Chapter 13

Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans

Josefa González and Dmitri A. Petrov

Abstract

Recent research is starting to shed light on the factors that influence the population and evolutionary dynamics of transposable elements (TEs) and TE life cycles. Genomes differ sharply in the number of TE copies, in the level of TE activity, in the diversity of TE families and types, and in the proportion of old and young TEs. In this chapter, we focus on two well-studied genomes with strikingly different architectures, humans and Drosophila, which represent two extremes in terms of TE diversity and population dynamics. We argue that some of the answers might lie in (1) the larger population size and consequently more effective selection against new TE insertions due to ectopic recombination in flies compared to humans; and (2) in the faster rate of DNA loss in flies compared to humans leading to much faster removal of fixed TE copies from the fly genome.

Key words: Transposable elements, Population dynamics, Population size, Ectopic recombination, Drosophila, Humans, Genetic drift, Genetic draft, Transposition rate, Deletion rate

1. Introduction

Transposable elements (TEs) are short DNA sequences, typically from a few hundred bp to ~10 kb long, that have the ability to move around in the genome by generating new copies of themselves. TEs are an ancient, extremely diverse and exceptionally active component of genomes. TEs have been found in virtually all organisms studied so far including bacteria, fungi, protozoa, plants, and animals (1, 2). The main TE groups, class I and class II, are present in all kingdoms, revealing their persistence over evolutionary time (1). These two classes of TEs differ in their transposition intermediates: while class I TEs transpose through RNA intermediates, class II TEs transpose directly as DNA (Fig. 1). TEs within each class are further classified into (1) different orders, based on their insertion mechanism, structure, and encoded proteins; (2) into

Maria Anisimova (ed.), *Evolutionary Genomics: Statistical and Computational Methods, Volume 1*,
Methods in Molecular Biology, vol. 855, DOI 10.1007/978-1-61779-582-4_13,

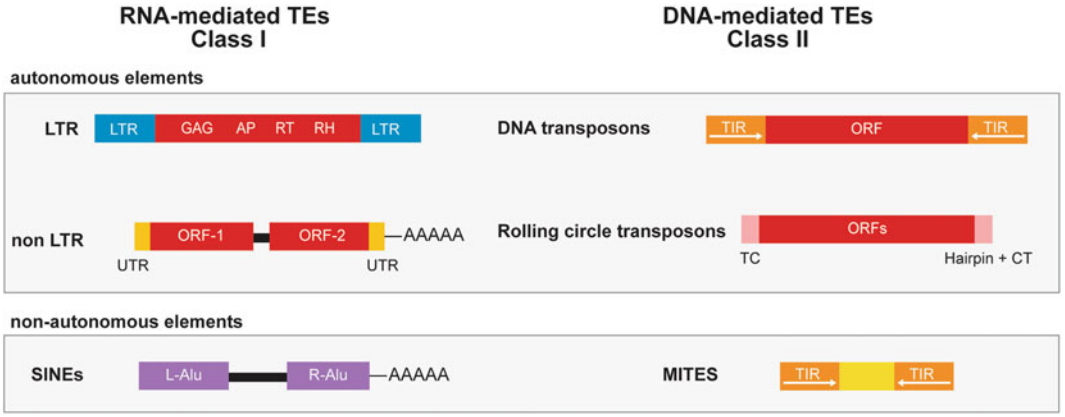

Fig. 1. The two main classes of TEs. Long terminal repeat (LTR) elements have several open reading frames (ORF): capsid protein (GAG), aspartic proteinase (AP), reverse transcriptase (RT), and RNase H (RH). These ORFs are flanked at both ends by LTRs with promoter capability. Non-LTR elements consist of a 5′ untranslated region (UTR) with promoter activity, two ORFs separated by a spacer, and a 3′ UTR with a poly-A tail. The Alu element, the most common short interspersed nuclear element (SINE) in the human genome, consists of two GC-rich fragments the left-Alu (L-Alu) and right-Alu (R-Alu) connected by an A-rich linker and ends in a poly-A tail. DNA transposons consist of an ORF flanked by short terminal inverted repeats (TIR). Rolling circle DNA transposons contain several ORFs and are flanked by a 5′ conserved TC dinucleotide and a 3′ conserved hairpin and CT dinucleotide. Miniature inverted repeat elements (MITEs) have no ORFs and are flanked by TIRs.

different superfamilies, based on their replication strategy; and (3) into different families, based on sequence conservation (1, 2).

TEs constitute a substantial albeit variable (from ~1% to greater than 80%) proportion of genomes (3, 4). In the human genome, for instance, ~40–45% of the genome is identifiable TEs, 5% is genes and other functional sequences (functional RNAs or regulatory regions), and the remaining ~50% of the genome has no identifiable origin (5). There is a reasonable chance that the unidentifiable 50% of the human genome is also ultimately of TE origin. In general, the TE-generated fraction of genomes is likely to be underestimated because methods for detecting TEs in genomic sequences are necessarily biased toward younger and more easily recognizable TEs. Even if we limit ourselves to detectable TEs, many genomes can be thought of as graveyards of TEs with genes sprinkled in between.

TEs are extremely active genomic denizens, at times generating a large proportion of all spontaneous visible mutations (e.g., 50–80% in Drosophila) (6) and importantly being able to generate mutations of a great diversity of types (7, 8). TE-induced mutations range from subtle regulatory mutations to gross genomic rearrangements and often have phenotypic effects of a complexity that is not achievable by point mutations. For example, TEs can affect the expression of nearby genes by adding new splice sites, adenylation signals, promoters, or transcription factor binding sites (9) and also by serving as the targets of epigenetic histone modifications that

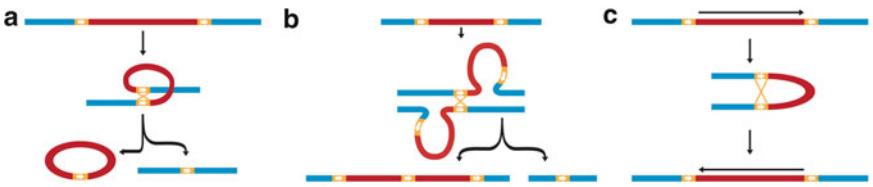

Fig. 2. TE-mediated rearrangements. Ectopic recombination between TE copies (*black*/*orange boxes*) in the same orientation can lead to deletions when recombination takes place between copies located on the same chromatid (**a**) or deletions and duplications when recombination takes place between copies in different chromosomes (**b**) (recombination between two nonhomologous chromosomes should lead to a translocation). Ectopic recombination between TE copies in opposite orientation leads to inversion of the DNA between the two TEs (**c**).

spread into adjacent genes (10). Moreover, there is evidence that insertions of multiple TE sequences containing a functional regulatory sequence near many genes at once might be instrumental in creating new regulatory networks and might contribute substantially to the evolution of novelty (9, 11–13). TEs can both remove DNA from the genome by generating target site deletions and add DNA through 3′ and, less frequently, through 5′ transduction (14, 15). TEs contribute to protein-coding regions both at the transcript and at the protein level (16–18) and TE-encoded proteins have been domesticated and are part of host genes (8). Additionally, ectopic recombination between TEs causes deletions, duplications, and sequence rearrangements (Fig. 2). TE-induced mutations are frequently deleterious, although a number of adaptive mutations have been described (19–22). We recently showed that TEs are a considerable source of adaptive mutations in Drosophila (23–26).

Given the abundance, ubiquity and the role of TEs in genome content, structure, and evolution, it is indisputable that a thorough understanding of TE population dynamics is essential for the understanding of the eukaryotic genome structure, function, and evolution.

2. Genomes Differ in Content, Diversity, and Activity of TEs

Genomes differ sharply in the number of TE copies in the genome, in their TE activity, in the diversity of TE families and types, and in the proportion of old and young TEs (8, 27). The information gleaned recently from the sequencing of multiple eukaryotic genomes is giving us a sense of the remarkable diversity of the intragenomic ecologies of TEs. The difficulty of TE annotation and the lack of TE polymorphism data for most organisms, however, have not allowed the field to progress as far as desired in understanding the evolutionary and population genetic forces acting on TEs in different genomes and lineages. In addition, as often is the case,

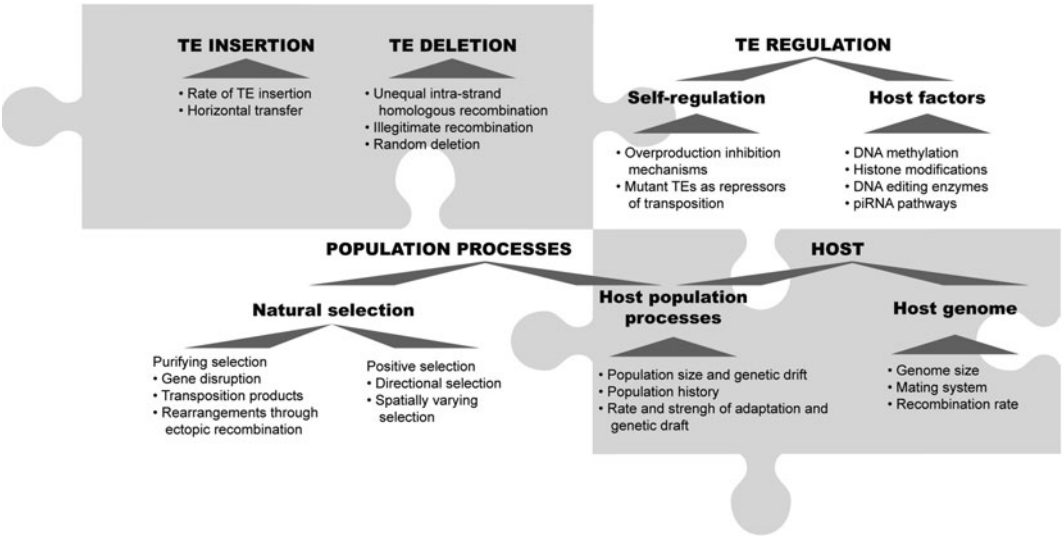

Fig. 3. Factors that influence the population and evolutionary dynamics of TEs. Our understanding of TE population and evolutionary dynamics is still incomplete. The different factors that affect TE population and evolutionary dynamics are interrelated and future research is likely to reveal existence of additional factors.

acquisition of more knowledge only exposes the true depths of our ignorance. Recent research is just starting to shed light on the factors that influence the population and evolutionary dynamics of TEs and TE life cycles and thus should contribute to the understanding of genomic patterns of TE diversity (Fig. 3).

Much of the detailed information on TE evolution still comes from two of the best-studied genomes: fruit flies (*Drosophila melanogaster*) and humans. Fortunately, these two genomes represent two extremes in terms of TE diversity and population dynamics and thus give a reasonably diverse picture of the TE evolution and dynamics. For the rest of this chapter, we focus primarily on these two genomes and will highlight the similarities and differences observed between them.

As mentioned above, the human genome has millions of TE copies, with at least 45% and possibly up to 90% of the genome derived from TE sequences (5). Two human retrotransposable element (Class I) families, LINE1 (L1) and Alu, account for 60% of all interspersed repeat sequences. The vast majority of the TEs in the human genome are fixed and most families are inactive. In fact, only one LINE1 subfamily, *L1Hs*, is active in humans. Consistent with their low level of activity, TEs are responsible for only ~0.2–0.3% of spontaneous mutations in humans (28, 29).

In contrast, the fruit fly *D. melanogaster* genome contains only thousands of individual TE copies (5,424 TE copies in Flybase R5.23) that account for only 5.3% of the euchromatin (30). *D. melanogaster* TEs belong to approximately 100 diverse families of both Class I and Class II elements (30, 31). Each family consists

of 1–304 copies with no dominant family corresponding to the majority of TEs. The only exception is INE-1 family that contains ~2,000 copies and has been inactive for the past ~3 million years (32, 33). The majority of TE families are active in Drosophila with individual TE copies generally polymorphic in the population and showing a high sequence similarity (30, 31, 34). Consistent with the abundance of young and polymorphic TE copies, TE families in *D. melanogaster* are often very active, responsible for >50% of spontaneous visible mutations (6).

Why do these two genomes differ so profoundly in content, diversity and activity of TEs? The answer must lie in different aspects of TE population dynamics within genomes and forces that lead to varying rates of TE family birth and extinction. In the rest of this review, we focus on the state of knowledge of different aspects of TE population dynamics and discuss aspects of TE family evolution. Specifically, we focus on rates of TE transposition, excision, fixation, or loss in human and *D. melanogaster* populations due to stochastic forces and natural selection for or against TE insertions, the persistence of TE sequences postfixation, and forces that affect coexistence of multiple TE families and the standing diversity of TE types (Fig. 3). We do not consider horizontal transfer because it is mostly restricted to closely related species and as such is not likely to contribute to the difference in population dynamics between Drosophila and humans (35–37).

3. Methodology Used to Study TE Population Dynamics

Three main approaches have been used to study TE population dynamics: mathematical modeling, computer simulations, and the analysis of empirical data. Most models attempt to determine conditions under which TE copy numbers stabilize in the genome despite the apparent ability of TEs to self-replicate and thus amplify uncontrollably through a positive feedback loop. Classical mathematical modeling approaches generally considered situations in which TEs were either subject to regulatory feedback that diminished the rate of transposition (or increased the rate of excision) or to stronger purifying natural selection when the copy numbers of TEs increased (38). More recent mathematical models take into account additional parameters, such as the specific features of the TE insertions or the mating system of the host species (39–42).

Mathematical models have traditionally considered the dynamics between the host and a homogeneous group of TEs in a homogeneous genome under constant population size and strength of selection. However, sequencing data demonstrated that only a small percentage of the TEs in a genome are full-length, potentially active copies (5, 31). Computer simulations allow more realistic

TE dynamics. For example, a recent simulation considered the variability of insertion effects, from deleterious to adaptive and also considered mutations that lead to the loss of transposition activity or to nonautonomous copies (43). Recently, computer simulations that take into account regulation of transposition by host factors have also been developed (44).

Analysis of empirical population data is often used to test predictions of these population dynamics models. Early studies of population variation were performed using Southern blot and in situ hybridization techniques (45, 46). Another popular technique—transposon display—was derived from Amplified Fragment Length Polymorphism (AFLPs) and allows detection of most individual TEs from a given family (47). TE insertions are identified by a ligation-mediated PCR that starts from within the TE insertion, and amplifies part of the flanking sequence up to a specific restriction site. The resulting PCR products are analyzed using high-resolution polyacrylamide gel systems and variation among individuals in the sizes of the PCR bands allows an assessment of the patterns of TE polymorphism in populations (48, 49).

However, these techniques have some important limitations. First, they are restricted to the known, well-studied families. Southern hybridization also has the drawback of not being able to identify individual TE insertions and gives no information about the location of the TE insertions. Variability in banding patterns in the transposon display technique could arise not only from the presence/absence of individual TEs, but also from variation in the presence/absence of the restriction sites flanking TEs and/or from polymorphisms at PCR primer-binding sites (50). On the other hand, although in situ hybridization does give immediate information about TE copy number and location, it unfortunately has a strong bias against short TE insertions. As a result, much of the data generated by in situ hybridization in Drosophila ended up limited to high copy number families that contain very long TEs. Unfortunately, such families behave very differently from families with low copy numbers containing short TEs (51). Specifically, long TE copies from high-copy number families tend to be rare while short TE copies from low-copy number families tend to be common (34, 51) (see Subheading 5.1).

The availability of the first whole genome sequences made it possible to develop PCR approaches in order to investigate TE population dynamics (51–55). Specific primers for individual TE insertions can be designed and used to test for the presence and/or absence of individual TE insertions in one or several different individuals. These approaches have been used to perform less biased population genomic analyses of TEs in *D. melanogaster* and *Arabidopsis thaliana* (34, 56). Although powerful, PCR approaches are time-consuming and fairly expensive. The availability of whole genome sequences from multiple individuals and software that can

perform automated annotation of TEs (30) as well as automated assessment of TE frequencies in the population from multiple resequenced strains (57–59) should significantly increase our ability to obtain and analyze empirical TE population genomics data.

4. Rates of Transposition

Below we briefly describe some of the empirical findings about the key aspects of TE dynamics in flies and humans. We first focus on estimates of the rates of transposition and then on the mechanisms that control transposition.

4.1. Empirical Estimates of the Rates of Transposition in Drosophila and Humans

In Drosophila, insertion rates have been empirically estimated by scoring TE insertions using in situ hybridization and rescoring them after several generations (60–62). The rationale behind those experiments is that since selection against a TE insertion in laboratory conditions is on average very weak, most transpositions will accumulate in laboratory populations as spontaneous neutral mutations. These estimates ranged from 10^{-5} to 10^{-3} events/copy/generation under standard conditions (62–65).

These rates are surrounded by considerable uncertainty, however, as they can vary over several orders of magnitude among different TE families, different strains (or mating between specific strains), and different environmental conditions (66–71). In fact, it appears that transposition rates for many TE families are typically very low if not at zero and that most transposition events take place in the strains where the repression of quiescent TEs breaks down. Hybrid dysgenesis in Drosophila, a sterility syndrome generated by very high rates of transposition of normally inactive TE families upon mating between (but not within) specific strains in *D. melanogaster* or *D. virilis*, is one of the clearest cases of such repression/derepression (66, 67, 69, 72, 73).

There is also evidence that variation at host loci can affect rates of transposition as has been discovered for the regulation of *gypsy* by the gene *flamenco* (74). In general, it is possible that in the population the rate of transposition for active families is primarily determined by the frequency of inactivating polymorphisms within host genes that normally prevent transposition from taking place (46). At the same time, some TE families may evolve active copies that escape regulation altogether and others are subject to such strict and redundant control that they are basically inactive in the population. Some tenuous evidence exists that TE families go through periods of high and low activities (34, 51) that might be a consequence of coevolution between active elements within a TE family that evolve to avoid repression and the host genes that reevolve to repress the TEs that get out of control.

In humans, several methodologies have been used to estimate the rate of transposition. The work naturally focused on the two dominant and still active TE families: L1 elements and Alus. Cell culture-based retrotransposition assays suggest that the rate of new insertions of L1 is one per two to 33 individuals (75, 76). Comparison of the frequency of disease-causing de novo L1 events compared to nucleotide mutations suggested that one new L1 retrotransposition event takes place every 10–20 births (77, 78). Deininger and Batzer (79) estimated the rate of new Alu insertions to be one in every 125 births based on the number of Alu insertions fixed during the last 5 million years, assuming that Alu insertions are predominantly neutral in their effect. Recently, two new methodologies have been applied to estimate TE insertion rates in humans. Ewing and Kazazian (80) used high-throughput sequencing techniques to analyze the genome of 25 individuals and estimated that the rate of new L1 insertions is one in 140 births. Briefly, they estimated Watterson's θ, a measure of genetic variation in a population, using segregating L1 insertion sites and then used this to estimate the rate of L1Hs retrotransposition per live birth assuming an effective population size of 10,000. Huang et al. (81) obtained a similar estimate (one L1 insertion in every 108 births) based on the analysis of 75 human genomes using genome-wide transposon insertion profiling by microarray. The authors identified the nonreference insertions in their dataset by comparing them to the reference genome. Because the reference genome is a haploid genome they estimate the ratio of homozygous to heterozygous insertions in order to obtain the total number of insertions in the haploid genome.

Both in Drosophila and humans, the current insertion rate estimates, one per one to 100 generations in flies and about one per 100 generations in humans, are subject to considerable error. Next-generation sequencing technology opens up new routes to obtaining much more accurate rates of transpositions by obtaining very deep population samples at a genome-wide level and focusing on TE copies present at very low population frequencies. Messer (82) showed that such data provide an accurate estimate of mutation rates (and thus transposition rates as well) in a way that is only weakly sensitive to effects of natural selection. Analyses of such deep population genomic data should elucidate whether Drosophila and humans do differ in the rates of TE insertions to the extent that previous estimates have suggested.

4.2. Transposition Control Mechanisms

The mechanisms of TE regulation are highly diverse and many are still shrouded in mystery. Here, we describe some basic findings about the way TEs self-regulate and the way they are regulated by the host factors. It is likely that the ways TEs are regulated will prove as diverse as, or even more diverse than, the ways in which gene expression is regulated in general.

4.2.1. TE Self-regulation

The occurrence of self-regulation has been described for some TE families in different species. For instance, P elements regulate their own transposition in a tissue-specific manner: the third intron of the transposase gene is removed to encode an active transposase protein in the germ line, but is not removed in somatic cells, producing an inactive transposase (83). This mechanism of regulation allows the element to be active in germ cells and to be passed to new genetic locations in future genomes, but to be stable in somatic cells where genetic damage could cause a reduction in host fitness.

There are also several mechanisms of regulation that control expression of active P elements. One of them, multimer poisoning, involves KP elements. KP elements are deletion-derived P elements that form inactive multimers with the transposase or with a host protein required for transposition (84). Multimer poisoning has also been shown to be involved in the regulation of mariner-like elements. Another common regulatory mechanism for P elements and mariner-like elements is transposase titration. This mechanism of regulation involves defective TE copies that retain their transposase-binding sites and regulate the transposition of full-length copies through titration of the active transposase (85, 86). Mariner elements are also regulated by an overproduction inhibition mechanism in which an excess of the wild-type transposase reduces the overall level of transposase activity likely by posttranslational interactions between protein subunits (87).

Most of the self-regulatory mechanisms described so far apply to DNA elements in organisms other than humans. Future research will determine whether self-regulatory mechanisms play an important role in human TE dynamics.

4.2.2. Regulation by Host Factors

Host genomes have evolved multiple strategies to control TE activity: DNA methylation, chromatin modification, DNA-editing enzymes, and RNAi pathways have all been implicated as ways of repressing TE activity. It is in fact possible that some or even most of these epigenetic control mechanisms evolved originally as means of controlling TEs.

DNA methylation is one such case—it is widely believed to have evolved primarily as a defense mechanism against TE insertions, although it obviously also plays an important role in regulation of host gene expression (88). Methylation of TEs by the host genome leads to suppression of transcription thus preventing further replication of TEs. Note that methylation is not a universal control mechanism of TE activity: while in humans, the majority of methylated cytosines occur in repetitive sequences, and methylation is likely responsible for repressing TEs, in Drosophila DNA methylation is rare, restricted to embryos and not present in germ line cells (89–91), and does not seem to play a role in TE control (92).

DNA-editing enzymes and histone modifications have also been suggested to play a role in controlling human TE insertions. For example, APOBEC3 proteins are known inhibitors of human exogenous retroviruses (93, 94). These proteins are cytidine deaminases that edit dC residues to dU during reverse transcription resulting in G to A hypermutations (93). Not surprisingly, APOBEC3 proteins are inhibitors of TE retrotransposition as well and, although initial reports suggested that inhibition was due to the DNA-editing capabilities of this enzyme, there is also an APOBEC3 deaminase-independent mechanism of TE repression whose action remains obscure at the moment (95, 96). Recently, APOBEC1 has been shown to employ both a deaminase-independent and a deaminase-dependent mechanism to reduce the mobility of L1 and LTR elements, respectively (97). A role of a DNA-editing enzyme in Drosophila TE population dynamics has not been described.

Posttranslation modifications of histones play a critical role in the assembly of heterochromatin and in gene expression. Traditionally, histone modifications have been considered to provide a molecular mechanism for TE silencing in plants, fungi, and mammals including humans (10, 98–100). However, a recent study in humans suggests that histone modifications may also represent an additional mechanism by which TEs can contribute to the regulatory functions of the host genome (101). In Drosophila, the relationship of histone modifications with TE expression has not been demonstrated (102).

Finally, TE silencing by small RNA pathways has been described both in humans (103, 104) and fruit flies (105, 106). RNAi is a mechanism in which double-stranded RNA (dsRNA) recognizes homologous mRNAs and causes sequence-specific degradation in a multistep process. The role of RNAi in TE control was discovered when nematodes deficient in RNAi pathways showed increased TE activity (107, 108). Two different small RNAs regulate TEs: small-interfering RNA (siRNA) and piwi-interacting RNA (piRNA). siRNAs are produced from dsRNA processed by the *Dicer* endoribonuclease while piRNAs are *Dicer*-independent (106). Repression of TEs by siRNAs and piRNAs takes place through transcriptional silencing by both DNA methylation and heterochromatin formation and through posttranscriptional silencing.

Recent work has demonstrated that maternally inherited small RNAs are the essential factor mediating germ line transposon silencing that underlies hybrid dysgenesis in Drosophila (69, 109, 110). Although small RNAs are produced in both sexes, they are only maternally loaded in embryos. This explains why TEs are derepressed in crosses between females lacking particular TEs and males carrying those TEs, causing the hybrid dysgenesis syndrome. The reciprocal cross, with females carrying particular TE insertions and males lacking those insertions, does not lead to hybrid

dysgenesis because, in this case, the females transmit to their progeny the small RNAs responsible for the repression of the TEs. Overall, small RNA pathways seem to play an important role in limiting TE proliferation.

5. Rate of Fixation and Frequency Distribution

In addition to the variation in the rate of transposition, it is essential to consider what happens to TEs in populations after they insert into the genome. Both natural selection and stochastic processes affect the probability that TEs will be lost or fixed in the population as well as the frequency distributions of the polymorphic TEs. Below we describe the current understanding of the population dynamics of TEs in Drosophila and humans.

5.1. Natural Selection Against TE Insertions

Natural selection against the deleterious effects of TE insertions has long been considered the dominant force limiting TE spread in populations (111). There are three not mutually exclusive hypotheses about the nature of selection acting against TE insertions. TE insertions may be deleterious because they affect the coding capacity or the regulation of their nearby genes ("deleterious insertion model") (112, 113). TEs could be deleterious because the translation of TE-encoded proteins or transcripts may be costly and these proteins/transcripts might generate deleterious effects by nicking chromosomes and disrupting cellular processes ("deleterious transposition model") (46). Finally, TE insertions belonging to the same family, and independently of their local effects, can provide substrates for ectopic recombination resulting in deleterious chromosomal rearrangements, ("ectopic recombination model") (114) (Fig. 2).

In humans, there is evidence that purifying selection does act against TE insertions, albeit weakly. Boissinot et al. (115) determined the population frequency of ~100 polymorphic L1 elements belonging to the active Ta1 family and a number of Alu elements. Full length elements, but not truncated elements or Alu insertions generated by Ta1 activity, were present at detectably lower frequency, indicating the action of natural selection against longer L1 elements. The strength of the selection was estimated to be on the order of $N_e s \sim -2$, corresponding to the selection coefficient of roughly 0.02%.

The reason for the deleterious effect of the longer L1 elements is not clear, but the data are more consistent with either the deleterious transposition or the ectopic recombination model than with the deleterious insertion model. Specifically, the effects of the regulatory sequences contained in full-length elements on host genes, TE-encoded proteins and/or RNAs, and the higher

propensity of full-length elements to ectopically recombine with other Ta1 elements, are among the likely explanations for the different behavior of full-length versus truncated L1 elements and Alu insertions.

Note that the ectopic recombination model explains naturally how detectably strong natural selection could be acting against new TE insertions while at the same time it explains how it is possible that the human genome could contain millions of TE copies. The possible reason for this is that ectopic recombination is more common for heterozygous TEs and thus selection would only operate against polymorphic TEs while allowing fixed TEs to be of little cost.

In Drosophila, the debate about which of the mechanisms of selection are dominant in controlling the spread of TEs has been going on for the last 30 years. Much of the recent evidence is pointing toward selection against insertions of TEs within genes being very strong such that TEs within genes are not observed even at low frequencies (31, 34). TE insertions outside of genes appear to be deleterious primarily due to ectopic recombination, with selection intensities ranging among different families from $N_e s \sim -2$ (as in humans) to $N_e s \sim -100$ (34). Below we briefly outline the evidence for these conclusions.

The ectopic recombination model predicts that areas of low recombination should accumulate TEs. Population studies of chromosomal distributions of several TE families in Drosophila using in situ hybridization have generally (51, 116–120) but not always (121) supported this prediction. One objection to these findings, other than being based on a small number of families with possibly idiosyncratic properties, was that other than reduced levels of ectopic recombination, areas of low recombination also experience less efficient selection due to the Hill–Robertson interference (122), which should also lead to accumulation of TEs in these regions. However, Dolgin and Charlesworth (123) performed extensive Monte Carlo simulations and determined that TEs should accumulate as a result of Hill–Robertson effects only in regions of extremely low recombination when excision is effectively absent. These authors argue that because DNA transposons do excise from the genome at appreciable rates (62, 124–127), this selective mechanism can probably be ruled out for DNA transposons.

Probably, the strongest evidence for the ectopic recombination model comes from our recent *D. melanogaster* study of 755 euchromatic TEs across 55 different families (including all the families with more than 20 copies in the reference genome) (34). We found evidence for all predictions of the ectopic recombination model: (1) TEs are subject to variable strength of selection depending on the family, but not superfamily, identity; (2) natural selection is stronger for longer TEs, (3) natural selection is stronger on TEs that belong to families with a larger number of copies and importantly a larger number of longer copies. Models that take only these

factors into account are able to explain more than 40% of the variance in the frequency estimates of different TEs (34). This indicates that the discovered rules apply to TEs independently of the details of their transposition and regulation mechanisms or specific proteins they encode. Only the ectopic recombination model appears consistent with these observations because it relies exclusively on the fact that TEs are repetitive sequences and not on the specific details of particular TEs.

It appears that TE families in Drosophila transpose at substantially variable rates and equilibrate at different copy numbers in the genome. The families that transpose frequently build up large copy numbers such that the rate of ectopic recombination among the TEs in such families and the attendant rate of generation of chromosomal abnormalities become sufficiently high that natural selection starts eliminating TEs from the population as fast as they are being added by transposition. These families end up containing relatively large copy numbers of TEs that are both young and present at low frequencies. The families that transpose less frequently end up with lower copy numbers with individual TEs being generally older and more frequent in the population. For these families, the rates of ectopic recombination are lower, and while the resulting purifying selection is weak it is nevertheless strong enough to counteract the low rates of transposition in these families.

Most of the TEs in Drosophila appear to be subject to sufficiently strong purifying selection that they are very unlikely to fix in the population. In contrast, TEs in the human genome fix at reasonable rates and accumulate in the genome with time. The reason for this difference might lie in the difference in effective population size of these organisms. Indeed, long-term effective population size in Drosophila is thought to be on the order of $N_e = 10^6$ (128–130) while in humans it is closer to $N_e = 10^4$ (131). The two orders of magnitude difference means that the 0.01% selective disadvantage estimated for the long L1 elements in humans, while it is weak and ineffective in human populations ($N_e s \sim -1$) would translate into very effective selection in Drosophila ($N_e s \sim -100$). It is possible that a low effective population size in humans is one of the reasons for the large numbers of fixed TEs in the human genome. Note that one way the total numbers of TEs could be reduced is via their loss by subsequent deletion (132, 133). We discuss the process of elimination of fixed TEs from the genome in more detail in Subheading 6.

5.2. Adaptation Generated by TE Insertions

Although it is likely that most TE insertions are deleterious, just as most mutations are, some might be adaptive under some conditions. Below we describe the current understanding of the role that TEs play in molecular adaptation.

5.2.1. Inference of Ancient TE Adaptations

TEs have been co-opted to play key organismic functions, such as the generation of antibody diversity in the vertebrate immune system (19) and maintenance of telomeres in Drosophila (134) and centromeres in plants (135, 136). It is also quite likely that epigenetic mechanisms, such as gene silencing through methylation or RNAi epigenetic mechanisms evolved as a means of genomic defense against TEs and later were used to regulate host genes.

TEs play a role in the generation of new coding sequences either by being domesticated as components of host transcripts (8, 16–18, 137–139) or by inducing duplication of host genes (14, 140). TEs can affect gene expression in several ways and some of these changes might be adaptive. For example, human L1 elements contain sense and antisense promoters in their 5′ UTR, which have occasionally been recruited as regulators of the transcription of host genes (141, 142).

Genome-wide assessment revealed that hundreds of TEs have been co-opted into regulatory regions of mammalian genes (143, 144). The authors compared nonexonic sequences conserved in diverse placental mammals and reported that a large number of them originated from repetitive elements undergoing strong purifying selection in mammals. They found that exapted TEs are extremely enriched for clustering near (within 1 Mb) developmental genes (e.g., *P-value* $= 8 \times 10^{-24}$ and 6×10^{-19} for GO terms "development" and "transcription regulator activity," respectively) and argue that this enrichment is not due to insertional bias but rather to a bias in retention, suggesting that TEs contributed substantially to regulatory elements (144). TEs also participate in the evolution of new and rewiring of old regulatory networks as first proposed decades ago by Britten and Davidson (9, 11). For example, human endogenous retroviruses (ERVs) have actively shaped the p53 transcriptional network in a species-specific manner (12). LTRs of these elements are in vivo-binding sites for p53 and account for more than 30% of the total number of p53-binding sites in the human genome. The authors show that the expressions of many genes located nearby these LTRs are regulated by p53, suggesting that ERVs have been exapted as regulatory sequences to expand the p53 network (12).

5.2.2. Ongoing TE-Induced Adaptation

We have recently performed a genome-wide screen for recent TE-induced adaptations in *D. melanogaster* (23, 26). *D. melanogaster* is originally from Africa and has only recently colonized the rest of the world (145, 146). This range expansion must have been accompanied by numerous adaptations to new habitats (129, 147, 148), which should still be detectable as selective sweeps (149). We focused specifically on identifying TE insertions that might have been adaptive to the out-of-Africa environments by looking for TE insertions that are present at low frequencies or absent in African populations and are present at high frequencies in North America.

The fact that long TEs in high copy number families are subject to purifying selection at the level of ectopic recombination and thus independently of the exact site of insertion makes our search for adaptive TEs much easier. Indeed, such long TEs in copious families should be rare in populations unless they cause an adaptive effect. They cannot simply drift to high frequencies. On the other hand, TEs that belong to families that are subject to relaxed purifying selection as a whole can drift to high frequencies and can serve as putatively neutral frequent TEs (150).

For a subset of the identified TE insertions, we showed that putatively adaptive but not putatively neutral TEs are associated with signatures of selective sweeps and with population differentiation patterns, suggesting that most if not all of the identified putatively adaptive TEs are indeed adaptive and play a significant role in adaptation to temperate environments (23, 26). Although genetic draft might play a role in the population dynamics of TEs, the detailed analyses of the regions surrounding several adaptive TEs suggested that the TE insertion was the causative mutation in each case (23). Preliminary analyses also showed that some of these TEs affect the expression of their nearby genes (23), and, for one of them, we further showed that the changes in gene expression are consistent with phenotypic differences in life-history traits (24). Overall, we showed that TEs are a considerable source of recently adaptive mutations in Drosophila—we estimate that as many as 50–100 recent adaptive mutations in *D. melanogaster* that are responsible for adaptation to out-of-Africa environments have been caused by TEs.

6. Rate of TE Loss

As we mentioned in Subheading 5.1, TEs reach fixation in the human populations at substantially higher rates than in Drosophila. In addition, the fixed TEs in humans remain in the genome for much longer periods of time, due to a much lower rate of DNA loss in humans compared to Drosophila. Indeed, deletions are more prevalent and almost eight times longer in *Drosophila* than in mammals, resulting in orders of magnitude faster rate of DNA elimination from Drosophila genomes (151). The average time to loss of 50% of nonfunctional DNA due to the preponderance of small deletions over small insertions is 14.3 million years in Drosophila compared with over 884 million years in mammals (132, 151).

In addition to the elimination of TEs by random deletion, TEs often suffer deletions between repetitive sequences that they contain. For instance, recombination between long terminal repeats (LTRs) in retrotransposable elements can reduce a ~10 kb TE

insertion to a solo LTR of a few hundred base-pairs. Additionally, recombination between the target site duplications flanking TE insertions sometimes results in the precise removal of the TE insertion (152). These deletion mechanisms vary as a function of meiotic recombination (153). Therefore, variation in recombination rates between organisms could lead to variation in deletion frequencies.

This high rate of DNA elimination might explain the absence of old TEs in the Drosophila genome, compared to human genomes, where even ancient TE insertions inserted 300 million years ago are still detectable (154). In general, variation in the rate of DNA loss in addition to the variation in the strength of purifying selection against TEs must be a key reason why some genomes accumulate TEs while others do not.

7. Conclusion

The past 40 years have revealed a staggering diversity of genomic architecture in eukaryotes, particularly in terms of genome size and the amounts and types of repetitive DNA. Much of that diversity is driven by the activity of TEs and it has become very clear that in order to understand how genomes evolve, how they are structured and how they function we will need to elucidate the evolutionary dynamics governing the activity and the impact of TEs.

In this chapter, we focused on two well-studied genomes with strikingly different architectures, humans, and Drosophila. The human genome contains millions of copies of TEs. These TEs are primarily fixed in the population, most are extremely old (some inserted prior to the split of mammals from reptiles), and active elements belong to only two families, L1 and Alu, both of which rely on reverse transcription for transposition. In contrast, there are only a few thousand TE copies in the Drosophila genome, the majority of these TEs are very young and in fact polymorphic in the population, and they belong to ~100 families from all major orders of TEs.

We have argued in this chapter that although we still do not understand the reasons for these sharply contrasting patterns of TE diversity, some of the answers are starting to emerge. It appears likely that some of the answers lie in the differences in the population sizes between humans and flies, leading to much more effective selection against new TE insertions due to ectopic recombination among TE copies within the same family. This is one reason for why TEs tend to reach fixation at substantially higher rates in humans than in Drosophila. This might also be a contributing reason for the evolution of high diversity of TE families in Drosophila—given the high strength of selection against TEs within a family there is a strong selective advantage for active TEs that generate TE copies

that do not recombine efficiently with TEs that are already present in the genome. Another reason for the differences might lie in the much slower rate of DNA loss in humans than in flies that allow TE sequences to persist for hundreds of millions of years in mammalian genomes and at most for a few million years in Drosophila.

We also described cursorily many other components of the TE lifestyles that need to be taken into account—variation in the rate of transposition, mechanisms of TE regulation, and the tendency of TEs to cause adaptations. Although we have made great strides toward an understanding of TE biology and genomic impact, we still know very little. The hope is that the new genomic data coming at ever increasing rates will allow us not only to observe TEs in various genomes (and we certainly will), but also to make more sense of their lifestyles and evolutionary dynamics.

8. Questions

1. Why does the fact that ectopic recombination takes place preferentially between heterozygous copies help allow accumulation of TEs in the human genome?
2. How differences in the rate of DNA loss can affect the evolutionary dynamics of TEs?

Acknowledgments

We thank Anna-Sophie Fiston-Lavier and all members of the Petrov lab for helpful discussions, Roberto Torres for figure design (www.torresdecomunicacion.org), and the three anonymous reviewers for comments on the manuscript. This work was supported by a Ramon y Cajal grant from the Spanish Ministry of Science and Innovation (MICINN: RYC-2010-07306) to J.G. and by a grant from the NIH (GM 089926) to D.A.P.

References

1. Wicker T, Sabot F, Hua-Van A, et al. (2007) A unified classification system for eukaryotic transposable elements, Nat Rev Genet **8**: 973–982.
2. Kapitonov VV, Jurka J. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase, Nat Rev Genet **9**: 411–412; author reply 414.
3. Cuomo C A, Guldener U, Xu J R, et al. (2007) The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization, Science **317**: 1400–1402.
4. Schnable P S, Ware D, Fulton R S, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics, Science **326**: 1112–1115.

5. Lander E S, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome, Nature **409**: 860–921.
6. Ashburner M., Golic, K.G., Hawley, R.S. (2005) *Drosophila: a laboratory handbook*, Cold Spring Harbour Laboratoy Press, New York.
7. Kidwell M G, Lisch D R. (2000) Transposable elements and host genome evolution, Trends Ecol Evol **15**: 95–99.
8. Feschotte C, Pritham E J. (2007) DNA transposons and the evolution of eukaryotic genomes, Annu Rev Genet **41**: 331–368.
9. Feschotte C. (2008) Transposable elements and the evolution of regulatory networks, Nat Rev Genet **9**: 397–405.
10. Lippman Z, Gendrel A V, Black M, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control, Nature **430**: 471–476.
11. Britten R J, Davidson E H. (1969) Gene regulation for higher cells: a theory, Science **165**: 349–357.
12. Wang T, Zeng J, Lowe C B, et al. (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53, Proc Natl Acad Sci U S A **104**: 18613–18618.
13. Bringaud F, Muller M, Cerqueira G C, et al. (2007) Members of a large retroposon family are determinants of post-transcriptional gene expression in Leishmania, PLoS Pathog **3**: 1291–1307.
14. Moran J V, DeBerardinis R J, Kazazian H H, Jr. (1999) Exon shuffling by L1 retrotransposition, Science **283**: 1530–1534.
15. Goodier J L, Kazazian H H, Jr. (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites, Cell **135**: 23–35.
16. Makalowski W, Mitchell G A, Labuda D. (1994) Alu sequences in the coding regions of mRNA: a source of protein variability, Trends Genet **10**: 188–193.
17. Gotea V, Makalowski W. (2006) Do transposable elements really contribute to proteomes? Trends Genet **22**: 260–267.
18. Wu M, Li L, Sun Z. (2007) Transposable element fragments in protein-coding regions and their contributions to human functional proteins, Gene **401**: 165–171.
19. Agrawal A, Eastman Q M, Schatz D G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system, Nature **394**: 744–751.
20. Pardue M L, DeBaryshe P G. (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres, Annu Rev Genet **37**: 485–511.
21. Daborn P J, Yen J L, Bogwitz M R, et al. (2002) A single p450 allele associated with insecticide resistance in Drosophila, Science **297**: 2253–2256.
22. Aminetzach Y T, Macpherson J M, Petrov D A. (2005) Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila, Science **309**: 764–767.
23. González J, Lenkov K, Lipatov M, et al. (2008) High rate of recent transposable element-induced adaptation in Drosophila melanogaster, PLoS Biol **6**: e251.
24. González J, Macpherson J M, Petrov D A. (2009) A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila melanogaster, Mol Biol Evol: 1949–1961.
25. González J, Petrov D A. (2009) The adaptive role of transposable elements in the Drosophila genome, Gene **448**: 124–133.
26. González J, Karasov T L, Messer P W, et al. (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila, PLoS Genet **6**: e1000905.
27. Biemont C, Vieira C. (2006) Genetics: junk DNA as an evolutionary force, Nature **443**: 521–524.
28. Kazazian H H, Jr. (1998) Mobile elements and disease, Curr Opin Genet Dev **8**: 343–350.
29. Callinan P A, Batzer M A (2006) *Retrotransposable elements and human disease*, Vol. 1, Karger, Basel.
30. Quesneville H, Bergman C M, Andrieu O, et al. (2005) Combined evidence annotation of transposable elements in genome sequences, PLoS Comput Biol **1**: 166–175.
31. Kaminker J S, Bergman C M, Kronmiller B, et al. (2002) The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective, Genome Biol **3**: RESEARCH0084.
32. Kapitonov V V, Jurka J. (2003) Molecular paleontology of transposable elements in the Drosophila melanogaster genome, Proc Natl Acad Sci U S A **100**: 6569–6574.
33. Singh N D, Petrov D A. (2004) Rapid sequence turnover at an intergenic locus in Drosophila, Mol Biol Evol **21**: 670–680.
34. Petrov D A, Fiston-Lavier A S, Lipatov M, et al. (2011) Population genomics of

transposable elements in Drosophila melanogaster, Mol Biol Evol **28**: 1633–1644.

35. Loreto E L, Carareto C M, Capy P. (2008) Revisiting horizontal transfer of transposable elements in Drosophila, Heredity **100**: 545–554.
36. Bartolome C, Bello X, Maside X. (2009) Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes, Genome Biol **10**: R22.
37. Schaack S, Gilbert C, Feschotte C. (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution, Trends Ecol Evol **25**: 537–546.
38. Charlesworth B, Charlesworth D. (1983) The population dynamics of transposable elements, Genetical Research **42**: 1–27.
39. Brookfield J F. (1991) Models of repression of transposition in P-M hybrid dysgenesis by P cytotype and by zygotically encoded repressor proteins, Genetics **128**: 471–486.
40. Wright S I, Schoen D J. (1999) Transposon dynamics and the breeding system, Genetica **107**: 139–148.
41. Morgan M T. (2001) Transposable element number in mixed mating populations, Genet Res **77**: 261–275.
42. Le Rouzic A, Deceliere G. (2005) Models of the population genetics of transposable elements, Genet Res **85**: 171–181.
43. Le Rouzic A, Boutin T S, Capy P. (2007) Long-term evolution of transposable elements, Proc Natl Acad Sci USA **104**: 19375–19380.
44. Lu J, Clark A G. (2010) Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila, Genome Res **20**: 212–227.
45. Strobel E, Dunsmuir P, Rubin G M. (1979) Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila, Cell **17**: 429–439.
46. Nuzhdin S V. (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number, Genetica **107**: 129–137.
47. Van den Broeck D, Maes T, Sauer M, et al. (1998) Transposon Display identifies individual transposable elements in high copy number lines, Plant J **13**: 121–129.
48. De Keukeleire P, Maes T, Sauer M, et al. (2001) Analysis by Transposon Display of the behavior of the dTph1 element family during ontogeny and inbreeding of Petunia hybrida, Mol Genet Genomics **265**: 72–81.
49. Behura S K. (2006) Molecular marker systems in insects: current trends and future avenues, Mol Ecol **15**: 3087–3113.
50. Badge R M, Alisch R S, Moran J V. (2003) ATLAS: a system to selectively identify human-specific L1 insertions, Am J Hum Genet **72**: 823–838.
51. Petrov D A, Aminetzach Y T, Davis J C, et al. (2003) Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila, Mol Biol Evol **20**: 880–892.
52. Carroll M L, Roy-Engel A M, Nguyen S V, et al. (2001) Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity, J Mol Biol **311**: 17–40.
53. Myers J S, Vincent B J, Udall H, et al. (2002) A comprehensive analysis of recently integrated human Ta L1 elements, Am J Hum Genet **71**: 312–326.
54. McCollum A M, Ganko E W, Barrass P A, et al. (2002) Evidence for the adaptive significance of an LTR retrotransposon sequence in a Drosophila heterochromatic gene, BMC Evol Biol **2**: 5.
55. Franchini L F, Ganko E W, McDonald J F. (2004) Retrotransposon-gene associations are widespread among D. melanogaster populations, Mol Biol Evol **21**: 1323–1331.
56. Hollister J D, Gaut B S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression, Genome Res **19**: 1419–1428.
57. Hormozdiari F, Hajirasouliha I, Dao P, et al. (2010) Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery, Bioinformatics **26**: i350–357.
58. Quinlan A R, Clark R A, Sokolova S, et al. (2010) Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome, Genome Res **20**: 623–635.
59. Fiston-Lavier A S, Carrigan M, Petrov D A, et al. (2011) T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data, Nucleic Acids Res **39**: e36.
60. Eggleston W B, Johnson-Schlitz D M, Engels W R. (1988) P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster, Nature **331**: 368–370.
61. Harada K, Yukuhiro K, Mukai T. (1990) Transposition rates of movable genetic elements in Drosophila melanogaster, Proc Natl Acad Sci U S A **87**: 3248–3252.

62. Nuzhdin S V, Mackay T F. (1995) The genomic rate of transposable element movement in Drosophila melanogaster, Mol Biol Evol **12**: 180–181.
63. Charlesworth B, Langley C H. (1989) The population genetics of Drosophila transposable elements, Annu Rev Genet **23**: 251–287.
64. Maside X, Bartolome C, Assimacopoulos S, et al. (2001) Rates of movement and distribution of transposable elements in Drosophila melanogaster: in situ hybridization vs Southern blotting data, Genet Res **78**: 121–136.
65. Maside X, Bartolome C, Charlesworth B. (2002) S-element insertions are associated with the evolution of the Hsp70 genes in Drosophila melanogaster, Curr Biol **12**: 1686–1691.
66. Kidwell M G, Kidwell J F, Sved J A. (1977) Hybrid dysgenesis in Drosophila melanogaster: A syndrome of aberrant traits including mutation, sterility and male recombination, Genetics **86**: 813–833.
67. Bingham P M, Kidwell M G, Rubin G M. (1982) The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family, Cell **29**: 995–1004.
68. Rubin G M, Kidwell M G, Bingham P M. (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations, Cell **29**: 987–994.
69. Petrov D A, Schutzman J L, Hartl D L, et al. (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis, Proc Natl Acad Sci U S A **92**: 8050–8054.
70. Vasilyeva L A, Bubenshchikova E V, Ratner V A. (1999) Heavy heat shock induced retrotransposon transposition in Drosophila, Genet Res **74**: 111–119.
71. Seleme M C, Busseau I, Malinsky S, et al. (1999) High-frequency retrotransposition of a marked I factor in Drosophila melanogaster correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes, Genetics **151**: 761–771.
72. Bucheton A, Paro R, Sang H M, et al. (1984) The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor, Cell **38**: 153–163.
73. Lozovskaya E R, Scheinker V S, Evgen'ev M B. (1990) A hybrid dysgenesis syndrome in Drosophila virilis, Genetics **126**: 619–623.
74. Prud'homme N, Gans M, Masson M, et al. (1995) Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster, Genetics **139**: 697–711.
75. Moran J V, Holmes S E, Naas T P, et al. (1996) High frequency retrotransposition in cultured mammalian cells, Cell **87**: 917–927.
76. Brouha B, Schustak J, Badge R M, et al. (2003) Hot L1s account for the bulk of retrotransposition in the human population, Proc Natl Acad Sci U S A **100**: 5280–5285.
77. Kazazian H H, Jr. (1999) An estimated frequency of endogenous insertional mutations in humans, Nat Genet **22**: 130.
78. Li X, Scaringe W A, Hill K A, et al. (2001) Frequency of recent retrotransposition events in the human factor IX gene, Hum Mutat **17**: 511–519.
79. Deininger P L, Batzer M A. (1993) Evolution of retroposons, In *Evolutionary Biology* (Hecht, M K, Ed.), Plenum Press, New York.
80. Ewing A D, Kazazian H H, Jr. (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes, Genome Res **20**: 1262–1270.
81. Huang C R, Schneider A M, Lu Y, et al. (2010) Mobile interspersed repeats are major structural variants in the human genome, Cell **141**: 1171–1182.
82. Messer P W. (2009) Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data, Genetics **182**: 1219–1232.
83. Laski F A, Rio D C, Rubin G M. (1986) Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing, Cell **44**: 7–19.
84. Andrews J D, Gloor G B. (1995) A role for the KP leucine zipper in regulating P element transposition in Drosophila melanogaster, Genetics **141**: 587–594.
85. Simmons M J, Bucholz L M. (1985) Transposase titration in Drosophila melanogaster: a model of cytotype in the P-M system of hybrid dysgenesis, Proc Natl Acad Sci U S A **82**: 8119–8123.
86. Hartl D L, Lohe A R, Lozovskaya E R. (1997) Regulation of the transposable element mariner, Genetica **100**: 177–184.
87. Lohe A R, Hartl D L. (1996) Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation, Mol Biol Evol **13**: 549–555.
88. Yoder J A, Walsh C P, Bestor T H. (1997) Cytosine methylation and the ecology of intragenomic parasites, Trends Genet **13**: 335–340.

89. Lyko F, Ramsahoye B H, Jaenisch R. (2000) DNA methylation in Drosophila melanogaster, Nature **408**: 538–540.
90. Lyko F. (2001) DNA methylation learns to fly, Trends Genet **17**: 169–172.
91. Kunert N, Marhold J, Stanke J, et al. (2003) A Dnmt2-like protein mediates DNA methylation in Drosophila, Development **130**: 5083–5090.
92. Mandrioli M, Borsatti F. (2006) DNA methylation of fly genes and transposons, Cell Mol Life Sci **63**: 1933–1936.
93. Mangeat B, Turelli P, Caron G, et al. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts, Nature **424**: 99–103.
94. Zhang H, Yang B, Pomerantz R J, et al. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA, Nature **424**: 94–98.
95. Schumann G G. (2007) APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition, Biochem Soc Trans **35**: 637–642.
96. Han J S. (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions, Mob DNA **1**: 15.
97. Ikeda T, Abd El Galil K H, Tokunaga K, et al. (2011) Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons, Nucleic Acids Res.
98. Gendrel A V, Lippman Z, Yordan C, et al. (2002) Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1, Science **297**: 1871–1873.
99. Kondo Y, Issa J P. (2003) Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells, J Biol Chem **278**: 27658–27662.
100. Martens J H, O'Sullivan R J, Braunschweig U, et al. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome, EMBO J **24**: 800–812.
101. Huda A, Marino-Ramirez L, Jordan I K. (2010) Epigenetic histone modifications of human transposable elements: genome defense versus exaptation, Mob DNA **1**: 2.
102. Vieira C, Fablet M, Lerat E. (2006) Intra- and transspecific clues to understanding the dynamics of transposable elements, In *Genome Dynamics and Stability* (Lankenau, D H, Volff, J N, Eds.), Springer-Verlag, Berlin.
103. Soifer H S, Zaragoza A, Peyvan M, et al. (2005) A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon, Nucleic Acids Res **33**: 846–856.
104. Yang N, Kazazian H H, Jr. (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells, Nat Struct Mol Biol **13**: 763–771.
105. Malone C D, Hannon G J. (2009) Molecular Evolution of piRNA and Transposon Control Pathways in Drosophila, Cold Spring Harb Symp Quant Biol: 225–234.
106. Blumenstiel J P. (2011) Evolutionary dynamics of transposable elements in a small RNA world, Trends Genet **27**: 23–31.
107. Tabara H, Sarkissian M, Kelly W G, et al. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans, Cell **99**: 123–132.
108. Ketting R F, Haverkamp T H, van Luenen H G, et al. (1999) Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD, Cell **99**: 133–141.
109. Blumenstiel J P, Hartl D L. (2005) Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis, Proc Natl Acad Sci U S A **102**: 15965–15970.
110. Brennecke J, Malone C D, Aravin A A, et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing, Science **322**: 1387–1392.
111. Charlesworth B, Sniegowski P, Stephan W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes, Nature **371**: 215–220.
112. Finnegan D J. (1992) Transposable elements, Curr Opin Genet Dev **2**: 861–867.
113. McDonald J F, Matyunina L V, Wilson S, et al. (1997) LTR retrotransposons and the evolution of eukaryotic enhancers, Genetica **100**: 3–13.
114. Montgomery E, Charlesworth B, Langley C H. (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.
115. Boissinot S, Davis J, Entezam A, et al. (2006) Fitness cost of LINE-1 (L1) activity in humans, Proc Natl Acad Sci U S A **103**: 9590–9594.
116. Langley C H, Montgomery E, Hudson R, et al. (1988) On the role of unequal exchange in the containment of transposable element copy number, Genet Res **52**: 223–235.

117. Montgomery E A, Huang S M, Langley C H, et al. (1991) Chromosome rearrangement by ectopic recombination in Drosophila melanogaster: genome structure and evolution, Genetics **129**: 1085–1098.
118. 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**: 103–114.
119. Charlesworth B, Lapid A, Canada D. (1992) 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 **60**: 115–130.
120. 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**: 926–937.
121. Hoogland C, Biemont C. (1996) Chromosomal distribution of transposable elements in Drosophila melanogaster: test of the ectopic recombination model for maintenance of insertion site number, Genetics **144**: 197–204.
122. Hill W G, Robertson A. (1966) The effect of linkage on limits to artificial selection, Genet Res **8**: 269–294.
123. Dolgin E S, Charlesworth B. (2008) The effects of recombination rate on the distribution and abundance of transposable elements, Genetics **178**: 2169–2177.
124. Suh D S, Choi E H, Yamazaki T, et al. (1995) Studies on the transposition rates of mobile genetic elements in a natural population of Drosophila melanogaster, Mol Biol Evol **12**: 748–758.
125. Vieira C, Biemont C. (1997) Transposition rate of the 412 retrotransposable element is independent of copy number in natural populations of Drosophila simulans, Mol Biol Evol **14**: 185–188.
126. Pasyukova E G, Nuzhdin S V, Filatov D A. (1998) The relationship between the rate of transposition and transposable element copy number for copia and Doc retrotransposons of Drosophila melanogaster, Genet Res **72**: 1–11.
127. Maside X, Assimacopoulos S, Charlesworth B. (2000) Rates of movement of transposable elements on the second chromosome of Drosophila melanogaster, Genet Res **75**: 275–284.
128. Kreitman M. (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster, Nature **304**: 412–417.
129. Li H, Stephan W. (2006) Inferring the demographic history and rate of adaptive substitution in Drosophila, PLoS Genet **2**: e166.
130. Thornton K, Andolfatto P. (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster, Genetics **172**: 1607–1619.
131. Takahata N. (1993) Allelic genealogy and human evolution, Mol Biol Evol **10**: 2–22.
132. Petrov D A, Lozovskaya E R, Hartl D L. (1996) High intrinsic rate of DNA loss in Drosophila, Nature **384**: 346–349.
133. Petrov D A, Sangster T A, Johnston J S, et al. (2000) Evidence for DNA loss as a determinant of genome size, Science **287**: 1060–1062.
134. Levis R W, Ganesan R, Houtchens K, et al. (1993) Transposons in place of telomeric repeats at a Drosophila telomere, Cell **75**: 1083–1093.
135. Ma J, Bennetzen J L. (2006) Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice, Proc Natl Acad Sci U S A **103**: 383–388.
136. Weber B, Schmidt T. (2009) Nested Ty3-gypsy retrotransposons of a single Beta procumbens centromere contain a putative chromodomain, Chromosome Res **17**: 379–396.
137. Cordaux R, Udit S, Batzer M A, et al. (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element, Proc Natl Acad Sci U S A **103**: 8101–8106.
138. Wang W, Zheng H, Fan C, et al. (2006) High rate of chimeric gene origination by retroposition in plant genomes, Plant Cell **18**: 1791–1802.
139. Baudry C, Malinsky S, Restituito M, et al. (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia, Genes Dev **23**: 2478–2483.
140. Esnault C, Maestre J, Heidmann T. (2000) Human LINE retrotransposons generate processed pseudogenes, Nat Genet **24**: 363–367.
141. Speek M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes, Mol Cell Biol **21**: 1973–1985.
142. Nigumann P, Redik K, Matlik K, et al. (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon, Genomics **79**: 628–634.

143. Bejerano G, Lowe C B, Ahituv N, et al. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon, Nature **441**: 87–90.

144. Lowe C B, Bejerano G, Haussler D. (2007) Thousands of human mobile element fragments undergo strong purifying selection near developmental genes, Proc Natl Acad Sci U S A **104**: 8005–8010.

145. David J R, Capy P. (1988) Genetic variation of Drosophila melanogaster natural populations, Trends Genet **4**: 106–111.

146. Lachaise D, Cariou, M-L, David, J R, Lemeunier, F, Tsacas F, et al (1988) Historical biogeography of the Drosophila melanogaster species subgroup, Evol Biol **22**: 159–225.

147. Glinka S, Ometto L, Mousset S, et al. (2003) Demography and natural selection have shaped genetic variation in Drosophila melanogaster: a multi-locus approach, Genetics **165**: 1269–1278.

148. Orengo D J, Aguade M. (2004) Detecting the footprint of positive selection in a european population of Drosophila melanogaster: multilocus pattern of variation and distance to coding regions, Genetics **167**: 1759–1766.

149. Przeworski M. (2002) The signature of positive selection at randomly chosen loci, Genetics **160**: 1179–1189.

150. González J, Macpherson J M, Messer P W, et al. (2009) Inferring the strength of selection in Drosophila under complex demographic models, Mol Biol Evol **26**: 513–526.

151. Petrov D A, Hartl D L. (1998) High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups, Mol Biol Evol **15**: 293–302.

152. van de Lagemaat L N, Gagnier L, Medstrand P, et al. (2005) Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates, Genome Res **15**: 1243–1249.

153. Katzourakis A, Pereira V, Tristem M. (2007) Effects of recombination rate on human endogenous retrovirus fixation and persistence, J Virol **81**: 10712–10717.

154. Arndt P F, Petrov D A, Hwa T. (2003) Distinct changes of genomic biases in nucleotide substitution at the time of Mammalian radiation, Mol Biol Evol **20**: 1887–1896.