The transposable elements of the *Drosophila serrata* reference panel

Zachary Tiedeman¹ and Sarah Signor^{*1}

*Corresponding author

sarah.signor@ndsu.edu

¹Department of Biological Sciences, North Dakota State University

Abstract: Transposable elements are an important element of the complex genomic ecosystem, proving to be both adaptive and deleterious - repressed by the piRNA system and fixed by selection. Transposable element insertion also appears to be bursty – either due to invasion of new transposable elements that are not yet repressed, de-repression due to instability of organismal defense systems, stress, or genetic variation in hosts. Here, we characterize the transposable element landscape in an important model *Drosophila*, *D. serrata*, and investigate variation in transposable element copy number between genotypes and in the population at large. We find that a subset of transposable elements are clearly related to elements annotated in *D. melanogaster* and *D. simulans*, suggesting they spread between species more recently than other transposable elements. We also find that transposable elements do proliferate in particular genotypes, and that often if an individual is host to a proliferating transposable element, it is host to more than one proliferating transposable element. In addition, if a transposable element is active in a genotype, it is often active in more than one genotype. This suggests that there is an interaction between the host and the transposable element, such as a permissive genetic background and the presence of potentially active transposable element copies. In natural populations an active transposable element and a permissive background would not be held in association as in inbred lines, suggesting the magnitude of the burst would be much lower. Yet many of the inbred lines have actively proliferating transposable elements suggesting this is an important mechanism by which transposable elements maintain themselves in populations.

Keywords: Transposable elements, *Drosophila serrata*, copy number, inbred lines

Introduction

Transposable elements (TEs) are short 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 non-hybridizing 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 *D. melanogaster* from *D. willistoni* in the 1950's, and its subsequent spread into *D. simulans* around 2010 (Daniels et al. 1990b; 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 which are generally depleted of variation, facilitating adaptation to novel environments (Schrader et al. 2019; Niu 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.

Despite the evidence for an adaptive role for TEs, most TE insertions are thought to be deleterious, and the host has a dedicated defense mechanism termed the PIWI-interacting (piRNA) system. piRNAs bind to PIWI-clade proteins, such as *Argonaute 3* 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). These piRNA clusters are large, and 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). These piRNA clusters are preferentially located in heterochromatic regions and usually have low recombination rates (Brennecke et al. 2007). This reduces the efficacy of purifying selection and may serve as ‘safe harbors’ for TEs to accumulate and develop into piRNA clusters (Brennecke et al. 2007; Kofler 2019; Zhang and Kelleher 2019). piRNA evolution is thought to be rapid enough that adaptation to a novel TE could occur within the lifetime of an individual (Khurana et al. 2011). The transposition rate of TEs is also controlled by other mechanisms, including regulation of promotor activity, chromatin structure, and splicing (Guerreiro 2019). In some cases the mechanism is unknown, such as the accumulation of *copia* in the genomes of inbred *D. melanogaster*, suggesting that there is still more to know about the regulation of TE copy number (Pasyukova 2004).

This apparent contradiction, between the existence of a dedicated repression machinery, and an apparent important role for TEs in adaptation, also complicates inferences about the

tempo and mode of TE transposition. 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). Or, it was not active TEs which 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 proliferated at approximately the rate that they were removed by selection. 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; Romero-Soriano and Garcia Guerreiro 2016; Guerreiro 2019). 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. Bursts of transposition would be expected upon invasion of a new TE, 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.

Recently an inbred panel of 110 genotypes was created for *D. serrata*, a member of the *montium* subgroup (Reddiex et al. 2018). The *montium* group contains 98 species and represents a significant fraction of known *Drosophila* species (Lemeunier et al. 1986; Reddiex et al. 2018). The *D. serrata* panel was sampled from a single large population within its endemic distribution in Australia (Reddiex et al. 2018). *D. 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). While 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 et al. 1990a). *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 will characterize the TE landscape in the *Drosophila serrata* Genetic Reference panel. This will have two goals: 1) To understand the TE content of *D. serrata* and its relationship to existing TE annotations 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 evolution in *D. serrata*, as well as provide another investigation into the proliferation of TEs in individual genetic backgrounds.

Methods

Fly lines and data

110 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. 104 genotypes were used for analysis. 4 genotypes were excluded based on unusually high relatedness, as described in (Reddiex et al. 2018), while the remaining 2 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 miniature inverted-repeat TEs (MITES). Class I elements are retrotransposons which use an RNA intermediate in their ‘copy and paste’ transposition. Class I can be divided into long terminal repeat (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, here we will only focus on LTRs in Class I, as benchmarking of software designed to detect non-LTRs is unreliable (Ou et al. 2019). Within the LTRs, there are two major superfamilies – *copia* and *gypsy* – which have distinct terminal sequences (Marlor et al. 1986). Class II elements are known as DNA transposons, or terminal inverted repeat transposons (TIR), and use DNA intermediates in a ‘cut and paste’ mechanism of transposition (McClintock 1984). Among the TIRs are also non-autonomous small DNA transposons such as miniature inverted-repeat TEs (MITES) (Fattash et al. 2013; Makiłowski et al. 2019). These can belong to any of the described TIR superfamilies, but they lack coding potential and rely on other autonomous DNA transposons for transposition. Lastly, the *Helitron* TIR was discovered in 2001 and has a different mechanism of transposition, referred to as a rolling circle, which frequently captures nearby genes or portions of them in the process (Kapitonov and Jurka 2001; Kapitonov and Jurka 2007).

Mapping and copy number estimation

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 the Extensive de-novo TE Annotator pipeline (EDTA) (Ou et al. 2019). This pipeline is intended to create a high quality non-redundant TE library based off of a

reference genome. Reads from the *D. serrata* reference panel were mapped to the genome and the TE library using bwa mem version 0.7.15 (Figure 1; Li 2015). Bam files were sorted and indexed with samtools v.1.9 and optical duplicates were removed using picard MarkDuplicates (<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). Using read coverage to determine copy number has been compared to other methods and is neither permissive nor conservative (Srivastav and Kelleher 2017). TE copy number was estimated using the average counts of reads mapping to the TE sequences and the genome with bedtools counts (Quinlan and Hall 2010; Hill et al. 2015). Then, 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.

SNPs and summary statistics

We called SNPs within the TEs using GATK Haplotypecaller (McKenna et al. 2010). SNPs were not filtered for missing calls given that not all individuals will share insertions. However, they were filtered for a coverage of at least 4 reads to be called in an individual. The site frequency spectrum (SFS) of SNPs in the TEs was estimated with VCFtools as the frequency of each SNP in the population, and then the frequency of the SNP frequencies was estimated in R (Danecek et al. 2011). The SFS was then folded in R, replacing any frequency i over 0.5 with $1-i$. This was done because we could not determine the derived allele.

Relationship to TEs in the EMBL TE library

The TE library from *D. serrata* was compared to TEs from the EMBL library using DFAM (Hubley et al. 2016). Hits were required to have a bit score of greater than 350. 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 TEs from the EDTA library were identified using ncbi blastn 2.8, with the minimum criteria being an alignment of greater than 400 bp for LTRs and TIRs, 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 .01. The consensus trees were generated using sumt conformat=simple. The resulting trees were displayed with the R package ape (Paradis et al. 2004).

Results

EDTA

197 EDTA identified 676 TEs in the *D. serrata* reference genome. The sequences of these TEs are
198 available at https://github.com/signor-molevol/serrata_transposable. The classification of the TEs
199 into superfamilies is broadly correct, and in many cases there is no clear relationship to an
200 existing TE. However, some errors are evident, for example, element 444 is classified as *copia*,
201 but aligns quite well with the 297 element in *D. melanogaster*, which is a member of the 17.6
202 clade/*gypsy* superfamily. In addition, some unknown elements such as 69 align well with

Figure 1: A. Mean and variance of copy number as well as folded SFS in transposable elements of Australian *D. serrata*. On the top, the majority of identified transposable elements have low mean copy number and variance within the population. The top three most variable elements are excluded from all graphs so as not to compress the y-axis. B. The 52 most variable transposable elements (top) with variance greater than 50 versus all of the transposable elements (bottom). Overall the number of LTRs represent fewer of the total identified TEs, however of those that are highly variable they are equally likely to be LTRs or TIRs.

existing *D. melanogaster*
annotations, in this case 17.6. In
all 6 elements 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 MITEs,
Helitrons, and TIRs are
distinguishable. This may be due
to deletion of canonical sequences,
nested insertions, or other
ambiguities of TEs.

Population frequency of TEs

An average of 17% of reads from individual *D. serrata* lines mapped to TE sequences. The average number of TEs 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 (Supplemental File 1). This sequence is classified as an LTR, though it does not share sequence similarity to other well characterized LTRs. This element shares a 36 bp segment with *D. melanogaster INE-1*, and may be misclassified given that *INE-1*s are generally very abundant. The next closest in copy number is a TIR with 541 copies, thus this is a significant outlier. 6 TEs identified in the reference are likely not present in this population. 2 of these are present as partial copies in a subset of individuals. Overall among the elements identified by EDTA approximately twice as many are TIRs compared to LTRs (Figure 1). However, the majority of the identified TEs have low copy number and variance. 390 of the identified elements have an average copy number of less than 3, and all but 2 of those have a variance of less than 1 (Figure 1, Supplemental file 1; the other two have variances of 3 and 4). Of those remaining, 148 have a variance of 3 or less (Figure 1). Therefore the vast majority of TEs in this

A

Element	Average	Variance	SFS
217 TIR	58.38	226.59	NA
368 LTR	36.63	41.26	NA
389 TIR	17.87	6.42	0.03
411 LTR Gypsy	102.80	155.98	NA
436 TIR	30.83	57.81	0.05
459 TIR	28.33	49.28	0.04
610 LTR Gypsy	5.89	4.50	NA
614 MITE	4.05	2.24	0.02
624 TIR	7.18	5.08	NA
638 LTR Gypsy	107.17	176.07	NA

B

Figure 2: A. A list of transposable elements with variance of at least 2 and either no SNPs or SFS of 0.05 or less. B. Variation in copy number in the population for two transposable elements (top) and the folded SFS (bottom).

population vary little in copy number (Figure 1). However, among those that do vary considerably, 52 elements have a variance in the population greater than 50. These represent a very different subset of TEs than those identified overall – an approximately equal number are LTRs and TIRs (Figure 1). This suggests that LTRs are more active in the population, which is consistent with other work on transposable elements that found that LTR insertions were generally of more recent origin than TIRs (Kofler et al. 2015).

Folded SFS

Overall ~6% of SNPs in TEs have a frequency of higher than 60% in this population. Rather than determine the ancestral state by adding an outgroup, we chose to fold the allele frequency spectrum. Overall the folded SFS is low (average of 0.13), however SNPs that are not in TEs also have excess of low frequency variants genome-wide according to genome-wide Tajima's *D* (Reddiex et al. 2018). There is not a clear relationship between the folded SFS and mean/variance (Figure 1A). A folded SFS of greater than 0.2 is associated with lower copy number and variance overall (Figure 1A). 16 elements have a folded SFS of 0.3 or higher, and the majority of these have low variance in copy number (2 or less) suggesting that they are not active and have been diverging. There are two obvious types of TEs that have likely been spreading recently in the population as a whole – those with no SNPs or low SFS, and with high variance and/or copy number. 5 TEs have no SNPs and therefore no folded SFS can be calculated (217, 411, 610, 624, 638, Table 1, Figure 2). This includes 610, a *gypsy* element which aligns to the internal sequence of *Dsim\ninja* suggesting it is distantly related, but more recently moved into *D. serrata* than TEs with no obvious relatives in related species. As with the population of TEs as a whole, most TEs with a low SFS (lower than .05) also have low copy number (less than 3) and low variance (less than 1). Of these 71 elements, 4 are an exception and

have a low SFS, higher copy number, and higher variance. This includes a MITE element and three TIRs. A lack of SNPs, or low SFS, along with higher copy number and variance, suggests that the TEs have been spreading recently in the population (Table 1).

However, TEs may also have been in the population for long enough to accumulate SNPs, but experience bursts of activity. Among the 52 TEs with a variance higher than 50, 17 TEs have an SFS of greater than 0.2 suggesting they have been in the population for longer but have had active transposition of divergent copies (Supplemental Table 1). This group consists of 15 LTRs, one MITE, and one TIR. This could be due to the presence of older copies accumulating SNPs compared to younger active copies, or divergence between different active copies in the population, or both.

Figure 3: An example of genotypes with an accumulation of transposable elements. In both panels the population average is 20-30, while individual genotypes have in excess of 150 copies.

*Outliers in individual
genotypes*

TEs tend to proliferate in
particular inbred genotypes.

Out of 104 genotypes, 73 have
no TEs with a number of
insertions that classify them as

outliers. 12 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 which share outliers for 4 of the 5. Many of these outliers are large, for example for element 5/2 the majority of the population has 20-30 copies, while a single individual has > 200 (Figure 3).

Relationship to existing TE annotations

Figure 4: The classification of transposable elements into clades which could be tied to annotated *D. melanogaster* elements. The two *Helitron* elements potentially related to those from *Heliconius* are not included.

123 of the 676 identified elements have a well-supported relationship to existing DFAM TE annotations (Figure 4). This includes, for example, 27 elements that are related to the *D. melanogaster* *Max-Element* and 10 elements that are related to the *D. simulans* *ninja* element. One of these is also among the most variable TEs (variance greater than

50), and is most closely related to the *Circe* element (*Osvaldo* family). These are likely to be TEs that moved between species more recently, and they are

almost exclusively LTRs. The exception being two TIRs from the *hobo* family, one *Helitron* from *D. melanogaster*, and two *Helitrons* most closely related to elements from *Heliconius*. No evidence of *P*-elements were 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 transposons. The overall phylogeny of the TEs is not what we wish to emphasize here, as the structure of TE classification changes frequently (for example 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

310 microinjection (Song et al. 1994; Kim et al. 1994). This makes them more likely to spread
 311 between species.
 312

Figure 5: Relatedness between groups of TEs annotated by EDTA. Posterior probabilities of each division is shown, however branch lengths are not meaningful given that TEs do not follow a standard substitution model. This does not represent only the degradation of old copies of TEs, different members of the TE families continue to likely be active, shown in bold.

313 Relationship between TEs annotated by EDTA

314 40 groups of 170 TEs annotated by EDTA are clearly related to one another (Supplemental File
 315 3). For example, 8 TEs (52, 60, 276, 346, 367, 424, 539, 601) share sequence similarity for the
 316 entirety of their length, but are separated by 39 deletions spread across the TE. In the largest
 317 related group of TEs, 23, most of the versions of this TE have low copy number and variance
 318 (Figure 4A, average copy number 2.3, average variance < 1). However, two members of the
 319 group are likely still active and have relatively high copy number and variance (376 and 672,
 320 copy number 27, 79; variance 10, 102). Note that the more active TEs do not group together
 321 (shown in bold), however because TEs cannot be assumed to follow a standard substitution
 322 model the branch lengths are not meaningful (Figure 4A). In another case, 3 members of the
 323 group are more distantly related, while 7 members are more closely related and form 2 clear

groups of origin (Figure 4B). Yet again, those which are active in the population, as evidenced by higher copy number and variance, are not the most closely related (Figure 4B, shown in bold)

The distance between TEs cannot be understood in the same way as genomic DNA, however in some cases the TEs are clearly related by a simple tree of SNPs and indels (as above), and in others it is more complicated. For example, in a group of 4 TEs (237, 225, 358, 468) the long terminal sequence of the TEs is nearly identical and allies them with *Max* elements, however member 468 appears to also carry an unshared insertion of a portion of a 297 element, among other complicated indels. 468 has a copy number of 5, indicating it continued to be active while carrying a portion of a 297 element, though it has low variance between individuals. Interestingly, in another case these relationships appear to describe the origin of 3 MITEs (399, 405, 472) from a parental TIR (660). The parental TIR has a high copy number in the population, with an average of 286 and a variance of more than 6,000. This is not intended to be an exhaustive accounting of relationships between these TEs, 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.

Discussion

There is an abundance of evidence from inbred lines that genotypes can vary considerably in copy number. 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, including polymorphisms in genes such as *Argonaute 3* and variation in the integration of TEs into piRNA clusters (Birchler et al. 1989; Péliisson et al. 1994; Csink 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).

While we cannot measure the likelihood of individual genotypes inheriting multiple active copies of TEs while fellow members of the population inherit none, the fact that multiple TEs are proliferating in individual genotypes 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 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 shows promise for determining whether we can detect polymorphisms in these individual genotypes for repressive alleles.

However, the fact that the TEs which 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 normal *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 is 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 ratchet may be responsible for the accumulation of insertions, even if they are deleterious (Muller 1932; 1964). 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.

Availability of data and materials

All raw data is available at NCBI SRA PRJNA410238.

Funding

This work was supported by the National Science Foundation Established Program to Stimulate Competitive Research (NSF-EPSCoR-1826834).

Competing interests

We declare that we have no competing interests.

Acknowledgements

S.S. would like to thank C. & S. Emery for insightful commentary on the manuscript.

Authors' 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.

Allen, S. L., E. K. Delaney, A. Kopp, and S. F. Chenoweth. 2017. Single-molecule sequencing of the *Drosophila serrata* genome. *G3* 7:781-788.

Aminetzach, Y. T., J. M. Macpherson, and D. A. Petrov. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309:764-767.

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

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

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

Birchler, J. A., J. C. Hiebert, and L. Rabinow. 1989. Interaction of the *mottler of white* with transposable element alleles at the *white* locus in *Drosophila melanogaster*. *Genes Dev.* 3:73-84.

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

431 Bourgeois, Y., and S. Boissinot. 2019. On the population dynamics of junk: A review on the
432 population genomics of transposable elements. *Genes* 10:419–23.

433 Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon.
434 2007. Discrete small RNA-generating loci as master regulators of transposon activity in
435 *Drosophila*. *Cell* 128:1089–1103.

436 Brookfield, J. F. 1991. Models of repression of transposition in P-M hybrid dysgenesis by P
437 cytotype and by zygotically encoded repressor proteins. *Genetics* 128:471–486.

438 Brookfield, J. F. Y. 1996. Models of the spread of non-autonomous selfish transposable elements
439 when transposition and fitness are coupled. *Genetics Research* 67:199–209.

440 Brunet, F., F. Godin, J. R. David, and P. Capy. 1994. The *mariner* transposable element in the
441 *Drosophilidae* family. *Heredity* 73:377–385.

442 Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden.
443 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1.

444 Cosby, R. L., J. Judd, R. Zhang, A. Zhong, N. Garry, E. J. Pritham, and C. Feschotte. 2020.
445 Recurrent evolution of vertebrate transcription factors by transposase capture. *bioRxiv*
446 15:e1008160.

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

449 Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil, S. Jeffers, N. Tijet, T. Perry, D.
450 Heckel, P. Batterham, R. Feyereisen, T. G. Wilson, and R. H. ffrench-Constant. 2002. A single
451 *p450* allele associated with insecticide resistance in *Drosophila*. *Science* 297:2253–2256.

452 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker,
453 G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis
454 Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

455 Daniels, S. B., A. Chovnick, and I. A. Boussy. 1990a. Distribution of *hobo* transposable elements
456 in the genus *Drosophila*. *Mol. Biol. Evol.* 7:589–606.

457 Daniels, S. B., K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, and A. Chovnick. 1990b.
458 Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species.
459 *Genetics* 124:339–355.

460 de Boer, J. G., R. Yazawa, W. S. Davidson, and B. F. Koop. 2007. Bursts and horizontal
461 evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics*
462 8:422–10.

463 Esnault, C., M. Lee, C. Ham, and H. L. Levin. 2019. Transposable element insertions in fission
464 yeast drive adaptation to environmental stress. *Genome Res.* 29:85–95.

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

467 Fattash, I., R. Rooke, A. Wong, C. Hui, T. Luu, P. Bhardwaj, and G. Yang. 2013. Miniature
468 inverted-repeat transposable elements: discovery, distribution, and activity. *Genome* 56:475–486.

469 Garcia Guerreiro, M. P. 2019. What makes transposable elements move in the *Drosophila*
470 genome? *Heredity* 108:461–468.

471 Wickham, H., François, R., Henry L., and Müller, K. (2020). dplyr: A Grammar of Data
472 Manipulation. R package version 0.8.4. <https://CRAN.R-project.org/package=dplyr>
473

474 Havecker, E. R., X. Gao, and D. F. Voytas. 2004. The diversity of LTR retrotransposons.
475 *Genome Biology* 5:225–6.

476 Hill, T., C. Schlötterer, and A. Betancourt. 2015. Hybrid dysgenesis in *Drosophila simulans* due
477 to a rapid global invasion of the *P*-element PLOS Genetics. 12(5): e1006058.
478

479 Hill T, Betancourt AJ. 2018. Extensive exchange of transposable elements in the *Drosophila*
480 *pseudoobscura* group. *Mobile DNA* 9:579–14.
481

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

484 Hubley R., R.D. Finn, J. Clements, S.R. Eddy, T.A. Jones, W. Bao, et al. 2016. The Dfam
485 database of repetitive DNA families. *Nucleic Acids Res.* 44:D81–89.

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

488 Josse, T., L. Teyssset, A.-L. Todeschini, C. M. Sidor, D. Anxolabéhère, and S. Ronsseray. 2007.
489 Telomeric trans-silencing: An epigenetic repression combining RNA silencing and
490 heterochromatin formation. *PLoS Genet.* 3:e158.

491 Kapitonov, V. V., and J. Jurka 2001. Rolling-circle transposons in eukaryotes. *Proc. Nat. Acad.*
492 *Sci. USA.* 98: 8714–8719.

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

495 Kapitonov, V. V., and J. Jurka. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons.
496 *Trends Genet.* 23:521–529.

497 Kaplan, N., T. Darden, and C. H. Langley. 1985. Evolution and extinction of transposable
498 elements in Mendelian populations. *Genetics* 109:459–480.

499 Khurana, J. S., J. Wang, J. Xu, B. S. Koppetsch, T. C. Thomson, A. Nowosielska, C. Li, P. D.
500 Zamore, Z. Weng, and W. E. Theurkauf. 2011. Adaptation to *P* element transposon invasion in
501 *Drosophila melanogaster*. *Cell* 147:1551–1563.

502 Kidwell, M. G. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*.
503 *Proc. Nat. Acad. Sci. USA* 80:1655–1659.

504 Kim, A., C. Terzian, P. Santamaria, A. Pélisson, N. Purd'homme, and A. Bucheton. 1994.
505 Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of
506 *Drosophila melanogaster*. *Proc. Nat. Acad. Sci. USA* 91:1285–1289.

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

509 Kofler, R., T. Hill, V. Nolte, A. J. Betancourt, and C. Schlötterer. 2015. The recent invasion of
510 natural *Drosophila simulans* populations by the *P*-element. *Proc. Nat. Acad. Sci. USA*
511 112:6659–6663.

512 Kramerov, D. A., and N. S. Vassetzky. 2019. Origin and evolution of SINEs in eukaryotic
513 genomes. *Heredity* 107:487–495.

514 Kramerov, D. A., and N. S. Vassetzky. 2011. SINEs. *WIREs RNA* 2:772–786.

515 Lee, Y. C. G., and C. H. Langley. 2012. Long-term and short-term evolutionary impacts of
516 transposable elements on *Drosophila*. *Genetics* 192:1411–1432.

517 Lee, Y. C. G., and C. H. Langley. 2010. Transposable elements in natural populations of
518 *Drosophila melanogaster*. *Philos. Trans. R. Soc. B* 365:1219–1228.

519 Levis, R., K. O'Hare, and G. M. Rubin. 1984. Effects of transposable element insertions on RNA
520 encoded by the *white* gene of *Drosophila*. *Cell* 38:471–481. Elsevier.

521 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
522 Durbin, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map
523 format and SAMtools. *Bioinformatics* 25:2078–2079.

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

527 Lim, J. K. 1988. Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila*
528 *melanogaster*. *Proc. Nat. Acad. Sci. USA* 85:9153–9157.

529 Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y.
530 Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barrón, C. Bess, K.
531 P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J.
532 C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S.
533 Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L.

534 Perales, L.-L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati,
535 K. C. Worley, Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman,
536 and R. A. Gibbs. 2012. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482:173–
537 178.

538 Magwire, M. M., F. Bayer, C. L. Webster, C. Cao, and F. M. Jiggins. 2011. Successive increases
539 in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a
540 duplication. *PLoS Genet.* 7:e1002337.

541 Makałowski, W., V. Gotea, A. Pande, and I. Makałowska. 2019. Transposable elements:
542 Classification, identification, and their use as a tool for comparative genomics. *Meth. Mol. Biol.*
543 1910:177–207.

544 Malone, C. D., J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, and G. J.
545 Hannon. 2009. Specialized piRNA pathways act in germline and somatic tissues of the
546 *Drosophila* ovary. *Cell* 137:522–535.

547 Marlor, R. L., S. M. Parkhurst, and V. G. Corces. 1986. The *Drosophila melanogaster* gypsy
548 transposable element encodes putative gene products homologous to retroviral proteins. *Mol.*
549 *Cell. Biol.* 6:1129–1134.

550 Maruyama, K., and D. L. Hartl. 1991. Evolution of the transposable element *mariner* in
551 *Drosophila* species. *Genetics* 128:319–329.

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

554 McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Nat. Acad. Sci.*
555 *USA* 36:344–355.

556 McClintock, B. 1984. The significance of responses of the genome to challenge. *Science*
557 226:792–801.

558 McGinnis, W., A. W. Shermoen, and S. K. Beckendorf. 1983. A transposable element inserted
559 just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell*
560 34:75–84.

561 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D.
562 Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a
563 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*
564 20:1297–1303.

565 Muller, H. J. 1932. Some genetic aspects of sex. *Am. Nat.* 66:118–138.

566 Muller, H. J. 1964. The relation of recombination to mutational advance *Mutat Res* 106:2–9.

567 Niu, X.-M., Y.-C. Xu, Z.-W. Li, Y.-T. Bian, X.-H. Hou, J.-F. Chen, Y.-P. Zou, J. Jiang, Q. Wu,
568 S. Ge, S. Balasubramanian, and Y.-L. Guo. 2019. Transposable elements drive rapid phenotypic
569 variation in *Capsella rubella*. *Proc. Nat. Acad. Sci. USA* 116:6908–6913.

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

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

575 Nuzhdin, S. V. 2000. Sure facts, speculations, and open questions about the evolution of
576 transposable element copy number. Pp. 129–137 in J. F. McDonald, ed. *Transposable Elements*
577 *and Genome Evolution*. Springer Netherlands, Dordrecht.

578 Nuzhdin, S. V., E. G. Pasyukova, and T. F. Mackay. 1997. Accumulation of transposable
579 elements in laboratory lines of *Drosophila melanogaster*. *Genetica* 100:167–175.

580 Okada, N., M. Hamada, I. Ogiwara, and K. Ohshima. 1997. SINEs and LINEs share common 3'
581 sequences: a review. *Gene* 205:229–243.

582 Ou, S., W. Su, Y. Liao, K. Chougule, J. R. A. Agda, A. J. Hellings, C. S. B. Lugo, T. A. Elliott,
583 D. Ware, T. Peterson, N. Jiang, C. N. Hirsch, and M. B. Hufford. 2019. Benchmarking
584 transposable element annotation methods for creation of a streamlined, comprehensive pipeline.
585 *Genome Biol.* 20: 275.

586 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in
587 R language. *Bioinformatics* 20:289–290.

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

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

592 Péliçon, A., S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton, and V. G. Corces. 1994.
593 *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the
594 tissue-specific control of the *Drosophila flamenco* gene. *The EMBO Journal* 13:4401–4411.

595 Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D. S. Brar, S.
596 Jackson, R. A. Wing, and O. Panaud. 2006. Doubling genome size without polyploidization:
597 Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild
598 relative of rice. *Genome Res.* 16:1262–1269.

599 Platt, R. N., II, L. Blanco-Berdugo, and D. A. Ray. 2016. Accurate transposable element
600 annotation Is vital when analyzing new genome assemblies. *Genome Biol. Evol.* 8:403–410.

601 Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing
602 genomic features. *Bioinformatics* 26:841–842.

603 Rahman, R., G.-W. Chirn, A. Kanodia, Y. A. Sytnikova, B. Brembs, C. M. Bergman, and N. C.
604 Lau. 2015. Unique transposon landscapes are pervasive across *Drosophila melanogaster*
605 genomes. *Nucleic Acids Res.* 43:10655–10672.

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

608 Reddiex, A. J., S. L. Allen, and S. F. Chenoweth. 2018. A genomic reference panel for
609 *Drosophila serrata*. *G3* 8:1335–1346.

610 Romero-Soriano, V., and M. P. Garcia Guerreiro. 2016. Expression of the retrotransposon
611 *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids. *PLoS ONE*
612 11:e0147903.

613 Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L.
614 Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic
615 inference and model choice across a large model space. *Systematic Biol.* 61:539–542.

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

619 Salinas, N. R., and D. P. Little. 2014. 2matrix: A utility for indel coding and phylogenetic matrix
620 concatenation. *App. Plant Sci.* 2:1300083.

621 Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L.
622 Fulton, T. A. Graves, P. Minx, A. D. Reily, L. Courtney, S. S. Kruchowski, C. Tomlinson, C.
623 Strong, K. Delehaunty, C. Fronick, B. Courtney, S. M. Rock, E. Belter, F. Du, K. Kim, R. M.
624 Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S. M. Jackson, B. Gillam, W. Chen, L.
625 Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K.
626 Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J.
627 Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S.
628 Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T.
629 Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M.
630 Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A.
631 Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C.
632 Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M. J.
633 Levy, L. McMahan, P. Van Buren, M. W. Vaughn, K. Ying, C.-T. Yeh, S. J. Emrich, Y. Jia, A.
634 Kalyanaraman, A.-P. Hsia, W. B. Barbazuk, R. S. Baucom, T. P. Brutnell, N. C. Carpita, C.
635 Chaparro, J.-M. Chia, J.-M. Deragon, J. C. Estill, Y. Fu, J. A. Jeddelloh, Y. Han, H. Lee, P. Li, D.
636 R. Lisch, S. Liu, Z. Liu, D. H. Nagel, M. C. McCann, P. SanMiguel, A. M. Myers, D. Nettleton,
637 J. Nguyen, B. W. Penning, L. Ponnala, K. L. Schneider, D. C. Schwartz, A. Sharma, C.
638 Soderlund, N. M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T. K. Wolfgruber,
639 L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J. L. Bennetzen, R. K. Dawe, J. Jiang, N. Jiang, G.
640 G. Presting, S. R. Wessler, S. Aluru, R. A. Martienssen, S. W. Clifton, W. R. McCombie, R. A.

641 Wing, and R. K. Wilson. 2009. The B73 maize genome: complexity, diversity, and dynamics.
642 Science 326:1112–1115.

643 Schrader, L., J. W. Kim, D. Ence, A. Zimin, A. Klein, K. Wyschetzki, T. Weichselgartner, C.
644 Kemena, J. S. O. Kl, E. Schultner, Y. Wurm, C. D. Smith, M. Yandell, J. U. R. Heinze, J. U. R.
645 Gadau, and J. Oettler. 2019. Transposable element islands facilitate adaptation to novel
646 environments in an invasive species. Nat. Comm. 5:5495.

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

650 Signor, S. 2020. Transposable elements in individual genotypes of *Drosophila simulans*. Ecol.
651 Evol. 130:499–11.

652 Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke, and V. G. Corces. 1994. An env-like
653 protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus.
654 Genes Dev. 8:2046–2057.

655 Srivastav, S. P., and E. S. Kelleher. 2017. Paternal induction of hybrid dysgenesis in *Drosophila*
656 *melanogaster* is weakly correlated with both *P*-element and *hobo* element dosage. G3 7:1487–
657 1497.

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

660 Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy,
661 M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A. H. Schulman. 2007. A unified
662 classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973–982.

663 Wilson, T. G. 1993. Transposable elements as initiators of insecticide resistance. J. Econ. Ent.
664 86:645–651.

665 Wright, E. 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. The R
666 Journal 8:352–359.

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

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

673 Zanni, V., A. Eymery, M. Coiffet, M. Zytnicki, I. Luyten, H. Quesneville, C. Vaury, S. Jensen.
674 2013. Distribution, evolution, and diversity of retrotransposons at the *flamenco* locus reflect the
675 regulatory properties of piRNA clusters. Proc. Natl. Acad. Sci. U.S.A. 110:19842–19847.

676 Zhang, S., and E. S. Kelleher. 2019. Rapid evolution of piRNA-mediated silencing of an
677 invading transposable element was driven by abundant de novo mutations. *bioRxiv* 5:252–40..

678