The Transposable Elements of the *Drosophila serrata* Reference Panel

Zachery Tiedeman and Sarah Signor *

Department of Biological Sciences, North Dakota State University, Fargo, USA

*Corresponding author: E-mail: sarah.signor@ndsu.edu

Accepted: 30 April 2021

Abstract

Transposable elements (TEs) are an important component of the complex genomic ecosystem. Understanding the tempo and mode of TE proliferation, that is whether it is in maintained in transposition selection balance, or is induced periodically by environmental stress or other factors, is important for understanding the evolution of organismal genomes through time. Although TEs have been characterized in individuals or limited samples, a true understanding of the population genetics of TEs, and therefore the tempo and mode of transposition, is still lacking. Here, we characterize the TE landscape in an important model *Drosophila*, *Drosophila serrata* using the *D. serrata* reference panel, which is comprised of 102 sequenced inbred genotypes. We annotate the families of TEs in the *D. serrata* genome and investigate variation in TE copy number between genotypes. We find that many TEs have low copy number in the population, but this varies by family and includes a single TE making up to 50% of the genome content of TEs. We find that some TEs proliferate in particular genotypes compared with population levels. In addition, we characterize variation in each TE family allowing copy number to vary in each genotype and find that some TEs have diversified very little between individuals suggesting recent spread. TEs are important sources of spontaneous mutations in Drosophila, making up a large fraction of the total number of mutations in particular genotypes. Understanding the dynamics of TEs within populations will be an important step toward characterizing the origin of variation within and between species.

Key words: transposable elements, *Drosophila serrata*, copy number, inbred lines.

Significance

Transposable elements (TEs) move about the genome increasing their copy number and causing large structural mutations, yet we currently know very little about their tempo and mode of transposition. Here we find that in inbred lines of *Drosophila serrata*, TE copy number varies due to transposition within particular genotypes. This suggests that TEs may undergo bursts of transposition when they encounter permissive genotypes, rather than increasing in copy number at a low steady rate.

Introduction

Transposable elements (TEs) are sequences of DNA that multiply within genomes despite potential deleterious impacts to the host (McClintock 1950). TEs are widespread across the tree of life, often making up a significant portion of the genome (Piegu et al. 2006; Schnable et al. 2009; Lee and Langley 2012). TEs also impose a severe mutational load on their hosts by producing insertions that disrupt functional sequences and mediate ectopic recombination (McGinnis et al. 1983; Levis et al. 1984; Lim 1988). TEs can spread through horizontal transfer between nonhybridizing species, allowing them to colonize new host genomes (Kidwell 1983; Kofler et al. 2015; Peccoud et al. 2017). For example, the spread of the *P*-element was documented in *Drosophila melanogaster* from

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Drosophila willistoni in the 1950s, and its subsequent spread into *Drosophila simulans* around 2010 (Daniels, Peterson, et al. 1990; Kofler et al. 2015).

TEs have also been implicated in adaptation. In *Drosophila*, insertion of TEs has been linked to resistance to pesticides and viral infection (Wilson 1993; Daborn et al. 2002; Aminetzach et al. 2005; Magwire et al. 2011; Mateo et al. 2014). In ants and *Capsella rubella*, TEs provide genetic diversity in invading populations that are generally depleted of genetic variation, facilitating adaptation to novel environments (Niu et al. 2019; Schrader et al. 2019). In fission yeast, TE activity was increased in response to stress and TE insertions were associated with stress response genes, supporting the supposition that TEs provide a system to modify the genome in response to stress (Esnault et al. 2019). There is also evidence from vertebrates that TEs provide the raw material for assembling new protein architectures through capture of their transposase domains (Cosby et al. 2020). In summary, there is extensive evidence that TEs provide genetic material for adaptation through a variety of mechanisms.

Early work on TE insertions concluded that on average, they are likely to be neutral or deleterious (Doolittle et al. 1980), and for a long time, active TEs variants were thought to be rare in natural populations (Kaplan et al. 1985; Ronsseray et al. 1991; Brookfield 1991, 1996; Nuzhdin et al. 1997). Alternatively, it was not active TEs that are rare but individuals with "permissive" genetic backgrounds, such that TEs would remain inactive until encountering a permissive genetic background and then proliferate (Nuzhdin 2000). Either way, these models assumed a transposition—selection balance such that TEs are removed by selection at approximately the rate that they proliferate. Since then, TEs have been observed to undergo bursts of activity, which could occur for multiple reasons such as colonization, hybridization, and stress (Vieira et al. 1999; Garcia Guerreiro 2012; Romero-Soriano and Garcia Guerreiro 2016). These bursts are documented in *Drosophila*, rice, fish, and other systems (Vieira et al. 1999; Piegu et al. 2006; de Boer et al. 2007; Bourgeois and Boissinot 2019; Signor 2020). In most cases, transposition bursts in *Drosophila* include few individuals and TEs (Biémont et al. 1987, 1990; Nuzhdin et al. 1997; Yang et al. 2006). The underlying explanation for this burstiness is unclear, including the potential role of burstiness in adaptation.

However, recent insights in the repression of TEs could also offer an alternative hypothesis for the burstiness of TE transposition. The host has a dedicated defense mechanism against TE activity termed the PIWI-interacting (piRNA) system (Brennecke et al. 2007). piRNAs bind to PIWI-clade proteins, such as the product of the *Argonaute 3* gene in *D. melanogaster*, and suppress transposon activity transcriptionally and post-transcriptionally (Brennecke et al. 2007). The majority of these piRNAs originate from genomic regions, which are enriched for TE fragments, termed piRNA clusters (Brennecke et al. 2007; Malone et al. 2009). There is some evidence that insertion of a TE into a piRNA cluster is enough to initiate piRNA-mediated silencing of the TE (Josse et al. 2007; Zanni et al. 2013). Therefore a newly invading TE would proliferate in the host until a copy jumps into a piRNA cluster, which then triggers piRNA silencing of the TE (Bergman et al. 2006; Malone and Hannon 2009; Zanni et al. 2013; Goriaux et al. 2014; Yamanaka et al. 2014; Ozata et al. 2019). This would be seen as a burst of transposition prior to silencing by the piRNA system. However, TEs also appear to become reactivated in response to stress, or potentially variation in the host suppression system.

The transposition rate of TEs is also controlled by other mechanisms, including regulation of promotor activity, chromatin structure, and splicing (Garcia Guerreiro 2012). Therefore the piRNA "trap" model as an explanation for burstiness is as yet still a hypothesis, and an understanding of the distribution of TEs within populations is still needed to understand the tempo and mode of TE evolution.

Recently an inbred panel of 110 genotypes was created for *D. serrata*, a member of the *montium* group (Reddiex et al. 2018). *Drosophila serrata* is a model system for understanding latitudinal clines and the evolution of species boundaries (Blows 1993; Jenkins and Hoffmann 1999; Hallas et al. 2002; Hoffmann and Shirriffs 2002; Liefting et al. 2009). The *montium* group contains 98 species and represents a significant fraction of known *Drosophila* species (6%, Lemeunier et al. 1986; Reddiex et al. 2018). For a long time, it was thought to be a subgroup of the *melanogaster* group, but has recently been reclassified as its own species group (Lemeunier et al. 1986; Da Lage et al. 2007; Yassin 2013). It split from the *melanogaster* group at least 40 Ma (Tamura et al. 2004). It has a broad geographic range from Papua New Guinea to South Eastern Australia (Jenkins and Hoffman 2001). The *D. serrata* panel was sampled from a single large population within its endemic distribution in Australia, and because of this also exhibits high nucleotide diversity (pi = 0.0079; Reddiex et al. 2018). Although the development of a panel represents a new opportunity for genomic investigation in the group, such as GWAS, very little work has been done understanding the landscape of repetitive elements in this group. For example, *D. serrata* was found to contain a domesticated *P*-element, though no evidence of active *P*-elements was noted (Nouaud and Anxolabéhère 1997; Nouaud et al. 1999). Screens for the presence of the *Drosophila hobo* element in the *montium* group were mixed, and inconclusive for *D. serrata* (Daniels, Chovnick, et al. 1990). *Copia* and *412* were not detected in *D. serrata*, though the DNA transposon *Bari-1* was (Biémont and Cizeron 1999), and evidence for the presence of the *mariner* element is equivocal (Maruyama and Hartl 1991; Brunet et al. 1994). Here we characterize the TE landscape in the *D. serrata* Genetic Reference panel. We have two goals: 1) To understand the TE content of *D. serrata* and its relationship to existing TE

annotations and 2) to understand variability in TE content between individuals in the population and how this relates to the tempo and mode of TE movement. This will provide the groundwork for understanding the role of TEs in the evolution of the *D. serrata* genome, as well investigate differences in the proliferation of TEs across genetic backgrounds.

Results

TEs in *D. serrata*

The Extensive de novo TE Annotator pipeline (EDTA v. 1.0) identified 676 TE families in the *D. serrata* reference genome (consensus sequences of related TEs) (Xu and Wang 2007; Ellinghaus et al. 2008; Xiong et al. 2014; Ou and Jiang 2018, 2019; Ou et al. 2019; Shi and Liang 2019; Zhang et al. 2019; supplementary file 1, Supplementary Material online and fig. 1*A*). The sequences of these TEs are have been deposited in Dfam (Hubley et al. 2016). The classification of the TE families into superfamilies is broadly correct, and in many cases, there is no clear relationship to an existing TE family. However, some errors are evident, for example, element *444* is classified as *copia*, but aligns quite well with the *297* element in *D. melanogaster*, which is a member of the *17.6* clade/*gypsy* superfamily. In addition, some unknown TE families such as *69* align well with existing *D. melanogaster* annotations, in this case *17.6*. In all six TE families that were classified as unknown or *copia* align well with members of the *gypsy* superfamily from *D. melanogaster.* Therefore, classification below the superfamily level is generally ambiguous, though miniature inverted-repeat TEs (MITEs), *Helitrons*, and other DNA transposons are distinguishable. This may be due to deletion of canonical sequences, nested insertions, or other ambiguities of TEs.

Relationship between TEs Found in *D. serrata* and TEs Annotated in Other Species

One hundred and twenty-three of the 676 identified TE families have a well-supported relationship to existing Dfam TE annotations (hmmer.org, Eddy 2008; Hubley et al. 2016, fig. 1*B* and supplementary file 2, Supplementary Material online). This includes, for example, 27 TE families that are related to the *D. melanogaster Max-Element* and 10 TE families that are related to the *D. simulans ninja* element. One of these is also among the most variable TEs and is most closely related to the *Circe* element (*Osvaldo* family). These are likely to be TE families that are younger and that moved between species more recently, and they are almost exclusively long terminal repeat (LTR) elements. The exception being two TIR transposons from the *hobo* family, one *Helitron* from *D. melanogaster*, and two *Helitrons* most closely related to elements from *Heliconius*. This result is expected as overall LTR elements are thought to be younger than non-LTR elements (Bergman and Bensasson 2007). No evidence of *P*-elements was found in the population of identified TEs. In addition, *jockey* elements (non-LTR retrotransposons) are not intended to be identified as a part of this pipeline but do appear to be the identity of two transposon families. The overall phylogeny of the TEs is not what we wish to emphasize here, as the structure of TE classification changes frequently (e.g., whether something is a clade or a family, etc.). In *Drosophila*, there is evidence that *gypsy* elements are infectious, as they can be transferred among strains through exposure or microinjection (Kim et al. 1994; Song et al. 1994).

Relationship between TE Families Annotated by ethylenediaminetetraacetic acid

Out of the 676 TEs annotated by ethylenediaminetetraacetic acid (EDTA), there are 170 TE families that fall into 40 groupings and appear to be closely related to one another (MrBayes 3.2.7, Ronquist et al. 2012, supplementary file 3, Supplementary Material online). Note that because TEs do not follow a standard substitution model, the branch lengths are not meaningful. For example, eight TE families (*52*, *60*, *276*, *346*, *367*, *424*, *539*, *601*) share sequence similarity for the entirety of their length, but are separated by 39 deletions spread across the consensus TE. In the largest-related group of TE consensus sequences, 23, most of the members of this TE group have low relative copy number and variance (fig. 2*A*, average relative copy number 2.3, average variance <1). However, two members of the group are likely still active and have relatively high relative copy number and variance (*376* and *672*, copy number 27, 79; variance 10, 102). Active TE families are not more likely to be related to each other than TEs without apparent activity (active TEs shown in bold, fig. 2). In another case, three members of the group are more distantly related, whereas seven members are more closely related and form two clear groups of origin (fig. 2*B*). Yet again, those which are active in the population, as evidenced by higher relative copy number and variance, are not the most closely related (fig. 2*B*, shown in bold). This is not intended to be an exhaustive accounting of relationships between these TE families, for example, at some point all members of the *roo* clade shared an ancestor. Rather, this is intended to describe recent divergence between members of a group within this species.

Population Frequency of TEs

Copy number of TEs was estimated as the normalized counts of reads mapping to each TE sequence (see Materials and Methods). An average of 17% of reads from individual *D. serrata* lines mapped to TE sequences. The average number of TE insertions per genome in this population of *D. serrata* is 19,909; however, almost 50% of that total (9,036) are from a single repetitive uncharacterized sequence (supplementary table 1, Supplementary Material online). This element shares ~100 bp of sequence similarity with *DNAREP1*, one of the

FIG. 1.—(*A*) The abundance of different types of TEs in *D. serrata* by broad classification type. Note that *gypsy*, *copia*, and unknown elements are LTR elements whereas *Helitrons*, DNA transposons, and MITEs are all different types of DNA transposon. (*B*) The classification of TEs into families that could be aligned to annotated *D. melanogaster* elements. The two *Helitron* elements potentially related to those from *Heliconius* are not included.

FIG. 2.—Relatedness among two TE families annotated by EDTA. Posterior probabilities of each division are shown, however, branch lengths are not meaningful given that TEs do not follow a standard substitution model. (*A*) Twenty-two LTRs (one annotated as a gypsy LTR), which are closely related. (*B*) Ten gypsy LTRs, which are closely related. Members of these relatedness clusters with high amounts of insertion polymorphism are shown in bold.

most abundant and ambiguous TEs in *Drosophila* (Kapitonov and Jurka 2003), suggesting it may be related to this element or be carrying an internal insertion of *DNAREP1*. The next closest in relative copy number is a DNA transposon with 541 copies, thus this is a significant outlier. Six TEs identified in the reference are likely not present in this population. Two of these are present as partial copies in a subset of individuals. Overall among the elements identified by EDTA approximately twice as many are DNA transposons compared with LTR elements. However, the majority of the identified TE families have low relative copy number and variance. Three hundred and ninety of the identified elements have an average relative copy number of less than 3, and all but 2 of those have a variance of less than 1 (supplementary table 1, Supplementary Material online; the other two have variances of 3 and 4). Of those remaining, 148 have a variance in relative copy number of 3 or less. Therefore the vast majority of TE families in this population vary little in relative copy number. Of TEs with an estimated relative copy number of greater than 10, there is still considerable variation in their apparent distribution in the population (fig. 3), with some having high relative copy number and variance (TE 592, TE 371) whereas many others vary

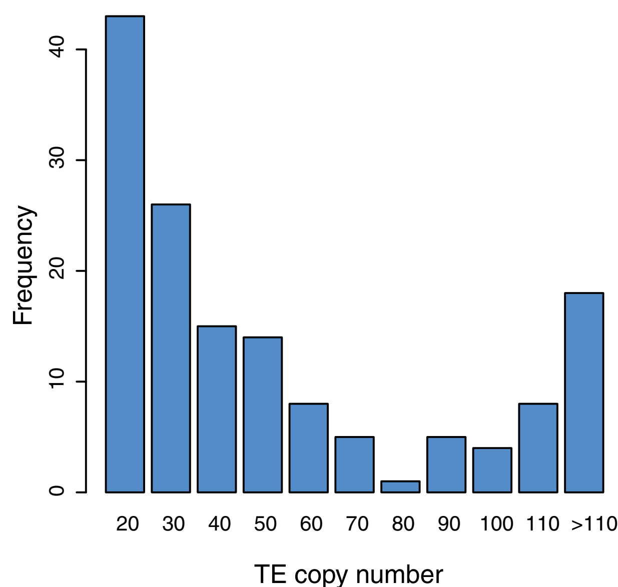


Fig. 3.—TEs vary in their copy number and the amount that this copy number varies between individuals in a population. Shown here is the frequency of TEs with copy number in each bin, that is the first bar is from 10 to 20 copies on average, and so on. Only TEs with a copy number of 10 or greater are shown.

little. Variance is dependent on the mean; therefore, TEs within higher relative copy number are also going to be more variable overall compared with lower relative copy number TEs with higher coefficients of variation (which is not dependent on the mean; supplementary table 1, Supplementary Material online).

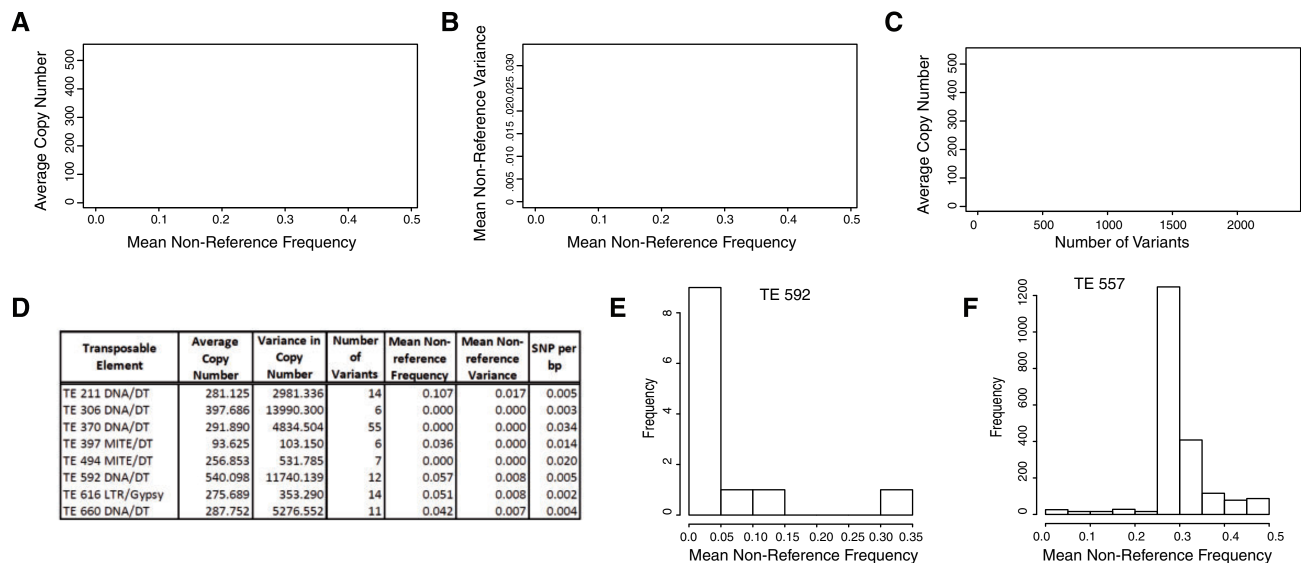
Of those TEs with high relative copy number (>100), two are able to be aligned to *D. melanogaster* elements—TE 63 and *micropia* and TE 616 and *Circe*. This suggests that these TEs invaded more recently than the other transposons, likely from other drosophilids, and that they were able to spread unchecked for some time. There is clearly a lot of variation in the population in susceptibility to TE transposition, as shown by differences in the standard deviation and relative copy number between strains. Higher relative copy number is not necessarily indicative of greater variation (TE 606) nor is relatively lower relative copy number indicative of less variation (TE 136) but both are more common.

Comparison to dnapipeTE

To measure TE abundance using an alternative approach and compare methods, we ran dnapipeTE on a subset of 11 individuals from our data set (v. 1.3, Goubert et al. 2015). dnapipeTE does not report copy number per se, but it does report the number of bases aligned to a given TE. We were most interested in the relative abundance of TEs compared with dnapipeTE; therefore, we compared the correlation between the proportion of total repetitive reads mapping to each element and our estimates of copy number. We limited this comparison to TEs with a copy number of greater than 4 and/or that had been evaluated by dnapipeTE, as it excludes low copy number elements as a part of its framework. Four hundred and thirty-seven consensus TEs remained in the data set. The correlation between the two methods was 0.69, suggesting that the two approaches are relatively concordant in their estimates of TE abundance.

Single Nucleotide Polymorphisms and Summary Statistics

In freebayes, we called single nucleotide polymorphisms (SNPs) within the TEs and allowed the number of copies of TEs to vary freely with the number estimated in this study, thus for example, a single individual for TE 51 could have up to 55 different reference/nonreference calls (v. 1.0, Garrison and Marth 2012). Although indels are often filtered out of SNP frequency data sets, we also chose to keep them here due to the high prevalence of indels in TEs. By averaging over individuals and then folding the frequency spectrum (as there is no way to polarize the direction of change), we then have a summary of the frequency of SNPs across the TEs. This can be examined visually as a sort of site frequency spectrum (SFS) or averaged to a mean frequency. The average frequency of nonreference variants at TEs varies from a low of 0.00 for TE 370 (DNA transposon, copy number 291) to a high of 0.5 for TE 449 (*DNA transposon*, copy number 1.37). This is dependent on copy number, however, as the lower the copy number, the more a nonreference SNP will weigh into the ratio, for example, 0/1 versus 0/0/0/0/0/0/0/1. However, this is also informative—if copy number is low and they have diverged at an SNP between just 1–2 copies, this suggests a long period between transposition events. Overall, TEs that have fewer variants, a higher copy number, and a lower mean nonreference frequency are more likely to be recent invaders who are not well controlled in the population. The best candidates for these criteria are TE 217, TE 370, TE 306, TE 397, TE 494, and TE 217 (fig. 4 and supplementary table 2, Supplementary Material online). Other good candidates having few SNPs and high copy number, but higher nonreference frequency, may have had more than one active copy bearing SNPs invading the population from the outset, including TE 211, TE 411, TE 592, TE 616, TE 638, and TE 660 (supplementary file 1, Supplementary Material online). TE 616 belongs to the *Osvaldo/Circe* family of TEs identified in *D. melanogaster*. Normalizing by the length of the TE does little to alter these results, though TE 411, TE 217, and TE 638 are quite short (90–139 bp) and therefore do have more SNPs per bp making them less likely candidates (supplementary table 2, Supplementary Material online). The variance for these candidates is also quite low, as in figure 4, it can be seen that with increasing nonreference allele frequency the variance between individuals also increases considerably. This could suggest slower invasion, during which SNPs are acquired en route and passed along to some genotypes but not others. Because



Transposable Element	Average Copy Number	Variance in Copy Number	Number of Variants	Mean Non-reference Frequency	Mean Non-reference Variance	SNP per bp
TE 211 DNA/DT	281.125	2981.336	14	0.107	0.017	0.005
TE 306 DNA/DT	397.686	13990.300	6	0.000	0.000	0.003
TE 370 DNA/DT	291.890	4834.504	55	0.000	0.000	0.034
TE 397 MITE/DT	93.625	103.150	6	0.036	0.000	0.014
TE 494 MITE/DT	256.853	531.785	7	0.000	0.000	0.020
TE 592 DNA/DT	540.098	11740.139	12	0.057	0.008	0.005
TE 616 LTR/Gypsy	275.689	353.290	14	0.051	0.008	0.002
TE 660 DNA/DT	287.752	5276.552	11	0.042	0.007	0.004

Fig. 4.—(*A*) The average relative copy number of TEs calculated from coverage data compared with the mean frequency of the nonreference allele in the population of TEs. (*B*) Variance in the frequency of the nonreference allele in the population of TEs compared with the mean of the nonreference allele frequency. (*C*) The number of SNPs within a TE population compared with the average relative copy number. (*D*) TEs with high relative copy number and few variants. Some also have a low nonreference frequency, whereas those that do not are presumed to have more than one active copy in the population differing by a few SNPs. (*E*) The SFS of TE 592, meaning the frequency of the mean frequency of nonreference alleles in the population of TEs. TE 592 is a candidate for recent spread. (*F*) The SFS of TE 557, meaning the frequency of the mean frequency of nonreference alleles in the population of TEs. TE 557 is not a candidate for recent spread, has more SNPs, and has a much higher nonreference allele frequency.

of the complexity of interpreting these data, the VCF files have been made available through dryad (https://datadryad.org/stash/share/QppclB4PpPqngDZcB5xyDYzJiBpRYHlRrQ08xnG9RCI).

Outliers in Individual Genotypes

TEs tend to proliferate in particular inbred genotypes. Out of 102 genotypes, 71 have no TEs with a number of insertions that classify them as outliers. Twelve genotypes contain a single TE with a copy number that is considered an outlier, and the remaining 19 contain 2 or more outliers. This includes 2 genotypes with 13 and 8 TEs with a copy number that is considered an outlier. This also tends to group by TE, as only 36 TEs have at least 1 genotype in which they are an outlier; however, for 18 of these, this is only in 1 genotype. For 5 genotypes, 5 TEs are shared as being outliers, with an additional 2 genotypes that share outliers for 4 of the 5. Many of these outliers are large, for example, for TE 512 the majority of the population has 20–30 copies, whereas a single individual has >200 (fig. 5).

Discussion

There is some evidence from inbred lines that genotypes can vary considerably in TE copy number (Nuzhdin et al. 1997; Pasyukova 2004; Rahman 2015; Signor 2020). The question remains—is it due to differences in the permissiveness of the genetic background, or inheritance of active TEs that segregate at low frequency in the population? In the former scenario, genes segregating in natural populations modify transcription and the rate of transposition of specific TEs. This could include polymorphisms in genes such as *Argonaute 3* and variation in the integration of TEs into piRNA clusters (Birchler et al. 1989; Csink et al. 1994; Pélisson et al. 1994; Lee and Langley 2010, 2012; Zhang and Kelleher 2019). Indeed, variation in the integration of TEs into piRNA clusters appears to be quite common, as Zhang and Kelleher (2019) documented 80 unique independent insertions of *P-elements* into piRNA clusters in the *Drosophila* Genetic Reference Panel (Mackay et al. 2012). If laboratory lines differ in these alleles, this can cause between line variability in transposition rates. In the latter scenario, different lines may have inherited copies of TEs with differences in the propensity to transpose (Ronsseray et al. 1991; Kim et al. 1994; Nuzhdin et al. 1997; Nuzhdin 2000).

Although we cannot measure the likelihood of individual genotypes inheriting multiple active copies of TEs whereas fellow members of the population inherit none, the fact that multiple TE families are proliferating in particular genotypes—that is proliferating TEs have some tendency to co-occur—this supports the idea that these individuals have polymorphisms in genes or other repressive structures that are more permissive to TE transposition. Were the genotypes with clear TE proliferation different for every TE family this would not support either scenario, however, it does seem more likely that these genotypes have a polymorphism, which

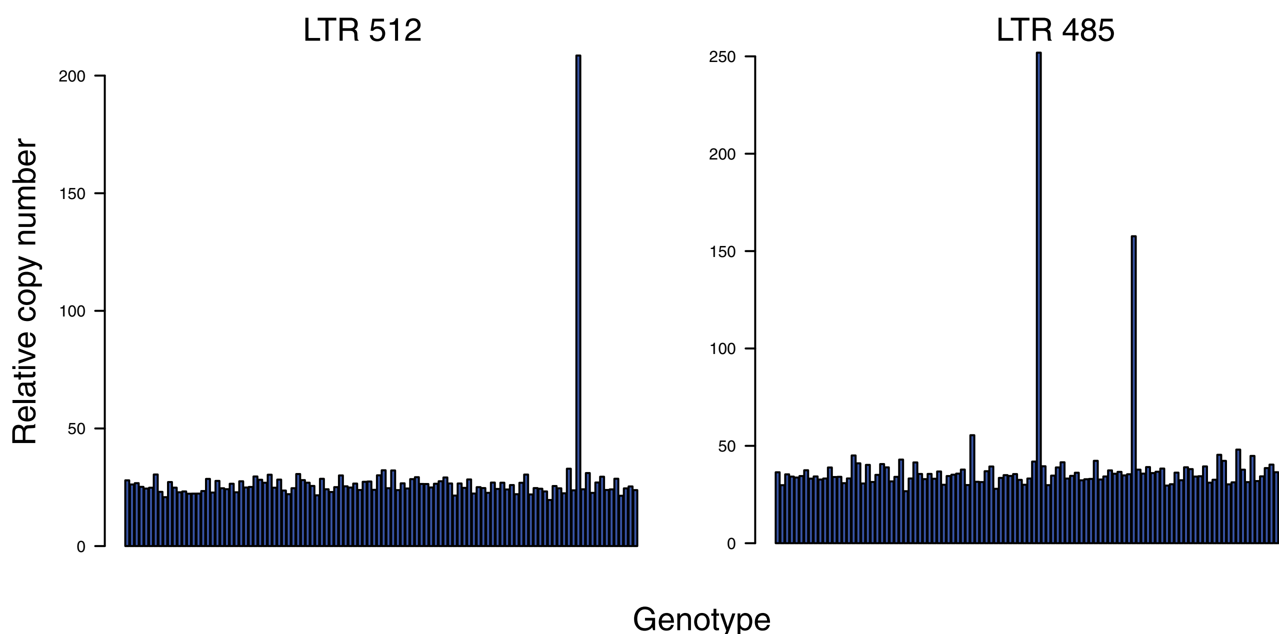


Fig. 5.—An example of genotypes with an accumulation of TEs. In both panels, the population average is 20–30, whereas individual genotypes have in excess of 150 copies.

fails to repress more than one type of TE, rather than that they preferentially inherited multiple active copies. We cannot at this time directly look for polymorphisms in repressive genes or complexes. Currently, we are unable to establish clear homologs of the *D. melanogaster* genes known to affect piRNA silencing in *D. serrata*, but as the *D. serrata* assembly improves this may be possible. In addition, the methods developed recently by Zhang and Kelleher (2019) to measure differences in piRNA cluster integration using small RNA libraries show promise for determining whether we can detect polymorphisms in these individual genotypes for repressive alleles.

However, the fact that the TEs that are proliferating do not appear to be a unique population suggests that there is interaction between potentially active TEs and genetic background—not all TEs are potentially active in all potentially permissive backgrounds. This suggests that the transposition rate of TEs in natural populations will be complex, depending upon differences in the inherited TE population and variation in the host genome. There is already a lot of evidence that there are multiple pathways and factors that control transposition in *Drosophila*. For example, in *D. melanogaster* strain *iso-1*, the piRNA pathway produces *hobo* and *I*-element-specific piRNAs, yet there is a high level of *hobo* and *I*-element transposition (Zakharenko et al. 2007; Shpiz et al. 2014). In *D. simulans*, there are large amounts of variation in piRNA pathway genes (Fablet et al. 2014). Therefore, there is abundant opportunity for variation in the host ability to suppress a TE and the ability of the TE to transpose.

Since the discovery of the piRNA repression system for TEs, the lifecycle of a TE in a host has been envisioned as three steps. First, the TE invades a novel population or species and amplifies unencumbered. TE proliferation is then slowed by segregating insertions in piRNA clusters, and finally inactivated by fixation of piRNA cluster insertions (Kofler 2019). However, clearly bursts, or activity, continues at some level within the population as many of the potentially active TEs in *D. serrata* have a high SFS. This indicates that the TEs have been in the population long enough to accumulate SNPs, potentially including copies with different SNPs continuing to proliferate in the population. It is true that suppression by piRNA cluster insertion may be unstable, but exactly why that is or how important it is for TE survival is not clear.

The accumulation of TEs in laboratory lines should be associated with fitness declines, and be eliminated by selection (Nuzhdin et al. 1997). However, accumulation of TE insertions in individual genotypes, or overall, in genotypes kept in small mass cultures appears to be the rule rather than the exception (Pasyukova 2004; Rahman et al. 2015; Signor 2020). Muller's (1932, 1964) rachet may be responsible for the accumulation of insertions, even if they are deleterious. As more work is done the tempo and mode of TE transposition it will be interesting to see the generality of these conclusions outside of *Drosophila*. What is clear is that TEs are important sources of spontaneous mutations in *Drosophila*, and that in laboratory lines, over time, they may make up a large fraction of the total number of mutations in particular genotypes.

Materials and Methods

Fly Lines and Data

One hundred and ten genotypes of *D. serrata* were collected from a wild population in Brisbane Australia in 2011 and inbred for 20 generations (Reddiex et al. 2018). The libraries were sequenced using 100 bp paired-end reads on an Illumina Hi-seq 2000. The raw reads were downloaded from NCBI SRA PRJNA410238. One hundred and two genotypes were used for analysis. Four genotypes were excluded based on unusually high relatedness, as described in Reddiex et al. (2018), whereas the remaining four genotypes were excluded based on library quality issues.

Classification of TEs

TEs are a diverse group, and the taxonomy of TEs is contentious and still developing (Wicker et al. 2007; Kapitonov and Jurka 2008; Platt et al. 2016). Here, we will rely only on broad classifications in Class I and Class II elements, including Helitrons and MITEs. Class I elements are retrotransposons that use an RNA intermediate in their "copy and paste" transposition. Class I can be divided into LTR and those that lack LTRs (SINEs and LINEs) (Okada et al. 1997; Havecker et al. 2004; Wicker et al. 2007; Kramerov and Vassetzky 2011, 2019). However, within Class I, we will only focus on LTR elements, as benchmarking of software designed to detect non-LTR elements is unreliable (Ou et al. 2019). Within the Class I LTR elements, there are three major superfamilies—copia, gypsy, and Bel/Pao—which have distinct terminal sequences (Marlor et al. 1986). Class II elements are known as DNA transposons and use DNA intermediates in a "cut and paste" mechanism of transposition (McClintock 1984). Among the Class II elements are also nonautonomous small DNA transposons such as MITEs (Fattash et al. 2013; Makałowski et al. 2019). These lack coding potential and rely on other autonomous DNA transposons for transposition. Lastly, *Helitron* transposons have a different mechanism of transposition from other DNA transposons. This is referred to as a rolling circle, which frequently captures nearby genes or portions of them in the process (Kapitonov and Jurka 2001, 2007).

Annotating TEs in *D. serrata*

The *D. serrata* 1.0 assembly available from the Chenoweth lab was used for genomic mapping and TE identification (http://www.chenowethlab.org/resources.html) (Allen et al. 2017). The TE library was constructed using EDTA (Xu and Wang 2007; Ellinghaus et al. 2008; Xiong et al. 2014; Ou and Jiang 2018, 2019; Ou et al. 2019; Shi and Liang 2019; Zhang et al. 2019). This pipeline is intended to create a high-quality nonredundant TE library based on a reference genome (supplementary file 1, Supplementary Material online).

Mapping

Reads from the *D. serrata* reference panel were mapped to the genome and the TE library using bwa mem version 0.7.15 (Li 2015). Bam files were sorted and indexed with samtools v.1.9 and optical duplicates were removed using picard MarkDuplicates v.2.25.2 (http://picard.sourceforge.net) (Li et al. 2009; McKenna et al. 2010). Reads with a mapping quality of below 15 were removed (this removes reads which map equally well to more than one location).

Relationship to TEs in the EMBL TE Library

The TE library from *D. serrata* was compared with TEs from the EMBL consensus sequence library using the Dfam database and hmmer similarity search (hmmer.org, Eddy 2008; Hubley et al. 2016). Hits were required to have a bit score of greater than 350. An hmmer bit score is the log of the ratio of the sequence's probability according to the homology hypothesis over the null model of nonhomology (hmmer.org, Eddy 2008). Multiple hits to the same TE were considered as a single hit, and if more than one EMBL TE was listed the best bit score was retained. In general, there were no TEs from the *D. serrata* library that had similar bit scores between different EMBL TEs.

Relationship between TEs Annotated by EDTA

Potentially related TE families from the EDTA library were identified using NCBI BlastN 2.8, with the minimum criteria being an alignment of greater than 400 bp for LTR elements and DNA transposons, and 200 bp for MITEs (Camacho et al. 2009). The sequences were aligned and oriented using the R package DECIPHER (Wright 2016). The fasta alignments were converted to nexus format, and indels were coded as binary characters, using the perl script 2matrix (Salinas and Little 2014). Trees were made if there were four or more related TEs using MrBayes 3.2.7 (Ronquist et al. 2012). The trees were built using a GTR substitution model and gamma distributed rate variation across sites. The Markov chain Monte Carlo chains were run until the standard deviation of split frequencies was below 0.01. The consensus trees were generated using sumt conformat = simple. The resulting trees were displayed with the R package ape (Paradis et al. 2004).

Relative Copy Number Estimation

Using read coverage to determine relative copy number has been compared with other methods and is neither permissive nor conservative (Srivastav and Kelleher 2017). Read coverage is preferable in this study to methods that rely on split read or split pair mapping, as decent accuracy for those methods requires at least 40× coverage, whereas some split pair methods require more than 90× coverage (Kofler et al. 2016; Vendrell-Mir et al. 2019). Further, we are interested only in relative copy number rather than the precise insertion points

of TEs, which are difficult to infer within heterochromatin. TE copy number was estimated using the average counts of reads mapping to the TE sequences and the genome with bedtools counts (v. 2.3, Quinlan and Hall 2010; Hill et al. 2015). Then, relative copy number of the TEs could be normalized using the average counts from a 7 MB contig from *D. serrata*, which corresponds to a portion of *D. melanogaster* 3L. This is one of the largest contigs in the *D. serrata* assembly. We calculated mean and variance for relative copy number of each TE family as well as the coefficient of variation to more accurately compare variation between TEs with different means. Many TEs have internal deletions or are present in fragments; therefore, this estimation of relative copy number can be thought of more generally as the overall genomic occupancy of each TE.

Comparison with Other dnapipeTE

Among the many TE-related software dnapipeTE (v.1.3) has the most similar overall detection goals to this study (Goubert et al. 2015). In dnapipeTE, trinity is used to produce contigs that represent all alternative contigs of all TEs (v.2.5.1, Grabherr et al. 2011). Then, these trinity contigs are annotated using RepeatMasker and our custom repeat library produced by EDTA (Smit 2013; Ou 2019). RepeatMasker (v 4.0.05) parameters are default, and only the best NCBI BLAST hit is kept. Then the reads are mapped back to this library of annotated contigs, and the number of aligned bases is reported. We ran dnapipeTE on a subset of 11 individuals from our data set. We aggregated the estimates of aligned bases per TE in R, such that if multiple contigs are reported for a TE, one final value would remain. We then normalized each estimate of aligned bases by the total number of bases aligned to TEs to gain a proportion, as we were most interested in comparing relative estimates of TE abundance. dnapipeTE does not include low copy number elements as they do not qualify as repetitive, therefore from our relative copy number estimates we excluded anything without an estimate in dnapipeTE and a copy number of less than four.

Outliers in Individual Genotypes

Outliers for TE relative copy number were identified as three times the third quartile of copy number in R.

SNPs and Summary Statistics

We called SNPs using freebayes v.1.0 (Garrison and Marth 2012). TEs with higher relative copy number will not have SNPs that are heterozygous or homozygous when all reads from multiple copies are pooled, as they are here. To estimate SNP frequencies for multicopy TEs, we instead emulated a pooled sample with freebayes, and relative copy number was allowed to vary for each individual for each TE using the –cnv-map option. The minimum support for an alternate allele was five reads. This allowed for the estimation of SNP frequency in the whole population of TEs within an individual. All of the following calculations were performed in RStudio v.1.0.143. To create summary statistics to more easily understand variation in SNP frequency, we calculated the number of nonreference alleles for each SNP compared with the total relative copy number of the SNP, for each individual. This was then averaged across individuals to create a form of SFS, though one in which the relative copy number varies. SFS is essentially the frequency of the frequency of nonreference alleles. Thus you can visualize whether SNPs are more commonly frequent or rare within the TE family. This is useful because if the SFS is low for all SNPs, this could indicate more recent spread in the population. The SFS for individual SNPS was folded in R, replacing any frequency i over 0.5 with 1-i. Folding the SFS means that any SNP with a frequency greater than 0.5 is assumed to actually be the reference SNP, as we cannot determine which SNP is derived or ancestral by comparing to the reference. Variance was also calculated for each SNP, as well as averaged across SNPs. We created histograms to visualize the SFS of each TE. Then the mean frequency of SNPs in each TE was compared with the number of SNPs, the average relative copy number, and the variance in SFS to determine which TEs might be actively proliferating in the population.

Supplementary Material

Supplementary data are available at https://github.com/signor-molevol/serrata_transposable.

Acknowledgments

S.S. would like to thank C. & S. Emery for insightful commentary on the manuscript. This work was supported by the National Science Foundation Established Program to Stimulate Competitive Research (NSF-EPSCoR-1826834).

Author Contributions

S.S. conceived of the study, performed bioinformatics, and drafted portions of the manuscript. Z.T. performed bioinformatics, interpreted the data, and contributed to the manuscript draft.

Data Availability

The data that support the findings of this study are openly available in the NCBI SRA at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA419238, reference number PRJNA419238.

Literature Cited

Allen SL, Delaney EK, Kopp A, Chenoweth SF. 2017. Single-molecule sequencing of the *Drosophila* serrata genome. G3 (Bethesda) 7:781–788.

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. Science 309(5735):764–767.

Bergman Casey M., Bensasson Douda. 2007. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in Drosophila melanogaster. Proceedings of the National Academy of Sciences 104 (27) 11340–11345; DOI: 10.1073/pnas.0702552104

Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biology. 2006;7(11):R112–21. PMCID: PMC1794594

Biémont C, Aouar A, Arnault C. 1987. Genome reshuffling of the *copia* element in an inbred line of *Drosophila* melanogaster. Nature 329(6141):742–744.

Biémont C, Cizeron G. 1999. Distribution of transposable elements in *Drosophila* species. Genetica 105:43–62.

Biémont C, Ronsseray S, Anxolabéhère D, Izaabel H, Gautier C. 1990. Localization of *P elements*, copy number regulation, and cytotype determination in *Drosophila melanogaster*. Genet Res. 56(1):3–14.

Birchler JA, Hiebert JC, Rabinow L. 1989. Interaction of the *mottler of white* with transposable element alleles at the *white* locus in *Drosophila melanogaster*. Genes Dev. 3:73–84.

Blows MW. 1993. The genetics of central and marginal populations of *Drosophila serrata* II. Hybrid breakdown in fitness components as a correlated response to selection for desiccation resistance. Evolution 47:1271–1285.

Bourgeois Y, Boissinot S. 2019. On the population dynamics of junk: a review on the population genomics of transposable elements. Genes 10(6):419–423.

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

Brookfield JF. 1991. Models of repression of transposition in P-M hybrid dysgenesis by P cytotype and by zygotically encoded repressor proteins. Genetics 128(2):471–486.

Brookfield JFY. 1996. Models of the spread of non-autonomous selfish transposable elements when transposition and fitness are coupled. Genet Res. 67(3):199–209.

Brunet F, Godin F, David JR, Capy P. 1994. The *mariner* transposable element in the Drosophilidae family. Heredity 73(4):377–385.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10(1):421.

Cosby RL, et al. 2020. Recurrent evolution of vertebrate transcription factors by transposase capture. bioRxiv 15:e1008160.

Csink AK, Linsk R, Birchler JA. 1994. *Mosaic suppressor*, a gene in *Drosophila* that modifies retrotransposon expression and interacts with *zeste*. Genetics 136:573–583.

Da Lage JL, et al. 2007. A phylogeny of Drosophilidae using the *Amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. J Zoolog Syst. 45(1):47–63.

Daborn PJ, et al. 2002. A single *p450* allele associated with insecticide resistance in *Drosophila*. Science 297(5590):2253–2256.

Daniels SB, Chovnick A, Boussy IA. 1990. Distribution of *hobo* transposable elements in the genus *Drosophila*. Mol Biol Evol. 7:589–606.

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124(2):339–355.

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 8:422–410.

Doolittle W, Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603. https://doi.org/10.1038/284601a0

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9(1):18.

Esnault C, Lee M, Ham C, Levin HL. 2019. Transposable element insertions in fission yeast drive adaptation to environmental stress. Genome Res. 29:85–95.

Fablet M, Akkouche A, Braman V, Vieira C. 2014. Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis. Gene 537:149–153.

Fattash I, et al. 2013. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. Genome 56:475–486.

Garcia Guerreiro MP. 2012. What makes transposable elements move in the *Drosophila* genome? Heredity 108(5):461–468.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN].

Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. 2014. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO reports. 15(4):411–8.

Goubert Clément, et al. 2015. De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). Genome Biol Evol. 7(4):1192–1205, https://doi.org/10.1093/gbe/evv050

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Hallas R, Schiffer M, Hoffmann A. 2002. Clinal variation in *Drosophila serrata* for stress resistance and body size. Genet Res. 79:141–148.

Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. Genome Biol. 5:225–226.

Hill T, Schlötterer C, Betancourt A. 2015. Hybrid dysgenesis in *Drosophila simulans* due to a rapid global invasion of the *P*- element. PLoS Genet. 12(5):e1006058.

Hoffmann AA, Shirriffs J. 2002. Geographic variation for wing shape in *Drosophila serrata*. Evolution 56:1068–1073.

Hubley R, et al. 2016. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44:D81–D89.

Jenkins NL, Hoffmann AA. 1999. Limits to the southern border of *Drosophila serrata*: cold resistance, heritable variation, and trade-offs. Evolution 53:1823–1834.

Jenkins N.L, Hoffmann AA. (2001), Distribution of Drosophila serrata Malloch (Diptera: Drosophilidae) in Australia with particular reference to the southern border. Australian Journal of Entomology, 40:41–48. https://doi.org/10.1046/j.1440-6055.2001.00197.x

Josse T, et al. 2007. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLoS Genet. 3(9):e158.

Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA. 98:8714–8719.

Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A. 100(11):6569-74.

Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 23:521–529.

Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet. 9:411–414.

Kaplan N, Darden T, Langley CH. 1985. Evolution and extinction of transposable elements in Mendelian populations. Genetics 109:459–480.

Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 80:1655–1659.

Kim A, et al. 1994. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc Natl Acad Sci USA. 91:1285–1289.

Kofler R. 2019. Dynamics of transposable element invasions with piRNA clusters. Mol Biol Evol. 36:1457–1472.

Kofler R, Gomez-Sanchez D, Schlotterer C. 2016. PoPoolationTE2: comparative population genomics of transposable elements using Pool-seq. Mol Biol Evol. 33:2759–2764.

Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. 2015. The recent invasion of natural *Drosophila simulans* populations by the *P*-element. Proc Natl Acad Sci USA. 112:6659–6663.

Kramerov DA, Vassetzky NS. 2011. SINEs. Wiley Interdiscip Rev RNA. 2:772–786.

Kramerov DA, Vassetzky NS. 2019. Origin and evolution of SINEs in eukaryotic genomes. Heredity 107:487–495.

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

Lee YCG, Langley CH. 2012. Long-term and short-term evolutionary impacts of transposable elements on *Drosophila*. Genetics 192:1411–1432.

Lemeunier F, David JR, Tsacas L, Ashburner M. 1986. The *melanogaster* species group. In: Ashburner M, Carson HL, Thompson JN, editors. The genetics and biology of Drosophila. Vol. 3. London: Academic. p. 147–256.

Levis R, O'Hare K, Rubin GM. 1984. Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. Cell 38(2):471–481.

Li H, et al. 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Li, H. 2015. Aligning sequence reads, clone sequences and assemble contigs with BWA-MEM arXiv 1–3.

Liefting M, Hoffmann AA, Ellers J. 2009. Plasticity versus environmental canalization: population differences in thermal responses along a latitudinal gradient in *Drosophila serrata*. Evolution 63:1954–1963.

Lim JK. 1988. Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 85:9153–9157.

Mackay TFC, et al. 2012. The *Drosophila* melanogaster genetic reference panel. Nature 482:173–178.

Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. 2011. Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication. PLoS Genet. 7:e1002337.

Makałowski W, Gotea V, Pande A, Makałowska I. 2019. Transposable elements: classification, identification, and their use as a tool for comparative genomics. Meth Mol Biol. 1910:177–207.

Malone CD, et al. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. Cell 137:522–535.

Malone CD, Hannon GJ. Molecular evolution of piRNA and transposon control pathways in Drosophila. Cold Spring Harbor Symposia on Quantitative Biology. 2010;74(0):225–34.

Marlor RL, Parkhurst SM, Corces VG. 1986. The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol Cell Biol. 6:1129–1134.

Maruyama K, Hartl DL. 1991. Evolution of the transposable element *mariner* in *Drosophila* species. Genetics 128(2):319–329.

Mateo L, Ullastres A, González J. 2014. A transposable element insertion confers xenobiotic resistance in *Drosophila*. PLoS Genet. 10:e1004560.

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA. 36:344–355.

McClintock B. 1984. The significance of responses of the genome to challenge. Science 226:792–801.

McGinnis W, Shermoen AW, Beckendorf SK. 1983. A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. Cell 34:75–84.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Muller HJ. 1932. Some genetic aspects of sex. Am Nat. 66(703):118–138.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 106:2–9.

Niu X-M, et al. 2019. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. Proc Natl Acad Sci USA. 116:6908–6913.

Nouaud D, Anxolabéhère D. 1997. *P* element domestication: a stationary truncated *P* element may encode a 66-kDa repressor-like protein in the *Drosophila montium* species subgroup. Mol Biol Evol. 14:1132–1144.

Nouaud D, Boëda B, Levy L, Anxolabéhère D. 1999. A *P* element has induced intron formation in *Drosophila*. Mol Biol Evol. 16(11):1503–1510.

Nuzhdin SV. 2000. Sure facts, speculations, and open questions about the evolution of transposable element copy number. In: McDonald JF, editor. Transposable elements and genome evolution. Dordrecht: Springer Netherlands. p. 129–137.

Nuzhdin SV, Pasyukova EG, Mackay TF. 1997. Accumulation of transposable elements in laboratory lines of *Drosophila* melanogaster. Genetica 100:167–175.

Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3' sequences: a review. Gene 205(1–2):229–243.

Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 176:1410–1422.

Ou S, Jiang N. 2019. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. Mob DNA. 10(1):48.

Ou S, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20(1):275.

Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nature Reviews Genetics. 2019;20(2):89–108.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.

Pasyukova EG. 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. J Heredity. 95:284–290.

Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci USA. 114:4721–4726.

Pélisson A, et al. 1994. *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. EMBO J. 13:4401–4411.

Piegu B, et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. 16:1262–1269.

Platt RN II, Blanco-Berdugo L, Ray DA. 2016. Accurate transposable element annotation is vital when analyzing new genome assemblies. Genome Biol Evol. 8:403–410.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rahman R, et al. 2015. Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. Nucleic Acids Res. 43(22):10655–10672.

Reddiex AJ, Allen SL, Chenoweth SF. 2018. A genomic reference panel for *Drosophila serrata*. G3 (Bethesda) 8:1335–1346.

Romero-Soriano V, Garcia Guerreiro MP. 2016. Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids. PLoS One. 11:e0147903.

Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61:539–542.

Ronsseray S, Lehmann M, Anxolabéhère D. 1991. The maternally inherited regulation of *P* elements in *Drosophila melanogaster* can be elicited by two *P* copies at cytological site 1A on the X chromosome. Genetics 129:501–512.

Salinas NR, Little DP. 2014. 2matrix: a utility for indel coding and phylogenetic matrix concatenation. Appl Plant Sci. 2(1):1300083.

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

Schrader L, et al. 2019. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun. 5:5495.

Shi J, Liang C. 2019. Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection. Plant Physiol. 180(4):1803–1815.

Shpiz S, Ryazansky S, Olovnikov I, Abramov Y, Kalmykova A. 2014. Euchromatic transposon insertions trigger production of novel pi- and endo-siRNAs at the target sites in the *Drosophila* Germline. PLoS Genet. 10:e1004138.

Signor S. 2020. Transposable elements in individual genotypes of *Drosophila simulans*. Ecol Evol. 130:499–411.

Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. 1994. An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev. 8:2046–2057.

Srivastav SP, Kelleher ES. 2017. Paternal induction of hybrid dysgenesis in *Drosophila* melanogaster is weakly correlated with both *P*-element and *hobo* element dosage. G3 (Bethesda) 7:1487–1497.

Tamura Koichiro, Subramanian Sankar, Kumar Sudhir. 2004. Temporal Patterns of Fruit Fly (Drosophila) Evolution Revealed by Mutation Clocks, Molecular Biology and Evolution, 21(1):36–44, https://doi.org/10.1093/molbev/msg236

Vendrell-Mir P, et al. (2019). A benchmark of transposon insertion detection tools using real data. Mobile DNA 10, 53. https://doi.org/10.1186/s13100-019-0197-9

Vieira C, Lepetit D, Dumont S, Biémont C. 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. Mol Biol Evol. 16:1251–1255.

Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.

Wilson TG. 1993. Transposable elements as initiators of insecticide resistance. J Econ Ent. 86(3):645–651.

Wright E. 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. R J. 8(1):352–359.

Xiong W, He L, Lai J, Dooner HK, Du C. 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci USA. 111: 10263–10268.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35:W265–W268.

Yamanaka S, Siomi MC, Siomi H. 2014. piRNA clusters and open chromatin structure. Mobile DNA 5, 22. https://doi.org/10.1186/1759-8753-5-22

Yang H-P, Hung T-L, You T-L, Yang T-H. 2006. Genome-wide comparative analysis of the highly abundant transposable Element DINE-1 suggests a recent transpositional burst in *Drosophila yakuba*. Genetics 173:189–196.

Yassin A. 2013. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the postgenomic era. Syst Entomol. 38(2):349–364.

Zakharenko LP, Kovalenko LV, Mai S. 2007. Fluorescence *in situ* hybridization analysis of *hobo*, *mdg1* and *412* transposable elements reveals genomic instability following the *Drosophila melanogaster* genome sequencing. Heredity 99:525–530.

Zanni V, et al. 2013. Distribution, evolution, and diversity of retrotransposons at the *flamenco* locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci USA. 110: 19842–19847.

Zhang R-G, Wang Z-X, Ou S, Li G-Y. 2019. TEsorter: lineage-level classification of transposable elements using conserved protein domains. bioRxiv 800177; doi: https://doi.org/10.1101/800177.

Zhang S, Kelleher ES. 2019. Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations. bioRxiv 5:252–240.

Associate editor: Sarah Schaack