piRNA-mediated silencing of an invading transposable element evolves rapidly through abundant beneficial *de novo* mutations

Shuo Zhang[1] and Erin S. Kelleher[1,*]

[1]Department of Biology and Biochemistry, University of Houston, 3455 Cullen Blvd. Suite 342, Houston, TX 77204

*Correspondence: eskelleher@uh.edu

ABSTRACT

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 20th 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* Genetic Reference Panel (DGRP) contain at least one *P*-element insertion in a piRNA cluster indicating repressor alleles are exceptionally common. Furthermore, in our sample of ~200 genomes, we uncovered no fewer than 84 unique *P*-element insertion alleles in at least 15 different piRNA clusters. Therefore, the ubiquitous repressive phenotype is underpinned by a plethora of alternate repressor alleles. Finally, we demonstrate that *P*-element insertions in piRNA clusters exhibit elevated polymorphic frequencies, consistent with positive selection. Our results highlight how the unique genetic architecture of piRNA production, in which numerous piRNA clusters can encode regulatory small RNAs upon transpositional insertion, facilitates the rapid evolution of repression. They furthermore provide one of the most striking examples of polygenic adaptation yet documented, in which alternative insertions into multiple piRNA clusters are targets of positive selection.

INTRODUCTION

Transposable elements (TEs) are widespread genomic parasites that increase their copy number by mobilizing and self-replicating within their host genomes. TEs impose a severe mutational load on their hosts by producing deleterious insertions that disrupt functional sequences (Levis et al. 1984; McGinnis et al. 1983), causing DNA damage through encoded endonucleases (Gasior et al. 2006), and mediating ectopic recombination leading to structural rearrangements (Lim 1988). 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 (Houwing et al. 2007; Brennecke et al. 2007; Aravin et al. 2007; Girard and Hannon 2008).

On evolutionary time scales, TEs are frequently horizontally transferred between non-hybridizing species, allowing TE families to colonize new host genomes (Thomas et al. 2010; Dotto et al. 2015; Peccoud et al. 2017). Although host regulation of endogenous TEs by piRNAs is ubiquitous, 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 inserts into a piRNA producing locus referred to as a piRNA cluster (Khurana et al. 2011; Girard and Hannon 2008). The existence of numerous alternative piRNA clusters (e.g., 142 loci or ~3.5% of assembled *D. melanogaster*

genome based on Brennecke et al. 2007) may facilitate the evolution of repression by increasing the mutation rate to generate repressors (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 bi-directional transcription and piRNA production (de Vanssay et al. 2012; Le Thomas 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, while 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 non-cluster insertions, although this is based on modest sample size and read depth (Lu and Clark 2010).

The recent invasion of both *Drosophila melanogaster* and *Drosophila simulans* by *P*-element DNA transposons (Kidwell 1983; Anxolabéhère et al. 1988; Kofler et al. 2015) provides a unique opportunity to study not only the contributions of *de novo* mutation to the evolution of piRNA-mediated silencing by resolving the location of piRNA clusters both before and after an invasion event, but also evolutionary dynamics of repressors when selection is most strong. Unlike most TEs that have colonized host genome for a long evolutionary time, *P*-elements invaded the *D. melanogaster* genome around 1950 by horizontal transfer from *D. willistoni* (Daniels et al. 1990; Kidwell 1983; Anxolabéhère et al. 1988). Similarly, *D. simulans* acquired *P*-elements from *D. melanogaster* around 2010 (Kofler et al. 2015). In response, many natural populations of *D. melanogaster* evolved piRNA-mediated repression in less than 50 years (Jensen et al. 2008; Brennecke et al. 2008; Kidwell 1983). However, numerous strains collected prior to both invasions are retained in laboratories and stock centers, providing a historical record of ancestral piRNA clusters that were active before the *P*-element invasion.

Here, we take advantage of ~200 fully sequenced *D. melanogaster* genomes, comprising the *Drosophila* Genetic Reference Panel (DGRP) (Mackay et al. 2012; Huang et al. 2014), to study the emergence and evolutionary dynamics of piRNA-mediated repressor alleles after the *P*-element invasion into *D. melanogaster* populations. To differentiate *de novo* insertions into ancestral piRNA clusters from

novel piRNA clusters, we identified piRNA clusters in *D. melanogaster* from 9 *P*-element free strains of *D. melanogaster* collected before invasion. Furthermore, to empower the identification of repressor alleles, we developed a novel approach to identify TE insertions in repetitive DNA. We show that more than 90% of DGRP lines have at least one *P*-element in an ancestral piRNA cluster, indicating *P*-element repressors are widespread in natural populations. Moreover, we detected no fewer than 84 independent *P*-element insertions in ancestral piRNA clusters, suggesting an exceptionally high *de novo* mutation rate for the formation of piRNA-mediated repressor alleles. Finally, we observed that *P*-element insertions in piRNA clusters segregate at higher frequency than putatively neutral insertions in similar genomic compartments, indicating they are targets of positive selection. Together, our results reveal a striking example of adaptation in a polygenic system, in which a plethora of *de novo* beneficial *P*-element insertions into piRNA clusters fueled the evolution of a ubiquitous repressive phenotype in <60 years.

RESULTS

Identification of ancestral piRNA clusters

We first sought to annotate ancestral piRNA clusters in the *D. melanogaster* genome. We took advantage of 27 small RNA sequencing libraries from 9 wild-type strains (Supplemental Table S1), which were isolated from nature prior to *P*-element invasion and are therefore devoid of genomic *P*-elements. Using proTRAC (Rosenkranz and Zischler 2012), we annotated piRNA clusters based on the density of mapped piRNAs. By varying the density of reads required to annotate a piRNA

cluster (pdens = 0.01, 0.05, and 0.10), we generated three sets of annotations, which contained 32, 159 and 497 piRNA clusters, and comprised 0.30%, 1.27 %, 3.68% of the assembled *D. melanogaster* genome, respectively (Fig. 1A; Supplemental Table S2, S3).

We identified some genomic loci that differ in their status as a piRNA cluster between genotypes, producing abundant piRNAs in some strains while remaining quiescent in others (Fig. 1A-B; Supplemental Table S2, S3). We therefore defined ancestral piRNA clusters as genomic regions that were annotated from at least one small-RNA library. Notably, major known piRNA clusters such as *flamenco* and *42AB* (Robert et al. 2001; Malone et al. 2009; Li et al. 2009; Brennecke et al. 2007) produced abundant piRNAs in all genotypes, and were annotated as piRNA clusters regardless of stringency (Fig. 1A; Supplemental Table S2, S3). 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.

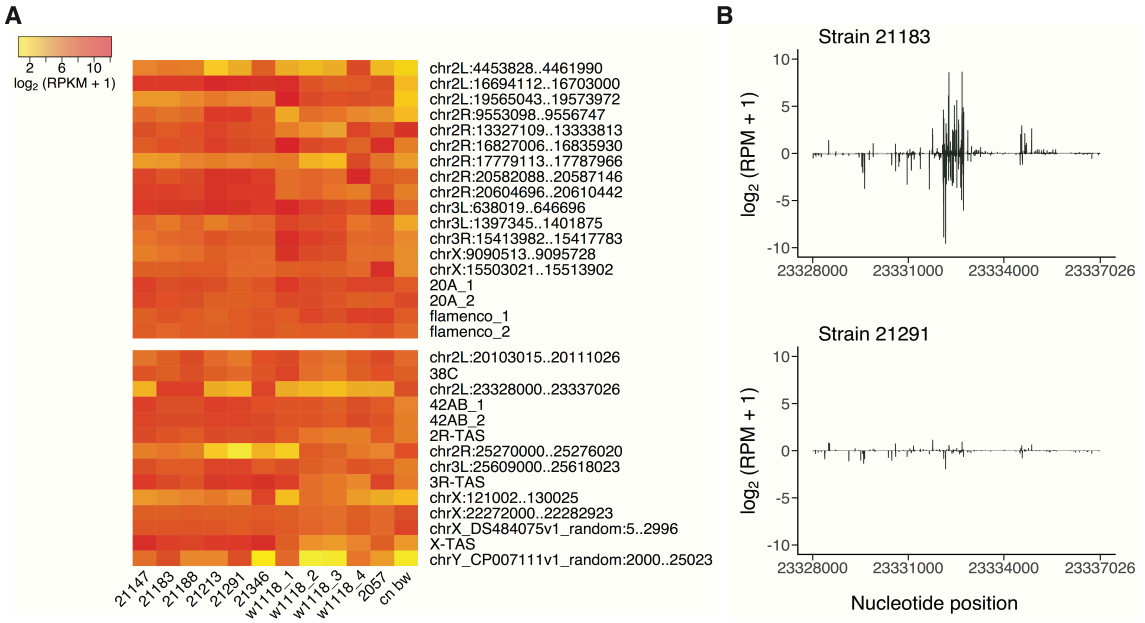

Figure 1. (A) Activity of 32 piRNA clusters in ancestral (*P*-element free) strains of *D. melanogaster*. Each column represents a small RNA sequencing library (biological replicates are combined) and each row represents a piRNA cluster annotated in at least one of these libraries. Coordinates of piRNA 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 ($\log_2$ (RPKM + 1)). Clusters above the white line are uni-strand piRNA clusters and ones 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 (23328000..23337026). 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 ($\log_2$ (RPM + 1)). Positive value represent piRNAs mapped to sense strand of the reference genome, and negative value represent piRNAs from antisense strand.

Most North American Genotypes have *P* elements in ancestral piRNA clusters

To identify *P*-element insertions in extant populations, we took advantage of the published genomes from the DGRP, which includes more than 200 fully sequenced inbred lines collected from North America in 2003 (Mackay et al. 2012). All DGRP genomes are known to harbor *P*-elements (Zhuang et al. 2014; Rahman et al. 2015) (Supplemental Fig. S1) and the majority of them are expected to repress *P*-

element activity (Kidwell et al. 1983; Anxolabéhère et al. 1988; Kidwell 1983). Although previous annotations suggest that less than 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 alignment to the reference genome prohibits the identification TE insertions in repeat-rich piRNA clusters. This is particularly problematic for identifying TE insertions in telomeric associated sequence (TAS), piRNA clusters comprised of subtelomeric satellite repeats (Yin and Lin 2007; Asif-Laidin et al. 2017). Indeed, ~50% of wild-derived genomes are believed to harbor a *P*-element insertion in *X*-TAS (Ajioka and Eanes 1989; Ronsseray et al. 1989; Biémont et al. 1990), making the identification of *P*-element insertions in this piRNA cluster of particular significance. We therefore developed an alternative approach to annotate *P*-elements among DGRP genomes.

First, we annotated *P*-element insertion sites based on high-quality alignments of split reads (mapping quality score, MAPQ ≥ 20), which are not necessarily unique, yet still support a particular genomic location with high confidence. Including high-quality non-unique alignments increases the number of annotated *P*-elements by 44% and 24% increase when compared to TEMP and TIDAL, respectively, two approaches that rely on unique alignments (Zhuang et al. 2014; (Rahman et al. 2015)Fig. 2A; Supplemental Table S4). While we did not validate these new insertions, the two additional insertions we identified in DGRP492 were also detected by previous study using hemi-specific PCR, indicating they are true insertions (Zhang and Kelleher 2017).

Despite relaxing the requirement for unique alignment, we still identified only 9 DGRP genomes (4.5%) with *P*-elements in *X*-TAS. High-quality alignments likely fail to provide a unique insertion site in TAS repeats because highly similar tandem satellite sequences provide multiple equivalent alignments (Fig. 3A). Therefore, we first sought to detect TAS insertions by identifying *P*-derived reads that aligned only to TAS repeats. 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), while only 3 DGRP genomes contained *P*-derived reads aligning to *2L* and *3L*-TAS.

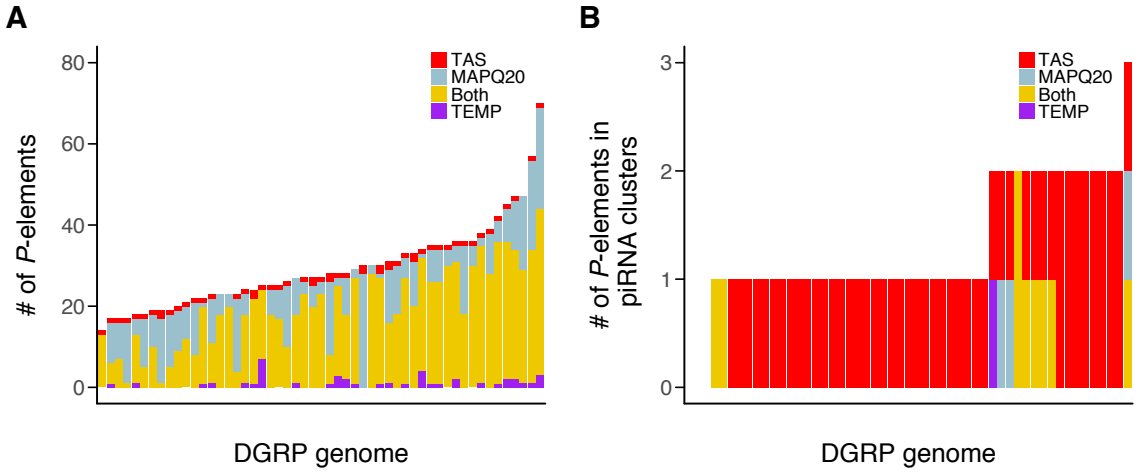

Figure 2. (A) Total number of *P*-elements and (B) number of *P*-elements in ancestral piRNA clusters annotated by different approaches. 53 DGRP genomes that were previously annotated by TEMP are compared. To identifying piRNA cluster insertions, the 32 cluster high confidence set was used. TEMP: insertions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

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 11 DGRP genomes that harbor no *P*-element insertions (6%, $Z < -1.96$), 137 DGRP genomes that harbor one *P*-element insertion (70%, $-1.96 < Z < 1.96$), and 47 genomes that carry two or more insertions into TAS arrays (24%, $1.96 < Z$) (Fig. 3B; Supplemental Table S5). Given that TAS arrays are ancestral piRNA clusters that are active in *P*-element free strains (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).

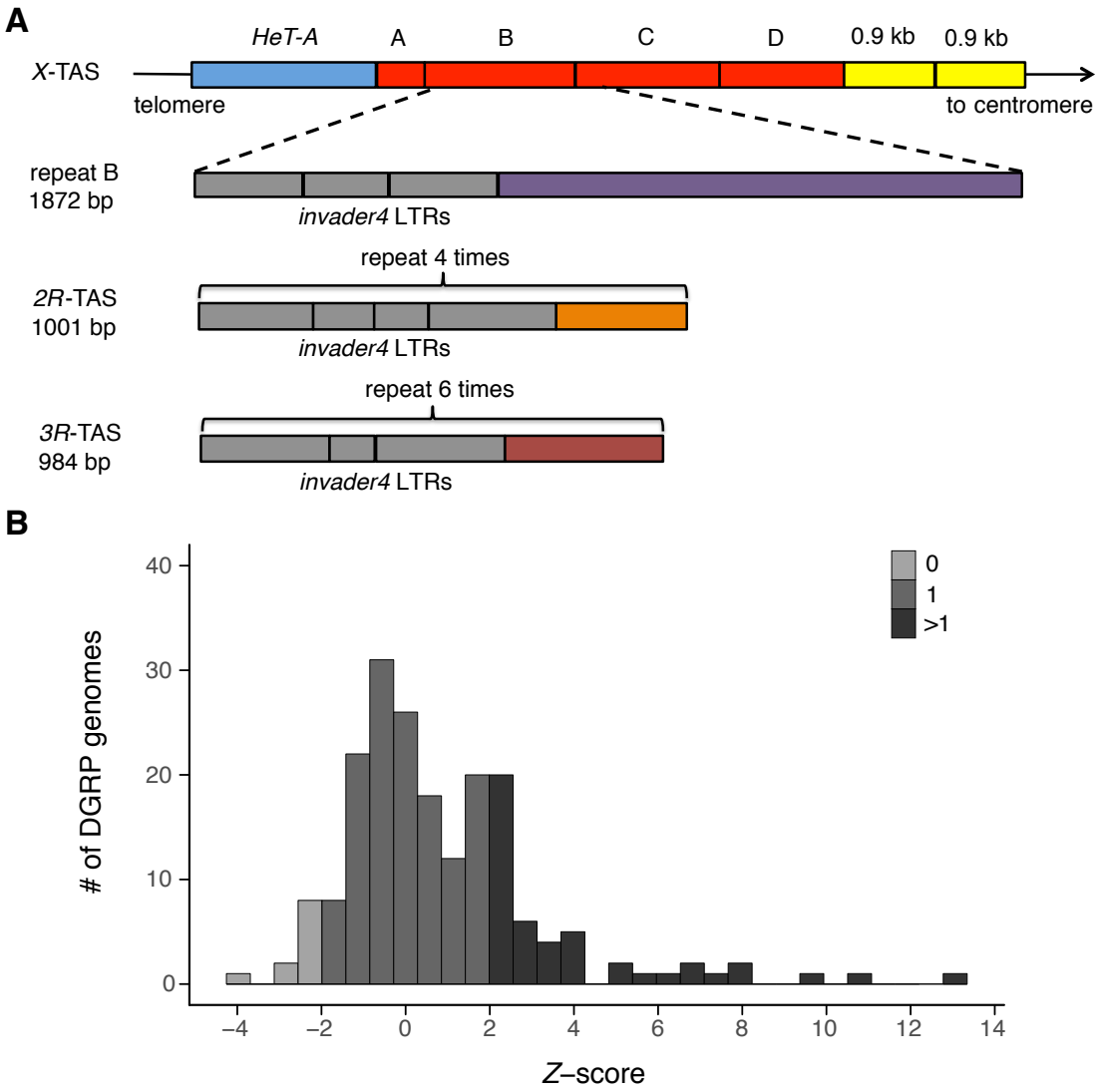

Figure 3. (A) The structure of TAS arrays (Asif-Laidin et al. 2017). *X*-TAS contains four tandem repeats (A, B, C and D in red) located between a *HeT-A* retrotransposon and two 0.9 kb repeats (Karpen and Spradling 1992). Repeat A is degenerated. Repeat B, C and D are ~1.8 kb in length and highly similar to each other (>95% identity). Repeat B (enlarged below) is compared to *2R*-chromosome TAS (*2R*-TAS) and *3R*-chromosome TAS (*3R*-TAS), which are represented by 4 and 6 copies in the assembled *D. melanogaster* genome, respectively. Each repeat of *2R*, *3R* and *X*-TAS also contains several subrepeats that are composed of *invader4* retrotransposon long terminal repeats (LTRs)(gray; Bergman et al. 2006). In addition to *invader4* LTRs, the *X, 2R* and *3R*-TAS repeats share other short homologous fragments (41-131 bp) with 909 bp being unique to the *X*-TAS repeat (not shown in the figure; Asif-Laidin et al. 2017). (B) The distribution of *Z*-scores among DGRP lines. 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.

Abundant repressors underpin repressive phenotype

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 read pairs whose best alignment supported an insertion in *X*, *2R* or *3R*-TAS (see methods). For the 82% of DGRP genomes for which we identified at least one candidate TAS array harboring a *P*-element insertion, we further identified the insertion site that was supported by the most read pairs (see methods). In addition, based on alternative breakpoints identified by alignments to TAS sequences, we also determined which of multiple alternate pseudo-genomes, containing *P*-element insertions into different sites, was supported by the most reads (see methods). Due to sequence homology among repeats within the same TAS array (>95% identity; Fig. 3A), we assumed all homologous insertion sites among tandem repeats corresponded to a single insertion event for these analyses.

Among 92 DGRP genomes, we found 102 *P*-element insertions into TAS where the best insertion site identified by reference genome and pseudo-genome alignments agreed, suggesting well-supported insertion sites (Supplemental Table S6). These corresponded to 43 unique insertion sites, 32 of which we were able to verify by site-specific PCR (74.4%). For the remaining 11 insertions, PCR revealed that two were located at different sites, PCR failed for seven sites, and PCR was not attempted for two sites. We further attempted PCR to determine the insertion sites in 14 DGRP genomes where the two computational methods did not agree, and 73 DGRP genomes where *P*-elements could not be assigned to a particular TAS or breakpoints could not be determined due to an absence of split reads. These PCRs

determined an additional 40 *P*-element insertion sites in 71 DGRP genomes (Supplemental Table S6).

In total, we identified 89 independent insertions of *P*-elements into TAS sequences (*2R*, *3R* or *X*-TAS), 84 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; Ronsseray et al. 1989; Biémont et al. 1990), we found that >50% of DGRP genomes had *P*-element insertions in *X*-TAS and ~20% DGPR genomes had *P*-elements in *2R* and *3R*-TAS (Table 1; Supplementary Table S6). Moreover, we discovered a multitude of insertion alleles in each TAS array: 20 in *2R*-TAS, 19 in *3R*-TAS and 50 in *X*-TAS (Table 1; Fig. 4A, B).

Table 1. *P*-element insertions in TAS

TAS array	# of genomes with insertions	# of alleles
2R-TAS	33	20 (20)
3R-TAS	38	19 (18)
X-TAS	133	50 (46)
2L and *3L*-TAS	1	1 (0)

Numbers in parentheses indicate PCR verified alleles.

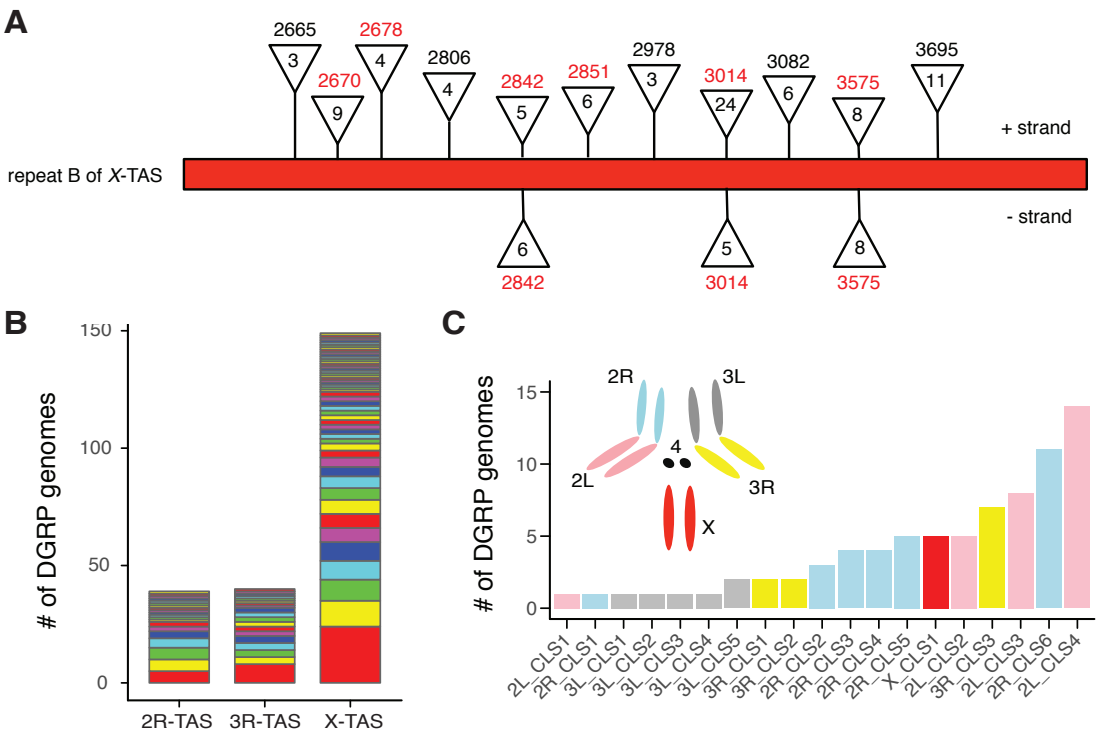

Figure 4. (A) Multiple *P*-element insertion sites were detected in *X*-TAS. Due to the high sequence similarity between repeats B, C and D of *X*-TAS, we arbitrarily assigned all *P*-element insertions in these repeats to repeat B. Only insertions present in more than 2 DGRP genomes are depicted. For each insertion, the breakpoint is labeled above the triangle and the number of DGRP genomes containing the insertion is indicated inside the triangle. Breakpoints positions in red correspond to insertion hotspots (Karpen and Spradling 1992). Sense *P*-elements, whose orientation is same as *X*-TAS were drawn above repeat B, whereas antisense *P*-elements were located below repeat B. (B) Multiple *P*-element insertion sites were detected in *2R*, *3R* and *X*-TAS. Each color represents a unique *P*-element insertion. (C) *P*-elements are present in non-TAS ancestral piRNA clusters across all major chromosome arms, based on our annotation set of 159 piRNA clusters.

P-elements preferentially insert into sequence-specific sites in *X*-TAS (Karpen and Spradling 1992), which are also found in *2R* and *3R*-TAS. We therefore wondered whether *P*-element insertions into these hotspots were unusually common in among TAS insertion alleles from natural populations. Indeed, we found that these sites were greatly enriched for *P*-element insertion alleles: 88.2% (15 out of 17) of hotspots had a *P*-element insertion allele, as compared to only 1.5% (58 out of 3840) of non-preferred sites (Pearson's Chi-square test, $\chi^2 = 639.65$, P-value <

2.20x10^{-16}). Additionally, hotspots were more likely to have two distinguishable insertion alleles, one in each strand, when compared to non-preferred sites (Fig. 5A; Pearson's Chi-square test, χ^2 = 18.92, P-value = 1.36x10^{-5}). Finally, individual *P*-element insertions in hotspots were more common among DGRP chromosomes than those occurring at non-preferred sites (Fig. 5B; Wilcoxon rank sum test, $W_{25, 64}$ = 1259, P-value = 5.88x10^{-6}), suggesting that recurrent insertion into these positions elevates the frequency of these insertion alleles. Taken together, these results suggest that the exceptional abundance of *P*-elements insertions in TAS arrays is at least partially explained by an insertion site preference.

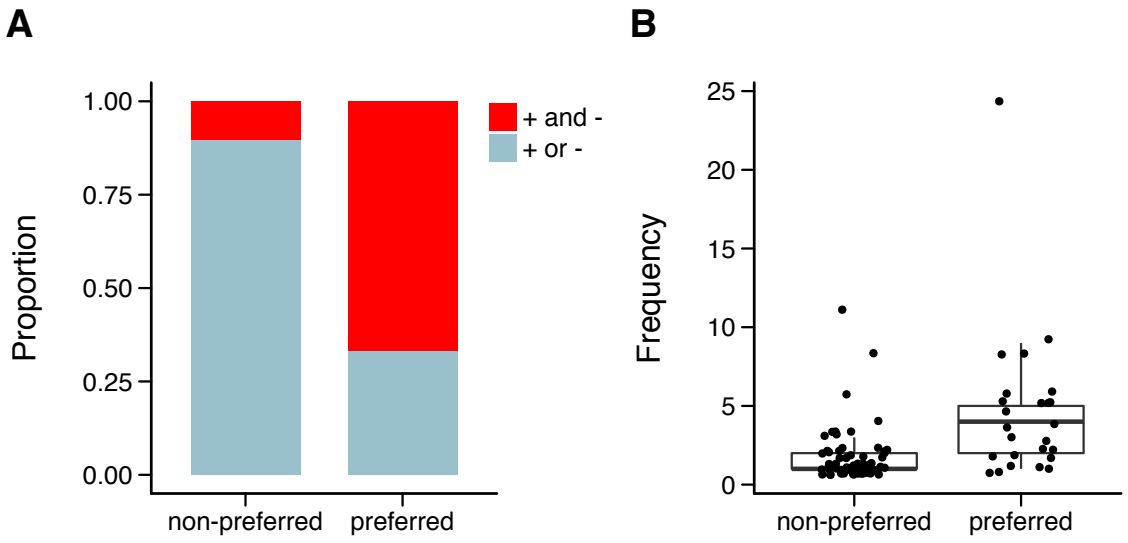

Figure 5. (A) The proportion of *P*-element insertions that insert into both sense and antisense strand of TAS (red), or single strand (blue) for non-preferred and preferred sites. (B) The comparison between frequencies of *P*-elements at non-preferred and preferred sites. *P*-elements inserted at same site but in opposite orientations were considered different insertions.

Cluster *P*-element insertions are targets of positive selection

Combining the TAS insertion alleles with those identified in non-TAS piRNA

clusters, we detected up to 193 *P*-element insertion events into at least 15 (up to 37) different ancestral piRNA clusters, which are located on all of the major chromosome arms of the *Drosophila* genome (Fig. 4C; Supplementary Figure S2; Supplemental Table S7). To determine whether *P*-element insertions in ancestral piRNA clusters are targets of positive selection, we considered their site frequency spectrum. Positive selection is expected to increase the frequency of beneficial alleles in natural populations when compared to neutral alleles (Nielsen 2005). However, this observation is potentially confounded by the recurrent insertion of *P*-elements into known insertion hotspots in TAS sequences, which elevates their frequencies (Fig. 5B). We therefore excluded *P*-element insertions in hotspots from our analysis of the site frequency spectrum. Consistent with positive selection, we found that *P*-element insertions in ancestral piRNA clusters segregate at higher frequencies than those outside of piRNA clusters (Fig. 6A).

piRNA clusters occur predominantly in heterochromatic regions of low recombination (Brennecke et al. 2007), meaning that differences in the site-frequency spectra of cluster and non-cluster *P*-element insertions are also potentially confounded by differences in nature and efficacy of selection in different genomic compartments (Hill and Robertson 1966; Haddrill et al. 2007). In particular, heterochromatic insertions are less likely to disrupt functional sequences or participate in ectopic recombination, making them less deleterious than those in euchromatin (Bartolomé and Maside 2004; Petrov et al. 2011; Kofler et al. 2012). However, when we restrict our comparison of cluster and non-cluster insertions to regions of low recombination where heterochromatin resides (≤1 cM/Mb) we

observe that the elevated frequency of cluster insertions becomes more pronounced (Fig. 6B; Supplemental Fig. S3B). This suggests that the relatively higher frequency of *P*-element insertions in ancestral piRNA clusters reflects positive selection, rather than reduced purifying selection against TE insertions in heterochromatin. Furthermore, the difference in frequency spectra between cluster and non-cluster *P* element insertions decreases when we include insertions in lower confidence piRNA clusters (Supplemental Fig. S3, S4), suggesting that the inclusion of false-positives (*i.e.* insertions into incorrectly annotated piRNA clusters) dampens the signature of positive selection. Alternatively, this pattern could result from stronger positive selection for highly-expressed piRNA clusters, which are over-represented among high-confidence piRNA clusters.

A second important distinction between selection in heterochromatin and euchromatin lies in the efficacy of positive selection, which is reduced in regions of low recombination by linked deleterious variation ('Hill-Robertson effects': Hill and Robertson 1966). Although the sample size of piRNA cluster insertion in regions of high recombination (>1 cM/Mb) is small (n = 20 for high confidence piRNA clusters), we did not observe that they segregate at higher frequencies than non cluster insertions (Fig. 6C; Supplemental Fig. S3C, S4C). While these observations are not consistent with Hill-Robertson effects, they are consistent with the theoretical observation that reduced recombination increases the efficiency of selection on TE repressors by maintaining linkage between repressor alleles and the chromosomal sites they've protected from mutational load (Charlesworth and Langley 1986).

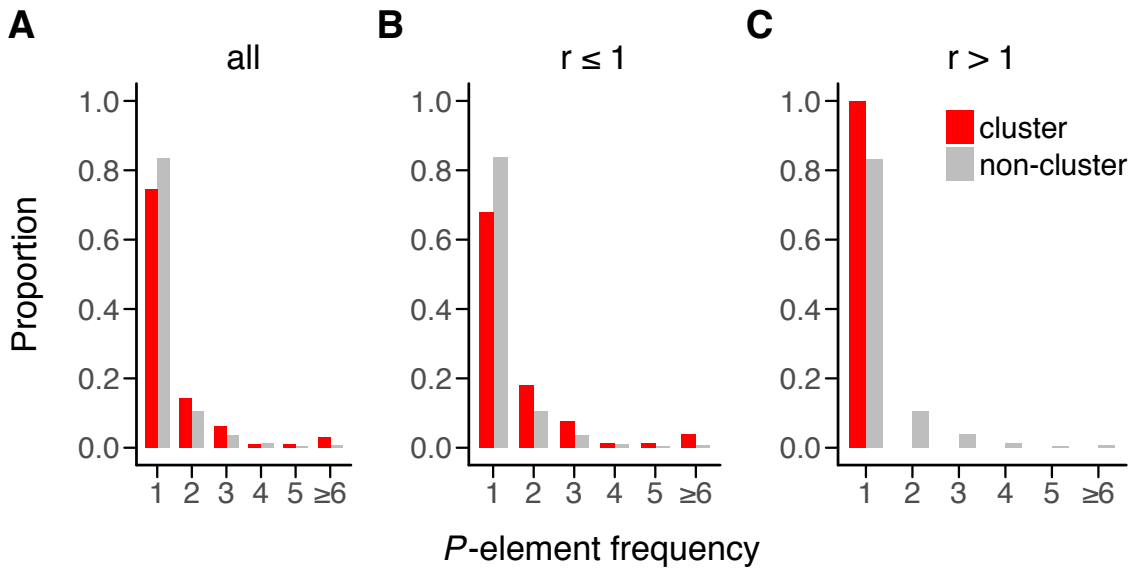

Figure 6. The frequency of *P*-elements in piRNA clusters (red) and the frequency of *P*-elements outside of clusters (gray) are compared for the high confidence set of 32 annotated piRNA clusters. The equivalent comparisons for lower stringency annotation sets are provided in supplementary figures S3 and S4. *P*-elements at hotspots were excluded from these comparisons. Insertions are compared in (A) all genomic regions ($W_{98, 4850}$ = 215720, P-value = 0.016), (B) regions of low recombination (≤1 cM/Mb, $W_{78, 1501}$ = 48966, P-value = 0.00019) and (C) high recombination (>1 cM/Mb, $W_{20, 3349}$ = 39130, P-value = 0.045).

DISCUSSION

In this study, we took advantage of the recent invasion of the *Drosophila* melanogaster genome by *P*-element DNA transposons to chronicle the evolution of piRNA-mediated repression. 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 pre-existing piRNA clusters, is the predominant mutational mechanism giving rise to piRNA-mediated silencing. Furthermore, we uncovered no fewer than 84 repressor alleles, which are targets of positive selection. Taken together, our results reveal that the common phenotype of *P*-element repression exhibited by North American *D. melanogaster* (Ogura et al. 2007;

Kidwell et al. 1983; Kidwell 1983) is underpinned by an unprecedented number of beneficial repressor alleles, which have arisen since the *P*-element invasion in the mid 20th century.

The existence of numerous segregating repressor alleles indicates that *P*-element repression didn't evolve through a classical "hard sweep", in which a single beneficial mutation arises and then goes into fixation (Maynard Smith and Haigh 1974). Rather, the evolution of *P*-element repression in *Drosophila* melanogaster occurred through a plethora of "soft sweeps" (Pennings and Hermisson 2006), in which numerous repressor alleles arose and increased in frequency simultaneously. Indeed, to our knowledge the evolution of *P*-element repression represents one of the most striking examples of soft-sweeps in a eukaryotic genome. By comparison, other well-known examples such as insecticide resistance in *D. melanogaster* (Menozzi et al. 2004; Karasov et al. 2010) and lactose tolerance in human populations (Enattah et al. 2002; Tishkoff et al. 2007), include only 4 and 5 adaptive mutations, respectively. The extreme soft-sweeps we observe in the evolution of *P*-element repression are at least partially a consequence of the unique genetic architecture of piRNA mediated silencing. The presence of multiple, functionally redundant piRNA clusters, which will enact repression when occupied by *P*-elements, provides an exceptionally large mutational target comprising at least 0.3% of the genome. Polygenic traits with large mutational targets are predicted to evolve via soft sweeps, because the overall beneficial mutation rate is increased (Pritchard et al. 2010; Messer and Petrov 2013; Pennings and Hermisson 2006; Karasov et al. 2010). Similarly, the per-site beneficial mutation rate within each

piRNA cluster is also high, owing to a very high genome-wide transposition rate of *P*-elements (~0.1 new insertions per element multiplied by genomic copy number (Eggleston et al. 1988; Berg and Spradling 1991; Kimura and Kidwell 1994)).

In addition to documenting an abundance of putatively beneficial alleles, we discovered that positive selection predominantly acts on *P*-element insertions in heterochromatic piRNA clusters. This is consistent with the theoretical prediction that reduced recombination enhances positive selection on TE repressors by maintaining linkage to the genomic regions they've protected from deleterious insertions (Charlesworth and Langley 1986). Indeed, enhanced selection on repressors in low recombination regions might explain why piRNA clusters are predominantly located in heterochromatic regions (Blumenstiel 2011). Interestingly, however, our results differ from those of Lu and Clark (2010), who show that piRNA cluster insertions of resident TE families that have occupied the genome for a long time segregate at higher frequency than those outside of piRNA clusters only in regions of high recombination (Lu and Clark 2010). The nature of the discrepancy is unclear, however, recent simulation models suggest that dynamics of piRNA mediated repressor alleles can differ significantly between recently invaded TEs and those that are at copy number equilibrium (Kelleher et al. 2018).

P-elements are not randomly distributed among piRNA clusters in the *D. melanogaster* genome. Rather, 72.4% of them (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; Brennecke et al. 2008). Our observations mirror those of a recent study of in laboratory populations of *D. simulans,* which demonstrated that *P*-element repression evolved by multiple independent insertions in piRNA clusters, particularly in the *3R*-TAS (Kofler et al. 2018). Furthermore, we discovered that *P*-elements are most commonly observed in previously identified insertion hotspots (Karpen and Spradling 1992), thereby demonstrating that this mutation bias shapes the distribution of *P*-element insertions in natural populations. 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.

In summary, *P*-element repression in *Drosophila melanogaster* evolved rapidly though abundant *de novo* mutations that arising from the transposition of *P*-elements into pre-existing piRNA clusters. These concurrent beneficial alleles are targets of positive selection, resulting a striking example of polygenic adaption. As piRNA-mediated silencing of TEs is conserved across animals, the model in which rapid adaptation to *P*-element invasion evolves through multiple beneficial *de novo* mutations applies to other TEs. Our observations reveal how the unique genetic architecture of piRNA-mediated silencing, in which insertion into multiple functionally redundant piRNA clusters results in a repressor allele, facilitates the evolution of repression of an invading TE.

MATERIALS and METHODS

DGRP stocks and genomes

All DGRP lines were ordered from the Bloomington Drosophila stock center.

piRNA cluster annotation

Ovarian small RNA sequencing libraries were downloaded from NCBI or were generated by our lab for another project (Lama and Kelleher *unpublished*, Supplemental Table S1). The latter libraries are available from SRA archive (SRP160954). 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 piRNA cluster annotation. Then, piRNA clusters were predicted separately in each library 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 detected less than 5 kb apart were considered a single cluster.

Detecting *P*-element insertions in DGRP genomes

DGRP whole genome sequencing reads were downloaded from the NCBI Sequence Read Archive (SRA study: SRP000694)(Mackay et al. 2012; Huang et al.

2014). 12 DGRP genomes were excluded from our analysis because 45 bp paired-end reads (DGPR357, DGRP379, DGRP427, DGRP486, DGRP786), or 75 bp single-end reads (DGRP153, DGRP237, DGRP28, 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 bowtie2 (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 down-stream analyses.

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 bowtie2. Reported alignments with mapping quality score greater than 20 and a mutational distance (sum of mismatches and gaps required to convert the read sequence to the reference) less than four were kept. Only the number of gaps was considered: we ignored gap extensions assuming that two adjacent nucleotide insertions or deletions were generated by one mutational event. 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 non-split *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 6 supporting read pairs (split or non-split) were required to

annotate a single *P*-element insertion in non-TAS regions. Identified *P*-element insertions are listed in Supplemental Table S8.

Detecting *P*-element Insertions in TAS

We divided the dm6 reference genome into two parts: TAS and non-TAS regions. TAS regions included full-length of *X*-TAS (9872 bp, L03284) (Karpen and Spradling 1992), 2*R*-TAS (chr*2R*:25258060..25261551, 3492 bp) and *3R*-TAS (chr*3R*:32073015..32079331, 6317 bp) (Yin and Lin 2007), *2L*-TAS (chr*2L*:1..5041, 5041 bp) and *3L*-TAS (chr*3L*:1..19608, 19608 bp)(Walter et al. 1995). The other genomic regions were categories as non-TAS.

To determine if *P*-derived reads that did not map to the non-TAS regions corresponded to insertions in TAS, they were aligned to the TAS reference using bowtie2 outputting all valid alignments (-a). A read pair was considered mapped to TAS if the mutational distance was fewer than six for paired-end reads and four for single-end reads. For each DGRP genome, we calculated a *Z*-score for TAS-aligned reads according to the 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 mutational distance for all

reported alignments of each read pair in that genome. We then assigned each read pair to the TAS array that it aligned to with lowest mutational 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 supporting reads were at least 2 times greater than 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 assay 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, as their repeats are indistinguishable from each other. 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 bp target site duplication

and 500 nt of adjacent TAS sequence. Paired-end reads were aligned to the constructed pseudo genomes (MAPQ > 10), and the breakpoint corresponding to the pseudo genome with the most reads aligned was identified. Identified *P*-element insertion sites are provided in Supplemental Table S6.

PCR verification of insertion sites

Genomic DNA was extracted using the QIAGEN DNeasy Blood & Tissue Kit (Cat. No. 69506) or a squish prep (Srivastav and Kelleher 2017). To determine the *P*-element insertion sites, a *P*-element specific and a TAS specific primer were used (Supplemental Table S9). As 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 (Cat. No. 28606), and sequenced to determine the breakpoint. PCR conditions are provided in the Supplemental Table S6.

Recombination rates

Recombination rates at *P*-element insertions sites were identified from the genome-wide map provided by Comeron *et al.* (Comeron 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 didn't have release 5 counterparts was assumed to 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 (http://www.uh.edu/cacds/resources/hpc/). All statistical analyses were performed in R (version 3.3.1)(R Core Team 2016). Graphs were made in RStudio (RStudio Team 2015) with R packages ggplot2 (version 2.2.1)(Wickham 2017b), gplots (version 3.0.1)(G. W. Warnes, B. Bolker 2016), reshape2 (version 1.4.3)(Wickham 2017a), and cowplot (version 0.7.0)(Wilke 2017).

Acknowledgements

Shuo Zhang, Erin Kelleher, and this research were supported by a National Science Foundation Division of Environmental Biology (NSF-DEB) award to E.S.K. (NSF-DEB #1457800).

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.

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**: 252–69. http://www.ncbi.nlm.nih.gov/pubmed/2838720 (Accessed December 13, 2013).

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science (80-)* **316**. http://science.sciencemag.org/content/316/5825/744.full (Accessed March 28, 2017).

Asif-Laidin A, Delmarre V, Laurentie J, Miller WJ, Ronsseray S, Teysset L. 2017. Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in Drosophila. *DNA Res*. http://www.ncbi.nlm.nih.gov/pubmed/28459978 (Accessed June 28, 2017).

Bartolomé C, Maside X. 2004. The lack of recombination drives the fixation transposable elements on the fourth chromosome of Drosophila melanogaster. *Genet Res*.

Berg CA, Spradling AC. 1991. Studies on the rate and site-specificity of P element transposition. *Genetics* **127**: 515–24. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1204379&tool=pmcentrez&rendertype=abstract (Accessed July 1, 2015).

Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogastergenome. *Genome Biol*.

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. http://www.ncbi.nlm.nih.gov/pubmed/2172082 (Accessed March 31, 2015).

Blumenstiel JP. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet* **27**: 23–31. http://www.ncbi.nlm.nih.gov/pubmed/21074888 (Accessed March 14, 2012).

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**: 1089–103. http://www.ncbi.nlm.nih.gov/pubmed/17346786 (Accessed March 1, 2012).

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**: 1387–92. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2805124&tool=pmcentrez&rendertype=abstract (Accessed March 9, 2012).

Charlesworth B, Langley CH. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* **112**: 359–83.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1202706&tool=pmcentrez&rendertype=abstract (Accessed February 22, 2013).

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in Drosophila melanogaster. *PLoS Genet* **8**: e1002905. http://www.ncbi.nlm.nih.gov/pubmed/23071443 (Accessed July 29, 2016).

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**: 339–55. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203926&tool=pmcentrez&rendertype=abstract (Accessed December 13, 2013).

de Vanssay A, Bougé A-L, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S. 2012. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature* **490**: 112–5. http://www.ncbi.nlm.nih.gov/pubmed/22922650 (Accessed July 13, 2014).

Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, Brown NH, Kaufman T, et al. 2015. FlyBase: Introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res*.

Dotto BR, Carvalho EL, Silva AF, Duarte Silva LF, Pinto PM, Ortiz MF, Wallau GL. 2015. HTT-DB: Horizontally transferred transposable elements database. *Bioinformatics*.

Eggleston WB, Schlitz DMJ, Engels WR. 1988. P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster. *Nature*.

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat Genet*.

G. W. Warnes, B. Bolker LB et al. 2016. gplots. *R Packag version301*.

Gasior SL, Wakeman TP, Xu B, Deininger PL. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**: 1383–1393.

Girard A, Hannon GJ. 2008. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol*.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biol*.

Hermant C, Boivin A, Teysset 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*.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.

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**: 445–58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4352887&tool=pmcentrez&rendertype=abstract (Accessed March 5, 2016).

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov D V., 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*.
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**: 1193–208. http://www.ncbi.nlm.nih.gov/pubmed/24714809 (Accessed July 11, 2018).
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**: 1785–93. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2516058&tool=pmcentrez&rendertype=abstract (Accessed July 15, 2014).
Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genet* **6**: e1000924. http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000924#pgen.1000924-Gillespie1 (Accessed March 15, 2016).
Karpen GH, Spradling AC. 1992. Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* **132**: 737–53. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1205211&tool=pmcentrez&rendertype=abstract (Accessed June 18, 2015).
Kelleher ES. 2016. Reexamining the P-Element Invasion of Drosophila melanogaster Through the Lens of piRNA Silencing. *Genetics* **203**: 1513–31. http://www.ncbi.nlm.nih.gov/pubmed/27516614 (Accessed August 18, 2016).
Kelleher ES, Azevedo RBR, Zheng Y. 2018. The Evolution of Small-RNA-Mediated Silencing of an Invading Transposable Element. *Genome Biol Evol*. http://www.ncbi.nlm.nih.gov/pubmed/30252073 (Accessed November 1, 2018).
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**: 1551–63. http://www.ncbi.nlm.nih.gov/pubmed/22196730 (Accessed March 9, 2012).
Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **80**: 1655–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=393661&tool=pmcentrez&rendertype=abstract (Accessed December 13, 2013).
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. http://www.ncbi.nlm.nih.gov/pubmed/8206365 (Accessed December 17, 2013).
Kofler R. 2019. Dynamics of transposable element invasions with piRNA clusters ed. J. Townsend. *Mol Biol Evol*. https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msz079/5435964 (Accessed April 10, 2019).
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. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3266889&tool=pmcentrez&rendertype=abstract (Accessed July 18, 2012).

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*.

Kofler R, Senti K-A, Nolte V, Tobler R, Schlötterer C. 2018. Molecular dissection of a natural transposable element invasion. *Genome Res*.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322381&tool=pmcentrez&rendertype=abstract (Accessed July 10, 2014).

Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen Y-CA, 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**: 1667–80. http://genesdev.cshlp.org/content/28/15/1667.long (Accessed July 14, 2015).

Lee YCG, Langley CH. 2012. Long-term and short-term evolutionary impacts of transposable elements on Drosophila. *Genetics* **192**: 1411–32. http://www.ncbi.nlm.nih.gov/pubmed/22997235 (Accessed March 17, 2013).

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

Li C, Vagin V V, 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**: 509–21. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2768572&tool=pmcentrez&rendertype=abstract (Accessed April 1, 2012).

Lim JK. 1988. Intrachromosomal rearrangements mediated by hobo transposons in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **85**: 9153–7.

Linheiro RS, Bergman CM. 2012. Whole Genome Resequencing Reveals Natural Target Site Preferences of Transposable Elements in Drosophila melanogaster ed. J.E. Stajich. *PLoS One* **7**: e30008. http://www.ncbi.nlm.nih.gov/pubmed/22347367 (Accessed June 9, 2017).

Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. *Genome Res* **20**: 212–27. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2813477&tool=pmcentrez&rendertype=abstract (Accessed March 16, 2012).

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**: 173–8. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3683990&tool=pmcentrez&rendertype=abstract (Accessed July 10, 2014).

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**: 522–35. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2882632&tool=p

mcentrez&rendertype=abstract (Accessed March 22, 2012).
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**: 1841–54. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461209&tool=pmcentrez&rendertype=abstract (Accessed July 18, 2014).
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favorable gene. *Genet Res* **23**: 23–35.
McGinnis W, Shermoen AW, Beckendorf SK. 1983. A transposable element inserted just 5′ to a Drosophila glue protein gene alters gene expression and chromatin structure. *Cell* **34**: 75–84.
Menozzi P, Shi MA, Lougarre A, Tang ZH, Fournier D. 2004. Mutations of acetylcholinesterase which confer insecticide resistance in Drosophila melanogaster populations. *BMC Evol Biol*.
Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet*.
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.
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**: 479–87. http://www.ncbi.nlm.nih.gov/pubmed/18270438 (Accessed July 28, 2014).
Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci* **114**: 4721–4726.
Pennings PS, Hermisson J. 2006. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol* **23**: 1076–84. http://www.ncbi.nlm.nih.gov/pubmed/16520336 (Accessed April 18, 2016).
Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in Drosophila melanogaster. *Mol Biol Evol* **28**: 1633–44. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3080135&tool=pmcentrez&rendertype=abstract (Accessed March 26, 2012).
Rahman R, Chirn G-W, Kanodia A, Sytnikova YA, Brembs B, Bergman CM, Lau NC. 2015. Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. *Nucleic Acids Res* **43**: 10655–72. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4678822&tool=pmcentrez&rendertype=abstract (Accessed November 23, 2015).
Robert V, Prud'homme N, Kim A, Bucheton A, Pélisson A. 2001. Characterization of the flamenco region of the Drosophila melanogaster genome. *Genetics* **158**: 701–13. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461675&tool=pmcentrez&rendertype=abstract (Accessed June 19, 2015).
Ronsseray S, Lehmann M, Anxolabehere 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.

http://www.genetics.org/content/129/2/501.abstract?ijkey=e54c3fa1b6357ea4f76fc3d3c21398cb3f18fa1b&keytype2=tf_ipsecsha (Accessed May 20, 2015).

Ronsseray S, Lehmann M, Anxolabéhère D. 1989. Copy number and distribution of P and I mobile elements in Drosophila melanogaster populations. *Chromosoma* **98**: 207–14. http://www.ncbi.nlm.nih.gov/pubmed/2555115 (Accessed June 18, 2015).

Rosenkranz D, Zischler H. 2012. proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* **13**: 5. http://www.ncbi.nlm.nih.gov/pubmed/22233380 (Accessed September 12, 2017).

Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–72. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2661848&tool=pmcentrez&rendertype=abstract (Accessed February 11, 2016).

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 Genes, Genomes, Genet* **7**.

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**: 1641–54. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1462367&tool=pmcentrez&rendertype=abstract (Accessed June 11, 2015).

Team RDC, R Development Core Team R. 2016. R: A Language and Environment for Statistical Computing. *R Found Stat Comput*.

Team Rs. 2015. RStudio: Integrated Development for R. *[Online] RStudio, Inc, Boston, MA URL http//www rstudio com*.

Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol* **2**: 656–64. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2997563&tool=pmcentrez&rendertype=abstract (Accessed May 1, 2015).

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet*.

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*.

Wickham H. 2017a. Flexibility Reshape Data: A Reboot of the Reshape Package. *R CRAN Proj*.

Wickham H. 2017b. ggplot2: Elegant Graphics for Data Analysis. *Journeal Stat Softw*.

Wilke CO. 2017. *Cowplot: streamlined plot theme and plot annotations for ggplot2*.

Yin H, Lin H. 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* **450**: 304–8. http://www.ncbi.nlm.nih.gov/pubmed/17952056 (Accessed July 20, 2012).

Zhang S, Kelleher ES. 2017. Targeted identification of TE insertions in a Drosophila genome through hemi-specific PCR. *Mob DNA* **8**: 10.

http://mobilednajournal.biomedcentral.com/articles/10.1186/s13100-017-0092-1 (Accessed July 28, 2017).

Zhuang J, Wang J, Theurkauf W, Weng Z. 2014. TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res* **42**: 6826–38. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4066757&tool=pmcentrez&rendertype=abstract (Accessed July 9, 2014).