Published in final edited form as:
Gene. 2009 December 15; 448(2): 124–133. doi:10.1016/j.gene.2009.06.008.

The Adaptive Role of Transposable Elements in the Drosophila Genome

Josefa González and **Dmitri A. Petrov**
Department of Biology, 371 Serra St. Stanford University, Stanford, CA 94305-3020

Abstract

Transposable elements (TEs) are short DNA sequences with the capacity to move between different sites in the genome. This ability provides them with the capacity to mutate the genome in many different ways, from subtle regulatory mutations to gross genomic rearrangements. The potential adaptive significance of TEs was recognized by those involved in their initial discovery although it was hotly debated afterwards. For more than two decades, TEs were considered to be intragenomic parasites leading to almost exclusively detrimental effects to the host genome. The sequencing of the *Drosophila melanogaster* genome provided an unprecedented opportunity to study TEs and led to the identification of the first TE-induced adaptations in this species. These studies were followed by a systematic genome-wide search for adaptive insertions that allowed for the first time to infer that TEs contribute substantially to adaptive evolution. This study also revealed that there are at least twice as many TE-induced adaptations that remain to be identified. To gain better understanding of the adaptive role of TEs in the genome we clearly need to (i) identify as many adaptive TEs as possible in a range of Drosophila species as well as (ii) carry out in-depth investigations of the effects of adaptive TEs on as many phenotypes as possible.

Keywords

transposable elements; adaptation; Drosophila; *in situ*; bottlenecks; selfish DNA

1. What is a transposable element?

Transposable Elements (TEs) are short DNA sequences, typically 1-10kb, capable of inserting copies of themselves into new genomic locations. TEs are an ancient, abundant and diverse component of eukaryotic genomes. They are present in virtually all eukaryotic species investigated so far where they represent from 3% to 80% of the total DNA (Hua-Van et al., 2005; Biemont and Vieira, 2006; Piegu et al., 2006). Most TEs can be assigned to one of two main classes defined according to their mode of transposition (Figure 1). Class I or retrotransposons use a “copy and paste” mechanism that involves an RNA intermediate. Retrotransposons are further subdivided into those that have long terminal repeats (LTRs) and those that do not (long and short interspersed nuclear elements (LINEs and SINEs)). Class II or DNA transposons utilize a “cut and paste” mechanism that involves a DNA intermediate. Inside these two classes, TEs are further classified into superfamilies according to features

---

Corresponding author: Josefa González, Department of Biology, 371 Serra St., Stanford University, Stanford CA 94305-3020, e-mail: jgonzalp@stanford.edu, phone: 650-7362240.

such as the presence or absence and the size of target site duplications (a short direct repeat that is generated on both sides of a TE when it inserts) and into families according to DNA sequence conservation.

TE insertions generate a wide variety of mutations which often have phenotypic effects of a complexity that is not achievable by a small number of point mutations. In addition, the presence of interspersed repetitive sequences introduced by the TE activity is a key source of genomic structural rearrangements such as chromosomal inversions, deletions, duplications, and translocations. It is clear that the abundance and mutagenic activity of TEs make them a key player in the function and the evolution of genomes (Kazazian, 2004; Miller and Capy, 2004; Biemont and Vieira, 2006; Muotri et al., 2007; Goodier and Kazazian, 2008).

2. From “controlling elements” to “selfish” DNA

TEs were first considered to play a largely beneficial role in the evolution of genomes. Initially they were described as “controlling elements” that have the ability to regulate gene expression and to restructure chromosomes (McClintock, 1950; McClintock, 1956). As soon as repetitive DNA was discovered to make up a large fraction of the genome of higher organisms, it was speculated that repetitive sequences in general would ultimately be found to be important to the phenotype (Britten and Kohne, 1968). Some believed that TEs could be essential in the evolution of complex regulatory systems by distributing binding sites for transcription factors (if the TE sequences contain such sites) to many chromosomal locations making it possible for many genes to be drawn into the same regulatory circuit (Britten and Davidson, 1969). This initial view of TEs as benevolent and essential contributors to the function and evolution of genomes was seriously questioned later on. The observation that the number and chromosomal location of TEs differed extensively among four phenotypically indistinguishable strains of Drosophila melanogaster suggested that TEs could not be as functionally important as initially thought (Strobel et al. 1979). Some authors proposed that the ubiquity of TEs could be explained solely by their ability to replicate within the genome, which in sexual organisms should lead to their spread even in the presence of a significant deleterious effect (Hickey, 1982). This idea served as a basis of the so-called “selfish DNA” hypothesis which postulates that TEs are selfish DNA parasites that essentially live and spread within genomes, generating excesses of DNA, introducing mutations and rearranging genomes (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Although their presence in the genome was generally considered deleterious, it was accepted that TEs could occasionally lead to beneficial effects. Some cases of TE domestication were discovered, such as the case of TEs protecting the ends of linear chromosomes in Drosophila by transposing to the chromosomal ends (Biessmann et al., 1992; Sheen and Levis, 1994; Pardue et al., 2005).

Even though the “selfish DNA” theory was predominant for more than two decades, all through this time, the commonality of TE domestication and the key role of TEs in genome function and adaptation were vigorously defended by a number of authors (Brosius, 1991; McDonald, 1993; McDonald, 1995; Shapiro, 1999; Kidwell and Lisch, 2001).

3. TEs in Drosophila

Studies of TEs in Drosophila have been key in understanding the evolutionary dynamics and effects of TEs in eukaryotes. Some of the first insights came from the population data gathered using *in situ* hybridization and restriction map surveys of TEs in Drosophila. These results generally supported the “selfish DNA” theory (Charlesworth and Langley, 1989; Charlesworth et al., 1994). TEs were found to be at low frequencies in most genomic locations and their maintenance in populations was considered to be the result of a balance between transpositional increase in copy number and the counter-effects of natural selection. Three distinct but not mutually exclusive hypotheses about the nature of selection acting against individual TE copies were proposed (Nuzhdin, 1999). TEs could be deleterious because they (i) disrupted genes, affecting their coding capacity (Finnegan, 1992) or their regulation (McDonald et al., 1997), (ii) recombined with each other inducing deleterious rearrangements (Montgomery et al., 1987), and (iii) translation of TE-encoded proteins could be costly and their activity could be harmful for the organism (Nuzhdin et al., 1997).

The observation that element frequencies were nearly always very low at particular chromosomal sites was taken as evidence against TE insertions frequently inducing beneficial mutations. Such insertions should quickly become fixed by selection and such cases were both very rare and found in low recombination regions where selection is expected to be less effective (see below; (Charlesworth et al., 1992a). This view was also consistent with the Neutral Theory of Molecular Evolution (Kimura, 1983) that suggested that adaptation in general is extremely rare. Thus it was not surprising that so few if any TE insertions were adaptive.

4. A new perspective on TEs: lessons from the *D. melanogaster* genome sequence

The sequencing and a very thorough annotation of the eukaryotic portion in a single *D. melanogaster* strain allowed for a quantum leap in our understanding of the *D. melanogaster* TEs (Kaminker et al., 2002; Quesneville et al., 2005). The analysis of the TE content largely validated the previous understanding of the TE abundance and distribution suggesting that the sequenced strain is typical in terms of its TE composition (Kaminker et al., 2002). However, it also revealed that *in situ* studies were only giving a partial picture of the population dynamics of these elements. Petrov et al (Petrov et al., 2003) analyzed population behavior of four non-LTR families and discovered that two of them contained many copies at high population frequencies suggesting that they were subjected to weak purifying selection.

Why hadn't we seen more fixed and/or high frequency TEs before? First, previous population analysis of Drosophila TEs were based on a limited number of primarily LTR families. Petrov et al (2003) provided evidence that selection against chromosome rearrangements caused by ectopic recombination limits the spread of some TEs as previously proposed by other authors (Montgomery et al., 1987; Langley et al., 1988; Charlesworth et al., 1992b; Bartolome et al., 2002). The ectopic recombination model suggests that selection should act family by family, since TE copies of any particular family can only recombine with other copies from the same family. It also predicts that the strength of selection should be stronger in the families with more numerous and longer copies. *In situs* were traditionally carried out with exactly such families (*297*, *412*, *copia*, *roo* or *jockey*; Strobel et al., 1979; Montgomery et al., 1987; Biemont and Terzian, 1988; Charlesworth et al., 1994; Dominguez and Albornoz, 1996; Vieira and Biemont, 1996) because these families were among the first ones to be discovered and because *in situ* with longer TEs from high-copy-number families generates better and more numerous signals. The extrapolation of the results based on these families to all TEs led to somewhat biased conclusions about the population dynamics of TEs in general.

The bias introduced by the *in situ* technique can be particularly severe for the discovery of adaptively important TE insertions. Such adaptive TEs are likely to be either present at high frequency or fixed and thus may be on average older and shorter than newly transposed deleterious copies (Petrov et al., 1996; Petrov and Hartl, 1998).

5. Individual cases of adaptive TEs in Drosophila

The sequencing of the *D. melanogaster* genome simplified and accelerated the search for adaptive TE insertions. The argument underlying these studies was that if a particular TE insertion had contributed to adaptive evolution, such an insertion was expected to be present at high frequencies or fixed in populations and/or species. Adaptive TEs were also identified as a result of investigations that were not specifically looking for adaptive insertions. A detailed account of all these studies is given below and it is summarized in Table 1.

5.1. Analysis of fixed TE insertions

As mentioned before, only a few examples of apparently fixed TE insertions in the *D. melanogaster* genome had been described prior to the sequencing of the Drosophila genome (Charlesworth et al., 1992a). Most of these insertions were located in regions of low recombination where selection was expected to be less efficient in removing them due to (i) lower frequency of ectopic exchange (Charlesworth et al., 1992b; Bartolome et al., 2002) and (ii) the genetic interference between linked sites subject to selection (Hill and Robertson, 1966). Therefore, it remains quite possible that many of the fixed TEs located in regions of low recombination reached fixation neutrally. Fixed TEs in regions of high recombination are likelier candidates for playing or having played an adaptive role.

Maside et al. (2002) reported the first study of a TE insertion apparently fixed in a highly recombining region of the genome: an *S-element* inserted upstream of gene *Hsp70Bb*. However, an updated estimation of the recombination rates in the Drosophila genome indicates that this TE is indeed located in a 0.02 cM/Mb recombining region (Singh et al., 2005; Table 1). The authors reported that the inverted terminal repeat sequences of this element were subject to purifying selection which led them to propose that this TE may be affecting the expression of their neighboring genes. Later on, they found that other *S-elements* were also fixed in the genome suggesting that other members of this family might be adaptive (Bartolome and Maside, 2004). In a follow-up study, Maside et al. (2005) analyzed the other four fixed insertions reported by Charlesworth et al. (1992a). Two of them were artifacts of the *in situ* technique caused by the presence of genomic DNA flanking the TEs and the other two correspond to true fixations located in low recombination regions of the genome. No evidence for the adaptive significance of these two TEs was found.

Another fixed TE located in a highly recombining region of the *D. melanogaster* genome was reported by Marsano et al. (2005): a *Bari-1* element inserted in the 3′ end of the *cyp12a4* gene. In flies with the insertion, the transcript of this gene includes 18 nucleotides of the transposon, it is shorter and it is over-expressed compared to the transcript of flies lacking the insertion. This result suggests that the insertion is likely to have a functional effect although the phenotypic effect is unknown.

More recent studies have identified a substantial number of fixed TEs in regions of low and high recombination in the *D. melanogaster* genome (Petrov et al., 2003; Bartolome and Maside, 2004; González et al., 2008). However, the potential adaptive significance of these insertions has not yet been assessed.

5.2. TE-gene association studies

A different approach aimed at detecting putatively adaptive insertions was followed by J. F. McDonald and collaborators (McCollum et al., 2002; Franchini et al., 2004). The finding that TEs are components of the regulatory and/or coding regions of a surprisingly large number of genes in different organism led these authors to analyze TE-gene associations in *D. melanogaster*. They identified 25 TEs located inside or near genes and scored their population frequency in 12-18 strains. Only three TEs were present in all the analyzed strains. The regions surrounding two of them showed reduced levels of sequence variation suggesting their adaptive significance (Table 1).

Lipatov et al. (2005) investigated chimeric gene-TE transcripts and reported that less than 1% of Drosophila genes produce mRNAs that included *bona fide* TE sequences indicating that chimeric TE insertions were generally strongly deleterious. Population genetics analysis of the observed chimeric transcripts suggested that (i) a small proportion of the observed chimeric TEs were fixed and therefore are unlikely to commonly contribute to the origin of new genes and (ii) the TEs that were parts of chimeric transcripts were not subject to unusually strong purifying or positive selection.

5.3. Unusually frequent TE insertions

As mentioned above, (Petrov et al., 2003) reported that in two out of four non-LTR *D. melanogaster* families studied, the majority of TEs were present at high frequency. However, despite being present at high frequency, these TEs are unlikely to be adaptive. It is more plausible that such TE families are subject to relaxed purifying selection as a whole (Petrov et al., 2003; González et al., 2009a). In the other two families, the majority of TEs were found at a low frequency suggesting that these families were subject to strong purifying selection. Therefore, the few TEs found at a high frequency in these families are likely to be enriched for adaptive TEs. Petrov et al. (2003) identified one such unusually frequent TE: one *Doc* element that was fixed in the North American (NA) populations and polymorphic in the African (AF) populations analyzed. This variation in allele frequency between AF and non-AF populations suggests that this insertion could represent an adaptation to the out-of-Africa environments. Indeed, *D. melanogaster* is believed to have originated in sub-Saharan Africa and expanded its population worldwide very recently (David and Capy, 1988; Lachaise, 1988). This expansion appears to have resulted in numerous adaptations to the new habitats (Harr et al., 2002; Glinka et al., 2003; Orengo and Aguade, 2004). Furthermore, the analysis of the insertion site of this TE revealed that it apparently truncated a conserved gene *CHKov1* suggesting that it was likely to have a selective effect. Aminetzach et al. (2005) further analyzed this insertion and found that the regions flanking the insertion showed signatures of a partial selective sweep that decayed at increasing distances from the TE suggesting that this insertion was indeed adaptive. The authors confirmed that the TE was truncating the *CHKov1* gene and based on the putatively function of this gene first hypothesized and then confirmed that it conferred resistance to organophosphate pesticides.

The Doc-containing allele of *CHKov-1* is quite divergent and thus likely to be old (~90,000 years old) suggesting that it evolved prior to the human usage of pesticides and thus its original adaptive function is unlike to be related to pesticides. Nevertheless, Aminetzach et al. (2005) estimated that the spread of this allele occurred recently, suggesting that its recent expansion might be an adaptive response to the introduction of pesticides in the mid-20th century (Aminetzach et al., 2005). Further investigation cast doubt on this scenario, however, as it appears that this allele was already present at high frequencies in the out-of-Africa populations of *D. melanogaster* prior to the introduction of pesticides, making its recent spread a possible example of an exaptation (Aminetzach, Karasov, Petrov, personal communication).

5.4. Other approaches

The investigation of the natural variation underlying thermotolerance (Michalak et al., 2001; Zatsepina et al., 2001; Lerman et al., 2003; Lerman and Feder, 2005; Chen et al., 2007), the transcription analysis of all the identified P450 genes (Daborn et al., 2002) and the analysis of the haplotype structure around the *Sr-CII* locus (Schlenke and Begun, 2004) identified several other examples of individual adaptive TE insertions reported so far in Drosophila.

The analysis of the *D. melanogaster* strains that showed differences in thermotolerance led to the discovery of different naturally occurring TE insertions in the promoters of *Hsp* genes. Different TE insertions associated with decreased gene expression and in some cases with changes in thermotolerance and in female reproductive success have been described in the promoter of *Hsp70Ba* (Michalak et al., 2001; Zatsepina et al., 2001; Lerman et al., 2003) *Hsp70Bb* (Lerman and Feder, 2005) and *Hsp26* genes (Walser et al., 2006; Chen et al., 2007). The in depth analysis of the insertions located in the *Hsp70* genes demonstrated that the effect of the TEs on gene expression is due to the spatial disruption of the promoter (Lerman and Feder, 2005).

Daborn et al. (2002) found an adaptive insertion of an *Accord* element in a screen for P450 alleles involved in DDT resistance in *D. melanogaster*. These authors demonstrated that over-transcription of the *Cyp6g1* gene was both necessary and sufficient for pesticide resistance. The sequencing of the resistant alleles revealed that they carried an insertion of the terminal direct repeat of an *Accord* element in the 5′ end of the gene. Catania et al. (2004) showed that the *Cyp6g1* allele carrying the *Accord* insertion swept to a high frequency in populations around the world with the frequency being higher in derived compared to ancestral populations. This result further suggested that this allele conferred an adaptive advantage in derived populations. In a follow up study, Chung et al. (2007) demonstrated that cis-regulatory sequences included in the *Accord* sequence drove the increase in expression of the *Cypg1* gene in a tissue-specific manner.

Schlenke and Begun (2004) identified a 100 kb region of the *D. simulans* genome with extremely reduced heterozygosity in NA but not in AF populations. The analysis of this region revealed that the insertion of a *Doc* element in the 5′ region of *Cyp6g1* was the most likely cause of the selective sweep. The insertion was associated with increased transcript abundance. However, although the insertion of an Accord element in the 5′ of this gene is associated with resistance to pesticides in *D. melanogaster*, only a weak support for added resistance to pesticides due to the *Doc* insertion near *Cyp6g1* was found in *D. simulans*. The sweep around *Cyp6g1* gene in NA but not AF populations of both *D. melanogaster* and *D. simulans* species, the insertion of different TEs in the 5′ regulatory region of this gene, and the associated transcriptional up-regulation provides a possible example of parallel evolution in these two species.

6. The first genome-wide scan for recent TE-induced adaptations

Based solely on the individual examples of putatively adaptive TEs reported, it seems that TEs could be adaptive fairly often. The evidence for the adaptive role of some of these insertions should be considered only as preliminary, however. For example, in several cases, reduced polymorphism in and/or around the insertion compared to the neutral expectations in the panmictic population was regarded as evidence for the putatively adaptive role of these TEs (Maside et al., 2002; McCollum et al., 2002; Franchini et al., 2004; Table 1). However, we know that analyzing patterns of polymorphism without taking into account the demographic history of the populations can lead to spurious inference of positive selection (Kreitman, 2000; Andolfatto and Przeworski, 2001; Teshima et al., 2006; Thornton et al., 2007; Macpherson et al., 2008). Therefore, although these results hinted that beneficial TE insertions were not as rare as previously estimated (Charlesworth et al., 1994), no reliable conclusions about the contribution of TEs to adaptive evolution could be reached at that point.

6.1. Identifying recent putatively adaptive TE insertions

In order to gain a more comprehensive understanding of the role of TEs in adaptation we performed a genome-wide screen for recent TE-induced adaptations in Drosophila (González et al., 2008). We used the annotated TEs in Release 3 of the *D. melanogaster* genome as the starting point for our search.

Our goal was to specifically identify TEs that may have contributed to the adaptation of *D. melanogaster* to the out-of-Africa environments (David and Capy, 1988; Lachaise, 1988). Therefore, we focused on identifying TEs that were rare or absent in Africa and frequent or fixed in North America. First, using a pooled-PCR strategy, we estimated the frequency of the majority of TEs in the Release 3 of the *D. melanogaster* genome. We used two sets of PCRs per TE such that we could determine both presence and absence of each TE and could identify whether the TE was absent, polymorphic or fixed in the DNA pool (see Figure 2). This strategy allowed us to identify 38 TEs that were likely to be present at high frequency in NA and not fixed in AF. These 38 TEs were located in regions of high recombination. We discarded TEs present at high frequency in regions of low recombination since they were more likely to have reached high frequencies neutrally due to the reduced efficacy of selection against insertions in these regions. Second, for each of these 38 TEs we determined their population frequency by performing PCRs with individual strains. We filtered out TEs that were present in <30% of the NA strains or in >30% of the AF strains assayed. And finally, we divided the remaining 21 TEs into two sets: a set of 13 putatively adaptive TEs and a set of eight putatively neutral TEs based on their family identity. Specifically, 13 putatively adaptive TEs came from families where most of the TEs are present at low population frequencies, whereas the eight putatively neutral TEs came from families where many of the TEs are found at high population frequencies. In summary, based exclusively on the population frequency of individual TEs and their families, we identified 13 TEs out of the initial set of 902 that were more likely to have played a role in the out-of-Africa adaptation (Table 1).

6.2. Evidence for the adaptive role of the identified TEs

There are alternative explanations for the high frequency of these 13 TEs other than the adaptive hypothesis. First, it is possible that these TEs reached high frequency via random genetic drift. Since selection against ectopic recombination among TE copies is likely to be one of the main forces controlling TE population dynamics, it is also possible that such frequent TEs do not in fact recombine ectopically very often either because they have an unusually short or divergent sequence (Petrov et al., 2003) or because they are located in a genomic region that is protected from ectopic recombination for some reason. The analysis of the sequence of these TEs showed that they were not unusually short compared to other members of the same family and that they were all young insertions (González et al., 2008). Also, these 13 TEs are evenly distributed among the different chromosomal arms and within a chromosomal arm they do not cluster but rather dispersed across multiple locations. These two observations make the neutral explanation less likely, but, because we understand the determinants of ectopic recombination poorly, not impossible.

We can test the adaptive hypothesis further by looking for evidence of a rapid, adaptive increase in population frequency revealed by the signature of a selective sweep in the pattern of polymorphism in the flanking sequences (Smith and Haigh, 1974; Kaplan et al., 1988; Kaplan et al., 1989). In *D. melanogaster*, the detection of selection is severely complicated by the likely population bottleneck experienced by this species during the expansion out of Africa (Kreitman, 2000; Andolfatto and Przeworski, 2001; Teshima et al., 2006; Thornton et al., 2007; Macpherson et al., 2008; González et al., 2009a). To avoid the spurious inference of selection, we constructed a null model that incorporated the demographic scenario specified in (Thornton and Andolfatto, 2006). We also incorporated our ascertainment bias, given that sampling TEs that are present at a high frequency in derived populations should by itself generate signatures of apparent selective sweeps (Macpherson et al., 2008). We found sweep signatures in the flanking regions of all five of the putatively adaptive TEs analyzed. In contrast the polymorphism patterns surrounding four of the putatively neutral TEs did not differ significantly from those expected under the null model.

Is the evidence for a selective sweep in the regions flanking a putatively adaptive mutation conclusive of adaptive evolution? Although identification of a selective sweep provides considerable evidence for positive selection (Glinka et al., 2003; Orengo and Aguade, 2004; DuMont and Aquadro, 2005; Ometto et al., 2005; Glinka et al., 2006; Pool et al., 2006; Hutter et al., 2007; Orengo and Aguade, 2007; Beisswanger and Stephan, 2008), they are not entirely conclusive of adaptive evolution for several reasons. First, there is uncertainty about the correct demographic model for *D. melanogaster*. There are two different demographic scenarios based on European population data (Li and Stephan, 2006; Thornton and Andolfatto, 2006), but both are extremely simplified. The demographic scenario for the NA populations is even less well established (David and Capy, 1988; Caracristi and Schlotterer, 2003; Baudry et al., 2004). Second, other factors such as purifying selection acting against the TE insertion or reduction of recombination could also affect the inference of positive selection although bottlenecks have been shown to have the strongest effect (Macpherson et al., 2008). Third, it is possible that a mutation located further away from the sequenced region is associated with the sweep. The polymorphism pattern around the five insertions that we analyzed is consistent with the TE being the cause of the sweep. However, further sequencing would be required to completely discard the existence of a polymorphism other than the TE linked to the sweep.

The observation that putatively adaptive TEs give stronger signals of adaptation than similarly frequent but putatively neutral TE insertions is suggestive of their adaptive increase in frequency. This contrast between the putatively adaptive and the putatively neutral TEs in many ways controls for the possible confounding effects of the factors mentioned above. However, we performed an additional, independent test of the adaptive role of these elements. We tested whether the frequencies of these TEs are higher in more temperate compared to more tropical out-of-Africa populations of *D. melanogaster*. This is what we expected if these TEs were indeed involved in the adaptation to the out-of Africa habitats. We found that the putatively adaptive TEs showed a significant heterogeneity in frequencies between two Australian populations while the putatively adaptive TEs did not. In all instances, the frequency in the temperate population was higher compared to the frequency in the tropical population, as predicted. We also controlled for the possible effect of inversions in these patterns (Hoffmann and Weeks, 2007). Once more, the contrast between the putatively adaptive and the putatively neutral TEs reinforced the hypothesis that these TEs were indeed adaptive.

7. What have we learned about adaptation in Drosophila?

Table 1 summarizes the data regarding all the putatively adaptive elements that have been identified so far both following the genome-wide screen approach and the analysis of individual insertions that had been identified previously. Only one of the 13 TEs identified in the genome-wide approach, the *Doc* element inserted into *CHKov1* gene, had been identified previously (Petrov et al., 2003;Aminetzach et al., 2005). Such a small overlap is not surprising – genome-wide screen ignored all fixed TEs and TEs that were not found in the highly recombining regions in the sequenced strain. This partial set of putatively adaptive insertions allows us to start making inferences about the adaptive process in Drosophila.

7.1. How frequent is TE-induced adaptation?

As mentioned above, no reliable conclusions about the rate of TE-induced adaptations could be reached based only on the individual examples of putatively adaptive TEs since the evidence for the adaptive role of some of them was not conclusive. Our systematic search for adaptive TEs allowed for the first time to estimate this rate (González et al., 2008). Based on the number of TEs identified as putatively adaptive and taking into account that the approach used to identify putatively adaptive TEs was conservative for many reasons, we estimated that the rate of TE-induced adaptation is high: one TE-induced adaptation every 200 to 1,250 years (González et al., 2008). This high rate of TE-induced adaptations appears incompatible with the observed number of fixed TEs in the *D. melanogaster* genome. If these adaptive TEs are destined to reach fixation, and considering only the TEs fixed within the past 1 million year, we should see 400-2,500 fixed TEs in euchromatic regions of high recombination. In sharp contrast we only see 25 (González et al., 2008). Why do we see so few fixed TEs in the *D. melanogaster* genome? There are several not mutually exclusive possibilities. For example, it is possible that the estimated high rate of TE-induced adaptations is only characteristic of this unusual evolutionary period in the history of this species. The high rate could reflect a burst in adaptations in response to the new challenges Drosophila faced in its out-of-Africa expansion. Another possibility is that these TEs are adaptive in some but not other environments and we indeed find evidence that eight of the 13 adaptive TEs identified appear to be adaptive to temperate climates. This result suggests that individual mutations could become adaptive for a period of time but eventually get lost. This would imply that some functional genetic variation within species could be due to ephemeral local adaptations. A third possibility is that the number of fixed TEs in the genome is higher than we think. Adaptive TEs might undergo fast sequence evolution driven by positive selection that might difficult its detection, if so, a more sensitive search for degenerate TE sequences in the *D. melanogaster* genome should help to identify them.

Our estimate of the rate of TE-induced adaptation is in general agreement with recent data suggesting that adaptation is common in Drosophila in general (Smith and Eyre-Walker, 2002; Bierne and Eyre-Walker, 2004; Andolfatto, 2005; Eyre-Walker, 2006; Andolfatto, 2007; Macpherson et al., 2007; Shapiro et al., 2007; Sella et al., 2009). These studies suggested that ~50% of substitutions at protein-altering sites (Fay et al., 2002; Smith and Eyre-Walker, 2002; Bierne and Eyre-Walker, 2004; Andolfatto, 2005; Welch, 2006) and ~30% of substitutions at regulatory sites (Andolfatto, 2005) in Drosophila are adaptive. This implies that the Drosophila genome undergoes an adaptation at a protein-coding site approximately every 45 years and at a regulatory site every 20 years. Genome-wide signatures of selective sweeps have been used to confirm these estimates (Andolfatto, 2007; Macpherson et al., 2007; Stephan and Li, 2007) and to further suggest that adaptation might often involve mutations of large selective effect (Macpherson et al., 2007). Adaptation may be similarly pervasive in *E.coli* (Charlesworth and Eyre-Walker, 2006) and HIV (Williamson, 2003) but possibly not in Arabidopsis (Bustamante et al., 2002), yeast (Liti et al., 2009) or human lineages (Bustamante et al., 2005; Nielsen et al., 2005). Our estimate of TE-induced adaptations is consistent with the high rate of adaptive evolution in Drosophila and suggests that TEs are a significant source of adaptive mutations.

7.2. Which genes or processes are involved in adaptation?

Several of the putatively adaptive insertions are located within or close to genes involved in response to stimulus: six are located in the promoter region or close to genes involved in response to heat, three are located within or close to genes involved in response to insecticide and another three within or close to genes involved in response to toxin, response to virus and olfactory learning (Table 1). The observation that TEs in general occur more often in genes related to external stimuli than in other gene classes has been reported previously (van de Lagemaat et al., 2003). We used FatiGO to look for evidence of over or under-representation of GO terms associated with the genes located close to putatively adaptive TEs compared to the rest of genes in the genome of *D. melanogaster* (Al-Shahrour et al., 2006). The terms “response to chemical stimulus” and “response to biotic stimulus” were over-represented in the genes close to putatively adaptive TEs (adjusted *P value* was 1.88e-7 and 1.54e-2 respectively). These results have to be taken with caution since the number of genes associated with putatively adaptive TEs is small and only 16 of them have been functionally annotated (Table 1).

Some of the adaptive TEs are located closed to genes involved in highly conserved pathways suggesting that they play a role in the fine-tuning of these processes (González et al., 2008). Among these TEs one particular *Bari1* element is inserted between two *Juvenile hormone epoxy hydrolase* (*Jheh*) genes involved in the Juvenile Hormone metabolism. Juvenile Hormone has major effects on various aspects of development and life history traits (Flatt et al., 2005) and *Bari1* is associated with the down-regulation in the expression of both *Jheh* genes (González et al., 2008; González et al., 2009b). Furthermore, we found subtle consequences of *Bari-Jheh* insertion on life history traits that are consistent with its effects of reduced expression of the *Jheh* genes. However, the adaptive effect of this insertion still remains to be elucidated (González et al., 2009b).

7.3. Are adaptations mostly regulatory or structural changes?

The analysis of the location of the putatively adaptive TEs gives insight into the relative contribution of protein-coding versus regulatory changes in adaptation. Most of the reported putatively adaptive TEs are either located in intergenic regions or in introns while only one TE disrupts a gene and another two are located in 3′ UTR regions (Table 1). This observation suggests that TEs are more often involved in regulatory changes. Five of the 13 putatively adaptive TEs reported in González et al. (2008) and eight of the previously reported putatively adaptive insertions have been associated with a change in the transcription of the nearby gene supporting the role of these TEs in the regulation of the adjacent genes (Daborn et al., 2002;Lerman et al., 2003;Schlenke and Begun, 2004;Marsano et al., 2005;Chen et al., 2007).

Although being located in introns or in intergenic regions, TEs can lead to structural changes if they happened to be incorporated into the transcript of the nearby gene. In our genome-wide screen for putatively adaptive TEs we looked for evidence of chimeric ESTs for all the putatively adaptive TEs identified and we only found evidence of chimeric transcripts for the two TEs located in the 3′UTR or the exon of a gene (González et al., 2008). This result reinforced the conclusion that most of the identified TEs were involved in gene regulation.

7.4. Do adaptations arise more often from standing variation or from *de novo* mutations?

Adaptation to novel or changing environments can happen through the selection on alleles already present in the ancestral populations or through selection on new mutations. Although the relative importance of these two sources of potentially beneficial alleles has not been determined, most of the theory on the genetics of adaptation has focused on adaptation from new mutations (Smith and Haigh, 1974; Kaplan et al., 1988; Kaplan et al., 1989). Only recently, models that consider selection from standing variation have been reported (Hermisson and Pennings, 2005; Przeworski et al., 2005). Several analyses with different organisms suggested that standing variation has an important role in facilitating rapid adaptation to novel environments (Feder et al., 1997; Feder et al., 2003; Colosimo et al., 2005; Pelz et al., 2005; Steiner et al., 2007; Tishkoff et al., 2007). This could also be the scenario in Drosophila since most of the TE insertions identified in a screen that was specifically looking for recent TE adaptive insertions involved in the out-of-Africa colonization have been found to be already present in ancestral African populations (González et al., 2008).

7.5. Is adaptation population/environment specific?

González et al. (2008) provided different lines of evidence suggesting that a big proportion of TE-induced adaptations represent local adaptations. First, none of the 13 putatively adaptive TEs identified were fixed in the out-of-Africa populations. Second, the frequencies of the 13 TEs were not consistently higher in recently sampled NA strains compared to M strains collected before the 1940's suggesting that these insertions were not on their way to fixation. And third, eight of the 13 TEs were present at higher frequency in a temperate compared to a tropical population suggesting that they were only adaptive in some, specifically temperate environments.

The contribution of fixed TEs to adaptive evolution is still unknown. Four of the identified putatively adaptive TEs are fixed in all the populations analyzed suggesting that they represent adaptations at the species level (Table 1). However, the evidence for the adaptive role of three of these TEs can be considered only preliminary since, as mentioned above, it was based on the reduced polymorphism in or around the insertion and the analysis was performed without taking into account the demographic history of the species (Maside et al., 2002;McCollum et al., 2002;Franchini et al., 2004). In our genome-wide screen for recent adaptive TE insertions, we identified several fixed TE insertions and suggested that a proportion of them might be adaptive (González et al., 2008). However, we did not analyze these insertions since fixed insertions are less likely to be recent and to have contributed to adaptation during or after the expansion of *D. melanogaster* out of Africa.

8. Conclusions and future prospects

Although being dismissed as “junk” DNA for two decades, TEs appear to be a significant source of adaptive mutations in Drosophila. Our population survey of the frequency of 902 TEs in the *D. melanogaster* genome confirmed that most of the TEs are present at low frequencies (González et al., 2008) suggesting that most of the insertions were deleterious and therefore subject to purifying selection as previously reported (Charlesworth and Langley, 1989; Charlesworth et al., 1994). However, this observation was not incompatible with TEs playing an important role in adaptation and indeed we found that TEs contribute significantly to the generation of recent adaptive changes (González et al., 2008). Over the past few years a number of studies suggested that adaptation might be much more common than had been believed previously (Eyre-Walker, 2006). The rate of TE-induced adaptations is similar to the rate of adaptive nucleotide substitutions in coding and non-coding regions suggesting that TEs contribute significantly to adaptive evolution in Drosophila.

Based on the analysis of the putatively adaptive TEs identified so far, we can conclude that adaptive TEs seem to be more often associated with genes involved in response to stimulus. Most of the adaptive TEs are located close but not inside protein coding regions and appear to affect the expression of these genes suggesting that they are more often involved in regulatory than in coding changes. Finally, recent TE-induced adaptations appear to arise more often from standing variants than from new mutations and to play a role in adaptation to temperate climates.

However, the number of identified adaptive TEs is still too small to draw any general conclusions about the TE-induced adaptive process. One of the future challenges is to identify a larger number of TE-induced adaptations. We identified 13 TE-induced adaptations and estimated that 25-50 of the TEs were likely to be involved in adaptation to the out of Africa environments (González et al., 2008). These TE-induced adaptations remain to be discovered. This should be possible as a large number of *D. melanogaster* strains are being sequenced at the moment (http://www.hgsc.bcm.tmc.edu/; Ayroles et al., 2009).

Once we have identified a more comprehensive set of TE-induced adaptations the next challenge will be to understand the functional relevance of these insertions. One possibility is to use the information about the functional identity of the nearby genes to construct plausible hypothesis about the phenotypic consequences of the insertion (Aminetzach et al., 2005; González et al., 2009b). However, as exemplified by the insertion of a *Doc* element in the *CHKov1* gene, showing that an adaptive TE has one specific functional effect predicted from the function of the genes does not ensure that this functional effect is the underlying reason for the adaptation. In that case, the insertion does lead to pesticide resistance to organophosphates as predicted from the molecular nature of CHKov1 gene, but population data strongly suggest that this was not the selective reason for its spread (Aminetzach, Karasov, Petrov, personal communication).

All mutations and adaptive mutations specifically are likely to have an array of pleiotropic molecular and phenotypic effects. Adaptive mutations need to be adaptive overall but some of their phenotypic effects might be neutral or even deleterious. Therefore, identifying the phenotypic trait on which selection is acting can be challenging even when we have clues about the potential phenotypes of interest. One promising venue of inquiry is to phenotype an array of adaptive TEs against a large number of traits. Effects in the traits that are associated repeatedly with adaptive TEs but not with neutral or deleterious ones might be good candidates for truly adaptive effects of adaptive mutations.

The Drosophila community is developing a genetic panel of 192 *D. melanogaster* strains that are currently being sequenced using high-throughput methodologies and phenotyped. 50 of them are already available at http://www.hgsc.bcm.tmc.edu/. The Drosophila Genetic Reference Panel will allow us to carry out association mapping of the adaptive and neutral TEs with many traits and hopefully gain insight into the nature of adaptation in Drosophila.

Acknowledgments

We thank the two anonymous reviewers for their constructive comments and useful suggestions. This work was supported by grants from the NIH (GM 077368) and the NSF (0317171) to D.A.P.

References

Al-Shahrour F, Minguez P, Tarraga J, Montaner D, Alloza E, Vaquerizas JM, Conde L, Blaschke C, Vera J, Dopazo J. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. Nucleic Acids Res 2006;34:W472–6. [PubMed: 16845052]

Aminetzach YT, Macpherson JM, Petrov DA. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science 2005;309:764–7. [PubMed: 16051794]

Andolfatto P. Adaptive evolution of non-coding DNA in Drosophila. Nature 2005;437:1149–52. [PubMed: 16237443]

Andolfatto P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the Drosophila melanogaster genome. Genome Res 2007;17:1755–62. [PubMed: 17989248]

Andolfatto P, Przeworski M. Regions of lower crossing over harbor more rare variants in African populations of Drosophila melanogaster. Genetics 2001;158:657–65. [PubMed: 11404330]

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in Drosophila melanogaster. Nat Genet 2009;41:299–307. [PubMed: 19234471]

Bartolome C, Maside X. The lack of recombination drives the fixation of transposable elements on the fourth chromosome of Drosophila melanogaster. Genet Res 2004;83:91–100. [PubMed: 15219154]

Bartolome 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. [PubMed: 12032249]

Baudry E, Viginier B, Veuille M. Non-African populations of Drosophila melanogaster have a unique origin. Mol Biol Evol 2004;21:1482–91. [PubMed: 15014160]

Beisswanger S, Stephan W. Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in Drosophila. Proc Natl Acad Sci U S A 2008;105:5447–52. [PubMed: 18381818]

Biemont C, Terzian C. Mdg-1 mobile element polymorphism in selected Drosophila melanogaster populations. Genetica 1988;76:7–14. [PubMed: 2852619]

Biemont C, Vieira C. Genetics: junk DNA as an evolutionary force. Nature 2006;443:521–4. [PubMed: 17024082]

Bierne N, Eyre-Walker A. The genomic rate of adaptive amino acid substitution in Drosophila. Mol Biol Evol 2004;21:1350–60. [PubMed: 15044594]

Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW, Pardue ML. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in Drosophila melanogaster. Mol Cell Biol 1992;12:3910–8. [PubMed: 1324409]

Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science 1969;165:349–57. [PubMed: 5789433]

Britten RJ, Kohne DE. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 1968;161:529–40. [PubMed: 4874239]

Brosius J. Retroposons--seeds of evolution. Science 1991;251:753. [PubMed: 1990437]

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG. Natural selection on protein-coding genes in the human genome. Nature 2005;437:1153–7. [PubMed: 16237444]

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. The cost of inbreeding in Arabidopsis. Nature 2002;416:531–4. [PubMed: 11932744]

Caracristi G, Schlotterer C. Genetic differentiation between American and European Drosophila melanogaster populations could be attributed to admixture of African alleles. Mol Biol Evol 2003;20:792–9. [PubMed: 12679536]

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C. World-wide survey of an Accord insertion and its association with DDT resistance in Drosophila melanogaster. Mol Ecol 2004;13:2491–504. [PubMed: 15245421]

Charlesworth B, Langley CH. The population genetics of Drosophila transposable elements. Annu Rev Genet 1989;23:251–87. [PubMed: 2559652]

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 1992a;60:103–14. [PubMed: 1334899]

Charlesworth B, Lapid A, Canada D. The distribution of transposable elements within and between chromosomes in a population of Drosophila melanogaster. II. Inferences on the nature of selection against elements. Genet Res 1992b;60:115–30. [PubMed: 1334900]

Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 1994;371:215–20. [PubMed: 8078581]

Charlesworth J, Eyre-Walker A. The rate of adaptive evolution in enteric bacteria. Mol Biol Evol 2006;23:1348–56. [PubMed: 16621913]

Chen B, Walser JC, Rodgers TH, Sobota RS, Burke MK, Rose MR, Feder ME. Abundant, diverse, and consequential P elements segregate in promoters of small heat-shock genes in Drosophila populations. J Evol Biol 2007;20:2056–66. [PubMed: 17714322]

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ. 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. [PubMed: 17179088]

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 2005;307:1928–33. [PubMed: 15790847]

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, ffrench-Constant RH. A single p450 allele associated with insecticide resistance in Drosophila. Science 2002;297:2253–6. [PubMed: 12351787]

David JR, Capy P. Genetic variation of Drosophila melanogaster natural populations. Trends Genet 1988;4:106–11. [PubMed: 3149056]

Dominguez A, Albornoz J. Rates of movement of transposable elements in Drosophila melanogaster. Mol Gen Genet 1996;251:130–8. [PubMed: 8668122]

Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature 1980;284:601–3. [PubMed: 6245369]

DuMont VB, Aquadro CF. Multiple signatures of positive selection downstream of notch on the X chromosome in Drosophila melanogaster. Genetics 2005;171:639–53. [PubMed: 16020794]

Eyre-Walker A. The genomic rate of adaptive evolution. Trends Ecol Evol 2006;21:569–75. [PubMed: 16820244]

Fay JC, Wyckoff GJ, Wu CI. Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature 2002;415:1024–6. [PubMed: 11875569]

Feder JL, Berlocher SH, Roethele JB, Dambroski H, Smith JJ, Perry WL, Gavrilovic V, Filchak KE, Rull J, Aluja M. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proc Natl Acad Sci U S A 2003;100:10314–9. [PubMed: 12928500]

Feder JL, Roethele JB, Wlazlo B, Berlocher SH. Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. Proc Natl Acad Sci U S A 1997;94:11417–21. [PubMed: 11038585]

Finnegan DJ. Transposable elements. Curr Opin Genet Dev 1992;2:861–7. [PubMed: 1335807]

Flatt T, Tu MP, Tatar M. Hormonal pleiotropy and the juvenile hormone regulation of Drosophila development and life history. Bioessays 2005;27:999–1010. [PubMed: 16163709]

Franchini LF, Ganko EW, McDonald JF. Retrotransposon-gene associations are widespread among D. melanogaster populations. Mol Biol Evol 2004;21:1323–31. [PubMed: 15014149]

Glinka EM, Edelweiss EF, Deyev SM. Eukaryotic expression vectors and immunoconjugates for cancer therapy. Biochemistry (Mosc) 2006;71:597–606. [PubMed: 16827650]

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. Demography and natural selection have shaped genetic variation in Drosophila melanogaster: a multi-locus approach. Genetics 2003;165:1269–78. [PubMed: 14668381]

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol 2008;6:e251. [PubMed: 18942889]

González J, Macpherson JM, Messer PW, Petrov DA. Inferring the strength of selection in Drosophila under complex demographic models. Mol Biol Evol 2009a;26:513–26.

González J, Macpherson JM, Petrov DA. A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila melanogaster. Mol Biol Evol. 2009b

Goodier JL, Kazazian HH Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 2008;135:23–35. [PubMed: 18854152]

Harr B, Kauer M, Schlotterer C. Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in Drosophilamelanogaster. Proc Natl Acad Sci U S A 2002;99:12949–54. [PubMed: 12351680]

Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 2005;169:2335–52. [PubMed: 15716498]

Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 1982;101:519–31. [PubMed: 6293914]

Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res 1966;8:269–94. [PubMed: 5980116]

Hoffmann AA, Weeks AR. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 2007;129:133–47. [PubMed: 16955331]

Hua-Van A, Le Rouzic A, Maisonhaute C, Capy P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet Genome Res 2005;110:426–40. [PubMed: 16093695]

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. Distinctly different sex ratios in African and European populations of Drosophila melanogaster inferred from chromosomewide single nucleotide polymorphism data. Genetics 2007;177:469–80. [PubMed: 17660560]

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M, Celniker SE. The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol 2002;3:RESEARCH0084. [PubMed: 12537573]

Kaplan NL, Darden T, Hudson RR. The coalescent process in models with selection. Genetics 1988;120:819–29. [PubMed: 3066685]

Kaplan NL, Hudson RR, Langley CH. The “hitchhiking effect” revisited. Genetics 1989;123:887–99. [PubMed: 2612899]

Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004;303:1626–32. [PubMed: 15016989]

Kidwell MG, Lisch DR. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 2001;55:1–24. [PubMed: 11263730]

Kimura, M. The neutral theory of molecular evolution. Cambridge Univ Press; Cambridge: 1983.

Kreitman M. Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet 2000;1:539–59. [PubMed: 11701640]

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas F, et al. Historical biogeography of the Drosophila melanogaster species subgroup. Evol Biol 1988;22:159–225.

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. [PubMed: 2854088]

Lerman DN, Feder ME. Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster. Mol Biol Evol 2005;22:776–83. [PubMed: 15574805]

Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME. Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. Mol Biol Evol 2003;20:135–44. [PubMed: 12519916]

Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet 2006;2:e166. [PubMed: 17040129]

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. [PubMed: 16283942]

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJ, van Oudenaarden A, Barton DB, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ. Population genomics of domestic and wild yeasts. Nature 2009;458:337–41. [PubMed: 19212322]

Macpherson JM, González J, Witten DM, Davis JC, Rosenberg NA, Hirsh AE, Petrov DA. Nonadaptive explanations for signatures of partial selective sweeps in Drosophila. Mol Biol Evol 2008;25:1025–42. [PubMed: 18199829]

Macpherson JM, Sella G, Davis JC, Petrov DA. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila. Genetics 2007;177:2083–99. [PubMed: 18073425]

Marsano RM, Caizzi R, Moschetti R, Junakovic N. Evidence for a functional interaction between the Bari1 transposable element and the cytochrome P450 cyp12a4 gene in Drosophila melanogaster. Gene 2005;357:122–8. [PubMed: 16076534]

Maside X, Assimacopoulos S, Charlesworth B. Fixation of transposable elements in the Drosophila melanogaster genome. Genet Res 2005;85:195–203. [PubMed: 16174338]

Maside X, Bartolome C, Charlesworth B. S-element insertions are associated with the evolution of the Hsp70 genes in Drosophila melanogaster. Curr Biol 2002;12:1686–91. [PubMed: 12361573]

McClintock B. The origin and behaviour of mutable loci in maize. Proc Natl Acad Sci U S A 1950;36:344–355. [PubMed: 15430309]

McClintock B. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol 1956;21:197–216. [PubMed: 13433592]

McCollum AM, Ganko EW, Barrass PA, Rodriguez JM, McDonald JF. Evidence for the adaptive significance of an LTR retrotransposon sequence in a Drosophila heterochromatic gene. BMC Evol Biol 2002;2:5. [PubMed: 11914129]

McDonald JF. Evolution and consequences of transposable elements. Curr Opin Genet Dev 1993;3:855–64. [PubMed: 8118210]

McDonald JF. Transposable elements: possible catalysts of organismic evolution. Trends Ecol Evol 1995;10:123–126.

McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ, Miller WJ. LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica 1997;100:3–13. [PubMed: 9440254]

Michalak P, Minkov I, Helin A, Lerman DN, Bettencourt BR, Feder ME, Korol AB, Nevo E. Genetic evidence for adaptation-driven incipient speciation of Drosophila melanogaster along a microclimatic contrast in "Evolution Canyon," Israel. Proc Natl Acad Sci U S A 2001;98:13195–200. [PubMed: 11687637]

Miller WJ, Capy P. Mobile genetic elements as natural tools for genome evolution. Methods Mol Biol 2004;260:1–20. [PubMed: 15020798]

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;49:31–41. [PubMed: 3032743]

Muotri AR, Marchetto MC, Coufal NG, Gage FH. The necessary junk: new functions for transposable elements. Hum Mol Genet 2007;16(Spec No 2):R159–67. [PubMed: 17911158]

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, J JS, Adams MD, Cargill M. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 2005;3:e170. [PubMed: 15869325]

Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica 1999;107:129–37. [PubMed: 10952206]

Nuzhdin SV, Pasyukova EG, Mackay TF. Accumulation of transposable elements in laboratory lines of Drosophila melanogaster. Genetica 1997;100:167–75. [PubMed: 9440270]

Ometto L, Glinka S, De Lorenzo D, Stephan W. Inferring the effects of demography and selection on Drosophila melanogaster populations from a chromosome-wide scan of DNA variation. Mol Biol Evol 2005;22:2119–30. [PubMed: 15987874]

Orengo DJ, Aguade M. Detecting the footprint of positive selection in a european population of Drosophila melanogaster: multilocus pattern of variation and distance to coding regions. Genetics 2004;167:1759–66. [PubMed: 15342514]

Orengo DJ, Aguade M. Genome scans of variation and adaptive change: extended analysis of a candidate locus close to the phantom gene region in Drosophila melanogaster. Mol Biol Evol 2007;24:1122–9. [PubMed: 17322555]

Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature 1980;284:604–7. [PubMed: 7366731]

Pardue ML, Rashkova S, Casacuberta E, DeBaryshe PG, George JA, Traverse KL. Two retrotransposons maintain telomeres in Drosophila. Chromosome Res 2005;13:443–53. [PubMed: 16132810]

Pelz HJ, Rost S, Hunerberg M, Fregin A, Heiberg AC, Baert K, MacNicoll AD, Prescott CV, Walker AS, Oldenburg J, Muller CR. The genetic basis of resistance to anticoagulants in rodents. Genetics 2005;170:1839–47. [PubMed: 15879509]

Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila. Mol Biol Evol 2003;20:880–92. [PubMed: 12716993]

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. [PubMed: 9501496]

Petrov DA, Lozovskaya ER, Hartl DL. High intrinsic rate of DNA loss in Drosophila. Nature 1996;384:346–9. [PubMed: 8934517]

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Res 2006;16:1262–9. [PubMed: 16963705]

Pool JE, Bauer DuMont V, Mueller JL, Aquadro CF. A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of Drosophila melanogaster. Genetics 2006;172:1093–105. [PubMed: 16322515]

Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. Evolution 2005;59:2312–23. [PubMed: 16396172]

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol 2005;1:166–75. [PubMed: 16110336]

Schlenke TA, Begun DJ. Strong selective sweep associated with a transposon insertion in Drosophila simulans. Proc Natl Acad Sci U S A 2004;101:1626–31. [PubMed: 14745026]

Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila genome. PLoS Genet 2009;5:e1000495. [PubMed: 19503600]

Shapiro JA. Transposable elements as the key to a 21st century view of evolution. Genetica 1999;107:171–9. [PubMed: 10952210]

Shapiro JA, Huang W, Zhang C, Hubisz MJ, Lu J, Turissini DA, Fang S, Wang HY, Hudson RR, Nielsen R, Chen Z, Wu CI. Adaptive genic evolution in the Drosophila genomes. Proc Natl Acad Sci U S A 2007;104:2271–6. [PubMed: 17284599]

Sheen FM, Levis RW. Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini. Proc Natl Acad Sci U S A 1994;91:12510–4. [PubMed: 7809068]

Singh ND, Arndt PF, Petrov DA. Genomic heterogeneity of background substitutional patterns in Drosophila melanogaster. Genetics 2005;169:709–22. [PubMed: 15520267]

Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res 1974;23:23–35. [PubMed: 4407212]

Smith NG, Eyre-Walker A. Adaptive protein evolution in Drosophila. Nature 2002;415:1022–4. [PubMed: 11875568]

Steiner CC, Weber JN, Hoekstra HE. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol 2007;5:e219. [PubMed: 17696646]

Stephan W, Li H. The recent demographic and adaptive history of Drosophila melanogaster. Heredity 2007;98:65–8. [PubMed: 17006533]

Strobel E, Dunsmuir P, Rubin GM. Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 1979;17:429–39. [PubMed: 110464]

Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps. Genome Res 2006;16:702–12. [PubMed: 16687733]

Thornton K, Andolfatto P. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. Genetics 2006;172:1607–19. [PubMed: 16299396]

Thornton KR, Jensen JD, Becquet C, Andolfatto P. Progress and prospects in mapping recent selection in the genome. Heredity 2007;98:340–8. [PubMed: 17473869]

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 2007;39:31–40. [PubMed: 17159977]

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 2003;19:530–6. [PubMed: 14550626]

Vieira C, Biemont C. Geographical variation in insertion site number of retrotransposon 412 in Drosophila simulans. J Mol Evol 1996;42:443–51. [PubMed: 8642613]

Walser JC, Chen B, Feder ME. Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. PLoS Genet 2006;2:e165. [PubMed: 17029562]

Welch JJ. Estimating the genomewide rate of adaptive protein evolution in Drosophila. Genetics 2006;173:821–37. [PubMed: 16582427]

Williamson S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol 2003;20:1318–25. [PubMed: 12777505]

Zatsepina OG, Velikodvorskaia VV, Molodtsov VB, Garbuz D, Lerman DN, Bettencourt BR, Feder ME, Evgenev MB. A Drosophila melanogaster strain from sub-equatorial Africa has exceptional thermotolerance but decreased Hsp70 expression. J Exp Biol 2001;204:1869–81. [PubMed: 11441029]

Abbreviations

AF	African
DR	direct repeat
ERVs	Endogenous Retrovirus
LINEs	Long Interspersed Nuclear Elements
LTRs	Long terminal repeats
MITEs	Miniature inverted repeat elements
NA	North American
ORFs	Open Reading Frames
SINEs	Short Interspersed Nuclear Elements
TIR	terminal inverted repeats
TEs	Transposable elements
UTR	untranslated regions

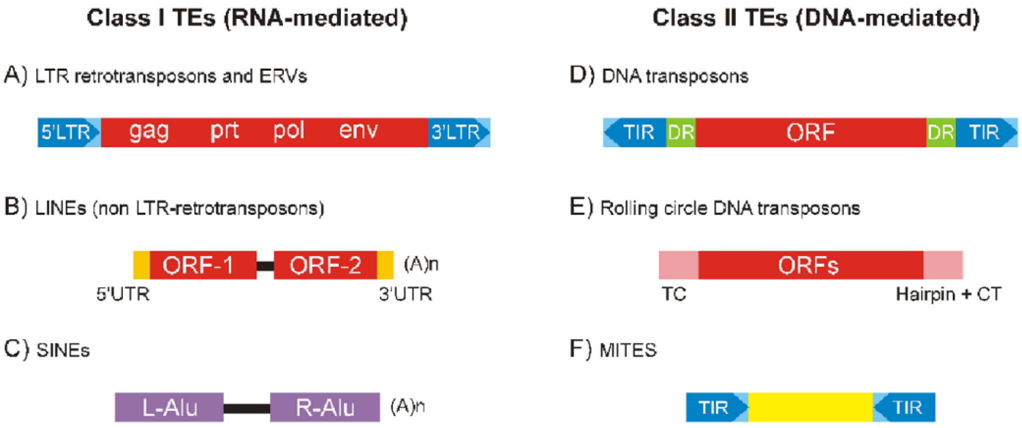

Figure 1.

Characteristics of the main types of TEs. A) Long Terminal Repeat (LTR) retrotransposons and Endogenous Retrovirus (ERVs) have partly overlappping Open Reading Frames (ORFs): group specific antigen (gag), protease (ap), an ORF encoding the reverse transcriptase and the integrase (pol) and envolope (env). The ORFs are flanked in both ends by LTRs with promoter capability. B) Long interspersed nuclear elements (LINEs) or non-LTR retrotransposons consist of a 5′ untranslated regions (UTR) with promoter activiy, two ORFs separated by a spacer and a 3′ UTR with a poly-A tail: (A)n. C) The Alu element, the most common Short Interspersed Nuclear Element (SINE) in the human genome consists of two GC-rich fragments the left (L-Alu) and right (R-Alu) connected by a A-rich linker and end in a poly-A tail. D) DNA transposons consist of an ORF that contains a DNA recognition and binding domain and a catalytic domain flanked by short terminal inverted repeats (TIR) an may also include direct repeats (DR). E) Rolling circle DNA transposons encode several ORFs including helicase, replication initiator protein (Rep) and single-stranded DNA binding protein (SSB) and are flanked by a 5′ conserved TC dinucleotide and a 3′ conserved hairpin and CT dinucleotide. F) Miniature inverted repeat elements (MITEs) have no ORFs and are flanked by TIR.

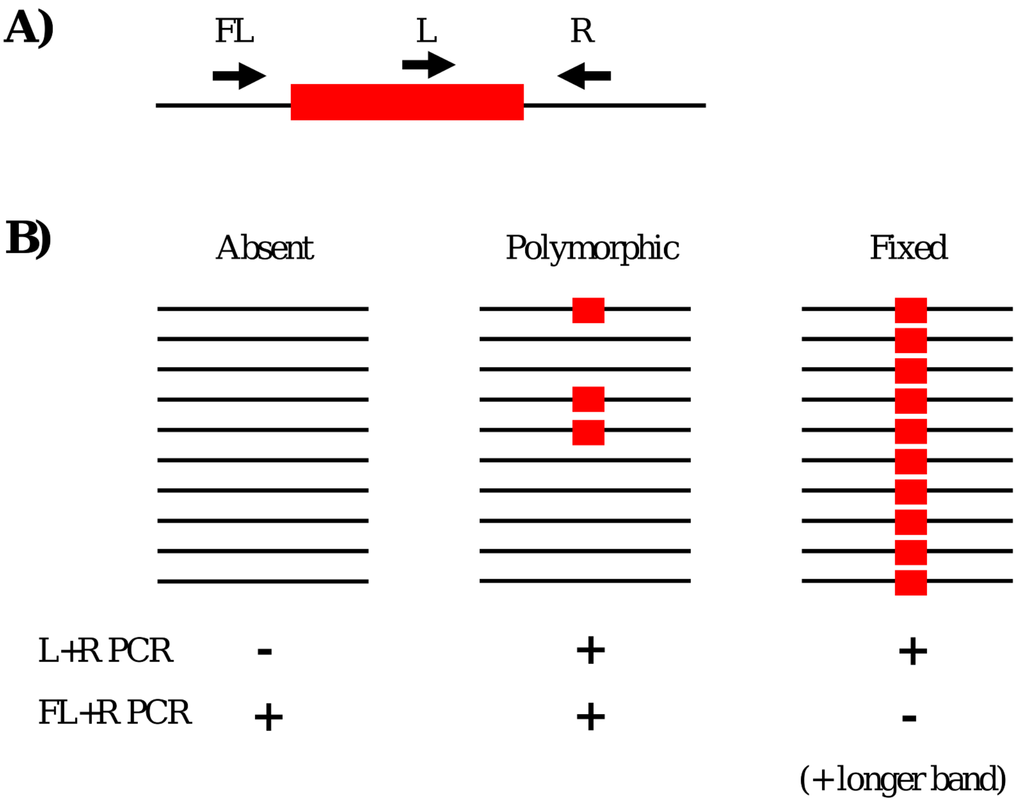

Figure 2.

Schematic representation of the pooled-PCR approach. A) The black line represents the genome, the red rectangle the TE and the arrows the different primers: L left primer, R right primer and FL flanking primer. B) Success of PCR with L and R primers indicates the presence and with FL and R primers the absence of a TE.

Table 1

Characteristics of the putatively adaptive TEs.

TE	Class	Rcb rate (cM/Mb)^{a}	Location	GO (biological process) of closest gene	Evidence of selection	Reference
S-element	TIR	0.02	5′ of *Hsp70Bb*	Response to heat	**1** Fixed in a highly recombining region **2** Reduced polymorphism in the inverted repeat	Maside et al. 2002
Bari-1	TIR	1.04	3′ UTR of *cyp12a4*	Response to insecticide	**1** Fixed insertion **2** Increased gene expression	Marsano et al. 2005
Antonia	LTR	0	intron of *Cht3*	Cuticle chitin catabolic process	**1** Fixed insertion **2** Reduced polymorphism in and around the insertion	McCollum et al. 2002
Quasimodo	LTR	3.19	5′ of *CTCF*	Negative regulation of transcription	**1** Fixed insertion **2** Reduced polymorphism in and around the insertion	Franchini et al. 2004
Doc	LINE	1.65	exon of *CHKov1*	RNA-dependent DNA replication	**1** High frequency in non-AF and low frequent in AF populations **2** Signatures of selective sweep **3** Truncation of *CHKov1* **4** Resistance to pesticides	Petrov et al. 2003; Aminetzach et al. 2005; Macpherson et al. 2008; González et al. 2008
Jockey	LINE	0.01	5′ of *Hsp70Ba*	Response to heat	**1** Decreased gene expression	Zatsepina et al. 2001; Lerman et al. 2003
P-element	TIR	0.01	5′ of *Hsp70Ba*	Response to heat	**1** Decreased gene expression **2** Increased female reproductive success	Michalak et al. 2001; Lerman et al. 2003
P-element	TIR	0.01	5′ of *Hsp70Ba*	Response to heat	**1** Decreased expression of *Hsp70Ba* **2** Increased female reproductive success **3** Increased thermotolerance	Lerman et al. 2003
P-element	TIR	0.02	5′ of *Hsp70Bb*	Response to heat	**1** Decreased gene expression	Lerman & Feder 2005
P-element	TIR	3.21	5′ of *Hsp26*	Response to heat	**1** Decreased gene expression **2** Increased thermotolerance in females **3** Increased fecundity	Chen et al. 2007
Accord	LTR	0.94	5′ of *Cyp6g1*	Response to insecticide	**1** Reduced polymorphism around the insertion **2** Increased gene expression **3** Resistance to pesticides	Daborn et al. 2002; Catania et al. 2004
Doc	LINE	--	5′ of *Cyp6g1*	Response to insecticide	**1** Reduced polymorphism around the insertion **2** Increased gene expression	Schlenke and Begun 2004
Bari1	TIR	2.05	3′ of *Jheh3*	defense response response to toxin	**1** Signatures of selective sweep	González et al. 2008

TE	Class	Rcb rate (cM/Mb)[a]	Location	GO (biological process) of closest gene	Evidence of selection		Reference
					2	Decreased gene expression	
					1	Signatures of selective sweep	
F	LINE	3.56	intron of *Kuz*	Notch signalling pathway CNS development	**2**	Population differentiation	González et al. 2008
					3	Increased gene expression	
pogo	TIR	3.31	3′ UTR of *Kmn1*	chromosome segregation	**1**	Signatures of selective sweep	González et al. 2008
					1	Signatures of selective sweep	
pogo	TIR	3.46	intron of *CG18210*	no info	**2**	Population differentiation	González et al. 2008
pogo	TIR	3.45	intron of *CG9413*	amino acid metabolism eye development			González et al. 2008
					1	Population differentiation	
S-element	TIR	0.28	intron of *rdx*	segment polarity determination	**2**	Decreased gene expression	González et al. 2008
invader4	LTR	0.63	intron of *sra*	Protein ubiquitination long-term memory olfactory learning regulation of female receptivity	**1**	Population differentiation	González et al. 2008
Doc	LINE	3.14	3′ of *Jon65Aiv*	proteolysis *defense response to virus*			González et al. 2008
					1	Population differentiation	
S-element	TIR	2.97	intron of Ago2	*RNA interference aumiddlehagic cell death*	**2**	Decreased gene expression	González et al. 2008
pogo	TIR	1.38	intron of *CG31163*	no info	**1**	Population differentiation	González et al. 2008
Rt1b	LINE	1.81	intron of *CG34353*	no info	**1**	Population differentiation	González et al. 2008
Rt1a	LINE	3.19	3′ of *CG6175*	no info	**1**	Population differentiation	González et al. 2008

[a] Recombination rate estimated using the Recombination Rate Calculator available at http://petrov.stanford.edu/software