill-Robertson interference and increases the efficiency of selection. Several lines of analysis of SNP data are consistent with increased efficiency of selection in the X chromosome (17, 81). Faster X evolution also implies that genes for reproductive isolation might have a higher probability of being X-linked, and in fact this is generally true.

However, TE density is higher in the X chromosome compared with autosomes (31, 68). When differences in the amount of low-recombining regions among chromosomes are taken into account, no differences between X and autosomes are observed (68, 97). Additionally, there is a clear family effect: Although some families show a higher number of insertions on the X compared with autosomes, other families show the opposite pattern (31). These observations are in contrast with the ectopic recombination model. To reconcile them, assumptions can be made: Some families may have an insertion bias toward the X chromosome, and/or meiotic recombination is not a good predictor of ectopic recombination. Vázquez and colleagues (113) found that indeed the transposition rate in the X chromosome was higher, but they analyzed only one specific TE family, *roo*. To date, there are no genome-wide transposition rate estimates that would allow us to shed light on this issue. Additionally, as mentioned above, the relationship between meiotic and ectopic recombination is not yet fully understood.

Overall, in our opinion, the negative correlation observed between TE population frequency and recombination rate, TE length, and family copy number, together with the observed main family effect, tips the scales in favor of ectopic recombination playing a relevant role in explaining TE dynamics in different *Drosophila* populations.

Transposition Burst Model

As mentioned above, one of the assumptions of the transposition-selection balance hypothesis is that transposition rate is constant over time. Some authors (7, 11) claim that this is unlikely to occur on the basis of evidence that some families can undergo periods of transposition bursts (34, 35, 66, 69, 80, 92, 104). Hence, a relaxation in transposition-selection equilibrium has been proposed to explain TE dynamics: the transposition burst model.

Most of the observed features of TE dynamics explained under the ectopic recombination hypothesis can also hold under the transposition burst model. First, the main effect of the family could be explained by bursts of transposition activity. Insertions from the same family tend to happen together in time (7, 14, 75, 92), and thus share a population frequency. This additionally generates a positive correlation between TE frequency and TE age that was not expected under the ectopic recombination hypothesis by itself, i.e., recently active families should be at low population frequencies and long-time inactive families should be fixed. A positive correlation between TE age, based on sequence diversity, and TE frequency is indeed observed for some families (68), although this is expected under any model of TE population dynamics. Second, the relationship between TE frequency and TE copy number or TE length is the same as under the ectopic recombination hypothesis. We expect that recent TEs will have larger family copy number and that older TEs will have lower copy number owing to the fact that some old TEs would have been

removed. Moreover, newly inserted TEs tend to be longer than old TEs because of the deletion bias observed in *Drosophila* (12, 98, 99). Hence, under this burst transposition activity model we also expect to observe a negative correlation between TE population frequency and TE length and TE family copy number, exactly as predicted by the ectopic recombination model. However, short but young TEs within a family are not expected to be more frequent, especially for families in which transposition itself commonly generates short TE copies (such as non-LTR elements), and this pattern is observed in the empirical data (96, 97). Furthermore, the observed negative correlation with recombination rate cannot be explained by this model.

Under the burst transposition model, recently inserted TE families have not had enough time to reach equilibrium and hence they are expected to be at low frequency even under a strictly neutral model. Older insertions can reach this equilibrium and be affected by negative selection, as explained under selection-transposition balance. Blumenstiel and colleagues (11) developed a method to test, based on insertion time, whether a strictly neutral model can explain the observed TE frequency pattern in the genome. TEs undergoing purifying selection are outliers of this model if they have lower frequency than expected under a neutral model, as are putatively adaptive TEs if they have higher frequency than expected under a neutral model. Blumenstiel et al. (11) analyzed in North American and African populations 190 LTR and non-LTR insertions previously shown to have a pseudogene-like, or unconstrained, sequence. Therefore, these analyzed TEs should evolve neutrally. However, in both populations a model that includes negative selection better fits the observed TE frequencies than a strictly neutral model. The authors argued that the bottleneck suffered by North American populations could be biasing their method, and when they corrected for demographic effect, their results suggested that young TEs (67 insertions) can be exclusively explained by a strictly neutral model, whereas middle-aged TEs (87 insertions) and old TE insertions (36 insertions) fit better to a model that includes purifying selection.

For these 190 pseudogene-like putatively neutral TEs, adding their age as a variable seems to be an important factor explaining ~72% of the observed TE frequency variation (11). Hence, at least for some families, TE age seems to be an important factor in explaining TE dynamics. TE age has also been considered as a variable to explain TE population dynamics of a random sample of 671 TEs (68). Together with other explanatory variables, such as recombination and TE length or distance to nearest genes, age can explain ~13.6% of the variance of TE frequency. However, these authors argue that the regression coefficient is not reliable owing to the fact that some variables are not independent among them. Moreover, interpretation of TE age can have a confounding effect because it is based on sequence identity. Having few mutations makes a TE look young, and its expected frequency will appear to be low. However, having fewer mutations also increases the probability of suffering ectopic recombination. By consequence such young TEs may tend to be at a lower frequency because of the deleterious effects of ectopic recombination. Hence, under both hypotheses (strict neutrality and negative selection via ectopic recombination model) the same observations are expected.

Cridland and colleagues (31) analyzed 6,613 relatively young TE insertions (>75% of the canonical length) in euchromatic regions of a North American population. The distribution of TE insertions showed an excess of rare variants compared with SNPs of small introns (<86 bp), which are considered the best candidates for neutral sites in the genome (52). Hence, these results suggest that negative selection is acting among TEs despite the effect of the bottleneck.

Moreover, the model of strict neutrality implicitly assumes that families with TE insertions at low frequency should have recently suffered a burst of activity in *Drosophila* populations. Moreover, the observation that most LTR families are at a lower frequency than non-LTR families (7, 11) implies that families in this order have coordinately invaded the *Drosophila* genome, and the other orders are older invaders. There is no known evidence suggesting such an extreme scenario.

Hence, overall, although the frequency distribution of some TE families can be explained without taking into account purifying selection, for the majority of TEs purifying selection is essential to fully explain their frequency distribution. Finally, although the transposition-selection equilibrium is a simplification, it is remarkable how well it can explain the observed patterns of TE frequencies in genomes. Although there are some families in which a recent burst of transposition activity could certainly explain their low frequency, evidence of purifying selection acting along genome-wide TE insertions is overwhelming. TE family characteristics, such as copy number and TE length and age, seem to be important factors in explaining the dynamics of TEs. Although the ectopic recombination model seems to explain the TE frequency distribution observed outside coding regions, it may not provide a complete picture. Further analysis and exploration are required to fully understand the family copy number, frequency distribution, and diversity, i.e., the dynamics of TEs. This will allow us to discard alternative models and to clarify observations that still remain to be explained, such as the higher, or similar, TE density observed in X versus autosomes.

FUNCTIONAL ROLES OF TRANSPOSABLE ELEMENT INSERTIONS

Although most of the TE insertions are deleterious or neutral, some TE insertions are expected to increase the fitness of the individual that carries them. Genome-wide studies of TE dynamics have made an attempt to list these putatively adaptive TEs using several distinct criteria. González and colleagues (47, 48) selected candidate TEs that are located in high-recombining regions, not located inside inversions, and present at low frequency in ancestral populations (Africa) and at high frequency in derived populations (North America and Australia). In addition, they applied a maximum likelihood approach that allowed them to identify TE families likely to evolve under strong purifying selection and families likely to evolve neutrally (48, 49).

Kofler and colleagues (68) selected insertions fixed in high-recombining regions and in the regions showing five percent lowest quantile genome-wide Tajima's D values (110) [One of the possible signatures of a sweep by positive selection is the generation of an excess of rare mutations (15) that can be detected by negative values of Tajima's D]. Finally, Blumenstiel and colleagues (11) selected insertions at high-recombining regions that have higher population frequency than expected according to their age. Using the previously characterized *Doc1420* (*FBti0019430*) in *CHKov1* gene insertion (3, 96), Blumenstiel and colleagues establish a cut-off value to detect new putatively adaptive TE insertions. The concordance of these findings is low, suggesting that each method detects different putatively adaptive TE insertions (**Table 2**).

To date, only a limited number of TE insertions in *Drosophila* have been unequivocally connected to their relevant fitness effects: an *Accord* LTR retrotransposon (26, 33, 105), a *Doc* non-LTR retrotransposon (3, 83), a *Bari1* transposon (48–50), and a *pogo* transposon (85).

Daborn and colleagues (33) identified an *Accord* element inserted in the 5′ regulatory region of *Cyp6g1*, a gene involved in the detoxification of multiple insecticides. The presence of this TE insertion was associated with increased *Cyp6g1* expression and increased resistance to insecticides such as DDT. Further analyses demonstrated that this *Accord* element carries regulatory sequences that specifically increase *Cyp6g1* expression in tissues important for detoxification (26). When several *D. melanogaster* strains were analyzed, it was discovered that successive mutations, including gene duplications and additional TE insertions, have occurred in the *Cyp6g1* locus (105). Interestingly, there is an increase in resistance level for at least three of the five mutant alleles described, suggesting that this allelic succession could have been driven by selection removing fitness costs associated with the preceding resistance allele (105).

Similarly, an allelic series affecting the gene region in which the adaptive *Doc1420* is inserted has also been reported (83). Aminetzach and colleagues (3) first described the putatively adaptive

Table 2 Putatively adaptive transposable element (TE) insertions identified in different genome-wide studies

Flybase ID	TE family/superfamily	Reference
FBti0018880	*Bari1*/TIR	47
FBti0019056	*pogo*/TIR	47
FBti0019065	*pogo*/TIR	47
FBti0019144	*Rt1b*/non-LTR	47
FBti0019164	*X*-element/non-LTR	47
FBti0019170	*F*-element/non-LTR	47
FBti0019372	*S*-element/TIR	47
FBti0019386	*Invader4*/LTR	47
FBti0019430	*Doc*/non-LTR	47, 68, 11
FBti0019443	*Rt1b*/non-LTR	47
FBti0019624	*Hopper*/TIR	47
FBti0019627	*pogo*/TIR	47
FBti0019679	1731/LTR	47
FBti0019747	*F*-element/non-LTR	47
FBti0020042	*Jockey*/non-LTR	47
FBti0020046	*Doc*/non-LTR	47, 11
FBti0020091	*Rt1a*/non-LTR	47
FBti0020119	*S*-element/TIR	47
FBti0019564	*Mdg1*/LTR	68
FBti0060479	*HMS-Beagle*/LTR	68
FBti0062283	*Ninja-Dsim*/LTR	68
FBti0019082	*Rt1b*/non-LTR	68
FBti0019655	*3S18*/LTR	68
FBti0060388	*S*-element/TIR	68
FBti0061742	*rooA*/LTR *Accord*/LTR *roo*/LTR	68
FBti0059793	*hobo*/TIR	68
FBti0063191	*gypsy12*/LTR	68
FBti0020329	*G5*/non-LTR	68
FBti0019200	*Doc*/non-LTR	11
FBti0020082	*412*/LTR	11
FBti0020086	*17.6*/LTR	11
FBti0020149	*BS*/non-LTR	11
FBti0019354	*17.6*/LTR	11
FBti0019199	*Doc*/non-LTR	11
FBti0020125	*BS*/non-LTR	11

Abbreviations: LTR, long terminal repeat; TIR, terminal inverted repeat.

insertion of this *Doc* element into the coding region of *CHKov1*. This insertion generates two sets of altered transcripts, and it is associated with increased resistance to an organophosphate insecticide (AZM). However, the authors estimated that the allele containing the *Doc1420* insertion was 90,000 years old. Because insecticides were first used only a few decades ago, the original reasons for the fast evolution and persistence in natural populations of the *Doc1420*-containing allele must be related to some other phenotypic effect. As anticipated by Aminetzach and colleagues (3), polymorphisms in the *CHKov1* gene region were later found to be associated with a different phenotype: resistance to viral infection (83). Magwire and colleagues (83) showed that although the truncation of *CHKov1* coding region by insertion of the *Doc1420* element confers resistance to the sigma virus, an allele containing two duplications resulting in three copies of the truncated allele of both *CHKov1* and *CHKov2* (one of which is also truncated) caused increased resistance.

Recently, a third TE insertion has been connected to its ecologically relevant fitness effect (50). A full-length *Bari1* insertion was identified as being putatively adaptive based on its population frequency and on the detection of a selective sweep in its flanking regions (48). This insertion, named *Bari-Jheh*, is located in the intergenic region of juvenile hormone epoxy hydrolase (*Jheh*) genes, and it was found to affect the level of expression of its two nearby genes (49). Phenotypic effects consistent with the reduced level of expression of *Jheh3* and *Jheh2* genes were also found. However, the phenotypic effects identified, reduced viability and increased developmental time, are likely to represent the cost of selection of this insertion, whose adaptive effect remains unknown. A detailed analysis of the *Bari-Jheh* sequence revealed that this TE adds extra antioxidant response elements (AREs) to the upstream regions of *Jheh2* and *Jheh1* genes. AREs are highly conserved sequences, found in organisms from flies to humans, which mediate response to stress by upregulating the expression of downstream genes. As expected, we found that flies with *Bari-Jheh* showed increased levels of expression of *Jheh2* and *Jheh1* under oxidative stress conditions and increased resistance to this stress (50). Furthermore, we also found that TEs other than *Bari-Jheh* add extra AREs to the upstream region of several genes, suggesting a more general role of *Bari* elements in response to oxidative stress (50).

Finally, a *pogo* transposon has been shown to mediate resistance to xenobiotics (85). This TE affects the polyadenylation signal choice of its nearby gene CG11699. As a result, only one of the two CG11699 transcripts is produced and the expression of CG11699 increases. Mateo et al. (85) further showed that increased CG11699 expression leads to increased aldehyde dehydrogenase III enzymatic activity that results in increased resistance to xenobiotic stress.

These four examples show the variety of molecular mechanisms underlying TE-driven adaptation: from adding tissue-specific or response-to-stress regulatory regions, to the generation of new transcripts, to inactivation of genes. However, the number of adaptive TEs whose adaptive phenotypic effects are known is still too small, and many more TEs need to be characterized to get a more general picture of the adaptive process.

Other than the adaptive effects of particular TE-induced mutations, evidence for the functional impact of TEs in a diversity of cellular processes is starting to accumulate in a range of organisms (19, 30, 37). Recently in *Drosophila*, substantiation has been provided for a role of TEs in (*a*) the establishment of dosage compensation (40), (*b*) heterochromatin assembly (106), and (*c*) brain genomic heterogeneity (95).

1. Dosage compensation in *Drosophila* is achieved by upregulating X-linked genes by approximately twofold in males. In *D. melanogaster*, the male-specific lethal (MSL) complex binds to MSL recognition elements (MREs) located on the male X chromosome, inducing a local change in the chromatin state that promotes the increase in gene expression. Ellison & Bachtrog (40) recently discovered that two related families of *Helitron* elements have

been independently domesticated to provide MRE sites in *Drosophila miranda* at two different evolutionary time points. In both cases, the acquisition of the TE was followed by fine-tuning mutations that refine the ability of the TEs to recruit the MSL complex and by the amplification of this modified TE on the X chromosome. Further secondary fine-tuning mutations and erosion of the TE sequences that are not required for binding MSL continued after the expansion. This secondary refinement and erosion may eventually lead to degradation of the signatures of the TE origins, suggesting that rewiring of regulatory networks by domesticated TEs may eventually be undetectable (40).

2. Heterochromatin assembly. It is known that TEs from the *1360* transposon family induce silencing of nearby genes by promoting the accumulation of Heterochromatin Protein 1a (HP1a) (55). In a recent work, Sentmanat & Elgin (106) narrowed down the specific region of the *1360* element responsible for heterochromatin assembly, leading to repression of transcription. It turns out that silencing does not require the terminal inverted repeats nor the internal transcription start sites, but the sequences that are bound by PIWI-interacting RNAs (piRNAs). Furthermore, the authors extended this observation to the *Invader4* LTR-retrotransposon family, suggesting that this silencing mechanism is not restricted to TEs from a single family and is likely a broadly applicable mechanism (106).
3. Brain genomic heterogeneity. Transposon expression has been found to be more abundant in αβ neurons of the mushroom body, a brain structure critical for olfactory memory, than in neighboring neurons (95). Perrat and colleagues further determined that the piRNA proteins Aubergine and Argonaute 3 were less abundant in αβ neurons and that increased expression of TEs in these neurons resulted in nonreference transposon insertions, some of which are located in memory-relevant loci. Perrat and colleagues (95) suggested that it is possible that mushroom body neurons differentially use piRNA to control the expression of memory-relevant loci and that the observed transposon mobilization is an associated cost. Although disruptive insertions may lead to neural decline and cognitive dysfunction (76), it is also possible that allowing transposition produces genetic variability that may contribute to normal brain function (95).

Overall, it is now clear that TEs are a major source of genetic variation. Although most TEs are present at low population frequencies, strongly suggesting that they are deleterious, a significant fraction of TEs have been recruited to perform cellular functions.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

J.G. is a Ramon y Cajal fellow (RYC-2010-07306). This work was supported by grants BFU-2011-24397 and PCIG-2011-293860 awarded to J.G. and grant R01GM089926 awarded to D.A.P.

LITERATURE CITED

1. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–95
2. Akagi K, Li J, Symer DE. 2013. How do mammalian transposons induce genetic variation? A conceptual framework: The age, structure, allele frequency, and genome context of transposable elements may define their wide-ranging biological impacts. *BioEssays* 35:397–407

3. Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309:764–67
4. Ashburner M, Bergman CM. 2005. *Drosophila melanogaster*: a case study of a model genomic sequence and its consequences. *Genome Res.* 15:1661–67
5. Bartolome C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* 19:926–37
6. Bellen HJ, Levis RW, He Y, Carlson JW, Evans-Holm M, et al. 2011. The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188:731–43
7. Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 104:11340–45
8. Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 99:13616–20
9. Biemont C. 1992. Population genetics of transposable DNA elements. A *Drosophila* point of view. *Genetica* 86:67–84
10. Biemont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, et al. 1994. Population dynamics of the *copia*, *mdg1*, *mdg3*, *gypsy*, and *P* transposable elements in a natural population of *Drosophila melanogaster*. *Genet. Res.* 63:197–212
11. Blumenstiel JP, Chen X, He M, Bergman CM. 2014. An age-of-allele test of neutrality for transposable element insertions. *Genetics* 196:523–38
12. Blumenstiel JP, Noll AC, Griffiths JA, Perera AG, Walton KN, et al. 2009. Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. *Genetics* 182:25–32
13. Boutin TS, Le Rouzic A, Capy P. 2012. How does selfing affect the dynamics of selfish transposable elements? *Mob. DNA* 3:5
14. Bowen NJ, McDonald JF. 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* 11:1527–40
15. Braverman JM, Lazzaro BP, Aguade M, Langley CH. 2005. DNA sequence polymorphism and divergence at the erect wing and suppressor of sable loci of *Drosophila melanogaster* and *D. simulans*. *Genetics* 170:1153–65
16. Burt A, Trivers R. 2006. *Genes in Conflict*. Cambridge, MA: Belknap Press
17. Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Mol. Biol. Evol.* 31:1010–28
18. Carr M, Bensasson D, Bergman CM. 2012. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLOS ONE* 7:e50978
19. Casacuberta E, Gonzalez J. 2013. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* 22:1503–17
20. Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* 130:113–46
21. Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. *Genet. Res.* 42:1–27
22. Charlesworth B, Jarne P, Assimacopoulos S. 1994. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin. *Genet. Res.* 64:183–97
23. Charlesworth B, Langley CH. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* 112:359–83
24. Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23:251–87
25. Charlesworth B, Lapid A, Canada D. 1992. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genet. Res.* 60:103–14
26. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, et al. 2007. *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* 175:1071–77

27. Comeron JM, Kreitman M. 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* 156:1175–90
28. Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genet.* 8:e1002905
29. Cooley L, Kelley R, Spradling A. 1988. Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* 239:1121–28
30. Cowley M, Oakey RJ. 2013. Transposable elements re-wire and fine-tune the transcriptome. *PLOS Genet.* 9:e1003234
31. Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol. Biol. Evol.* 30:2311–27
32. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, et al. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317:1400–2
33. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, et al. 2002. A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297:2253–56
34. Daniels SB, Chovnick A, Boussy IA. 1990. Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol. Biol. Evol.* 7:589–606
35. de la Chaux N, Wagner A. 2011. BEL/Pao retrotransposons in metazoan genomes. *BMC Evol. Biol.* 11:154
36. Deloger M, Cavalli FM, Lerat E, Biemont C, Sagot MF, Vieira C. 2009. Identification of expressed transposable element insertions in the sequenced genome of *Drosophila melanogaster*. *Gene* 439:55–62
37. de Souza FS, Franchini LF, Rubinstein M. 2013. Exaptation of transposable elements into novel *cis*-regulatory elements: Is the evidence always strong? *Mol. Biol. Evol.* 30:1239–51
38. Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* 178:2169–77
39. Dray T, Gloor GB. 1997. Homology requirements for targeting heterologous sequences during P-induced gap repair in *Drosophila melanogaster*. *Genetics* 147:689–99
40. Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* 342:846–50
41. Finnegan DJ. 1992. Transposable elements. *Curr. Opin. Genet. Dev.* 2:861–67
42. Fiston-Lavier AS, Barrón M, Petrov DA, González J. 2014. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *BioRxiv* doi: **http://dx.doi.org/10.1101/002964**
43. Fiston-Lavier AS, Carrigan M, Petrov DA, Gonzalez J. 2011. T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. *Nucleic Acids Res.* 39:e36
44. Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in de novo annotation approaches. *PLOS ONE* 6:e16526
45. Goldman AS, Lichten M. 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* 144:43–55
46. Goldman AS, Lichten M. 2000. Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. *Proc. Natl. Acad. Sci. USA* 97:9537–42
47. González J, Karasov TL, Messer PW, Petrov DA. 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLOS Genet.* 6:e1000905
48. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element–induced adaptation in *Drosophila melanogaster*. *PLOS Biol.* 6:e251
49. González J, Macpherson JM, Petrov DA. 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol. Biol. Evol.* 26:1949–61
50. Guio L, Barron MG, Gonzalez J. 2014. The transposable element *Bari-Jheh* mediates oxidative stress response in *Drosophila*. *Mol. Ecol.* 23:2020–30
51. Guzzardo PM, Muerdter F, Hannon GJ. 2013. The piRNA pathway in flies: highlights and future directions. *Curr. Opin. Genet. Dev.* 23:44–52
52. Haddrill PR, Charlesworth B, Halligan DL, Andolfatto P. 2005. Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol.* 6:R67

53. Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol.* 8:R18
54. Hartl DL, Lohe AR, Lozovskaya ER. 1997. Regulation of the transposable element mariner. *Genetica* 100:177–84
55. Haynes KA, Caudy AA, Collins L, Elgin SC. 2006. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr. Biol.* 16:2222–27
56. Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8:269–94
57. Hilton HR, Kliman M, Hey J. 1994. Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. *Evolution* 48:1900–13
58. Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, et al. 2014. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* 24:688–96
59. Jaillet J, Genty M, Cambefort J, Rouault JD, Auge-Gouillou C. 2012. Regulation of mariner transposition: the peculiar case of Mos1. *PLOS ONE* 7:e43365
60. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3:research0084.1–0084.20
61. Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. USA* 100:6569–74
62. Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* 9:411–12; author reply 4
63. Keane TM, Wong K, Adams DJ. 2013. RetroSeq: transposable element discovery from next-generation sequencing data. *Bioinformatics* 29:389–90
64. Kelley JL, Peyton JT, Fiston-Lavier A-S, Teets NM, Yee MC. 2014 Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nat. Commun.* 5:4611
65. Khurana JS, Theurkauf W. 2010. piRNAs, transposon silencing, and *Drosophila* germline development. *J. Cell Biol.* 191:905–13
66. Kidwell MG. 1983. Hybrid dysgenesis in *Drosophila melanogaster*: factors affecting chromosomal contamination in the P-M system. *Genetics* 104:317–41
67. King EG, Macdonald SJ, Long AD. 2012. Properties and power of the *Drosophila* synthetic population resource for the routine dissection of complex traits. *Genetics* 191:935–49
68. Kofler R, Betancourt AJ, Schlotterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLOS Genet.* 8:e1002487
69. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
70. Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* 52:223–35
71. Langley CH, Stevens K, Cardeno C, Lee YC, Schrider DR, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192:533–98
72. Le Rouzic A, Boutin TS, Capy P. 2007. Long-term evolution of transposable elements. *Proc. Natl. Acad. Sci. USA* 104:19375–80
73. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, et al. 2012. Landscape of somatic retrotransposition in human cancers. *Science* 337:967–71
74. Lee YC, Langley CH. 2010. Transposable elements in natural populations of *Drosophila melanogaster*. *Philos. Trans. R. Soc. Lond. Ser. B* 365:1219–28
75. Lerat E. 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity* 104:520–33
76. Li W, Prazak L, Chatterjee N, Gruninger S, Krug L, et al. 2013. Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* 16:529–31
77. Linheiro RS, Bergman CM. 2012. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLOS ONE* 7:e30008
78. Lipatov M, Lenkov K, Petrov DA, Bergman CM. 2005. Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. *BMC Biol.* 3:24
79. Loreto EL, Carareto CM, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 100:545–54

80. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. 2012. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* 29:1005–17
81. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482:173–78
82. Magwire ML. 2011. Addressing barriers to insulin therapy: the role of insulin pens. *Am. J. Ther.* 18:392–402
83. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. 2011. Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication. *PLOS Genet.* 7:e1002337
84. Makalowski W, Pande A, Gotea V, Makalowska I. 2012. Transposable elements and their identification. *Methods Mol. Biol.* 855:337–59
85. Mateo L, Ullastres A, González JA. 2014. A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLOS Genet.* 10(8):e1004560
86. Maumus F, Quesneville H. 2014. Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PLOS ONE* 9:e94101
87. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, et al. 2014. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly repetitive transposable elements. *PLOS ONE* 9(9):e106689
88. McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ, Miller WJ. 1997. LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* 100:3–13
89. Modolo L, Picard F, Lerat E. 2014. A new genome-wide method to track horizontally transferred sequences: application to *Drosophila*. *Genome Biol. Evol.* 6:416–32
90. Montgomery E, Charlesworth B, Langley CH. 1987. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet. Res.* 49:31–41
91. Montgomery EA, Huang SM, Langley CH, Judd BH. 1991. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* 129:1085–98
92. Nuzhdin SV. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* 107:129–37
93. Nuzhdin SV, Pasyukova EG, Mackay TF. 1997. Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. *Genetica* 100:167–75
94. Permal E, Flutre T, Quesneville H. 2012. Roadmap for annotating transposable elements in eukaryote genomes. *Methods Mol. Biol.* 859:53–68
95. Perrat PN, DasGupta S, Wang J, Theurkauf W, Weng Z, et al. 2013. Transposition-driven genomic heterogeneity in the *Drosophila* brain. *Science* 340:91–95
96. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol.* 20:880–92
97. Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol. Biol. Evol.* 28:1633–44
98. Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol. Biol. Evol.* 15:293–302
99. Petrov DA, Lozovskaya ER, Hartl DL. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384:346–49
100. Platzer A, Nizhynska V, Long Q. 2012. TE-locate: a tool to locate and group transposable element occurrences using paired-end next-generation sequencing data. *Biology* 1(2):395–410
101. Puig M, Caceres M, Ruiz A. 2004. Silencing of a gene adjacent to the breakpoint of a widespread *Drosophila* inversion by a transposon-induced antisense RNA. *Proc. Natl. Acad. Sci. USA* 101:9013–18
102. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, et al. 2005. Combined evidence annotation of transposable elements in genome sequences. *PLOS Comput. Biol.* 1:166–75
103. Robb SM, Lu L, Valencia E, Burnette JM 3rd, Okumoto Y, et al. 2013. The use of RelocaTE and unassembled short reads to produce high-resolution snapshots of transposable element generated diversity in rice. *G3* 3:949–57

104. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20:43–45
105. Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, et al. 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLOS Genet.* 6:e1000998
106. Sentmanat MF, Elgin SC. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc. Natl. Acad. Sci. USA* 109:14104–9
107. Shapiro JA, Huang W, Zhang C, Hubisz MJ, Lu J, et al. 2007. Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* 104:2271–76
108. Singh ND, Petrov DA. 2004. Rapid sequence turnover at an intergenic locus in *Drosophila*. *Mol. Biol. Evol.* 21:670–80
109. Sniegowski PD, Charlesworth B. 1994. Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* 137:815–27
110. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95
111. Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J. 2011. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* 3:219–29
112. Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* 24:969–81
113. Vázquez JF, Albornoz J, Dominguez A. 2007. Direct determination of the effects of genotype and extreme temperature on the transposition of *roo* in long-term mutation accumulation lines of *Drosophila melanogaster*. *Mol. Genet. Genomics* 278:653–64
114. Venner S, Feschotte C, Biemont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* 25:317–23
115. Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, et al. 2013. The genome sequence of the colonial chordate, *Botryllus schlosseri*. *Elife* 2:e00569
116. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973–82
117. Yang HP, Nuzhdin SV. 2003. Fitness costs of Doc expression are insufficient to stabilize its copy number in *Drosophila melanogaster*. *Mol. Biol. Evol.* 20:800–4
118. Zhang Z, Parsch J. 2005. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Mol. Biol. Evol.* 22:1945–47