



Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations

Shuo Zhang, Beverly Pointer and Erin S. Kelleher

Genome Res. 2020 30: 566-575 originally published online April 1, 2020

Access the most recent version at doi:[10.1101/gr.251546.119](https://doi.org/10.1101/gr.251546.119)

References This article cites 71 articles, 27 of which can be accessed free at:
<http://genome.cshlp.org/content/30/4/566.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations

Shuo Zhang, Beverly Pointer, and Erin S. Kelleher

Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA

The regulation of transposable element (TE) activity by small RNAs is a ubiquitous feature of germlines. However, despite the obvious benefits to the host in terms of ensuring the production of viable gametes and maintaining the integrity of the genomes they carry, it remains controversial whether TE regulation evolves adaptively. We examined the emergence and evolutionary dynamics of repressor alleles after *P*-elements invaded the *Drosophila melanogaster* genome in the mid-twentieth century. In many animals including *Drosophila*, repressor alleles are produced by transpositional insertions into piRNA clusters, genomic regions encoding the Piwi-interacting RNAs (piRNAs) that regulate TEs. We discovered that ~94% of recently collected isofemale lines in the *Drosophila melanogaster* Genetic Reference Panel (DGRP) contain at least one *P*-element insertion in a piRNA cluster, indicating that repressor alleles are produced by de novo insertion at an exceptional rate. Furthermore, in our sample of approximately 200 genomes, we uncovered no fewer than 80 unique *P*-element insertion alleles in at least 15 different piRNA clusters. Finally, we observe no footprint of positive selection on *P*-element insertions in piRNA clusters, suggesting that the rapid evolution of piRNA-mediated repression in *D. melanogaster* was driven primarily by mutation. Our results reveal for the first time how the unique genetic architecture of piRNA production, in which numerous piRNA clusters can encode regulatory small RNAs upon transpositional insertion, facilitates the nonadaptive rapid evolution of repression.

[Supplemental material is available for this article.]

Transposable elements (TEs) are widespread genomic parasites that increase their copy number by mobilizing and self-replicating within their host genomes. TE expression and proliferation are therefore strictly regulated, particularly in germline cells where TEs are exceptionally active and resulting mutations are transmitted to offspring. In the germline of most metazoans, TEs are controlled by a conserved small-RNA-mediated pathway, in which Piwi-interacting RNAs (piRNAs), in complex with Argonaute proteins, silence TEs in a sequence-specific manner (Aravin et al. 2007; Brennecke et al. 2007; Houwing et al. 2007; Girard and Hannon 2008). Although host regulation of endogenous TEs by piRNAs is ubiquitous, TEs are frequently horizontally transferred between nonhybridizing species, allowing TE families to colonize new host genomes (Thomas et al. 2010; Dotto et al. 2015; Peccoud et al. 2017). How the host evolves repression to novel TEs invading the genome remains poorly understood.

After invasion, repressor alleles are proposed to arise through de novo mutation, when an invading TE copy randomly inserts into a piRNA-producing locus referred to as a piRNA cluster (Girard and Hannon 2008; Khurana et al. 2011). The existence of numerous alternative piRNA clusters—for example, 142 loci or ~3.5% of assembled *Drosophila melanogaster* genome based on Brennecke et al. 2007—may facilitate the evolution of repression by increasing the probability that a random insertion occurs in a piRNA-producing site (Kelleher 2016; Kelleher et al. 2018; Kofler 2019). However, the technical challenge of annotating polymorphic TE insertions in repeat-rich piRNA clusters has limited the identification and study of these repressor alleles. Furthermore,

for most TE families it is impossible to distinguish repressor alleles that arose via de novo insertion into existing piRNA clusters from the reciprocal: de novo piRNA clusters that arose at existing TE insertions. In particular, recent studies suggest that novel piRNA clusters may emerge frequently via epigenetic mutation, when a change in chromatin state triggers bidirectional transcription and piRNA production (de Vanssay et al. 2012; Le Thomas et al. 2014; Shpiz et al. 2014; Hermant et al. 2015).

The role of selection in the evolution of host TE repression, through piRNA-mediated silencing or otherwise, also remains controversial. In sexually reproducing organisms, the selective advantage of a repressor allele is limited by recombination, which separates the repressor from the DNA it has protected from deleterious mutation (Charlesworth and Langley 1986). Additionally, although selection for repression may be strong when the genome is invaded by a new TE family, it is unclear whether it is sustained for a sufficient number of generations to enact meaningful changes in repressor allele frequency (Lee and Langley 2012). On the other hand, forward simulation models suggest that piRNA-mediated repressor alleles are targets of positive selection, especially when transposition rates are high and TEs are highly deleterious (Lu and Clark 2010; Kelleher et al. 2018; Kofler 2019). Moreover, an early population genomic analysis of *D. melanogaster* suggests that TE insertions in piRNA clusters may segregate at higher frequency than noncluster insertions, although this is based on modest sample size and read depth (Lu and Clark 2010).

Corresponding author: eskelleher@uh.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.251546.119>.

© 2020 Zhang et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

The recent invasion of *P*-element DNA transposons into both *D. melanogaster* and *D. simulans* (Kidwell 1983; Anxolabéhère et al. 1988; Kofler et al. 2015; Hill et al. 2016) provides a unique opportunity to study the evolution of piRNA-mediated repression. Unlike most TE families that are long-term constituents of their host genome, *P*-elements appeared in the *D. melanogaster* genome around 1950 (Kidwell 1983; Anxolabéhère et al. 1988; Daniels et al. 1990) and in *D. simulans* around 2010 (Kofler et al. 2015). Many natural populations of *D. melanogaster* and *D. simulans* rapidly and recently evolved piRNA-mediated repression (Kidwell 1983; Brennecke et al. 2008; Jensen et al. 2008; Hill et al. 2016), providing the opportunity to reconstruct the history of piRNA-producing alleles, as well as examine their evolutionary dynamics.

P-element insertions in piRNA clusters have a well-established role in the evolved repression of *P*-elements in both *D. melanogaster* and *D. simulans*. Of particular significance are *P*-element insertions in subtelomeric piRNA clusters located in telomeric associated sequence (TAS). Seven unique wild-derived *P*-element insertions into X-TAS have been shown to confer maternal repression of *P*-element transposition (Ronsseray et al. 1996; Marin et al. 2000; Stuart et al. 2002), three of which have been independently shown to produce *P*-element-derived piRNAs (Brennecke et al. 2008). Similarly, in laboratory populations of *D. simulans*, evolved repression is associated with the insertion of *P*-elements into the TAS piRNA cluster on Chromosome 3R (Kofler et al. 2018). Finally, in *D. melanogaster*, *P*-element insertions into non-TAS piRNA clusters on Chromosomes 3 and 4 have also been shown to confer piRNA-mediated repression (Khurana et al. 2011; Moon et al. 2018), supporting a general role for piRNA clusters in controlling *P*-elements.

Here, we examine approximately 200 fully sequenced *D. melanogaster* genomes from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay et al. 2012; Huang et al. 2014), which were recently sampled from a repressive population, to identify *P*-element insertion alleles. We furthermore take advantage of piRNA libraries generated from multiple strains collected before the *P*-element invasion, which provide a historical record of piRNA cluster locations in the ancestral population. By combining these data sets, we reveal the de novo insertion of *P*-elements into ancestral piRNA clusters and examine these candidate repressor alleles for signatures of recent positive selection.

Results

North American strains strongly repress *P*-elements

Both recent and historic samples suggest that *P*-elements are robustly repressed in North American populations of *D. melanogaster* (Kidwell 1983; Itoh et al. 2007). To confirm that this is also true for the DGRP, collected in North Carolina in 2003, we assayed *P*-element repression in dysgenic crosses between DGRP females and Harwich males. In the absence of maternally deposited piRNAs, offspring of such crosses are sterile, showing atrophied ovaries (Kidwell et al. 1977; Brennecke et al. 2008; Kelleher 2016). We observed that for 97.6% of DGRP maternal genotypes we sampled (41 of 42), F1 offspring were fertile, suggesting the presence of maternally deposited piRNAs (Supplemental Fig. S1A). The single strain that did not show strong repression (DGRP531) was also very difficult to maintain in the laboratory, suggesting that infertility may be unrelated to *P*-element activity. We also looked directly at the production of *P*-element-derived piRNAs among the DGRP using a previously published set of ovarian small-RNA libraries

from 16 DGRP genomes (Song et al. 2014). We discovered that 100% (16 of 16) of these strains produce a robust number of *P*-element-derived piRNAs (Supplemental Fig. S1B), consistent with the repressive phenotypes we observed. Taken together these observations suggest that maternal piRNA-mediated repression is prevalent, if not ubiquitous, among DGRP genomes.

Identification of ancestral piRNA clusters

To uncover the genetic basis of piRNA-mediated repression, we first sought to annotate ancestral piRNA clusters in the *D. melanogaster* genome, which acted as source loci for piRNAs before the introduction of *P*-elements. We took advantage of 27 small-RNA sequencing libraries from nine wild-type strains (Supplemental Table S1), which were isolated from nature before *P*-element invasion and are therefore devoid of genomic *P*-elements. We annotated piRNA clusters based on the density of mapped piRNAs using proTRAC (Rosenkranz and Zischler 2012). We varied the stringency of piRNA cluster identification to generate three sets of annotations, which contained 32, 159, and 497 piRNA clusters, and composed 0.30%, 1.27%, and 3.68% of the assembled *D. melanogaster* genome, respectively (Fig. 1A; Supplemental Tables S2, S3).

Some genomic loci are polymorphic in their status as a piRNA cluster: producing abundant piRNAs in some strains while remaining quiescent in others (Fig. 1A,B; Supplemental Tables S2, S3). We therefore defined ancestral piRNA clusters as genomic regions that were annotated from at least one small-RNA library. In light of clear examples of polymorphism, our annotations should not be considered a comprehensive list of the piRNA clusters segregating in ancestral populations, but rather a representative sample that includes most clusters segregating at high frequency or fixed at the time of invasion. Major known piRNA clusters such as *flamenco*, *42AB*, and TAS (Robert et al. 2001; Brennecke et al. 2007; Li et al. 2009; Malone et al. 2009) produced abundant piRNAs in all genotypes and were annotated as piRNA clusters regardless of stringency (Fig. 1A; Supplemental Tables S2, S3).

Most DGRP genomes have *P*-elements in ancestral piRNA clusters

Although previous annotations suggest that <20% DGRP genomes have *P*-elements in ancestral piRNA clusters (based on 32 annotated piRNA clusters) (Zhuang et al. 2014; Rahman et al. 2015), we suspected that this was a gross underestimate, because the common requirement for unique read or read-pair alignment to the reference genome prohibits the identification of TE insertions in repeat-rich piRNA clusters. We therefore annotated *P*-element insertion sites throughout the genome based on high-quality, split-read alignments (mapping quality score, MAPQ \geq 20), which are not necessarily unique yet still support a particular insertion site with high confidence. We further removed potential false positives and false negatives based on the realignment of reads to pseudo-genomes corresponding to each proposed *P*-element insertion (Methods; Supplemental Fig. S2). Including high-quality non-unique alignments increases the number of annotated *P*-elements by 71% and 66% when compared to TEMP and TIDAL, respectively, two approaches that rely on unique alignments (Fig. 2A; Supplemental Table S4; Zhuang et al. 2014; Rahman et al. 2015). Although we did not validate these new insertions, six of seven additional insertions we identified in DGRP492 were also detected by previous study using hemispecific PCR, indicating they are true insertions (Zhang and Kelleher 2017).

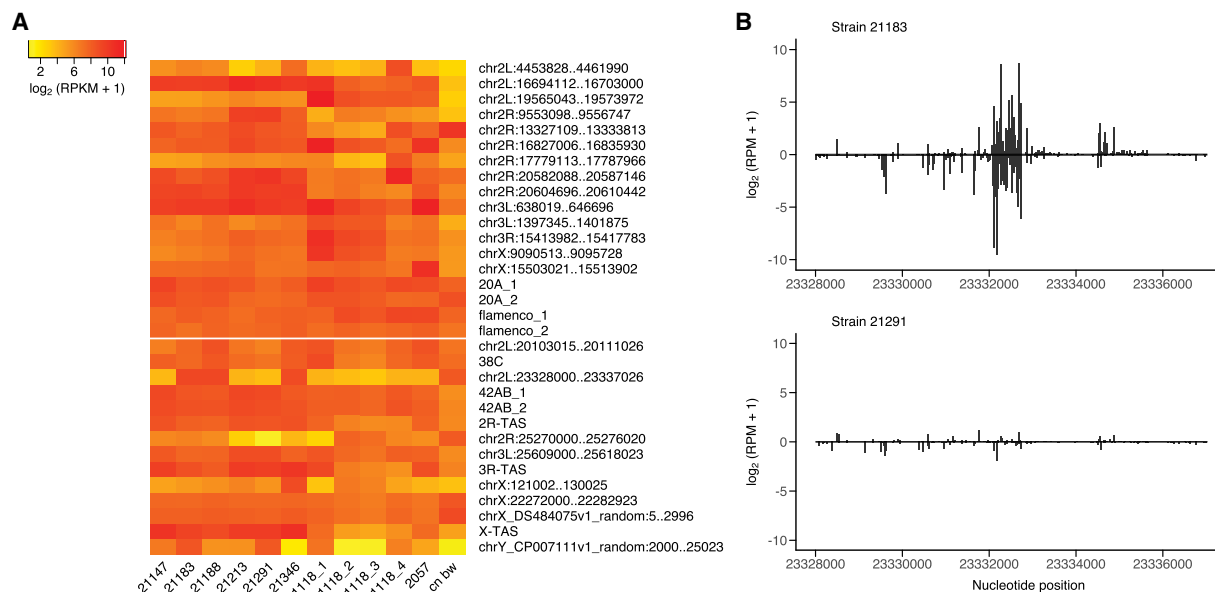


Figure 1. piRNA cluster annotation in *P*-element free strains. (A) Activity of 32 piRNA clusters in ancestral (*P*-element free) strains of *D. melanogaster*. Each column represents a small-RNA sequencing library (biological replicates are combined), and each row represents a piRNA cluster annotated in at least one of these libraries. Coordinates of piRNA clusters are based on the *D. melanogaster* release 6 assembly (dm6) (Hoskins et al. 2015). piRNA cluster expression levels are estimated by reads per kilo base per million mapped reads (RPKM) and transformed to \log_2 scale [$\log_2(\text{RPKM} + 1)$]. Clusters above the white line are unistrand piRNA clusters, and those below the white line are dual-strand piRNA clusters. Details on small-RNA library prep, which may be related to differences in annotated piRNA clusters between libraries from *w1118*, are provided in Supplemental Table S1. (B) An example of a polymorphism in piRNA cluster activity in an annotated cluster on Chromosome 2L (23,328,000–23,337,026). Abundant piRNAs are detected from strain 21183, whereas strain 21291 produces few piRNAs. Only uniquely mapping piRNAs are considered. piRNA density is measured in reads per million mapped reads (RPM) and transformed to \log_2 scale [$\log_2(\text{RPM} + 1)$]. Positive values represent piRNAs mapped to sense strand of the reference genome, and negative values represent piRNAs from antisense strand.

Despite relaxing the requirement for unique alignment, we still identified only 11 DGRP genomes (5.6%) with *P*-elements in X-TAS. This is radically different from previous studies using hybridization-based approaches, which estimate that ~50% of wild-derived genomes harbor a *P*-element insertion in X-TAS alone (Ajioka and Eanes 1989; Ronsseray et al. 1989; Biémont et al. 1990). High-quality alignments likely fail to provide a unique insertion site in TAS repeats because these highly similar tandem satellite sequences allow for multiple equivalent alignments (Fig. 3A; Karpen and Spradling 1992; Yin and Lin 2007; Asif-Laidin et al. 2017). Therefore, we relaxed the requirement for a single “best” alignment; we rather identified *P*-derived reads and read pairs that also aligned to TAS repeats but aligned nowhere else in the reference genome. We found that the majority of DGRP genomic libraries contain *P*-derived read pairs that align to X, 2R, or 3R-TAS (Supplemental Table S5), whereas only three DGRP genomes contained *P*-derived reads aligning to 2L and 3L-TAS.

To estimate the number (0, 1, >1) of *P*-elements in X, 2R, and 3R-TAS for each DGRP line, we took advantage of the distribution of the number of read pairs supporting individual insertions outside of TAS from the same genome. We then calculated a Z-score for the number of

P-derived reads mapped to TAS. Using this approach we identified 12 DGRP genomes that harbor no *P*-element insertions in TAS (6%, $Z < -1.96$), 126 DGRP genomes that harbor one *P*-element insertion in TAS (65%, $-1.96 < Z < 1.96$), and 57 genomes that carry two or more insertions into TAS arrays (29%, $1.96 < Z$) (Fig. 3B; Supplemental Table S5). Given that TAS arrays are ancestral piRNA clusters that are active in all *P*-element free strains we

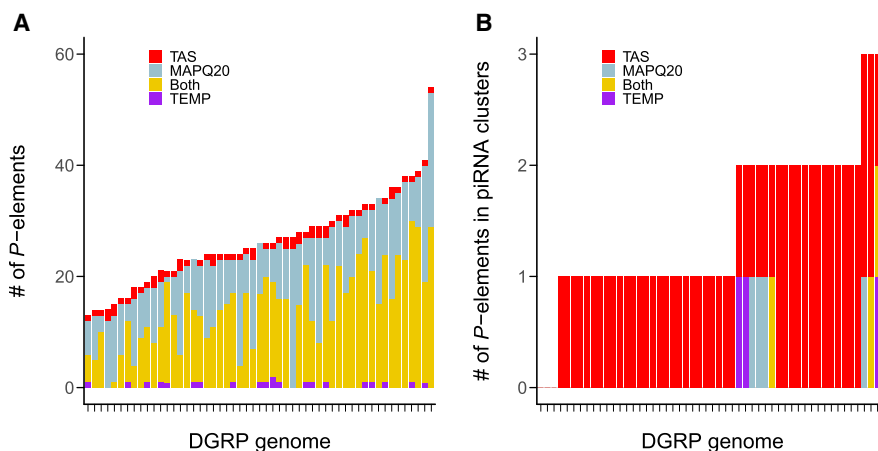


Figure 2. Total number of *P*-elements (A) and number of *P*-elements in ancestral piRNA clusters (B) annotated by different approaches for 53 DGRP genomes that were previously annotated by TEMP. For B, the 32 high-confidence cluster set was used. (TEMP) Insertions only found by TEMP; (MAPQ20) insertions only found based on high-quality mapping; (Both) insertions found by TEMP and MAPQ20; and (TAS) insertions only found when homologous TAS sequences were treated as a single locus.

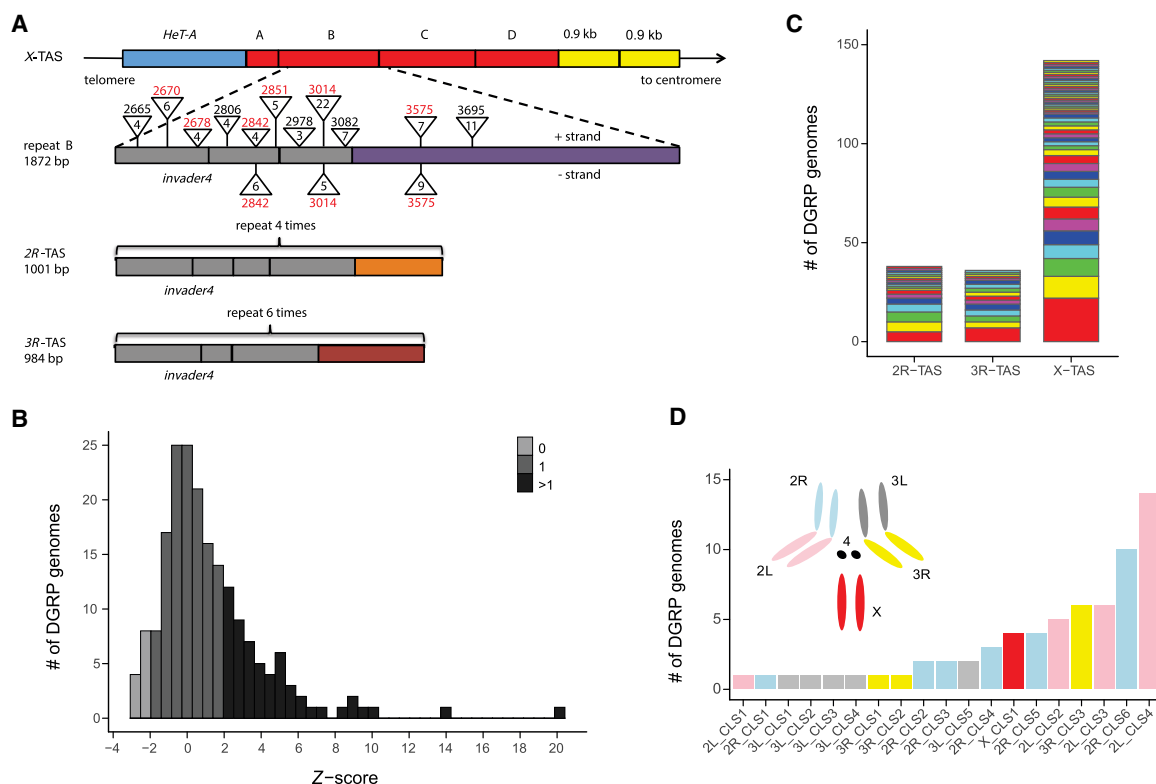


Figure 3. *P*-element insertions in TAS and non-TAS piRNA clusters. (A) The structure of TAS arrays (modeled after Asif-Laidin et al. 2017). X-TAS contains four tandem repeats (A–D in red) located between a *HeT-A* retrotransposon array and two 0.9-kb repeats (Karpen and Spradling 1992). Repeat A is degenerated. Repeats B–D are ~1.8 kb in length and are highly similar to each other (>95% identity). Repeat B is compared to repeats from 2R and 3R-TAS. Each repeat of 2R, 3R, and X-TAS contains several *invader4* retrotransposon long terminal repeats (LTRs) (gray) (Bergman et al. 2006), as well as other short homologous fragments (41–131 bp), with 909 bp being unique to the X-TAS repeat (Asif-Laidin et al. 2017). *P*-element insertions located in X-TAS and present in more than two DGRP genomes are depicted as triangles, with the numeric value above corresponding to the breakpoint, and the numeric value inside corresponding to the number of DGRP genomes containing the insertion. Breakpoints in red correspond to insertion hotspots (Karpen and Spradling 1992). (B) The distribution of Z-scores for *P*/TAS-derived read counts among DGRP libraries. DGRP genomes with $Z < -1.96$, $-1.96 < Z < 1.96$, and $Z > 1.96$ were estimated to have 0, 1, and >1 *P*-element, respectively. (C) Multiple *P*-element insertion sites were detected in 2R, 3R, and X-TAS. Each color represents a unique *P*-element insertion. (D) The frequency of *P*-element insertions in non-TAS ancestral piRNA clusters (based on our annotation set of 159 piRNA clusters) across all major chromosome arms.

sampled (Fig. 1A; Brennecke et al. 2007; Yin and Lin 2007), our observations reveal that the majority of DGRP genomes carry repressor alleles that arose by de novo insertion into existing piRNA clusters (Fig. 2B).

Numerous repressor alleles reveal a high mutation rate to repression

We next sought to isolate individual repressor alleles that arose via de novo insertion into TAS arrays. First, we identified the candidate TAS array(s) containing *P*-element insertions in each DGRP genome, based on proportion of *P*-derived/TAS-derived read pairs whose best alignment occurred in X, 2R, or 3R-TAS (Methods; Supplemental Table S6). We further identified the specific insertion site that was supported by the most read pairs, based on alignment to the reference genome as well as pseudo-genomes we constructed with *P*-element insertions in alternate sites (Methods). We assumed all homologous insertion sites among tandem repeats corresponded to a single insertion event for these analyses.

We found 40 unique *P*-element insertions into TAS, represented in 80 DGRP genomes, where the best insertion site identified by reference genome and pseudo-genome alignments agreed, suggesting well-supported insertion sites (Supplemental

Table S6). We were able to verify 27 of these by site-specific PCR (68%), seven insertions were revealed to reside in different sites by PCR (17.5%), two could not be amplified, and four were not attempted. We further used PCR to determine TAS insertion sites in an additional 94 DGRP genomes, uncovering an additional 43 *P*-element insertion sites in 77 DGRP genomes (Supplemental Table S6). Therefore, in total we identified 85 independent insertions of *P*-elements into TAS sequences (2R, 3R, or X-TAS), 80 of which were verified by PCR in at least one DGRP genome (Table 1; Supplemental Table S6).

Consistent with previous studies (Ajioka and Eanes 1989; Ronssey et al. 1989; Biémont et al. 1990), we found that >50%

Table 1. *P*-element insertions in TAS

TAS array	Number of genomes with insertions	Number of alleles
2R-TAS	32	19 (19)
3R-TAS	34	16 (16)
X-TAS	127	50 (45)
2L and 3L-TAS	0	0 (0)

Numbers in parentheses indicate PCR-verified alleles.

of DGRP genomes had *P*-element insertions in *X*-TAS, and ~17% DGRP genomes had *P*-elements in *2R* and *3R*-TAS (Table 1; Supplemental Table S6). Moreover, we discovered a multitude of insertion alleles in each TAS array: 19 in *2R*-TAS, 16 in *3R*-TAS, and 50 in *X*-TAS (Table 1; Fig. 3A,C; Supplemental Table S6). The large number of alleles, all of which arose after invasion and over the course of 50 yr, reveals an exceptionally high insertion rate of *P*-elements into TAS piRNA clusters.

P-elements in preferred insertion sites show elevated polymorphic frequencies

P-elements have known insertion hot points in *X*, *2R*, and *3R*-TAS (Fig. 8 of Karpen and Spradling 1992), as well as at euchromatic sites around the genome (Spradling et al. 2011). We therefore wondered whether *P*-element insertion into these hot points explains the large number of TAS insertions. Indeed, we found that hot points were greatly enriched for *P*-element insertion alleles: 88.2% (15 of 17) of hot points in TAS arrays had a *P*-element insertion allele compared to only 1.4% (55 of 3840) of nonpreferred sites (Fisher's exact test P -value $< 10^{-15}$) (Fig. 4A). Similarly, for non-TAS regions, 57.1% (16 of 28) of hot points have an insertion allele compared to 0.003% of nonpreferred sites (3642 of 143,691,516; Fisher's exact test P -value $< 10^{-15}$) (Fig. 4A). Hot points were also more likely to have two distinguishable insertion alleles, one in each strand when compared to nonpreferred sites (TAS: Fisher's exact test P -value = 1.51×10^{-5} ; non-TAS: Fisher's exact test P -value $< 10^{-15}$) (Fig. 4B).

Recurrent insertion into hot points potentially also elevates the population frequency of insertion alleles in the event of recurrent insertion. Indeed, even after separating insertions occurring in the same genomic position but on opposite strands, individual insertion alleles in hot points show higher population frequencies than those occurring at nonpreferred sites (TAS: Wilcoxon rank-sum test Z -approximation = 3.71, P -value = 2.05×10^{-5} ; non-TAS: Wilcoxon rank-sum test Z -approximation = 8.36, P -value $< 10^{-15}$) (Fig. 4C). Taken together, our observations suggest that the exceptional number of *P*-element insertions in *X*-TAS arrays, as well as their unusually high frequency, is at least partially explained by an insertion site preference.

No evidence of positive selection on repressor alleles

Combining the TAS insertion alleles with those identified in non-TAS piRNA clusters, we detected up to 170 *P*-element insertion events into at least 15 (up to 33) different ancestral piRNA clusters, which are located on all of the major chromosome arms of the *Drosophila* genome (Fig. 3D; Supplemental Fig. S3; Supplemental Table S7). *P*-element insertions into ancestral piRNA clusters are proposed to benefit the host by preventing the accumulation of additional deleterious insertions, suppressing dysgenic sterility, and potentially establishing heterochromatin formation at *P*-element loci that suppresses their participation in ectopic recombination (Charlesworth and Langley 1986; Lee and Langley 2012; Kelleher et al. 2018). To detect a signature of positive selection on *P*-element insertion alleles in ancestral piRNA clusters, we compared the frequencies of these putatively beneficial alleles to those of neutral or deleterious *P*-element insertions that do not establish repression. If piRNA-mediated repression is a target of positive selection, elevated frequencies are predicted among insertions in piRNA clusters (Nielsen 2005).

Regardless of how stringently we defined ancestral piRNA clusters, we observed that *P*-element insertion alleles in piRNA clusters are significantly more common among DGRP genomes than those in other genomic regions (Fig. 5A–C). Although the elevated frequency of cluster *P*-elements might suggest positive selection, the observation is confounded by two factors. First, recurrent insertion into hot points, which occur disproportionately in TAS piRNA clusters, elevates the frequency of those insertion alleles (Fig. 4C). Second, TE insertions rise to higher frequency in regions of low recombination, where piRNA clusters reside, most likely because of reduced purifying selection against ectopic recombination (Charlesworth and Langley 1989; Dolgin and Charlesworth 2008; Petrov et al. 2011; Kofler et al. 2012). To disentangle the potential impact of positive selection on cluster insertions from the confounding effects of reduced purifying selection and recurrent insertion, we fit a multiple regression

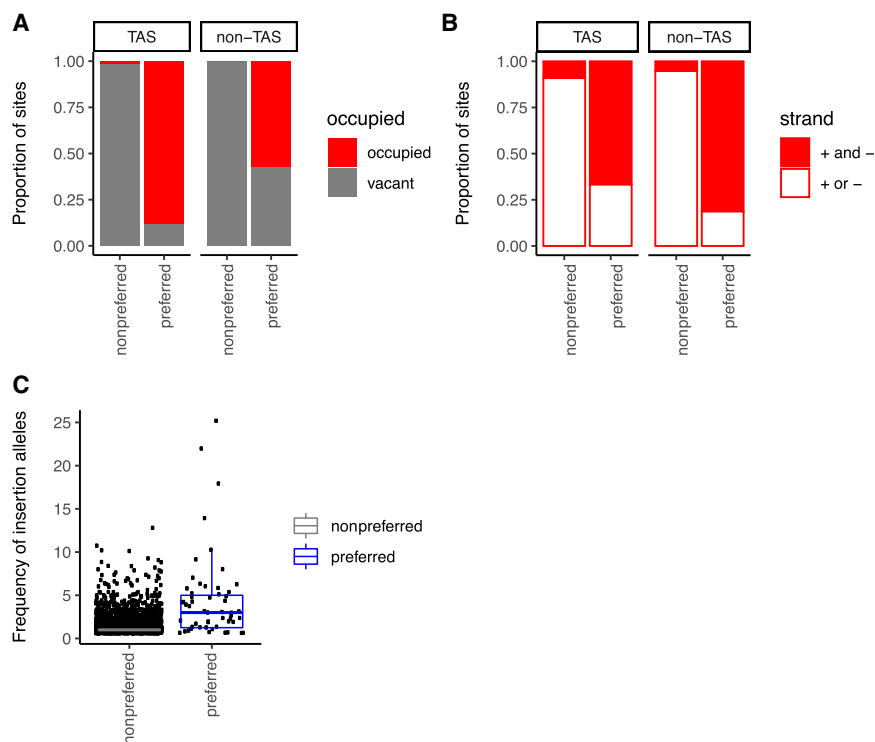


Figure 4. Insertion hot points in TAS arrays contribute to the large number of high-frequency *P*-element insertion alleles. (A) The proportion of nonpreferred and preferred (hot point) sites that are occupied by a *P*-element insertion in at least one DGRP genome or vacant in all DGRP genomes. (B) The proportion of occupied sites containing *P*-element insertions in both the sense and antisense strands (red), or a single strand (white), for nonpreferred and preferred sites. (C) The comparison between frequencies of *P*-elements at nonpreferred and preferred sites. *P*-elements inserted at the same site but in opposite orientations were considered different insertions.

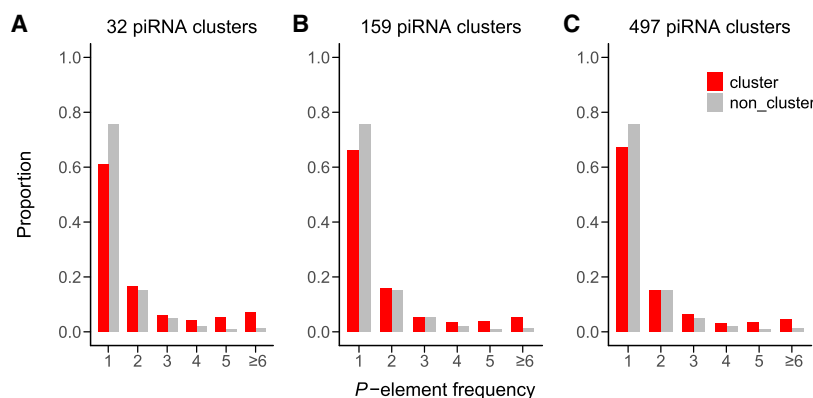


Figure 5. The frequency of *P*-elements in piRNA clusters (red) and the frequency of *P*-elements outside of clusters (gray) are compared for three sets of annotated piRNA clusters: (A) 32, (B) 159, and (C) 497 clusters.

model that predicted the frequency of each *P*-element insertion as a function of its recombination rate, whether or not it occurs in a hot point, and its location inside or outside of a piRNA cluster, for all three sets of cluster annotations. Recombination rate ($F_{1,3942} = 8.66$, P -value = 0.0033) and insertion preference ($F_{1,3942} = 404.49$, P -value < 10^{-15}) were both strongly associated with the polymorphic frequency of *P*-element insertion alleles. However, after accounting for these two confounding variables, we were unable to detect a difference in the polymorphic frequency of insertions inside and outside of piRNA clusters (pden = 0.01: $F_{1,3942} = 0$, P -value = 1; pden = 0.05: $F_{1,3942} = 0.0009$, P -value = 0.98; pden = 0.1: $F_{1,3942} = 0$, P -value = 1). We therefore find no evidence that positive selection has elevated the polymorphic frequencies of *P*-element insertions into ancestral piRNA clusters among DGRP genomes.

Discussion

In this study, we took advantage of the recent invasion of the *D. melanogaster* genome by *P*-element DNA transposons to chronicle the evolution of piRNA-mediated repression. We reveal that the common phenotype of *P*-element repression (Kidwell 1983; Kidwell et al. 1983; Ogura et al. 2007) is underpinned by an unprecedented number of candidate repressor alleles, which have arisen since the *P*-element invasion in the mid-twentieth century. We uncovered 115 unique candidate repressor alleles, which are independent insertions of *P*-elements into high-confidence piRNA clusters, 80 of which we verified by PCR. Furthermore, we found no evidence that positive selection has increased the frequency of these insertions, suggesting that mutation alone is responsible for the rapid evolution of the repressive phenotype in less than ~30 yr (Kidwell 1983).

Our observations represent, to our knowledge, the first demonstration that mutation rates can be sufficiently high to drive a rapid evolutionary change. Except in cases of extreme mutation limitation, the contribution of mutation rate to the rate of evolutionary change is thought to be negligible, because the mutation rate per site is comparatively slow when compared against the action of selection. However, the exponential increase in transposition rate that occurs as TE copies accumulate, and the large numbers of functionally redundant piRNA clusters that establish repression when carrying an insertion allele, result in a mutation rate to piRNA-mediated repressor alleles that is exceptionally high (Kelleher et al. 2018). Indeed, forward simulations have previously shown that piRNA-mediated repression may

evolve even in the absence of any fitness cost to TEs (Kofler 2019), which mirrors the absence of a footprint of positive selection on *P*-element insertions in piRNA clusters.

P-element invasions may be particularly conducive to the mutation-dependent evolution of piRNA-mediated silencing. *P*-elements show an unusually high transposition rate (10^{-1} to 10^{-3} new insertions per element per generation) (Eggleston et al. 1988; Berg and Spradling 1991; Kimura and Kidwell 1994), which exceeds the single base-pair mutation rate for *Drosophila* by at least six orders of magnitude (10^{-9} new nucleotide substitutions per site per generation) (Haag-Liautard et al. 2007). The

presence of multiple insertion hot points in the TAS piRNA clusters (Karpen and Spradling 1992), will further accelerate this already high mutation rate to repressor alleles. Indeed, 73.9% of *P*-element insertions into ancestral piRNA clusters (based on 32 annotated piRNA clusters) occurred in TAS regions, consistent with previous studies that detected *P*-elements insertions using hybridization-based approaches (Ronsseray et al. 1991; Marin et al. 2000; Stuart et al. 2002). Furthermore, we discovered that *P*-elements are most commonly observed in previously identified insertion hot points (Fig. 4A; Karpen and Spradling 1992; Spradling et al. 2011), thereby demonstrating that mutation bias shapes the distribution of *P*-element insertions even within the TAS clusters. TE insertions in TAS were likely not detected among DGRP genomes previously because the reliance on unique alignments excludes read pairs supporting insertions in satellite arrays (Linheiro and Bergman 2012; Zhuang et al. 2014; Rahman et al. 2015). Therefore, allowing for multiple mapping within highly homologous satellite repeats represents a powerful method for annotating TEs in these regions from short, paired-end reads. Our observations echo those of a recent study in laboratory populations of *D. simulans*, which showed that *P*-element repression evolved by multiple independent insertions in piRNA clusters, particularly in the 3R-TAS (Kofler et al. 2018).

Finally, we found that ~94% *D. melanogaster* genomes have at least one *P*-element in an ancestral piRNA cluster, suggesting de novo mutation, in which *P*-elements transpose into preexisting piRNA clusters, is the predominant mutational mechanism giving rise to piRNA-mediated silencing. Our conclusions are minimally dependent on the assumption that ancestrally polymorphic piRNA clusters were active in the genotype where *P*-element insertion occurred, because only eight of 115 annotated insertions in high-confidence piRNA clusters occur in clusters that are polymorphic in historic collections (Fig. 1A; Supplemental Tables S7, S8). Nevertheless we cannot exclude a potential role of epigenetic mutations in the evolution of piRNA-mediated *P*-element repression. Indeed, of six strains we examined that do not contain insertions in ancestral piRNA clusters, five are strong repressors of *P*-element hybrid dysgenesis (Supplemental Fig. S4). Although these strains may contain insertions into piRNA clusters that we were unable to identify (false negatives), they also may contain natural epialleles: *P*-element insertions that have been converted into heritable piRNA clusters through changes in epigenetic state (de Vanssay et al. 2012; Le Thomas et al. 2014; Hermant et al. 2015). If epigenetic mutation occurs in natural populations, it would provide

an even greater increase to the already high mutation rate to piRNA-mediated silencing, further accelerating population-level transition to a repressive state.

In summary, *P*-element repression in *Drosophila melanogaster* evolved rapidly through abundant de novo mutations: the transposition of *P*-elements into preexisting piRNA clusters. These potentially beneficial alleles show no signature of positive selection, representing a heretofore-unique example of rapid evolutionary change that emerges from mutation alone. Our observations reveal how the unique genetic architecture of piRNA-mediated silencing, in which transposition into multiple functionally redundant piRNA clusters results in a repressor allele, facilitates the evolution of repression of an invading TE, thereby removing the requirement for natural selection.

Methods

DGRP stocks and genomes

All DGRP lines were ordered from the Bloomington *Drosophila* Stock Center.

Assays of dysgenic sterility

Virgin DGRP females were crossed to males from the reference *P* strain Harwich at 29°C. Three- to five-day-old F1 female offspring were assayed for ovarian development using a squash prep, as described in Srivastav and Kelleher (2017).

piRNA cluster annotation

Ovarian small-RNA sequencing libraries were downloaded from NCBI (Supplemental Table S1) or were generated by our laboratory (see Data access). For each library, adapters were trimmed using cutadapt (version 1.9.1) (Martin 2011). Trimmed reads with 23–29 nt (typical size of piRNAs in *Drosophila*) were kept for further analysis.

piRNA clusters were predicted separately for each library from an ancestral *P*-element free strain using proTRAC (Rosenkranz and Zischler 2012), which identifies genomic loci corresponding to piRNA clusters based on the density of mapped piRNAs. We considered different values of the proTRAC pdens parameter (0.01, 0.05, 0.1), with lower pdens values corresponding to annotation sets that include a smaller number of higher-confidence piRNA clusters. Annotated piRNA clusters <5 kb apart were considered a single cluster.

To identify *P*-element-derived piRNAs in ovarian small-RNA libraries from 16 DGRP genomes examined in Song et al. (2014), piRNAs were aligned to both the *P*-element consensus and assembled *D. melanogaster* genome (dm6) (Hoskins et al. 2015). *P*-element-derived piRNA abundance was estimated as reads per million mapped piRNA reads (RPM).

Detecting *P*-element insertions in DGRP genomes

DGRP whole-genome sequencing reads were downloaded from the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra/>; study: SRP000694) (Mackay et al. 2012; Huang et al. 2014). Twelve DGRP genomes were excluded from our analysis because 45-bp paired-end reads (DGRP357, DGRP379, DGRP427, DGRP486, DGRP786), or 75-bp single-end reads (DGRP153, DGRP237, DGRP28, DGRP313, DGRP325, DGRP386, DGRP41, DGRP730) were too short to allow for identification of *P*-element insertion sites. To identify read pairs that include *P*-element sequence in the remaining genomes, individual reads were separately and locally aligned to full-length *P*-element consensus

(O'Hare and Rubin 1983) using Bowtie 2 (v2.1.0) (Langmead and Salzberg 2012) with default parameters. *P*-element sequences were then trimmed from mapped reads using a custom Perl script. Trimmed reads longer than 30 bp were kept and used for downstream analyses. A flow chart of the annotation strategy for non-TAS insertions is provided in Supplemental Figure S2.

For each DGRP genome, the *P*-derived trimmed reads were first aligned to the *D. melanogaster* release 6 reference genome (dm6) (Hoskins et al. 2015) as well as *X*-TAS (Karpen and Spradling 1992) using Bowtie 2. Reported alignments with mapping quality score greater than 20 and edit distance (sum of mismatches and gaps required to convert the read sequence to the reference) less than four were kept. To isolate breakpoints corresponding to *P*-element insertion sites, we took advantage of split reads, in which one segment aligned to the *P*-element consensus and the remainder aligned to the reference genome. After breakpoints were located, all nonsplit *P*-derived read pairs (i.e., one read aligns to *P*-element, its mate to the reference genome) within 500 bp were identified. At least six supporting read pairs (split or nonsplit) were required to identify a candidate *P*-element insertion.

For each of 6528 candidate *P*-element insertions, we constructed a pseudo insertion allele containing 500 nt of genomic sequence on either side of insertions, an 8 nt target site duplication, and the full-length *P*-element as consensus as in Zhang and Kelleher (2017). To identify potential false positives, all paired-end reads from each DGRP genome were realigned to a library of pseudo insertions and reference alleles for all insertion sites annotated in the given genome, requiring more than six read pairs (MAPQ ≥ 20, edit distance < 4). We then calculated the frequency of each insertion as the fraction of reads supporting the insertion allele. We removed 239 annotated insertions with fewer than six supporting reads in an individual genome, and 900 singleton insertions with estimated frequencies < 80% as likely false positives (Supplemental Fig. S2).

After false positives were removed we sought to identify false negative insertions, which were not identified in a given DGRP genome owing to an absence of split reads, but were annotated in another DGRP genome. To this end we constructed a combined reference genome including dm6, *X*-TAS, and pseudo-genomes from the combined set of all 3861 candidate *P*-element insertions identified in any DGRP genome. Reads from all DGRP genomes were then aligned to this combined reference. We identified 156 false negative insertions that were supported by more than six read pairs from the Illumina library from a given strain (MAPQ ≥ 20, edit distance < 4), but were not annotated in our original alignments owing to absence of split reads. The complete list of *P*-element insertions, including their estimated frequencies in each DGRP genome, are provided in Supplemental Table S8.

Detecting *P*-element insertions in TAS

We divided the dm6 reference genome into TAS regions including *X*-TAS (9872 bp, L03284) (Karpen and Spradling 1992), *2R*-TAS (Chr 2R: 25,258,060–25,261,551, 3492 bp) and *3R*-TAS (Chr 3R: 32,073,015–32,079,331, 6317 bp) (Yin and Lin 2007), *2L*-TAS (Chr 2L: 1–5041, 5041 bp) and *3L*-TAS (Chr 3L: 1–19,608, 19,608 bp) (Walter et al. 1995), and non-TAS regions (everywhere else). To determine if *P*-derived reads that did not align to non-TAS regions corresponded to insertions in TAS, they were aligned to the TAS reference using Bowtie 2 outputting all valid alignments (-a). A read was considered aligned to TAS if the edit distance was fewer than four. For each DGRP genome, we calculated a *Z*-score for TAS-aligned reads according to the following formula: $Z = (x - \mu) / \sigma$, where *x* is the number read pairs aligned to *X*, *2R*, or *3R*-TAS; μ

is the average number of reads supporting individual non-TAS *P*-element insertions in a given genome; and σ is the standard deviation for reads supporting non-TAS insertions. A significance level $\alpha = 0.05$ ($Z = \pm 1.96$) was used to estimate the number of *P*-elements in TAS in each DGRP genome (Supplemental Table S5).

To determine which TAS arrays (*X*, *2R*, or *3R*-TAS) contained a *P*-element insertion in each DGRP genome, we first calculated the edit distance for all reported alignments of each read pair in that genome. We then assigned each read pair to the TAS array to which it aligned with lowest edit distance. For DGRP genomes with one *P*-element in TAS ($-1.96 < Z < 1.96$), the insertion was predicted to occur in the TAS array whose number of supporting reads $>2\times$ the reads supporting the other two TAS arrays. For DGRP genomes with more than one *P*-element in TAS ($1.96 < Z$), we sought to determine the locations of two *P*-elements. The first insertion was predicted to occur in the TAS array supported by the highest number of reads. Then, we subtracted the average number of reads supporting a non-TAS *P*-element insertion in the given DGRP genome from the reads supporting the first TAS insertion. The second insertion was predicted the same way as DGRP genomes with one *P*-element. The predicted *P*-element locations are provided in Supplemental Table S6.

Localizing insertion sites of *P*-element insertions in TAS

A read pair may be equally well aligned to several homologous satellite repeats within a TAS array. Therefore, for *2R* and *3R*-TAS, we assigned *P*-elements to consensus sequences. Similarly for *X*-TAS, we were unable to determine whether a given insertion occurred in Repeat B, C, or D, so we arbitrarily assigned all insertions to Repeat B. We then identified the insertion breakpoint supported by the most split reads.

As an alternative approach, we also constructed pseudo-genomes for each alternative TAS insertion site in a given DGRP genome, which included the *P*-element consensus sequence flanked at each end by an 8-nt target site duplication and 500 nt of adjacent TAS sequence. Paired-end reads were aligned to the constructed pseudo-genomes (MAPQ ≥ 20), and the breakpoint corresponding to the pseudo-genome with the most reads aligned was identified. Corresponding *P*-element insertion sites are provided in Supplemental Table S6.

PCR verification of insertion sites

A *P*-element-specific and a TAS-specific primer were used (Supplemental Table S9) to amplify TAS insertions from DGRP genomic DNA. Because multiple bands were generally produced, owing to alternative annealing of the TAS primer to multiple repeats, the main band was purified by gel extraction using the QIAGEN MinElute Gel Extraction Kit (Qiagen 28606) and sequenced to determine the breakpoint. PCR conditions are provided in the Supplemental Table S6.

Recombination rates

Recombination rates at *P*-element insertion sites were identified from the genome-wide map provided by Cameron et al. (2012). Because these rates were based on the release 5 of *D. melanogaster* reference genome, we converted our annotated *P*-element insertions in release 6 coordinates to release 5 on the FlyBase (<http://flybase.org>). The recombination rate of insertions that did not have release 5 counterparts was assumed to be 0, because the major improvement of release 6 relative to release 5 is the assembly of heterochromatin regions (Dos Santos et al. 2015; Hoskins et al. 2015).

Data analysis

Annotating piRNA clusters and identifying *P*-element insertions were powered by the high-performance computing resources from the Center for Advanced Computing and Data Science (CACDS) at the University of Houston (<https://uh.edu/rcdc/resources/hpc/>). All statistical analyses were performed in R (version 3.3.1) (R Core Team 2016). Graphs were made in RStudio (<https://www.rstudio.com>) with R packages ggplot2 (version 2.2.1; <https://ggplot2-book.org/>), gplots (version 3.0.1; <https://cran.r-project.org/web/packages/gplots/index.html>), reshape2 (version 1.4.3; <https://rdrr.io/cran/reshape2/>), and cowplot (version 0.7.0; <https://wilkelab.org/cowplot/>).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI BioProject database (BioProject; <https://www.ncbi.nlm.nih.gov/bioproject>) under accession number PRJNA490147. All scripts required to implement the analysis described in this manuscript can be found in the Supplemental Code, as well as GitHub (non-TAS insertions: <https://github.com/szhang32/PIDFE>; TAS insertions: https://github.com/szhang32/TAS_P_insertion).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

S.Z., B.P., and E.S.K. and this research were supported by a National Science Foundation Division of Environmental Biology (NSF-DEB) award to E.S.K. (NSF-DEB #1457800). We thank members of the Kelleher laboratory for helpful discussion and three anonymous reviewers for their comments on this manuscript.

References

- Ajioka JW, Eanes WF. 1989. The accumulation of *P*-elements on the tip of the *X* chromosome in populations of *Drosophila melanogaster*. *Genet Res* **53**: 1–6. doi:10.1017/S0016672300027798
- Anxolabéhère D, Kidwell MG, Periquet G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Mol Biol Evol* **5**: 252–269. doi:10.1093/oxfordjournals.molbev.a040491
- Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747. doi:10.1126/science.1142612
- Asif-Laidin A, Delmarre V, Laurentie J, Miller WJ, Ronsseray S, Teyssset L. 2017. Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in *Drosophila*. *DNA Res* **24**: 459–472. doi:10.1093/dnares/dsx017
- Berg CA, Spradling AC. 1991. Studies on the rate and site-specificity of *P* element transposition. *Genetics* **127**: 515–524.
- Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol* **7**: R112. doi:10.1186/gb-2006-7-11-r112
- 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**: 3–14. doi:10.1017/S0016672300028822
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**: 1089–1103. doi:10.1016/j.cell.2007.01.043
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392. doi:10.1126/science.1165171

- Charlesworth B, Langley CH. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* **112**: 359–383.
- Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet* **23**: 251–287. doi:10.1146/annurev.ge.23.120189.001343
- Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet* **8**: e1002905. doi:10.1371/journal.pgen.1002905
- 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**: 339–355.
- de Vanssay A, Bougé AL, Boivin A, Hermant C, Teyssset L, Delmarre V, Antoniewski C, Ronsseray S. 2012. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* **490**: 112–115. doi:10.1038/nature11416
- Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**: 2169–2177. doi:10.1534/genetics.107.082743
- Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, the FlyBase Consortium. 2015. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res* **43**: D690–D697. doi:10.1093/nar/gku1099
- Dotto BR, Carvalho EL, Silva AF, Duarte Silva LF, Pinto PM, Ortiz MF, Wallau GL. 2015. HTT-DB: horizontally transferred transposable elements database. *Bioinformatics* **31**: 2915–2917. doi:10.1093/bioinformatics/btv281
- Eggleston WB, Schlitz DMJ, Engels WR. 1988. P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. *Nature* **331**: 368–370. doi:10.1038/331368a0
- Girard A, Hannon GJ. 2008. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* **18**: 136–148. doi:10.1016/j.tcb.2008.01.004
- Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**: 82–85. doi:10.1038/nature05388
- Hermant C, Boivin A, Teyssset L, Delmarre V, Asif-Laidin A, Van Den Beek M, Antoniewski C, Ronsseray S. 2015. Paramutation in *Drosophila* requires both nuclear and cytoplasmic actors of the piRNA pathway and induces cis-spreading of piRNA production. *Genetics* **201**: 1381–1396. doi:10.1534/genetics.115.180307
- Hill T, Schlötterer C, Betancourt AJ. 2016. Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the *P*-element. *PLoS Genet* **12**: e1005920. doi:10.1371/journal.pgen.1005920
- Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. 2015. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res* **25**: 445–458. doi:10.1101/gr.185579.114
- Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**: 69–82. doi:10.1016/j.cell.2007.03.026
- Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, et al. 2014. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res* **24**: 1193–1208. doi:10.1101/gr.171546.113
- Itoh M, Takeuchi N, Yamaguchi M, Yamamoto M-T, Boussy IA. 2007. Prevalence of full-size *P* and *KP* elements in North American populations of *Drosophila melanogaster*. *Genetica* **131**: 21–28. doi:10.1007/s10709-006-9109-2
- Jensen PA, Stuart JR, Goodpaster MP, Goodman JW, Simmons MJ. 2008. Cytotype regulation of *P* transposable elements in *Drosophila melanogaster*: repressor polypeptides or piRNAs? *Genetics* **179**: 1785–1793. doi:10.1534/genetics.108.087072
- Karpen GH, Spradling AC. 1992. Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome *Dp1187* by single *P* element insertional mutagenesis. *Genetics* **132**: 737–753.
- Kelleher ES. 2016. Reexamining the *P*-element invasion of *Drosophila melanogaster* through the lens of piRNA silencing. *Genetics* **203**: 1513–1531. doi:10.1534/genetics.115.184119
- Kelleher ES, Azevedo RBR, Zheng Y. 2018. The evolution of small-RNA-mediated silencing of an invading transposable element. *Genome Biol Evol* **10**: 3038–3057. doi:10.1093/gbe/evy218
- Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to *P* element transposon invasion in *Drosophila melanogaster*. *Cell* **147**: 1551–1563. doi:10.1016/j.cell.2011.11.042
- Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci* **80**: 1655–1659. doi:10.1073/pnas.80.6.1655
- Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **86**: 813–833.
- Kidwell MG, Frydrk T, Novy J. 1983. The hybrid dysgenesis potential of *Drosophila melanogaster* from diverse temporal and geographic origins. *Drosoph Inf Serv* **59**: 63–69.
- Kimura K, Kidwell MG. 1994. Differences in *P* element population dynamics between the sibling species *Drosophila melanogaster* and *Drosophila simulans*. *Genet Res* **63**: 27–38. doi:10.1017/S0016672300032055
- Kofler R. 2019. Dynamics of transposable element invasions with piRNA clusters. *Mol Biol Evol* **36**: 1457–1472. doi:10.1093/molbev/msz079
- Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet* **8**: e1002487. doi:10.1371/journal.pgen.1002487
- 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* **112**: 6659–6663. doi:10.1073/pnas.1500758112
- Kofler R, Senti KA, Nolte V, Tobler R, Schlötterer C. 2018. Molecular dissection of a natural transposable element invasion. *Genome Res* **28**: 824–835. doi:10.1101/gr.228627.117
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923
- Lee YCG, Langley CH. 2012. Long-term and short-term evolutionary impacts of transposable elements on *Drosophila*. *Genetics* **192**: 1411–1432. doi:10.1534/genetics.112.145714
- Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Toth KF, et al. 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* **28**: 1667–1680. doi:10.1101/gad.245514.114
- Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509–521. doi:10.1016/j.cell.2009.04.027
- Linheiro RS, Bergman CM. 2012. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS One* **7**: e30008. doi:10.1371/journal.pone.0030008
- Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res* **20**: 212–227. doi:10.1101/gr.095406.109
- Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* **482**: 173–178. doi:10.1038/nature10811
- Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**: 522–535. doi:10.1016/j.cell.2009.03.040
- Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabéhère D, Ronsseray S. 2000. *P*-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric *P* copy. *Genetics* **155**: 1841–1854.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10–12. doi:10.14806/ej.17.1.200
- Moon S, Cassani M, Lin YA, Wang L, Dou K, Zhang ZZ. 2018. A robust transposon-endogenizing response from germline stem cells. *Dev Cell* **47**: 660–671.e3. doi:10.1016/j.devcel.2018.10.011
- Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet* **39**: 197–218. doi:10.1146/annurev.genet.39.073003.112420
- Ogura K, Woodruff RC, Itoh M, Boussy IA. 2007. Long-term patterns of genomic *P* element content and P-M characteristics of *Drosophila melanogaster* in eastern Australia. *Genes Genet Syst* **82**: 479–487. doi:10.1266/ggs.82.479
- O'Hare K, Rubin GM. 1983. Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34**: 25–35. doi:10.1016/0092-8674(83)90133-2
- Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci* **114**: 4721–4726. doi:10.1073/pnas.1621178114
- Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol* **28**: 1633–1644. doi:10.1093/molbev/msq337
- Rahman R, Chirn GW, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, Lau NC. 2015. Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. *Nucleic Acids Res* **43**: 10655–10672. doi:10.1093/nar/gkv1193
- R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>.

- Robert V, Prud'homme N, Kim A, Bucheton A, Péliisson A. 2001. Characterization of the *flamenco* region of the *Drosophila melanogaster* genome. *Genetics* **158**: 701–713.
- Ronsseray S, Lehmann M, Anxolabéhère D. 1989. Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations. *Chromosoma* **98**: 207–214. doi:10.1007/BF00329685
- 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.
- Ronsseray S, Lehmann M, Nouaud D, Anxolabéhère D. 1996. The regulatory properties of autonomous subtelomeric P elements are sensitive to a *Suppressor of variegation* in *Drosophila melanogaster*. *Genetics* **143**: 1663–1674.
- Rosenkranz D, Zischler H. 2012. proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* **13**: 5. doi:10.1186/1471-2105-13-5
- 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. doi:10.1371/journal.pgen.1004138
- Song J, Liu J, Schnakenberg S, Ha H, Xing J, Chen KC. 2014. Variation in piRNA and transposable element content in strains of *Drosophila melanogaster*. *Genome Biol Evol* **6**: 2786–2798. doi:10.1093/gbe/evu217
- Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci* **108**: 15948–15953. doi:10.1073/pnas.1112960108
- 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. doi:10.1534/g3.117.040634
- Stuart JR, Haley KJ, Swedzinski D, Lockner S, Kocian PE, Merriman PJ, Simmons MJ. 2002. Telomeric P elements associated with cytotype regulation of the P transposon family in *Drosophila melanogaster*. *Genetics* **162**: 1641–1654.
- Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol* **2**: 656–664. doi:10.1093/gbe/evq050
- Walter MF, Jang C, Kasravi B, Donath J, Mechler BM, Mason JM, Biessmann H. 1995. DNA organization and polymorphism of a wild-type *Drosophila* telomere region. *Chromosoma* **104**: 229–241. doi:10.1007/bf00352254
- Yin H, Lin H. 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* **450**: 304–308. doi:10.1038/nature06263
- Zhang S, Kelleher ES. 2017. Targeted identification of TE insertions in a *Drosophila* genome through hemi-specific PCR. *Mob DNA* **8**: 10. doi:10.1186/s13100-017-0092-1
- Zhuang J, Wang J, Theurkauf W, Weng Z. 2014. TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res* **42**: 6826–6838. doi:10.1093/nar/gku323

Received April 16, 2019; accepted in revised form March 24, 2020.