Genome-Wide Estimates of Transposable Element Insertion and Deletion Rates in *Drosophila Melanogaster*

Jeffrey R. Adrion[1,†], Michael J. Song[2,†], Daniel R. Schrider[3], Matthew W. Hahn[1,4], and Sarah Schaack[5,*]

[1]Department of Biology, Indiana University, Bloomington, IN

[2]Department of Integrative Biology, University of California, Berkeley, CA

[3]Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ

[4]School of Informatics and Computing, Indiana University, Bloomington, IN

[5]Department of Biology, Reed College, Portland, OR

†These authors contributed equally to this work.
*Corresponding author: E-mail: schaack@reed.edu.

Abstract

Knowing the rate at which transposable elements (TEs) insert and delete is critical for understanding their role in genome evolution. We estimated spontaneous rates of insertion and deletion for all known, active TE superfamilies present in a set of *Drosophila melanogaster* mutation-accumulation (MA) lines using whole genome sequence data. Our results demonstrate that TE insertions far outpace TE deletions in *D. melanogaster*. We found a significant effect of background genotype on TE activity, with higher rates of insertions in one MA line. We also found significant rate heterogeneity between the chromosomes, with both insertion and deletion rates elevated on the X relative to the autosomes. Further, we identified significant associations between TE activity and chromatin state, and tested for associations between TE activity and other features of the local genomic environment such as TE content, exon content, GC content, and recombination rate. Our results provide the most detailed assessment of TE mobility in any organism to date, and provide a useful benchmark for both addressing theoretical predictions of TE dynamics and for exploring large-scale patterns of TE movement in *D. melanogaster* and other species.

Key words: transposable elements, *Drosophila melanogaster*, transposition, insertion rate, deletion rate.

Introduction

Transposable elements (TEs) make up a significant portion of most multicellular eukaryotic genomes and can profoundly influence their evolution (Burns and Boeke 2012). Often considered genomic parasites, these discrete DNA sequences are capable of moving and replicating throughout the genome and have been found to comprise ~20% of the *Drosophila melanogaster* genome, and ~65 and ~85% of the human and maize genomes, respectively (Quesneville et al. 2005; Schnable et al. 2009; de Koning et al. 2011). TE abundance is highly variable among taxa, as is the spatial distribution and differential proliferation of TE types within species (Kidwell 2002; Feschotte and Pritham 2007). While there are numerous examples of beneficial TE insertions (reviewed in Casacuberta and González 2013), transposition events are expected to be deleterious on average (Pasyukova et al. 2004; Casacuberta and González 2013). Moreover, selection against the deleterious effects of TEs is expected to shape both the rates of TE activity (Charlesworth and Charlesworth 1983; Charlesworth and Langley 1989) and the spatial distribution of TEs along and among chromosomes (Duret et al. 2000; Bartolomé et al. 2002; Rizzon et al. 2002). However, features of the host genome and of the transposition process itself may contribute to the observed variation in TE abundance, diversity, and distributions. Natural selection has the potential to obscure these patterns, which may be harder to detect in natural populations. Thus, knowing the rates and distribution of TEs in the absence of selection is a critical component for understanding their role in genome evolution.

The earliest and most numerous studies on TE movement in metazoans have been performed in *D. melanogaster* (e.g. Engels 1983; Lewis and Brookfield 1987). The results of these kinds of landmark studies provided the data and insights to form a theoretical framework within which many subsequent studies investigating TE dynamics in other systems have been interpreted (e.g. Charlesworth and Langley 1989; Lee and Langley 2010). Rates of TE movement have been estimated empirically in both natural populations and in laboratory experiments, in many cases taking advantage of polytene chromosomes to perform *in situ* hybridization (see supplementary table S1, Supplementary Material online). However, none of these prior studies were able to examine the movement of all TEs in the genome simultaneously and instead relied on data from one or a few families to generalize patterns across TE families, despite major differences in transposition mechanisms (Rebollo et al. 2012). While useful, simply estimating the absolute rates of insertion or deletion for individual TE families is only a first step towards investigating the long-term dynamics of TEs in the genome. Instead, considering the relative rates of gains and losses genome-wide, as well as the spatial distribution of these events along the chromosomes, allows one to understand the global dynamics of TE movement.

Although selection against the deleterious effects of TEs undoubtedly contributes to the rate variation and spatial distribution of TEs along and among the chromosomes, non-uniform mutation is often overlooked as an explanation for much of this variation in nature. Indeed, TE insertion-site preference could in part shape this distribution, and such preferences are quite common across eukaryotes. For instance, Vazquez and et al. (2007) found that *roo* elements preferentially integrate into proximal and distal regions of autosomal arms and the X chromosome in *D. melanogaster*. *P* elements in *D. melanogaster* have also been shown to preferentially insert into specific sequences acting as origins of replication (Spradling et al. 2011). Most new *Ty5* insertions (~95%) in *Saccharomyces cerevisiae* occur either in heterochromatin at the telomeres or in the silent mating cassettes (Bushman 2003). TE insertion-site preference has also been described in many other organisms, including *D. willistoni* (Gonçalves et al. 2014), *Daphnia pulex* (Elliott et al. 2013), *Schizosaccharomyces pombe* (Singleton and Levin 2002), *Oryza* (Miyao et al. 2003), and mouse and human cell cultures (Yant et al. 2005). Transposition bias may contribute to the non-random spatial distribution of TEs in *D. melanogaster*, but TE insertion and deletion rates have yet to be investigated on a genome-wide scale in the absence of selection.

In order to estimate genome-wide rates and patterns of TE movement in the absence of natural selection, we took a whole-genome sequencing mutation-accumulation (MA) approach. We estimate the insertion and deletion rate for all known active TE superfamilies based on whole-genome sequence data from a set of eight *D. melanogaster* MA lines

derived from two different inbred founder genotypes. We use the term "line" when referring to either of the two founder genotypes, and "subline" when referring to the unit of replication within each founder line (i.e. there are four sublines within each line). We present both per-site and per-copy rate estimates for all superfamilies where either insertion or deletion events were detected. We examine the spatial distribution of new insertions and deletions, and test for associations between transposition activity and characteristics of individual TE superfamilies and of the local genomic environment. To our knowledge, this genome-wide analysis of TE mobility provides the most detailed assessment in any organism to date, and provides both a useful benchmark for addressing theoretical predictions of TE dynamics and for exploring large-scale patterns of TE movement in *D. melanogaster* and other species.

Materials and Methods

Mutation-Accumulation Lines

Two inbred lines (Line 33 and Line 39) originating from the IV laboratory population of flies captured in Massachusetts in 1975 (described in Houle and Rowe 2003) were used to establish eight sublines. Once founded, each subline was subjected to 145–149 generations of mutation accumulation (sublines are referred to by number: 33–45, 33–27, 33–55, 33–5 and 39–58, 39–67, 39–51, and 39–18, respectively). During mutation-accumulation, a single pair of flies is used to found each new generation. This reduces the efficacy of natural selection relative to the strength of genetic drift, and allows for deleterious mutations that may have otherwise been purged by selection to accumulate in each line. DNA was extracted from whole flies collected from each of these sublines after mutation-accumulation, and was multiplexed and sequenced using an Illumina Genome Analyzer II at the Indiana University Center for Genomics and Bioinformatics (see Schrider et al. 2013 for additional details). We obtained paired-end Illumina reads with 74 bp ends and an average insert size of 182 bp (see supplementary table S2, Supplementary Material online). We used cutadapt (Martin 2011) to trim adapters and low quality bases from both 5′ and 3′ ends until the minimum aggregate quality score was ≥ 20. Finally, we randomly subsampled reads to ensure that all eight sublines started with an equal number of paired-end reads prior to mapping (see supplementary table S2, Supplementary Material online).

TE Discovery

We used our custom TE identification program, TEFLoN (https://github.com/jradrion/TEFLoN), to discover the position and superfamily identity of all TEs present in the eight sublines. Briefly, TEFLoN creates a pseudo-reference genome with all known (i.e. reference annotated) TE sequences removed. It uses BWA-mem v.0.7.10 (Li and Durbin 2009) to align

paired-end reads to a database of full- and partial-length TEs annotated in the reference in addition to aligning reads to unique genomic locations in the pseudo-reference. TEFLoN characterizes the breakpoints and superfamily identity of both new and known elements by identifying paired-end reads where one end maps to a TE and the other end maps uniquely to the pseudo-reference (map quality ≥ 30). Next, TEFLoN catalogs all reads at the putative breakpoints as either "presence" reads (where the read is either soft-clipped at a breakpoint or has a mate aligning to a TE), "absence" reads (where the alignment spans the breakpoints), and uninformative reads (the aligned read does not satisfy either of the previous statements) and tallies these categories. We excluded the family *ine-1* from our analysis, as evidence suggests this family has been inactive for millions of years (Kapitonov and Jurka 2003). Much like other programs for identifying TEs using short-read data, TEFLoN is unable to discover or quantify nested TEs—those TEs located entirely within other TE sequence—making our estimates of starting copy-number, along with counts of insertions and deletions, a lower bound.

Estimating Rates of TE Activity

New insertions in a focal subline were scored if they satisfied three criteria: 1) ≥ 3 presence reads in the focal subline, 2) ≥ 3 absence reads and ≤ 1 presence reads in the three non-focal sublines, and 3) a ratio of presence reads to total reads in the focal subline of $\geq 70\%$. Likewise, three criteria were used to score deletions thought to have occurred during the experiment: 1) ≥ 5 absence reads and ≤ 1 presence read in the focal line, 2) ≥ 5 presence reads in the three non-focal lines, and 3) a ratio of presence reads to total reads in the three non-focal sublines of $\geq 70\%$. The asymmetry in read thresholds between insertions and deletions was discovered via simulation (described below). Allowing a single presence read when classifying an element as being absent corrected for small errors in the TE breakpoint estimation, especially in the case of a partial target site duplication (TSD) sharing sequence identity with the 5' or 3' end of a sequence in the TE database. We classified elements with ≥ 3 presence reads and a ratio of presence reads to total reads that was $\geq 70\%$ in all four sublines as being present in the ancestor of the four sublines (i.e. starting copies). Finally, elements detected as being present in two sublines and absent in the other two sublines were excluded from our analyses. Our filtering methods—which require a ratio of $\geq 70\%$ presence to absence reads—were used to filter any somatic TE mutations that may have occurred during the final generation of mutation accumulation, as somatic TE mutations should be present on fewer chromosome copies than germline TE mutations. Insertions and deletions were visually inspected and validated using the Integrative Genome Browser (Robinson et al. 2011).

Genome-wide rates of insertion and deletion were calculated as $\frac{N_g}{S_g \gamma}$, where N_g is the total number of insertions or deletions genome wide, S_g is the number of observable sites in all eight sublines, and γ is the number of generations of MA. We defined an observable site as any site in the genome for which the minimum number of reads required to identify an event (≥ 3 for insertions, ≥ 5 for deletions) were successfully mapped (mapping quality ≥ 30). Superfamily-specific insertion and deletion rates were calculated as $\frac{N_f}{N_s \gamma}$, where N_f is the number of new insertions or deletions for that superfamily, N_s is the starting copy-number of that superfamily, and γ is the number of generations of MA. Superfamily-specific insertion and deletion rates were estimated for all active TE superfamilies, and TEs were considered active if an insertion *or* deletion event was observed in either line. All 95% confidence intervals were calculated by a genome-wide bootstrap of 100 kb windows, calculating rates of activity 1000 times.

Estimating False Positive and False Negative Rates

To estimate false positive and negative rates (see supplementary table S3, Supplementary Material online), we simulated an MA experiment and analyzed these simulated data using our TEFLoN pipeline. We generated four unique chromosomes, representing four independently evolving sublines, by simulating single nucleotide polymorphisms (SNPs) in *D. melanogaster* chromosome 2R (r.5.57) using pIRS v1.1.0 (options: -d 0.0 -v 0.0; Hu et al. 2012). Next, we randomly inserted a set of 100 new TEs and removed a set of 100 reference TEs from all four sublines. Finally, we inserted 100 new TEs and removed 100 reference annotated TEs from one of the four sublines (the focal subline). This technique both mimics the differences between our starting lines and the *D. melanogaster* reference genome, and simulates insertions and deletions of both new (relative to the reference) and known (reference annotated) TEs.

Simulated insertions and deletions were restricted to lengths ≥ 500 bp, but were not restricted to full-length elements. We also simulated a TSD flanking each insertion, where the TSD length was randomly drawn from a Poisson distribution ($\lambda = 5$). The physical position and family identity of all simulated insertions and deletions was chosen randomly, with the caveat that we did not allow nested events. Finally, we independently simulated Illumina PE sequencing of the four chromosomes using pIRS (options: -l 74 -×17 -m 182) and used quality control and alignment methods identical to those described above. False positive rates were estimated independently for insertions and deletions as $\frac{FP}{FP+TN}$, where FP is the number of discovered TEs falsely inferred to be either insertions or deletions and TN is the number of pre-existing TE copies (i.e. discovered TEs not classified as either insertions or deletions). False negative rates (FNR) were estimated independently for insertions and deletions as $\frac{FN}{FN+TP}$, where FN is the number of simulated insertions or deletions that were not identified and TP is the number of simulated insertions or deletions that were correctly classified as either
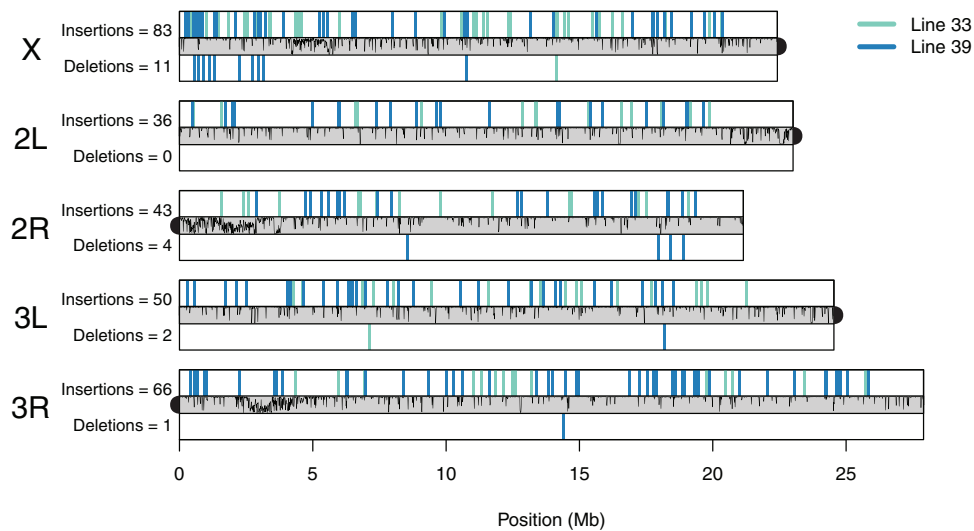
Fig. 1.—Genome-wide plot of transposable element insertions and deletions events discovered along chromosomes X, 2L, 2R, 3L, and 3R in *D. melanogaster* (r5.57). Counts represent events discovered in both Line 33 (light green) and Line 39 (blue). The fraction of observable sites in non-overlapping 10 kb windows is plotted in gray. Centromeres are shown with black semicircles.

insertions or deletions. We also estimated these rates for euchromatic and non-euchromatic regions of the genome separately, as we expect reduced power to detect events in non-euchromatic regions given biases in sequencing and aligning to these regions. The discovery of these simulated insertions and deletions provided the basis for the read-count threshold parameters used in our study.

Statistical Analysis

We used a negative binomial generalized linear model (nbGLM) to test for significant linear relationships between TE activity and features of the local genomic environment such as TE content, exon content, GC content, and recombination rate [formula: TE events per window ~ TE content + exon content + GC content + recombination rate]. The nbGLM used only genomic windows with >70% observable sites. TE insertion and deletion counts, TE content, exon content, and GC content were calculated for non-overlapping 10 kb windows using the *D. melanogaster* reference genome (FlyBase v.5.57). TE and exon contents were calculated as the fraction of bases in each window within annotated TEs or exons, respectively. Recombination rate data were acquired from Comeron et al. (2012).

We tested for non-random patterns in the spatial distribution of insertions and deletions between the chromosomes, between lines, and between chromatin state, using Fisher's exact tests. To control for unequal power to detect events across genomic regions (due to generally higher coverage in euchromatin), we standardized each region or chromosome by the number of observable sites. One column of the contingency table was comprised of the counts of observable sites, whereas the other column was comprised of insertion or

deletion counts. We tested for a proximity effect by randomly permuting our observed insertions 1,000 times to identify a distribution of distances to the nearest element of the same superfamily (calculated separately for DNA and RNA elements) and a distribution of counts for which we observe a new insertion and a pre-existing copy from the same superfamily within the a specified genomic window [1 kb, 10 kb, 100 kb, and 1,000 kb windows tested]. We obtained canonical TE lengths from the FlyBase set of full-length TEs (dos Santos et al. 2015) and chromatin states from Kharchenko et al. (2011; http://modencode.org). We used Bonferroni corrections when assessing the statistical significance of multiple tests; all statistical analyses were performed in R Development Core Team 2011.

Results

In total, we observed 280 insertion and 18 deletion events across all eight sublines of the MA experiment after 145–149 generations of mutation accumulation (fig. 1, table 1, see supplementary tables S6 and S7, Supplementary Material online). These observations were based on paired-end sequence data providing, on average, 17x coverage of the genome (see supplementary table S2, Supplementary Material online), which allowed us to obtain support for each event from multiple reads. We tested the performance of our TEFLoN pipeline and estimated false positive and FNR by simulating a MA line under conditions representative of the real MA experiment. To do this, we simulated four starting sublines—each derived from *D. melanogaster* (r.5.57) chromosome 2R—unique in their TE composition relative to the reference. We then inserted and removed TE sequence from one of those sublines and independently simulated Illumina

Table 1

Observed Insertion and Deletion Events for TEs in Eight Sublines of two *Drosophila melanogaster* MA Lines

	Subline	Insertions	Deletions
Line 33	33–45	19	0
(starting copy-number: 2311)	33–27	19	0
	33–55	33	1
	33–5	33	1
	Total	104	2
Line 39	39–58	46	2
(starting copy-number: 2231)	39–67	63	13
	39–51	45	1
	39–18	22	0
	Total	176	16

sequencing on all four sublines (see methods). Our estimate of FNR suggest that we have less power to detect TE deletion events (FNR = 0.3) than insertion events, (FNR = 0.16), likely because many TE deletions occur in heterochromatic regions of the genome—regions that are generally more repetitive and more difficult to sequence and map. False positive rates (FPR) were similar between insertions (FPR = 0.02) and deletions (FPR = 0.01) and were not dramatically different between euchromatic and non-euchromatic regions of the genome (see supplementary table S3, Supplementary Material online).

We found that 24 known TE superfamilies in *D. melanogaster* are active in these lines. We note that our methods are unable to distinguish between TE excisions (transposition events mediated by TE machinery) and TE deletions arising by other mechanisms; both are simply classified as deletions in this report. Further, because our method does not detect nested TEs and because estimated FNR are roughly an order of magnitude greater than false positive rates (see supplementary table S3, Supplementary Material online), our estimates provide a lower bound for both the rates of transposition and the starting copy-number of TEs in each line. Despite this limitation, our counts of the starting copy-number of all TEs (2311 and 2231 in Lines 33 and 39, respectively; table 1) are roughly consistent with the number of annotated TEs in the *D. melanogaster* reference genome (3170 after the exclusion of *ine-1* elements; FlyBase v5.57; dos Santos et al. 2015), although they are considerably lower than some recent reports of copy-number in natural populations of *D. melanogaster* (e.g. >23,000 copies [Cridland et al. 2013] and ~10,000 copies [Kofler et al. 2012]); these differences are likely due to the fact that many rare TEs are discovered in population studies, but may also reflect differences in the annotation methods used.

Genome-Wide Rates of Insertion and Deletion

We characterized TE activity by first estimating the genome-wide rate of insertion and deletion across all TE superfamilies

per-site per-generation. Sites were calculated as the total number of positions in the genome that met the thresholds of base quality, map quality, and read depth necessary to detect insertions or deletions (see Methods, fig. 1). The genome-wide rate of insertion (2.11×10^{-9} [95% CI = 1.87×10^{-9}–2.38×10^{-9}] per-site per-generation) was significantly elevated relative to the rate of deletion (1.37×10^{-10} [95% CI = 8.36×10^{-11}–2.06×10^{-10}] per-site per-generation) ($P = 2.2 \times 10^{-16}$; Fisher's exact test [FET]). Moreover, this difference persisted after correcting our TE counts for a higher FNR for deletions relative to insertions ($P_{FET} < 2.2 \times 10^{-16}$). The eight sublines were derived from two unrelated founder lines, allowing us to compare rates of TE activity between genotypes. We found that both insertion and deletion rates were significantly elevated in Line 39 relative to Line 33 ($P_{FET} < 7.23 \times 10^{-4}$ for both types of events). The estimated rates of insertion in Line 33 and 39 were 1.57×10^{-9} [95% CI = 1.30×10^{-9}–1.88×10^{-9}] per-site per-generation and 2.66×10^{-9} [95% CI = 2.26×10^{-9}—3.07×10^{-9}] per-site per-generation, respectively, while the estimated rates of deletion were 3.04×10^{-11} [95% CI = 0.0–7.61×10^{-11}] and 2.44×10^{-10} [95% CI = 1.37×10^{-10}–3.66×10^{-10}] per-site per-generation, respectively. The genome-wide deletion rate in Line 39 was strongly driven by deletions that occurred in a single subline (39–67), which accounted for 72% of all deletions observed in the experiment. Deletions rates were not significantly different between the lines after excluding deletions in subline 39–76.

Superfamily-Specific Rates of Insertion and Deletion

We also calculated superfamily-specific rates of insertion and deletion per-copy per-generation (with starting copy-number counted separately for each superfamily). Rates of superfamily-specific insertion and deletion were highly variable, and ranged from 0 to 5.13×10^{-3} per-copy per-generation for insertions and from 0 to 1.29×10^{-4} per-copy per-generation for deletions (fig. 2; see supplementary tables S4 and S5, Supplementary Material online). *Copia* insertions comprised 61% (107 out of 176) of the total insertion events in Line 39, while not a single *copia* insertion was detected in Line 33. This observation agrees with results reported in Houle and Nuzhdin (2004) based on *in situ* experiments using the same MA lines.

Superfamily-specific insertion and deletion rates were not significantly different between lines ($P > 0.11$ for both comparisons; Mann–Whitney *U* tests; fig. 2). Because of the exceptionally high rate of *copia* insertions in Line 39, we tested for a difference between rates after excluding *copia* elements and found superfamily-specific insertion rates in Line 33 were marginally elevated relative to Line 39 ($P_{MWU} = 0.053$). Similar superfamily-specific rates between the lines (measured per-copy) suggest that the higher genome-wide rate of insertions
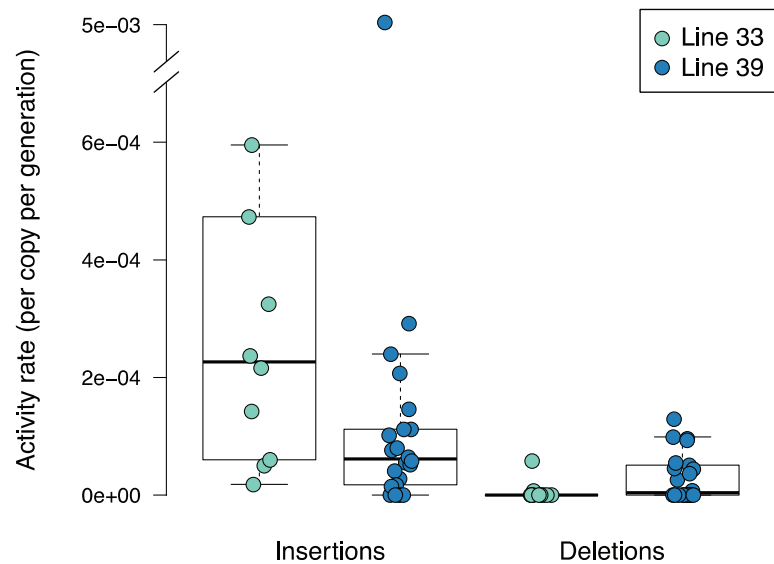
Fig. 2.—Superfamily-specific insertion and deletion rates for all active superfamilies in Line 33 (light green) and Line 39 (blue). Each dot represents the per-copy per-generation rate for an individual superfamily. *Copia* insertion rate in Line 39 is shown using an axis break.

(measured per-site) in Line 39 might be driven by *copia*. Indeed, the exclusion of *copia* elements reversed the pattern of higher insertion in Line 39 for genome-wide per-site per-generation rate estimates, causing Line 33 to have a higher rate of insertions when measured per-site per-generation ($P_{FET} = 0.010$). Notably, the elevated rate of genome-wide deletions (measured per-site per-generation) in Line 39 was not affected by the exclusion of *copia* ($P_{FET} = 0.002$). Most of the families for which estimates are available from earlier studies were found to be active in this experiment, and the rates we estimated are generally within the range of those previously reported (see supplementary table S1, Supplementary Material online).

We also tested for an effect of TE order (LTR, non-LTR, TIR), TE class (DNA, RNA), canonical sequence length, and starting copy-number on superfamily-specific rates of insertion and deletion (fig. 3). Rates of activity between orders were not significantly different for both insertions ($P_{ANOVA} = 0.32$) and deletions ($P_{ANOVA} = 0.46$). A similar pattern was seen for differences between DNA and RNA elements (class) for both insertions ($P_{MWU} = 0.46$) and deletions ($P_{MWU} = 0.75$). Superfamily-specific rates of both insertion and deletion were positively correlated with the canonical length of the superfamily and negatively correlated with starting copy-number, but these correlations were not statistically significant ($P_{Spearman's\ rho} > 0.5$ for all comparisons, fig. 3). Importantly, TEs in heterochromatin may be contributing to new insertions even though these donor copies would go undetected by TEFLoN—potentially influencing the association between insertion rates and copy-number. To estimate the extent of undetected TE donors relative to discovered copy-number, we associated the copy-number of each superfamily with its respective read coverage. The significant positive correlation between copy-number and coverage ($\rho = 0.7$, $P < 10^{-16}$; see supplementary fig. S3, Supplementary Material online) suggests that there is a positive correlation between the number of true TE donors in these lines and the superfamily copy-numbers discovered by TEFLoN.

It should be noted that a negative relationship between activity rate and copy-number is expected in the absence of any causative relationship between copy-number and counts of insertions or deletions, as our measure of superfamily-specific rate is not independent of copy-number (i.e. copy-number appears in the rate term). However, we assume that TE superfamilies with higher copy-numbers do have more opportunities to transpose relative to superfamilies with low copy-numbers. Therefore, a negative correlation between insertion rate and copy-number is consistent with theory that predicts the evolution of TE self-regulation or the evolution of host suppression (Charlesworth and Charlesworth 1983; Charlesworth and Langley 1989). These results run contrary to previous experiments linking increases in transposition rates with higher copy-number (Nuzhdin et al. 1996; Pasyukova et al. 1998).

The Local Genomic Environment Influences Patterns of TE Insertions and Deletions

We tested for rate heterogeneity between chromosome types by comparing counts of insertions and deletions on each chromosome arm (relative to the number of sites observable). We found significantly elevated rates of both insertions (0.86 fold increase) and deletions (5.97-fold increase) on the X chromosome relative to the autosomes ($P_{FET} < 7.15 \times 10^{-5}$ for both
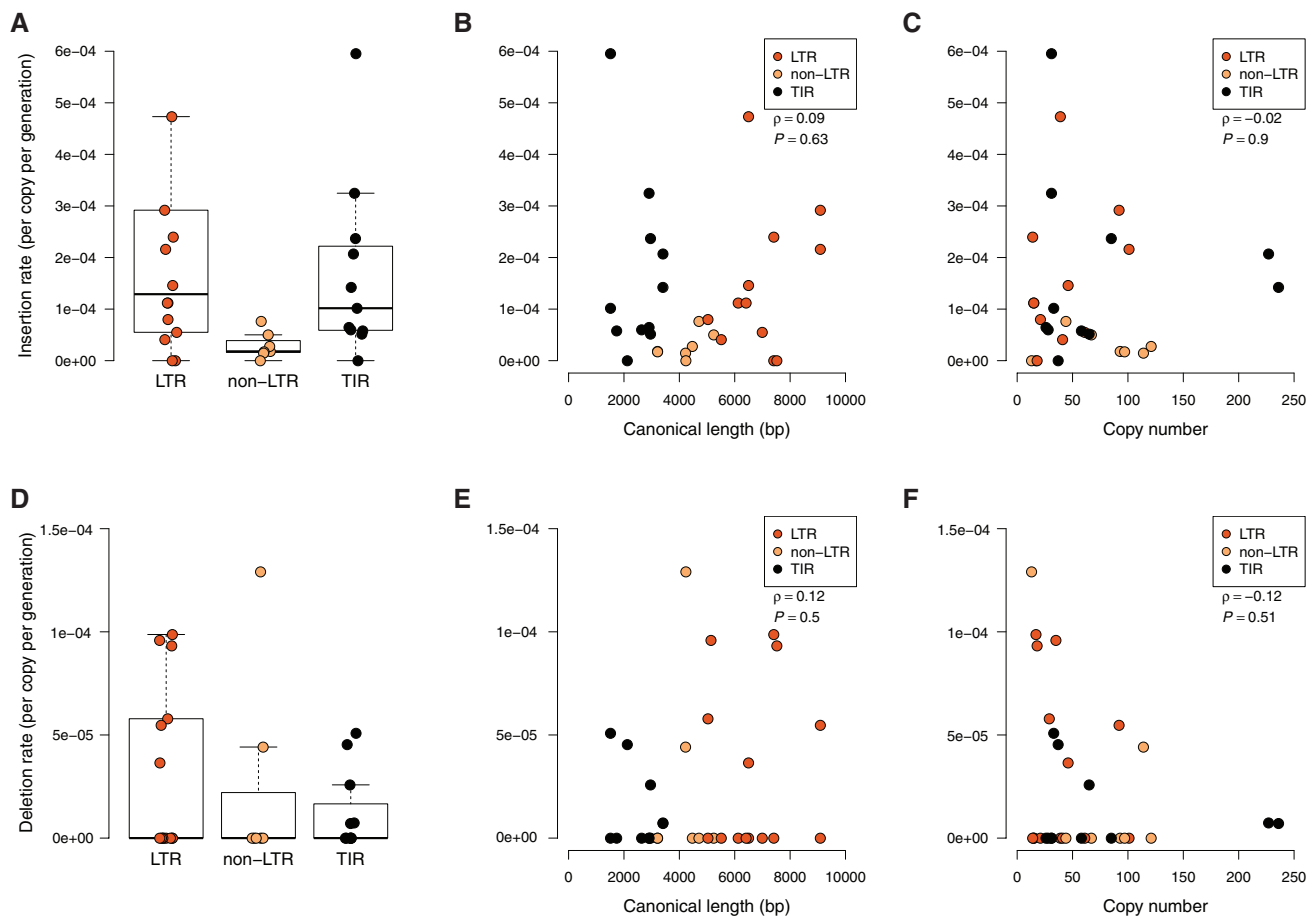
Fig. 3.—Comparison of superfamily-specific rates of insertion (*A*, *B*, and *C*) and deletion (*E*, *F*, and *G*) among TE orders (LTR, non-LTR, TIR), based on TE length, and relative to starting copy-number. Spearman's ρ and *P* values obtained by testing for a correlation between activity rate and either length or copy-number for all active superfamilies.

comparisons; see supplementary fig. S1, Supplementary Material online) and a significant reduction of the insertion rate (60% reduction) on chromosome 2L ($P_{FET} = 7.58 \times 10^{-3}$; see supplementary fig. S1, Supplementary Material online). Moreover, the exclusion of *copia* elements from these analyses strengthened the statistical significance and magnitude of chromosome-specific biases.

We tested for non-independence between TE activity and chromatin state based on data from two *D. melanogaster* cell lines—BG3 and S2 (Kharchenko et al. 2011)—by comparing counts of insertions and deletions in each of nine chromatin states relative to observable sites. We subdivided our data to individually test for an effect of chromatin state on the insertion rate of 1) all active TE superfamilies, 2) all superfamilies excluding *copia*, and 3) *copia* alone. Insertions of all TE superfamilies were biased to occur in regulatory chromatin (enhancers) ($P < 1.44 \times 10^{-4}$ for both cell lines; FET), however this pattern is strongly driven by *copia* insertions and is not statistically significant after excluding *copia* (see supplementary fig. S2, Supplementary Material online). There was no

significant relationship between any chromatin state and patterns of deletion activity ($P_{FET} > 0.06$ for all deletion tests). These results suggest that chromatin state may play an important role in shaping the spatial distributions of some TE families along the chromosomes, but that this role may be idiosyncratic to individual TE families. It should also be noted that the landscape of chromatin states identified in *D. melanogaster* cell lines may not be representative of the landscape in our experimental lines, although there are general consistencies found between the cell lines (Kharchenko et al. 2011).

We also used a generalized linear model to test for associations between insertion TE activity and additional features of the local genomic environment (i.e. TE content, GC content, exon content, and recombination rate). We found a weakly significant negative correlation between insertion activity and GC content and a suggestive negative correlation between deletion activity and exon content (table 2). The latter result is expected, as selection likely shaped the spatial distribution of TEs along the chromosomes in the founding population prior to the start of mutation accumulation. We did not find a

Table 2

Results from Negative Binomial Generalized Linear Models Characterizing the Effect of Local Genomic Features on TE Activity

	Coefficient [StdErr]	Test Statistic	P Value
Insertions			
TE content[a]	−3.44 [2.10]	−1.63	0.10
Exon content[b]	0.10 [0.26]	0.40	0.69
GC content[c]	−4.66 [2.23]	−2.09	0.04
Recombination rate[d]	−0.005 [0.03]	−0.17	0.86
Deletions			
TE content[a]	−6.93 [11.68]	−0.60	0.55
Exon content[b]	−2.08 [1.13]	−1.85	0.06
GC content[c]	12.31 [9.10]	1.35	0.18
Recombination rate[d]	0.06 [0.10]	0.64	0.52

Recombination rate estimates were acquired from Comeron et al. (2012). All other genomic features were estimated using non-overlapping 10 kb windows in the *D. melanogaster* reference genome (FlyBase v.5.57).

[a]% of window in annotated TE sequence.
[b]% of window in exons.
[c]% GC.
[d]cM/Mb.

significant correlation between TE activity and recombination rate. Importantly, we also did not find a significant correlation between insertions and exon content, consistent with little to no selection acting in our MA experiment. These results suggest little direct effect for recombination rate on the distribution of TE copies across the genome, but suggest that TE activity may be influenced by other factors of the local genomic environment, such as GC content.

Finally, we used permutation tests to test two proximity-effect hypotheses. First, we tested whether new insertions were more likely than expected to occur near pre-existing copies from the same superfamily. Second, we tested if DNA elements insert closer to pre-existing copies of the same superfamily than do RNA elements—which have to be reverse transcribed in the cytosol. We did not find a significant effect of proximity to pre-existing copies for either hypothesis (see supplementary table S8, Supplementary Material online).

Discussion

Despite the abundance of data characterizing transposable element dynamics in natural populations, there have been a limited number of experiments characterizing their mutation rates and mutational properties when selection is minimized. In *D. melanogaster*, previous experiments quantifying TE insertion and deletion rates using molecular techniques were indirect (e.g. de Boer et al. 2007; Petrov et al. 2011) or limited to one or a few TE families (e.g. Maside et al. 2000; Nuzhdin and Mackay 1994; Vázquez et al. 2007; see supplementary table S1, Supplementary Material online). Our MA survey provides direct estimates of the genome-wide rates and patterns of movement for all known TE superfamilies in *D. melanogaster*. Further, we were able to look at patterns of insertion and deletion with respect to features of TE superfamilies and

features of the host genomic environment in order to determine what, if any, non-selective factors determine the accumulation of TEs in certain regions of the genome.

We found that TE insertions were vastly more common than TE deletions, and also identified a strong interaction between TE activity and host genotype, as per-site per-generation insertion and deletion rates were significantly elevated in Line 39 relative to Line 33. These results hold even after taking into account the higher FNR for deletions relative to insertions. The elevated insertion rate in Line 39 was entirely driven by a burst of activity in a single family, *copia*, which had previously been shown to be highly active in this line using *in situ* methods (Houle and Nuzhdin 2004). However, the elevated rate of deletions in Line 39 was strongly driven by deletions that occurred in a single subline (39–67), which accounted for 72% of all deletions observed in the experiment, and therefore cannot be ascribed to the genetic background of Line 39. Comparative and population genetic data from *Drosophila* generally find a deletion bias among small indels (Petrov 2002), suggesting that the genome would be shrinking all other things being equal. Although our TEFLoN pipeline cannot distinguish between true TE excisions and spontaneous large deletions, visualizing the data using IGV suggests that many of the deletions we report are the products of complete excisions of the TE sequence that was present prior to MA (whether full or partial). The elevated rate of insertion compared with deletion of TEs reported here—coupled with the larger size of non-TE-associated duplications found previously (Schrider et al. 2013)—may therefore help explain the relative stability of TE numbers and genome size (*Drosophila* 12 Genomes Consortium 2007) in the face of deletion bias.

We identified a weakly significant negative relationship between GC content and overall rates of TE insertion (table 2).

Similar correlations have been identified between LINE elements and GC content in humans (Jin et al. 2012; Ovchinnikov et al. 2001), though this pattern is notably different for some other TE families (Jin et al. 2012; Hellen and Brookfield 2013). Based on their method of transposition, it might be expected that we should identify fewer deletions of RNA elements relative to DNA transposons, as the latter encode a transposase gene used in TE excision. Surprisingly, 13 out of the 18 TE deletions (72%) identified were deletions of RNA elements, suggesting mechanisms other than excision—such as deletion through non-allelic homologous recombination—may be a more common way of TE removal in these genomes. Moreover, we did not find significant differences in superfamily-specific rates of insertion or deletion between TE orders (LTR, non-LTR, TIR), though we did find many fewer active non-LTR superfamilies than LTRs or TIRs (fig. 3). Other studies in *Drosophila* found that non-LTRs tend to be older than LTRs and thus are expected to exhibit less recent activity than LTRs (Bergman and Bensasson 2007). We also did not find any evidence that new insertions occur closer to members of the same superfamily—as was recently shown for IS elements in *Escherichia coli* (Lee et al. 2016). Nor did we find that new DNA-element insertions were closer to TEs of the same superfamily than are new RNA element insertions, the latter requiring reverse transcription in the cytosol and thus, perhaps, more likely to insert farther from initial donor sites.

We identified a significantly elevated rate of TE deletion on the X chromosome relative to the autosomes (fig. 1, see supplementary fig. S1, Supplementary Material online). An elevated deletion rate on the X is consistent with the absence of a homologue-dependent DNA repair mechanism—excised or deleted TEs might not be restored in hemizygous males. However, very little is known about the precise mechanism of TE excision repair, and it is generally thought that excised elements are repaired from the sister chromatid during the replication cycle (reviewed in Burt and Trivers 2006; Hickman and Dyda 2015), thus rates of deletion on hemizygous chromosomes may not be expected to increase. We also found a significantly higher rate of new TE insertions on the X chromosome (fig. 1, see supplementary fig. S1, Supplementary Material online). Recent studies using natural populations of *D. melanogaster* have also described conflicting patterns of accumulation on X chromosome: higher densities of TEs on the X relative to the autosomes (Cridland et al. 2013), lower TE densities on the X (before controlling for recombination rate), or no effect (after controlling for recombination rate; Kofler et al. 2012). Male hemizygosity for the X chromosome—in concert with the lack of male recombination in *Drosophila*—means that recombination rates are, on average, higher on the X than on the autosomes (Comeron et al. 2012). Our results suggest that selection against the deleterious effects of TEs might be stronger on the X chromosome than on the autosomes, consistent with data suggesting more

effective selection on the X chromosome overall (Charlesworth et al. 1987; Langley et al. 2012; Charlesworth and Campos 2014). Indeed, there are significantly fewer reference-annotated TEs on the X chromosome than on the autosomes in the *D. melanogaster* reference genome ($P_{FET} = 3.46 \times 10^{-5}$), though for some TE families population frequency does not appear to be different between the X chromosome and autosomes (Petrov et al. 2011).

Numerous studies have focused on the strong association between TEs and other hemizygous sex chromosomes—TEs are greatly overrepresented on the Y and W chromosomes in many animal lineages (Clinton and Haines 1999; Charlesworth and Charlesworth 2000; Graves 2006; Steinemann and Steinemann 2005; Bachtrog 2013; Chalopin et al. 2015). These patterns have typically been attributed to the lack of recombination on the Y and W chromosomes (Charlesworth and Charlesworth 2000; Steinemann and Steinemann 2005), particularly in *Drosophila* where males do not recombine. However, elevated insertion rates on the X chromosome could also be driven by an effect of heterochromatin, as a larger fraction of the X chromosome is heterochromatic relative to the autosomes (Hilliker et al. 1980). A recent study in teleost fishes (Chalopin et al. 2015) demonstrates that TEs accumulate not only on the Y and W chromosomes, but also in young sex-determining regions of the X and Z chromosomes—chromosomes expected to freely recombine during female meiosis. Moreover, specific classes of TEs have recently proliferated in these regions, suggesting that the biased recruitment of certain TE types may be playing an active role in sex chromosome differentiation (Chalopin et al. 2015) and that associations between TEs and sex chromosomes in the early stages of differentiation may be independent of recombination. We also found a significant association between TE insertions and regulatory chromatin—based on experimentally determined heterochromatic marks (Kharchenko et al. 2011), although this pattern was limited to *copia* elements. Our results therefore provide some support for the hypothesis that heterochromatin may play a bigger role than recombination in shaping TE accumulation.

Going forward, it will be essential to characterize the various routes by which insertion and deletion are facilitated or impeded at the molecular level. In particular, the special role of small RNAs is only beginning to be investigated (reviewed in Lee and Langley 2010). Recent evidence for the suppression of transposition, especially through piRNA-mediated epigenetic silencing (Lee 2015), suggests the potential for biased TE recruitment into piRNA clusters—discrete genomic loci comprised of nested TE fragments that generate piRNA primary transcripts (Brennecke et al. 2007). The recruitment of TEs into piRNA clusters (many of which lie in heterochromatic regions) could be facilitated through heterochromatin binding proteins, such as *Drosophila* HP1 (reviewed in Vermaak and Malik 2009) or its homolog, Rhino, that specifically binds piRNA clusters in *D. melanogaster* (Zhang et al. 2014).

A similar integration preference has been observed in *S. cerevisiae*, where nearly all new *Ty5* insertions occur in heterochromatin at the telomeres (Bushman 2003), and this integration preference is driven by an interaction between the *Ty5* integrase and a yeast heterochromatin binding protein (Xie et al. 2001; Zhu et al. 2003).

In addition to the mutational biases reported here, selection against the deleterious effects of TEs is likely to be a substantial contributor to the patterns of distribution across the genome. Previous studies using natural populations of *D. melanogaster* have identified a negative correlation between insertion-site frequency and recombination rate (Petrov et al. 2011; Kofler et al. 2012), consistent with more efficient purifying selection in areas of higher recombination. However, this association disappears after excluding pericentromeric regions of the genome (Kofler et al. 2012). Selection could also shape the spatial landscape of TEs by favoring the recruitment of TEs into piRNA clusters. Consistent with this hypothesis, simulations have shown that piRNA-generating TEs should be selectively advantageous, as their integration thus represses the transposition of other elements (Lu and Clark 2010). Together, these results suggest that insertion and deletion biases, in addition to the effects of selection, are likely contributing to the non-random spatial distribution of TEs.

Comparative analyses of TE insertion and deletion rates between the germline and soma, between the sexes, and among sister taxa are also needed to fully understand TE dynamics (e.g. Keightley et al. 2009; Diaz-Gonzalez et al. 2011). Evidence for differences in male versus female germline transposition rates exists—for example, *R2* rates are higher in females (Zhang et al. 2008) and *roo* rates are higher in males (Vázquez et al. 2007)—but the extent and consequences of heterogeneous rates over long time-scales is unknown. Our genome-wide estimates of the rates and patterns of TE movement provide an opportunity to test key assumptions about the behaviors of TEs in a well-studied model system. Additional in-depth analyses of transposable element mobility in an experimental framework with and without selection will help explain the impact of this dynamic component of the genome over longer time-scales.

Acknowledgments

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 14:113–124.

Bartolomé 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.

Bergman CM, Bensasson D. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 104:11340–11345.

de Boer JG, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8:422.

Brennecke J, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128:1089–1103.

Burns KH, Boeke JD. 2012. Human transposon tectonics. Cell 149:740–752.

Burt A, Trivers R. 2006. Genes in conflict. Cambridge (MA): Belknap Press.

Bushman FD. 2003. Targeting survival: Integration site selection by retroviruses and LTR-Retrotransposons. Cell 115:135–138.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. Mol Ecol. 22:1503–1517.

Chalopin D, Volff J-N, Galiana D, Anderson JL, Schartl M. 2015. Transposable elements and early evolution of sex chromosomes in fish. Chromosom Res. doi:10.1007/s10577-015-9490-8.

Charlesworth B, Campos JL. 2014. The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. Annu Rev Genet. 48:383–403.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci. 355:1563–1572.

Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet Res. 42:1–27.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130:113–146.

Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. Annu Rev Genet. 23:251–287.

Clinton M, Haines LC. 1999. An overview of factors influencing sex determination and gonadal development in birds. Cell Mol Life Sci. 55:97–115.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8:e1002905.

Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. Mol Biol Evol. 30:2311–2327.

Diaz-Gonzalez J, Vazquez JF, Albornoz J, Dominguez A. 2011. Long-term evolution of the *roo* transposable element copy number in mutation accumulation lines of *Drosophila melanogaster*. Genet Res. 93:181–187.

Drosophila 12 Genomes Consortium 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203–218.

Duret L, Marais G, Biemont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. Genetics 156:1661–1669.

Elliott TA, Stage DE, Crease TJ, Eickbush TH. 2013. In and out of the rRNA genes: characterization of *Pokey* elements in the sequenced *Daphnia* genome. Mob DNA. 4:20.

Engels WR. 1983. The *P* family of transposable elements in *Drosophila*. Annu Rev Genet. 17:315–344.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 41:997-331–368.

Gonçalves JW, Valiati V, Delprat A, Valente VLS, Ruiz A. 2014. Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. BMC Genomics 15:792.

Graves JAM. 2006. Sex chromosome specialization and degeneration in mammals. Cell 124:901–914.

Hellen EH, Brookfield JF. 2013. *Alu* elements in primates are preferentially lost from areas of high GC content. Peer J. 1:e78.

Hickman AB, Dyda F. 2015. Mechanisms of DNA transposition. Microbiol Spectr. 3:1–22.

Hilliker AJ, Appels R, Schalet A. 1980. The genetic analysis of *D. melanogaster* heterochromatin. Cell 21:607–619.

Houle D, Nuzhdin SV. 2004. Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. Genet Res. 83:7–18.

Houle D, Rowe L. 2003. Natural selection in a bottle. Am Nat. 161:50–67. [12650462]

Hu X, et al. 2012. pIRS: Profile-based Illumina pair-end reads simulator. Bioinformatics 28:1533–1535.

Jason de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 7:e1002384.

Jin P, et al. 2012. Evolutionary rate of human tissue-specific genes are related with transposable element insertions. Genetica 140:513–523.

Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci U S A. 100:6569–6574.

Keightley PD, et al. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res. 19:1195–1201.

Kharchenko PV, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature 471:480–485.

Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49–63.

Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genet. 8:e1002487.

Langley CH, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192:533–598.

Lee H, Doak TG, Popodi E, Foster PL, Tang H. 2016. Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. Nucleic Acids Res. gkw647. doi:10.1093/nar/gkw647.

Lee YCG. 2015. The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. PLoS Genet. 11:e1005269.

Lee YCG, Langley CH. 2010. Transposable elements in natural populations of *Drosophila melanogaster*. Philos Trans R Soc B Biol Sci. 365:1219–1228.

Lewis AP, Brookfield JFY. 1987. Movement of *Drosophila melanogaster* transposable elements other than *P* elements in a P-M hybrid dysgenic cross. Mol Gen Genet. 208:506–510.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. Genome Res. 20:212–227.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17. doi:10.14806/ej.17.1.200.

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.

Miyao A, et al. 2003. Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780.

Nuzhdin SV, Mackay TFC. 1994. Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*. Genet Res. 63:139–144.

Nuzhdin SV, Pasyukova EG, Mackay TF. 1996. Positive association between *copia* transposition rate and copy number in *Drosophila melanogaster*. Proc Biol Sci. 263:823–831.

Ovchinnikov I, Troxel AB, Swergold GD. 2001. Genomic characterization of recent human *LINE-1* insertions: evidence supporting random insertion. Genome Res. 11:2050–2058.

Pasyukova EG, Nuzhdin SV, Morozova TV, Mackay TFC. 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. J. Hered. 95:284–290.

Pasyukova EG, Nuzhdin SV, Filatov DA. 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.

Petrov DA. 2002. DNA loss and evolution of genome size in *Drosophila*. Genetica 115:81–91.

Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. Mol. Biol. Evol. 28:1633–1644.

Quesneville H, et al. 2005. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput. Biol. 1:e22.

R Core Development Team. 2011. R: A language and environment for statistical computing. Vienna (Austria): The R Foundation for Statistical Computing.

Rebollo R, Farivar S, Mager DL. 2012. C-GATE - catalogue of genes affected by transposable elements. Mob DNA. 3:9.

Rizzon C, Marais G, Gouy M, Biémont C. 2002. Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. Genome Res. 12:400–407.

Robinson JT, et al. 2011. Integrative genomics viewer. Nat Biotechnol. 29:24–26.

dos Santos G, et al. 2015. FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 43:D690–D697.

Schnable PS, et al. 2009. The B73 *Maize* genome: complexity, diversity, and dynamics. Science 326:1112–1115.

Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. Genetics 194:937–954.

Singleton TL, Levin HL. 2002. A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukaryot Cell. 1:44–55.

Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila P* elements preferentially transpose to replication origins. Proc Natl Acad Sci U S A. 108:15948–15953.

Steinemann S, Steinemann M. 2005. Y chromosomes: born to be destroyed. BioEssays 27:1076–1083.

Vázquez JF, Albornoz J, Domínguez A. 2007. Direct determination of the effects of genotype and extreme temperature on the transposition of roo in long-term mutation accumulation lines of *Drosophila melanogaster*. Mol Genet Genomics. 278:653–664.

Vermaak D, Malik HS. 2009. Multiple roles for Heterochromatin Protein 1 genes in *Drosophila*. Annu Rev Genet. 43:467–492.

Xie W, et al. 2001. Targeting of the yeast *Ty5* retrotransposon to silent chromatin is mediated by interactions between integrase and *Sir4p*. Mol Cell Biol. 21:6606–6614.

Yant SR, et al. 2005. High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol. 25:2085–2094.

Zhang X, Zhou J, Eickbush TH. 2008. Rapid *R2* retrotransposition leads to the loss of previously inserted copies via large deletions of the rDNA locus. Mol Biol Evol. 25:229–237.

Zhang Z, et al. 2014. The *HP1* homolog *Rhino* anchors a nuclear complex that suppresses piRNA precursor splicing. Cell 157:1353–1363.

Zhu Y, Dai J, Fuerst PG, Voytas DF. 2003. Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci U S A. 100:5891–5895.

Associate editor: Emmanuelle Lerat