The Evolution of Small RNA-Mediated Silencing of an Invading Transposable Element

Erin S. Kelleher[*,1], **Ricardo B. R. Azevedo**[*] **and Yichen Zheng**[*,†]

[*]Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA, [†]Senckenberg Gesellschaft für Naturforschung, Biodiversität und Klima Forschungszentrum,Senckenberganlage 25, Frankfurt am Main 60313, Germany

ABSTRACT Transposable elements (TEs) are genomic parasites that impose fitness costs on their hosts by producing deleterious mutations and disrupting gametogenesis. Host genomes avoid these costs by regulating TE activity, particularly in germline cells where new insertions are heritable and TEs are exceptionally active. However, the capacity of different TE-associated fitness costs to select for repression in the host, and the role of selection in the evolution of TE regulation more generally, remain controversial. In this study, we use individual-based simulations to examine the evolution of TE regulation through small RNA-mediated silencing. Small RNA silencing is a common mechanism for TE regulation employed by both prokaryotes and eukaryotes. We observed that even under conservative assumptions, where small RNA-mediated regulation reduces transposition only, repression evolves rapidly and adaptively after the genome is invaded by a new TE. We further show that the spread of repressor alleles is greatly enhanced by two additional TE-imposed fitness costs: dysgenic sterility and ectopic recombination. Finally, we demonstrate that the mutation rate to repression (*i.e.*, the size of the mutational target) is a critical parameter that influences both the evolutionary trajectory of host repression and the associated proliferation of TEs. Our findings suggest that adaptive evolution of TE regulation may be stronger and more prevalent than previously suggested, and complement recent empirical observations of positive selection on small RNA-mediated repressor alleles.

KEYWORDS intra-genomic conflict; TE repression, piRNA pathway; *Drosophila melanogaster*

Transposable elements are intragenomic parasites. Although some TE families or individual TE insertions have been domesticated to perform important functions (Kunarso *et al.* 2010; Lynch *et al.* 2011; Silva-Sousa *et al.* 2012) the presence and mobilization of TEs generally reduce the fitness of their host genome. TEs introduce deleterious mutations by causing DNA damage during mobilization and by inserting into functional sequences (Spradling *et al.* 1999; Dupuy *et al.* 2001). Ectopic recombination events between TE insertions in different genomic locations further produces large structural rearrangements, duplications, and deletions, which are overwhelmingly deleterious (reviewed in Hedges and Deininger 2007). In addition to these mutational impacts, TE activity itself can have drastic fitness consequences by causing dysgenic sterility, a condition in which germline DNA damage prohibits the production of viable gametes (Bingham *et al.* 1982; Orsi *et al.* 2010; Josefsson *et al.* 2006). Despite these fitness costs, nearly all genomes are populated by TEs (reviewed in Huang *et al.* 2012), and TEs frequently colonize new genomes through horizontal transfer (reviewed in Schaack *et al.* 2010).

Both prokaryotic and eukaryotic genomes minimize the fitness costs of TEs by controlling their activity through small RNA mediated silencing (reviewed in Blumenstiel 2011). In eukaryotic germlines, small RNA mediated silencing of TEs is enacted by Argonaute proteins that are found in complex with small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). Regulatory siRNAs and piRNAs are produced from anti-sense TE transcripts, and identify TE-derived mRNAs by base complementarity. Complexed Argonaute proteins then silence the targeted TE transcriptionally by inducing heterochromatin formation, or post-transcriptionally by degrading the transcript. Consistent with prevention of germline DNA damage, small RNA-mediated silencing suppresses TE-associated dysgenic sterility in *Drosophila* (Blumenstiel and Hartl 2005; Brennecke *et al.* 2008; Chambeyron *et al.* 2008; Rozhkov *et al.* 2010) and *Arabidopsis* (Josefsson *et al.* 2006). Transcriptional silencing

Copyright © 2017 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Tuesday 9th May, 2017%
[1]3455 Cullen Blvd. Suite #342. Houston, TX 77204. eskelleher@uh.edu

may further reduce the rate of ectopic recombination between TE insertions by inducing heterochromatin formation at TE loci (Haynes *et al.* 2006; Sentmanat and Elgin 2012; Sienski *et al.* 2012).

When a genome is invaded by a new TE, the repertoire of regulatory small RNAs must be expanded to recognize the new TE sequence. One mechanism by which this occurs is transposition of the invading TE into a genomic region that produces small regulatory RNAs, thereby leading to the production of new RNA species that target the invading element for silencing (Brennecke *et al.* 2008; Khurana *et al.* 2011). Consistent with this model, TE insertions in small RNA producing sites exhibit signatures of positive selection in both *Drosophila* (Lu and Clark 2010) and human genomes (Lukic and Chen 2011).

However, the degree to which repression of ectopic recombination, dysgenic sterility, and transposition select for the evolution of repression remains unclear. Selection on repression of transposition is thought to be particularly weak in sexually reproducing organisms because recombination decouples repressor alleles from the chromosomal regions that they protect from TE-induced mutation load (Charlesworth and Langley 1986). Nevertheless, a recent simulation study found that small RNA-mediated repression of transposition can evolve adaptively, at least if certain assumptions about the fitness effects of TE insertions are met (Lu and Clark 2010). The extent to which repression of dysgenic sterility or ectopic recombination enhances the adaptive evolution of small RNA-mediated repression has not previously been considered.

Here we describe a forward simulation model of the evolution of small RNA-mediated repression after genome invasion by a novel TE. To design our model and parameterize our simulations, we harnessed the wealth of empirical data from the *P*-element invasion of the *Drosophila melanogaster* genome (reviewed in Kelleher 2016). The status of *D. melanogaster* as a premier genetic model for the study of piRNA-mediated TE silencing (reviewed in Senti and Brennecke 2010), and the *P*-element as a vector for genetic modification (reviewed in Rubin and Spradling 1982), allowed us to develop a modeling framework that closely matches the known biology of transposition and TE regulation. We demonstrate that despite predictions of weak selection, small RNA-mediated silencing of transposition alone evolves rapidly and adaptively in response to genome invasion. We further show that positive selection on TE insertions in small RNA producing sites is greatly enhanced in models that incorporate dysgenic sterility or ectopic recombination. Finally, we demonstrate that the abundance of small RNA producing sites in the host genome is a critical parameter that influences both the proliferation of the TE and the adaptive behavior of repressor alleles. Our study provides novel insights into selective forces that influence the evolution of repression, as well as an updated framework for modeling TE invasion, which is broadly applicable to other systems.

Model

We employ an individual-based Wright-Fisher model of evolution, with a selection-reproduction-transposition life cycle, constant population size, and discrete, non-overlapping generations. See Table 1 for a list of parameters.

Genome

The genome is modeled after that of *D. melanogaster*. Each individual is diploid with a haploid complement of 3 chromosomes: one sex chromosome, containing 226 sites, and two autosomes, containing 442 and 527 sites respectively (total: 1195 sites). The numbers of sites per chromosome are based on the lengths of *D. melanogaster* chromosomes X, 2 and 3, with one site every $\sim 10^5$ base pairs (we ignore chromosome 4, which accounts for less than 1% of the genome). Likewise, the recombination rates in all sites are set to reflect the actual *D. melanogaster* recombination rates within each genomic window (Comeron *et al.* 2012). There are four kinds of sites in the genome: small RNA sites, pseudo-small RNA sites, functional sites and neutral sites.

Small RNA sites: 35 sites (2.9% of sites in the genome) establish small RNA silencing when occupied by a TE and represent the piRNA producing regions in the *D. melanogaster* genome known as piRNA clusters (Brennecke *et al.* 2007). Accordingly, the locations of small RNA sites are randomly drawn from the known locations of piRNA clusters in the *D. melanogaster* genome (Brennecke *et al.* 2007). Because piRNA clusters are generally found in heterochromatic regions where chromatin compaction restricts transposition and recombination (Brennecke *et al.* 2007), TE insertions in small RNA sites do not transpose and do not participate in ectopic recombination in our model.

Pseudo-small RNA sites: 35 sites (2.9% of sites in the genome) do not contribute to small RNA silencing but are otherwise indistinguishable from small RNA sites. Pseudo-small RNA sites are placed near telomeres or centromeres, which mirrors the distribution of piRNA clusters in the *D. melanogaster* genome. Each chromosome arm contains three and four pseudo-small RNA sites at the proximal and distal ends, respectively, skipping sites that are already assigned as either small RNA or functional.

Functional sites: Theoretical work in this area often assumes that TE insertions show negative epistasis for fitness (*e.g.*, Dolgin and Charlesworth 2008; Lu and Clark 2010; Lee and Langley 2012). This assumption appears to be required to achieve TE copy number equilibrium in the absence of repression (Charlesworth and Charlesworth 1983). Although it is proposed that the aneuploid progeny that result from ectopic exchange could produce negative epistasis between TE insertions (Langley *et al.* 1988), there is as yet no empirical evidence to support this fitness model.

Here, we take a different approach, allowing the fitness effects of TE insertions to vary between sites. We modeled these fitness effects based on the known transposition bias of *P*-elements (Bellen *et al.* 2004, 2011; Spradling *et al.* 2011) and the distribution of estimated fitness effects of individual *P*-element insertions (Cooley *et al.* 1988; Mackay *et al.* 1992; Lyman *et al.* 1996).

To determine the number of functional sites in our theoretical genome, we relied on results from extensive *P*-element mediated mutagenesis screens, which indicate that $\gtrsim$ 80% of new *P*-element insertions occur within gene bodies, within 500 bp of a transcription start site (Bellen *et al.* 2004, 2011). Accordingly, 80% (956) of sites in our theoretical genome are functional, and cause a multiplicative change in viability when occupied by a TE (*i.e.*, there is no epistasis). We furthermore assume that the effects on viability of TE insertions in functional sites are completely recessive, as has been observed empirically for *P*-elements (Mackay *et al.* 1992). In some simulations, however, TE insertions in functional sites contribute dominantly to the production of inviable gametes by ectopic recombination (see below).

When homozygous, the effect on viability of a TE insertion at a functional site can be lethal (selection coefficient, $s = -1$),

Table 1 Model parameters

Parameter	Description
s	Effect of a TE insertion on viability (Table 2; Equation 6)
u	Transposition rate per genome per generation (Equation 1)
n_a	Number of TE insertions in functional and neutral sites (Equations 1 and 4)
R	Strength of small RNA-mediated repression (Equations 1–4)
F	Fertility (Equations 2 and 5)
D	Maximum possible fertility cost caused by dysgenesis (Equation 2)
n	Total number of TE insertions (Equation 2)
n_m	Number of TE insertions in the maternal genome (Equation 3)
ε	Fitness cost of ectopic recombination (Equations 4 and 5)
E_0	Baseline ectopic recombination rate (Equation 4)
$W_{♂}$	Male fitness (Equation 5)
$W_{♀}$	Female fitness (Equation 5)
V	Viability (Equations 5 and 6)
N	Population size

Table 2 Viability selection coefficients (s) of TE insertions in the 1,195 sites of different kinds.

Type	s	Number	Frequency (%)
Lethal	−1.00	45	3.8
Strongly Deleterious	−0.25	325	27.2
Moderately Deleterious	−0.05	350	29.3
Slightly Deleterious	−0.01	224	18.7
Neutral, small RNA and pseudo-small RNA	0.00	239	20.0
Slightly Beneficial	0.01	12	1.0

strongly deleterious ($s = -0.25$), moderately deleterious ($s = -0.05$), slightly deleterious ($s = -0.01$), or slightly beneficial ($s = 0.01$).

The frequency of each functional site class in the theoretical genome is described in Table 2. Mutagenesis suggests that the frequency of recessive lethal mutations produced by *P*-element insertions is in the range 3.8–10% (Cooley *et al.* 1988; Mackay *et al.* 1992). Therefore, we conservatively assigned 3.8% of sites in the theoretical genome to behave as recessive lethals when occupied by a TE. Similarly, the average homozygous fitness cost of a TE insertion in the theoretical genome, $\bar{s} = 12.2\%$, matches an empirical estimate from mutagenesis (Mackay *et al.* 1992).

Neutral sites: 169 sites (14.1% of sites in the genome) have no effect on viablity when occupied by a TE. However, neutral sites, like functional sites, contribute to the production of inviable gametes through ectopic recombination in some models (see below).

Transposition, Excision, and Inactivation

The transposition rate per genome per generation is given by

$$u = \mu\, n_a \left(\frac{k-R}{k}\right) \quad , \tag{1}$$

where μ is the transposition rate per element per generation of unrepressed TEs, n_a is the number of TEs in the genome that can be mobilized in the presence of transposase (*i.e.*, those in functional and neutral sites, but not in small RNA or pseudo-small RNA sites), R is the strength of small RNA-mediated repression (see Equation 3), and k is a constant. We set $\mu = 0.1$ following the empirical estimate for *P*-elements (Eggleston *et al.* 1988). We also set $k = 1.01$, limiting the reduction in transposition rate caused by small RNA-mediated silencing to two orders of magnitude, as is suggested by observed differences in *P*-element activity in the presence and absence of regulation (Eggleston *et al.* 1988).

Precise *P*-element excisions occur with an estimated frequency of 0.15% (per element/genome) when the insertion is homozygous and ∼13.5% when the insertion is heterozygous (Engels *et al.* 1990). We therefore assume that the excision rate is 0.01 per element per generation, or 10% of the transposition rate. Mutational loss of transposase-encoding capacity, which occurs predominantly through internal deletions that are acquired during transposition (Staveley *et al.* 1995), occurs at a rate of 0.001 per element per generation.

Dysgenic Sterility

When *Drosophila* males bearing a transpositionally active TE family are crossed to females who do not produce piRNAs that regulate that TE family, their F1 offspring can suffer from dysgenic sterility (Bingham *et al.* 1982; Bucheton *et al.* 1984; Blackman *et al.* 1987; Evgen'ev *et al.* 1997; Yannopoulos *et al.* 1987; Brennecke *et al.* 2008; Hill *et al.* 2016). A similar sterility syndrome has been associated with activity of paternally inherited TEs in *Arabidopsis* (Josefsson *et al.* 2006). We therefore modeled dysgenic sterility as a function of the number of paternally inherited TEs, whose unregulated activity are thought to induce the sterility phenotype (equation 2). Indeed, in the case of *P*-element induced dysgenic sterility, the proportion of offspring that are sterile has been correlated with the *P*-element germline excision rate, consistent with *P*-element activity as the cause of sterility (Kocur *et al.* 1986).

Bingham *et al.* (1982) described the quantitative relationship between the number of paternally inherited *P*-elements and the incidence of dysgenic sterility among F1 female offspring. The paternal genotypes that induced dysgenesis in these crosses were otherwise isogenic, such that differences in dysgenic sterility among F1 females could be exclusively attributed to *P*-element copy number. They observed that $n = 4$ *P*-element copies produced ∼50% dysgenic offspring, whereas $n \gtrsim 9$ copies produced $\gtrsim$ 80% dysgenic offspring.

We matched these empirical observations by expressing the fertility, F, of an individual as a function of its total diploid TE copy number (n)

$$F = \begin{cases} 1 - D\,(1-R)\left(\dfrac{n-3}{n-2}\right) \quad , & \text{if } n \geqslant 4 \\ 1 \quad , & \text{otherwise} \quad , \end{cases} \tag{2}$$

where D is a constant that represents the maximum fertility cost, and can be varied from 0 (no dysgenic sterility) to 1 (complete dysgenic sterility); R is the strength of small RNA-mediated repression (see Equation 3). F can vary from 0 (complete infertility) to 1 (complete fertility).

Small RNA-Mediated Silencing

piRNA-mediated silencing represses both germline TE activity and dysgenic sterility in *Drosophila* (Brennecke *et al.* 2008; Khurana *et al.* 2011). The production of regulatory piRNAs requires the presence of at least one homologous TE insertion in a piRNA cluster, and is amplified by the presence of additional TE copies elsewhere in the genome (Brennecke *et al.* 2008; Zanni *et al.* 2013). The enhancement of piRNA production by TE insertions outside of piRNA clusters is explained by the feed-forward biogenesis of piRNAs through transcriptional silencing of sense TE transcripts (Brennecke *et al.* 2007; Gunawardane *et al.* 2007). piRNAs are maternally transmitted, such that piRNA production in the maternal genotype determines the strength of silencing in the offspring (Brennecke *et al.* 2008).

Empirical data regarding the quantitative relationship between TE copy numbers inside and outside of piRNA clusters and reductions in transposition rate are sparse for *P*-elements, and entirely lacking for other TE families. We therefore harnessed the extensive data on the the repression of dysgenic sterility (Ronsseray *et al.* 1991, 1996, 1998; Marin *et al.* 2000; Stuart *et al.* 2002; Simmons *et al.* 2007, 2012, 2014) (Supplemental Material Table S1, Figure S1). Based on these observations, we describe the strength of small RNA-mediated silencing, R, conferred by maternal genotypes with at least one occupied small RNA site as

$$R = 1 - e^{a - b n_m} \quad , \tag{3}$$

where n_m is the total number of TE copies in the maternal genome, and a and b are constants. We fit Equation 3 by nonlinear least squares to the data in Table S1 and obtained: $a = 0.8 \pm 0.5$ and $b = 1.0 \pm 0.4$ (estimates and 95% confidence intervals). R can vary from 0 (no repression) to 1 (complete repression).

Ectopic Recombination

Ectopic recombination occurs between TE insertions in different sites, and produces inviable gametes in our model. We assume that the fitness cost of ectopic recombination is proportional to the number of pairwise combinations of TEs in the genome, excluding insertions in small RNA and pseudo-small RNA sites.

The fitness consequences of ectopic recombination are analogous to the negative epistasis between TE insertions that is assumed by previous models (Dolgin and Charlesworth 2006, 2008; Lu and Clark 2010; Lee and Langley 2012).

Small RNA-mediated silencing acts as a multiplicative modifier of ectopic recombination, which reflects the potential suppression of recombination at TE loci due to heterochromatin formation (Haynes *et al.* 2006; Klenov *et al.* 2007). The fitness cost of ectopic recombination (ε) is described by the following expression:

$$\varepsilon = n_a(n_a - 1)(1 - R)E_0 \quad , \tag{4}$$

where E_0 is the baseline ectopic recombination rate; n_a is the number of TE insertions that can contribute to ectopic recombination (*i.e.*, do not occur in piRNA or pseudo-piRNA sites); and R is the strength of small RNA-mediated repression.

Natural Selection

Male and female fitness are given by

$$W_{♂} = VF \quad \text{and} \quad W_{♀} = VF(1 - \varepsilon) \quad , \tag{5}$$

respectively, where

$$V = \prod_i (1 + s_i) \tag{6}$$

is viability; s_i is the viability selection coefficient of the TE insertion into the i locus in a hemizygous or homozygous state; F is fertility (Equation 2); and ε is the fitness cost of ectopic recombination (Equation 4).

Evolution

Each population is defined by the following parameters: population size (N), baseline ectopic recombination rate (E_0), and the maximum possible fertility cost caused by dysgenesis (D). In the initial population, the sex of each individual is decided randomly, so that the expected sex ratio is 1:1. 5% of the population harbors one TE insertion at a randomly chosen neutral site on chromosome 2. The remaining 95% of the population harbors no TE insertions.

We simulated reproduction by randomly selecting one male and one female, each, with probability proportional to their fitness (Equation 5). After an individual parent is selected to reproduce, the number of new TE copies to transpose within its germline is drawn from a Poisson distribution with parameter u (Equation 1). The new TE copies are distributed randomly among unoccupied sites and are only used to produce one specific offspring; if a parent produces multiple offspring, transposition is repeated independently for each offspring produced.

Haploid gametes are produced by meiosis, simulated by choosing one chromosome per pair at random. Crossing over occurs in females only, as is true for *D. melanogaster* (Morgan 1914). Diploid progeny are formed by fusion of one gamete from each parent. Once N offspring are produced, the parents are discarded.

P-elements invaded natural populations of *D. melanogaster* worldwide in less than 40 years (Kidwell 1983; Anxolabéhère *et al.* 1988). Assuming the conventional 10 generations/year for natural populations of *D. melanogaster*, this corresponds to an estimated 400 generations. We therefore ran forward simulations for 500 generations, and considered only populations in which 99% of genomes contained at least one *P*-element at the end of simulated evolution. For all combinations of parameters we considered, we continued to run simulations until 100 successful invasions were achieved, or two weeks of computing time had passed on our high-performance computing cluster, whichever came first.

Statistical Analyses

All statistical analyses were conducted with R version 3.3 (R Core Team 2016). All graphs were produced in ggplot2 (Wickham 2009).

Data Availability

The software used to run all simulations was written in Perl and will be available at https://github.com/... at the time of publication.

Results

Host repression evolves rapidly and stabilizes TE copy number

We began by considering a conservative model, in which TEs impose fitness costs by occupying functional sites in the host genome, and repression benefits the host by preventing the occurrence of these deleterious insertions. Dysgenic sterility and ectopic recombination were not incorporated in the model ($D = 0$, $E_0 = 0$). Starting from populations in which 5% of individuals harbor a single autosomal TE insertion at a neutral site, the TE successfully invaded 100% of simulations (out of 100 runs; Figure 1A, red). Invasion was rapid, with 100% of genomes harboring at least one TE copy within less than 100 generations. This rapid spread of genomic TEs is accompanied by an exponential increase in copy number for ~80 generations, which asymptotes at an equilibrium copy number of ~50 elements per diploid genome (Figure 1B).

Copy number equilibrium of a TE requires a balance between the production of new insertions via transposition, and the removal of existing insertions through excision or negative selection (Charlesworth and Charlesworth 1983). In the absence of repression that equalizes the transposition and excision rates, negative epistasis between TE insertions is required to achieve transposition-selection balance (Charlesworth and Charlesworth 1983). Because our minimal model assumes that the fitness effects of TE insertions are independent, the copy number equilibrium that occurs after invasion must be the result of evolved small RNA-mediated repression. Indeed, we observed that small RNA-mediated repression evolved rapidly after genome invasion, approaching complete repression after only 200 generations (Figure 1A, blue). By contrast, when repression was not allowed to evolve (*i.e.*, $R = 0$ for all individuals, regardless of genotype), 100% of simulated populations went extinct within 173 generations (802 runs) because all individuals of one sex were homozygous or hemizygous for a lethal TE insertion. For 795 of the runs, no viable males remained, suggesting hemizygous X-chromosomes carrying recessive lethals contributed disproportionately to population extinction.

Selection against deleterious TE insertions generates weak positive selection on occupied small RNA sites

To examine the role of positive selection in the evolution of TE repression, we considered the site frequency spectrum of TE insertions that produce regulatory small RNAs (*i.e.*, insertions in small RNA sites) after 500 generations of simulated evolution. We compared these sites to control pseudo-small RNA sites, neutral sites that do not produce small RNAs when occupied by

TEs but otherwise are matched to small RNA sites in their genomic locations and fitness effects (see Model). Positive selection will increase the abundance of high frequency derived alleles in the population as compared to a neutral model (Fay and Wu 2000; Nielsen 2005). Consistent with positive selection on small RNA silencing, occupied small RNA sites segregate at higher frequency than their neutral pseudo-small RNA counterparts (Figure 1C). Furthermore, the site frequency spectra of occupied small RNA producing sites is significantly different than that of pseudo-small RNA sites (Kolmogorov-Smirnov test, $N = 2000$: $D = 0.13$, $p < 2.2 \times 10^{-16}$).

Although occupied small RNA sites segregated at higher frequency than occupied pseudo-small RNA sites, it was unusual for them to reach high frequency in simulated populations ($> 50\%$ of chromosomes). Only 5% of 100 replicate simulated populations exhibited a high frequency of occupied small RNA sites after 500 generations. This indicates that positive selection on host repression is weak, as has been proposed previously (Charlesworth and Langley 1986). Further supporting weak positive selection, we observed that the the difference between the site frequency spectra of occupied small RNA sites and pseudo-small RNA sites was attenuated when we reduced the size of the simulated populations to $N = 500$ individuals, revealing the enhanced influence of genetic drift (Figure 1D; Kolmogorov-Smirnov test, $N = 500$: $D = 0.09$, $p = 1.10 \times 10^{-5}$). Notably, occupied small RNA and pseudo-small RNA sites segregated at a higher frequency on average at $N = 500$, since new mutations are more likely to fix in smaller populations.

The contrast between the prevalence of repression at the end of simulated evolution (Figure 1A, blue) and the low frequency of individual occupied small RNA sites (Figure 1C, red), reveals a common phenotype that is produced by many alternative genotypes. The genetic architecture of small RNA production, in which ~3% of the genome produces small RNAs when occupied by a TE insertion (35 sites in our theoretical genome, see Model), facilitates this common phenotype-rare genotype scenario. On average, there were 30.83 different small RNA sites that were segregating for a TE insertion in a given simulated population after 500 generations. This value is only slightly higher than the number of pseudo-small RNA sites that were segregating for an insertion (29.51, $t_{198} = 3.76$, $p = 0.0002$). The plethora of segregating pseudo-small RNA sites reveals a high genome-wide transposition rate, and therefore a high mutation rate to occupied sites. Furthermore, the modest difference in the average number of segregating small RNA sites as compared to pseudo-small RNA sites is consistent with weak selection for the retention of small RNA producing insertions.

Size of the mutational target influences the strength of positive selection on occupied small RNA sites, but not the rapid evolution of the repressive phenotype

To explore how the rate at which repressor alleles are produced influences the invasion process and the evolution of small RNA-mediated repression, we considered genomes with different numbers of small RNA sites. Differences in the size of the mutational target correspond to different mutation rates for the production of repressor alleles by transposition. We observe that regardless of the size of the mutational target, repression evolves rapidly, within 200 generations of TE invasion. However, the evolutionary trajectory of repression depends strongly on the abundance of small RNA sites. When small RNA sites are common, repressive individuals arise earlier in the invasion, but increase in abundance more slowly. In contrast, when there are few small RNA producing sites, the first repressive individuals occur later in the invasion on average, but the population transitions from permissive to repressive much more rapidly (Figure 2A), suggesting stronger selection for the repressive phenotype.

Consistent with stronger positive selection, the more rapid transition to a repressive population that occurs with smaller mutational targets is accompanied by increased abundance of high frequency ($\geq 50\%$ of chromosomes) small RNA producing sites after 500 generations of simulated evolution (Figure 2D). However, the increased abundance of high frequency repressor alleles in part reflects adaptation that is more mutation-limited. When small RNA sites are rare, the probability that a TE inserts into a second unoccupied small RNA site before the first occupied small RNA site reaches fixation decreases, allowing for a greater number of lower frequency repressor alleles to co-occur in the same population. To disentangle differences in mutation-limitation from positive selection, we considered the cumulative frequency of individuals who harbor at least one insertion in a small RNA producing site over the course of simulated evolution (Figure 2B). In this way we consider the increase in frequency of the repressor allele class as a whole, independently of the number of alleles. We observed that the cumulative frequency of individuals with at least one repressor allele increases more rapidly when the mutational target to repression is small, suggesting stronger positive selection.

Stronger selection on repressor alleles in genomes with a small mutational target for small RNA mediated repression is explained by the accumulation of a greater number of TE copies (Figure 2C). Higher TE copy numbers increase the benefits of repression by increasing the deleterious mutation rate, as well as the likelihood that deleterious insertions are homozygous and therefore exposed to selection. Indeed, the accumulation of TE copies in genomes with smaller mutational targets is associated with decreased mean population fitness and increased variance in fitness , indicating stronger selection (Supplemental Material Figure S2).

Dysgenic sterility impacts TE invasion success and copy number

In *Drosophila*, dysgenic sterility occurs among F1 offspring when males bearing transpositionally active elements mate with females that do not produce regulatory piRNAs (Blumenstiel and Hartl 2005; Josefsson *et al.* 2006; Brennecke *et al.* 2008; Chambeyron *et al.* 2008; Rozhkov *et al.* 2010). We examined how dysgenic sterility could impact TE invasion by performing simulations in which different fitness costs were imposed on individual offspring of dysgenic matings. Five different maximum fitness costs of dysgenesis were considered: $D = 0$, 0.1, 0.25, 0.5, and 1. In the absence of small RNA repression, the enhanced purifying selection imposed by moderate dysgenic sterility ($D = 0.1$ or 0.25), was unable to oppose TE proliferation, and 100% of populations went extinct due to mutation load (Figure 3A). On the other hand, when dysgenic sterility is severe ($D = 0.5$ or 1), the TE was unable to invade the population within 500 generations (Figure 3A).

When individuals were permitted to enact small RNA mediated repression of dysgenic sterility as well as transposition, the TE successfully invaded 100% of simulated populations when the fitness cost of dysgenic sterility was moderate ($D = 0.1$ or 0.25). However, when more severe fitness costs were considered ($D \geq 0.5$), dysgenic sterility remained a significant barrier to

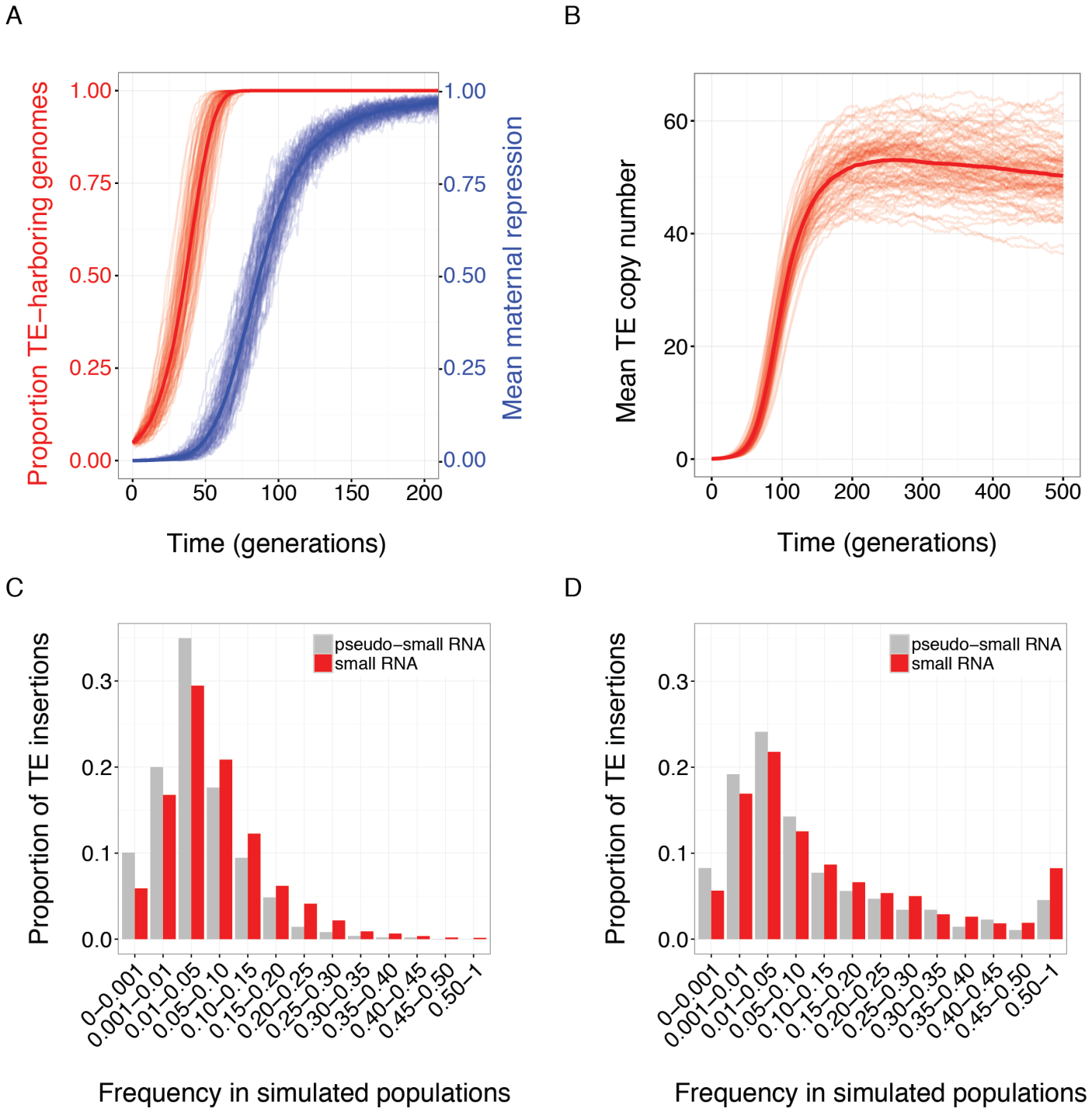

Figure 1 Invasion is followed by rapid and adaptive evolution of host repression. A) Evolution of the proportion of individuals with at least one genomic TE (red) and of the mean maternal repression (R) over 200 generations in 100 replicate simulated populations of $N = 2000$ individuals. B) Evolution of mean diploid TE copy number (n) over 500 generations in 100 replicate simulated populations of $N = 2000$ individuals. Thick lines represent the average responses across all replicate populations. C) and D) The site frequency spectra of occupied small RNA producing sites (red), which enact maternal repression of transposition, and pseudo-small RNA producing sites (grey) which do not repress transposition $N = 500$ (C) and $N = 2000$ (D) individuals. The spectra were calculated by pooling the data from 100 replicate simulated populations after 500 generations of evolution. Site frequency bins include sites whose frequencies are higher than the lower bound, and less than or equal to the upper bound. Individuals in all simulations (A–D) experienced neither dysgenic sterility nor ectopic recombination $(D = 0\,,\ E_0 = 0)$.

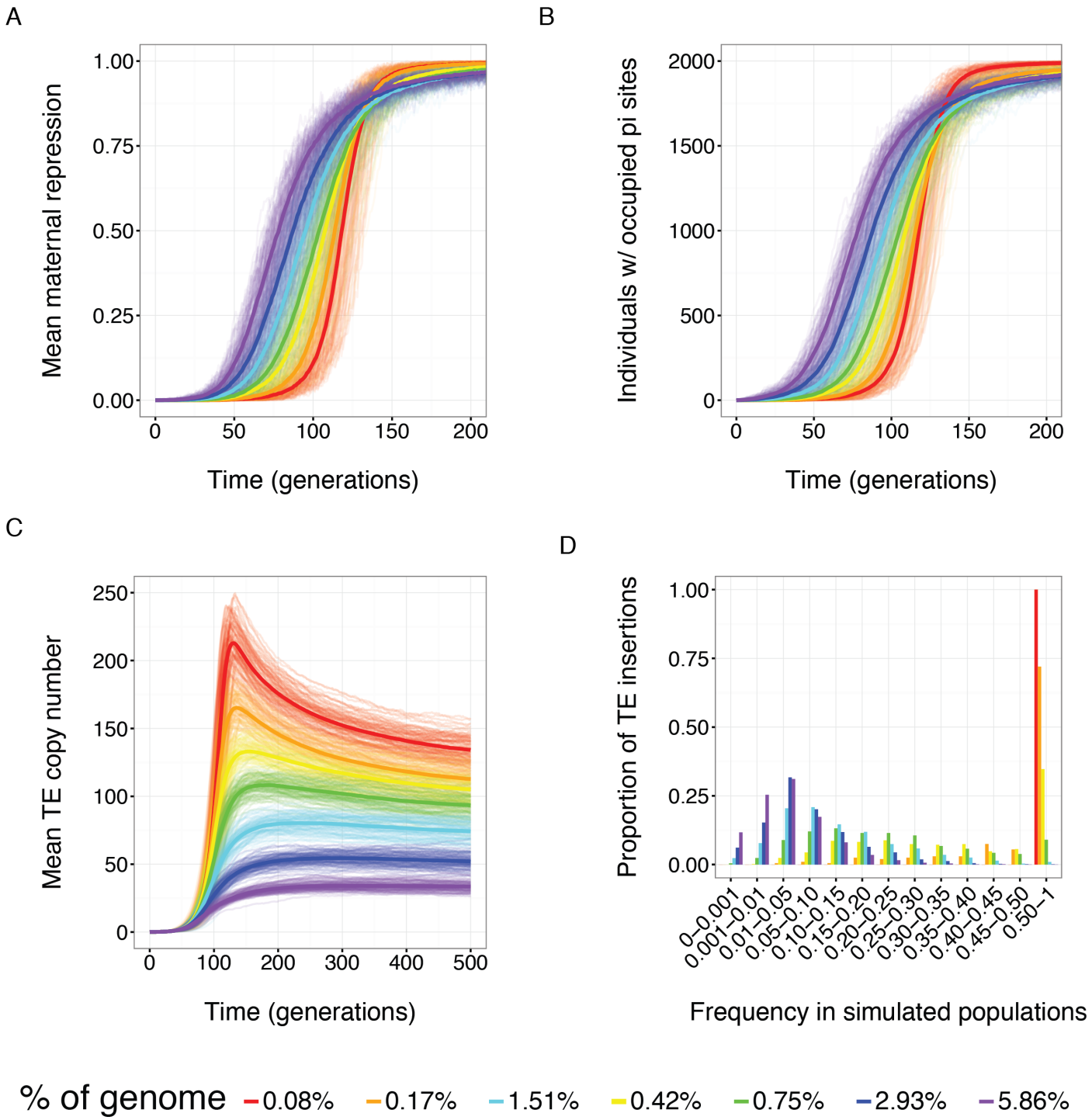

Figure 2 The mutation rate to repressor alleles determines the strength of positive selection. A) and B) Evolution of the mean maternal repression (R) (A) and cumulative frequency of individuals with at least one occupied small RNA (B) site over 200 generations in replicate simulated populations. C) Mean diploid TE copy number (n) over 500 generations in simulated populations. The thick line represents the average response across 100 replicate populations. D) The site frequency spectra of occupied small RNA producing sites after 500 generations of simulated evolution. Site frequency bins include sites whose frequencies are higher than the lower bound, and less than or equal to the upper bound. The site frequency spectra were calculated by pooling the data from all 100 replicate simulated populations. All simulations included $N = 2000$ individuals, who experienced neither dysgenic sterility nor ectopic recombination $(D = 0,\ E_0 = 0)$. Colors indicate simulations that involved genomes with different fractions of small RNA producing sites.

TE invasion. Only 26% (100 of 381) of simulated populations exhibited successful invasion after 500 generations for $D = 0.5$. In the remaining 74% of simulated populations, the TE persisted after 500 generations but the invasion was not considered successful, because the TE was not present in $\geq 99\%$ of individual genomes (Supplementary Material Figure S3, see Methods). Under complete dysgenic sterility ($D = 1$), the TE failed to invade the simulated population 100% of the time (377 runs). Dysgenic sterility had a similarly dramatic effect on TE proliferation, with increasing values of D corresponding to sequentially lower TE copy numbers at equilibrium (Figure 4A).

Hybrid dysgenesis increases the strength of positive selection on occupied small RNA sites

Dysgenic sterility has potentially opposing effects on the evolution of small RNA-mediated repression. The lower TE copy numbers that are observed in the presence of dysgenic sterility (Figure 4A) correspond to a lower rate at which repressor alleles are produced by transposition into small RNA sites, which could delay the evolution of repression. On the other hand, minimizing the fitness costs of dysgenic sterility represents an added selective advantage that could enhance positive selection on occupied small RNA sites when they arise.

When dysgenic sterility is absent or moderate ($D < 0.5$), we detect these opposing effects. Although repression evolves rapidly, within $\sim$100 generations of invasion, the time to emergence of repressive individuals in simulated populations is longer with higher values of D (Figure 4B, Supplementary Material Figure S4). Greater fitness costs of dysgenic sterility also corresponded to enhanced positive selection on occupied small RNA sites. The site-frequency spectrum of occupied small RNA sites after 500 generations of simulated evolution included more common variants (Figure 3C), with 55% of simulated populations exhibiting a high frequency occupied small RNA site ($\geq$ 50% of chromosomes) when $D = 0.25$, as compared to 5% of populations when dysgenic sterility does not occur ($D = 0$) (100 replicates for each value of D). Enhanced positive selection on repression is also evident from the increased number of segregating small RNA sites, as compared to pseudo-small RNA sites in simulated populations (Figure 3D).

We observed a different selective regime when the costs of dysgenic sterility are high ($D = 0.5$), in which TE proliferation was limited by purifying selection against dysgenesis to such an extent that selection for repression was weakened. In these simulated populations, repression evolved in a slower, more linear fashion (Figure 4B, blue). Consistent with attenuated positive selection, we did not observe a greater abundance of occupied small RNA sites segregating at high frequency ($\geq$ 50% of chromosomes) at the end of 500 generations when $D = 0.5$, as compared to lower $D = 0.25$ (Figure 4C, blue). Furthermore, the site frequency spectra of small RNA mediated repressors is likely affected by an ascertainment bias when $D = 0.5$ that artificially increases the abundance of occupied small RNA sites. Because the TE fails to invade in 74% of simulations when $D = 0.5$, and small RNA-mediated silencing facilitates invasion by decreasing the fitness cost of TEs, simulated populations that acquired a high-frequency occupied small RNA sites are disproportionately represented among invasion successes. Indeed, when we consider the full set of simulations for $D = 0.5$ (both successful and failed invasions) we observe a reduced proportion of high frequency insertions as compared to $D = 0.25$ (Figure 4C, purple).

The average number of segregating small RNA and pseudo small RNA sites also suggests that positive selection is weaker at $D = 0.5$ when compared to $D = 0.25$ (Figure 4D). Although the mean number of segregating pseudo-small RNA sites does not differ between the two groups of simulated populations after 500 generations ($t_{198} = 0.14$, $p = 0.89$), larger numbers of segregating small RNA sites are observed in when $D = 0.25$ (mean = 18.67) than when $D = 0.5$ (mean = 15.63, $t_{198} = 7.09$, $p = 2.35 \times 10^{-11}$). Therefore, selection for retention of occupied small RNA sites appears stronger when $D = 0.5$ as compared to $D = 0.25$. Taken together, these observations suggest that when fitness costs imposed by dysgenic sterility are high, they may limit not only TE invasion success, but also the evolution of repression.

Ectopic recombination facilitates TE invasion in the absence of repression

We lastly considered the impact of fitness costs imposed by ectopic recombination on invasion success and selection for repression. In our model, the probability of an ectopic recombination event in an individual's germline increases faster than linearly with TE copy number (Equation 4), effectively producing negative epistasis between TE insertions. We therefore explored whether ectopic recombination could produce a transposition-selection balance in the absence of repression, as has been reported previously (Dolgin and Charlesworth 2008). We observed that regardless of baseline ectopic recombination rate (E_0), severe dysgenic sterility ($D \geq 0.5$) prevents TE invasion of the population in 100% of simulations (Figure 4A). However, when dysgenic sterility is absent or moderate ($D < 0.5$), moderate to high levels of ectopic recombination ($E_0 \gtrsim 10^{-4}$) can produce sufficiently strong negative selection to prevent TE over-proliferation and population extinction. Ectopic recombination, therefore, facilitates TE invasion and genomic maintenance in the absence of repression (Figure 4A). Although our simulated populations did not appear to arrive at transposition-selection balance within 500 generations, the rate of accumulation of TE copies slowed (Figure 4B), suggesting that longer simulations could achieve copy number equilibrium.

Ectopic recombination increases the strength of positive selection on occupied small RNA sites

Finally, we explored how ectopic recombination impacts the evolution of small RNA mediated silencing. In these simulations, small RNA silencing repressed transposition and suppressed ectopic recombination, and individuals did not experience hybrid dysgenesis ($D = 0$). Similar to previous simulations that include small RNA silencing (Figures 1A, 2C, 3A), we observed an initial exponential increase in TE copy number that tapered off in less than 200 generations, followed by a gradual approach to copy number equilibrium (Figure 5A). Higher baseline ectopic recombination rates (E_0), were associated with lower TE copy numbers throughout the simulation (Figure 5A), as expected from the higher fitness cost for each TE.

Fitness costs imposed by ectopic recombination were also associated with enhanced positive selection on occupied piRNA sites. The site frequency spectra for occupied small RNA sites at the end of the 500 generations of simulated evolution included a greater proportion of high-frequency variants, with 45% of 100 populations exhibiting a high-frequency ($\geq$ 50%) occupied small RNA site at the highest baseline ectopic recombination rate we considered ($E_0 = 0.01$, Figure 5B, purple). Similarly,

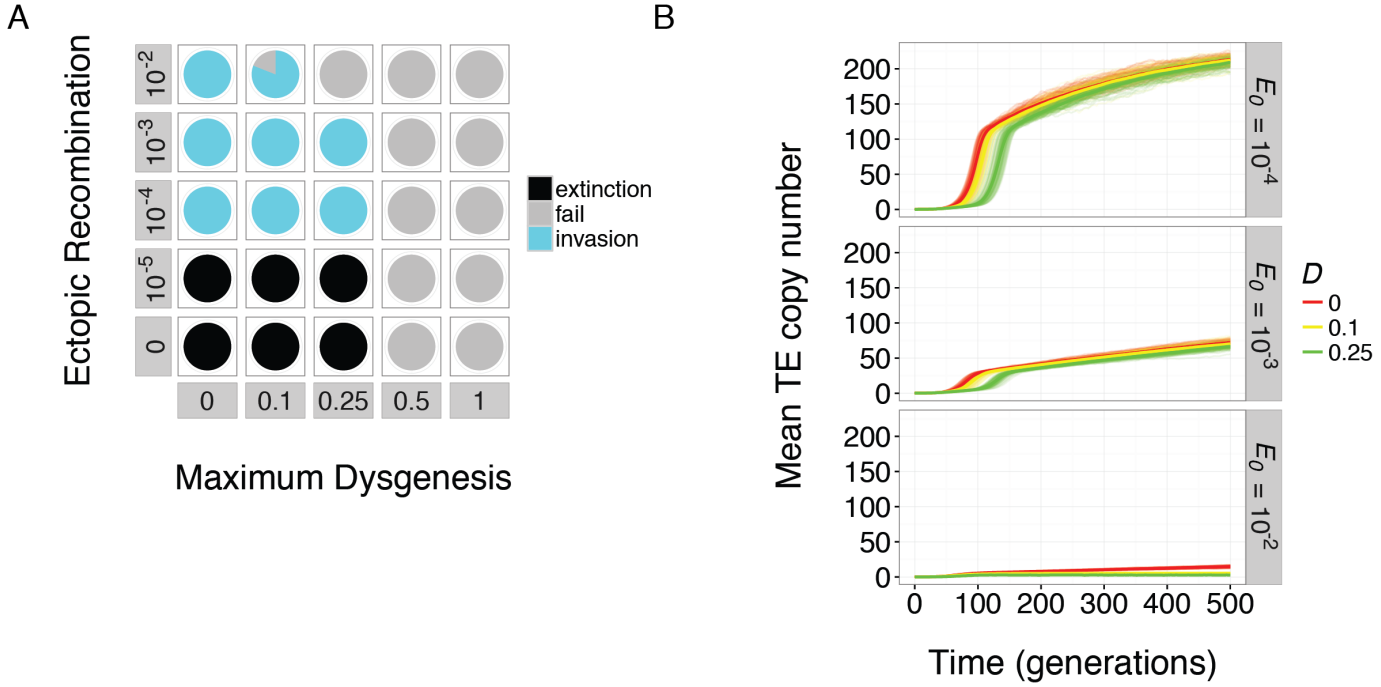

Figure 3 Ectopic recombination stabilizes TE copy numbers in the absence of repression. A) All possible combinations of 5 different baseline rates of ectopic recombination (E_0) and maximum dysgenic sterility (D) were considered. After 500 generations, four different outcomes were observed for the aggregate of ≥ 100 simulations for each combination of parameters. "Fail" indicates that $< 99\%$ of the genomes in a simulated population harbored at least one TE insertion. 'Invasion" indicates at least 1 TE copy was present in $\geq 99\%$ of genomes in a simulated population. "Extinction" indicates that the population went extinct due to an absence of viable males or females, within 500 generations of simulated evolution. B) Mean diploid copy number in 100 replicate populations is shown over 500 generations of simulated evolution. Different panels include different baseline ectopic recombination rates ($E_0 = 10^{-4}, 10^{-3}$, and 10^{-2}), while different colors compare different values of maximum dysgenesis ($D = 0$, 0.1, and 0.25). The solid line represents the average across all replicate simulated populations for the same D value. Small-RNA mediated repression was absent ($R = 0$, for all individuals, regardless of genotype) and $N = 2000$ in all simulations (A and B).

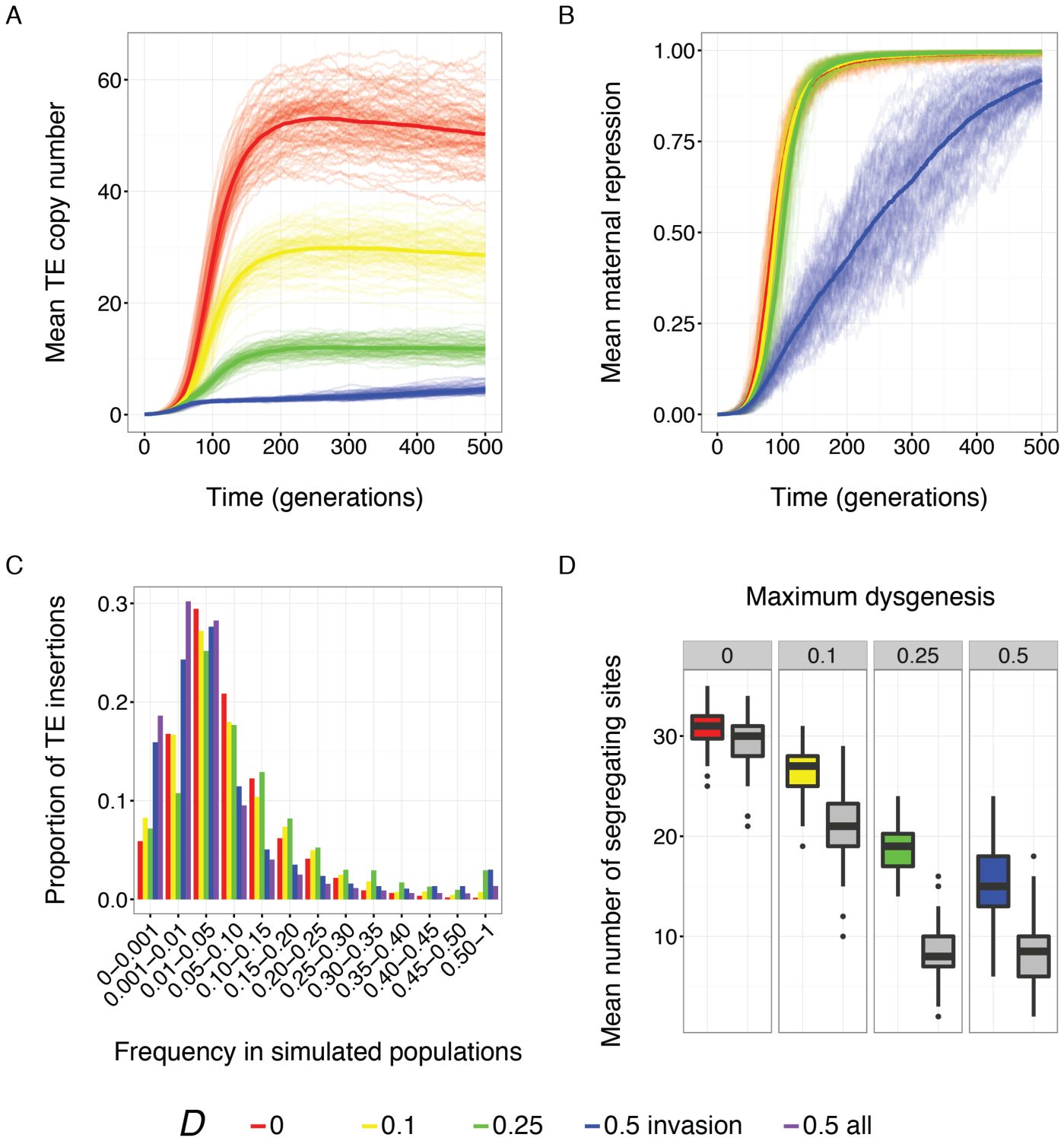

Figure 4 Positive selection on repressor alleles is enhanced by increasing costs of hybrid dysgenesis. Evolution of A) mean diploid TE copy number (n) and B) maternal repression (R) over 500 generations of simulated evolution. The thick line represents the average response across all replicate populations. C) The site frequency spectra of occupied small RNA producing sites after 500 generations of simulated evolution. The spectra were calculated by pooling the data from all replicate simulated populations. Site frequency bins include sites whose frequencies are higher than the lower bound, and less than or equal to the upper bound. D) The average number of occupied small RNA and pseudo-small RNA sites in each replicate simulated population after 500 generations. A site is considered occupied if it carries a TE insertion in at least one individual in a population. Population sizes were $N = 2000$ for all simulations, and individuals experienced no ectopic recombination $(E_0 = 0)$. Colors indicate simulations with different values for maximum hybrid dysgenesis (D). For $D = 0.5$, blue indicates only simulated populations that exhibited successful invasion while purple indicates all simulated populations.

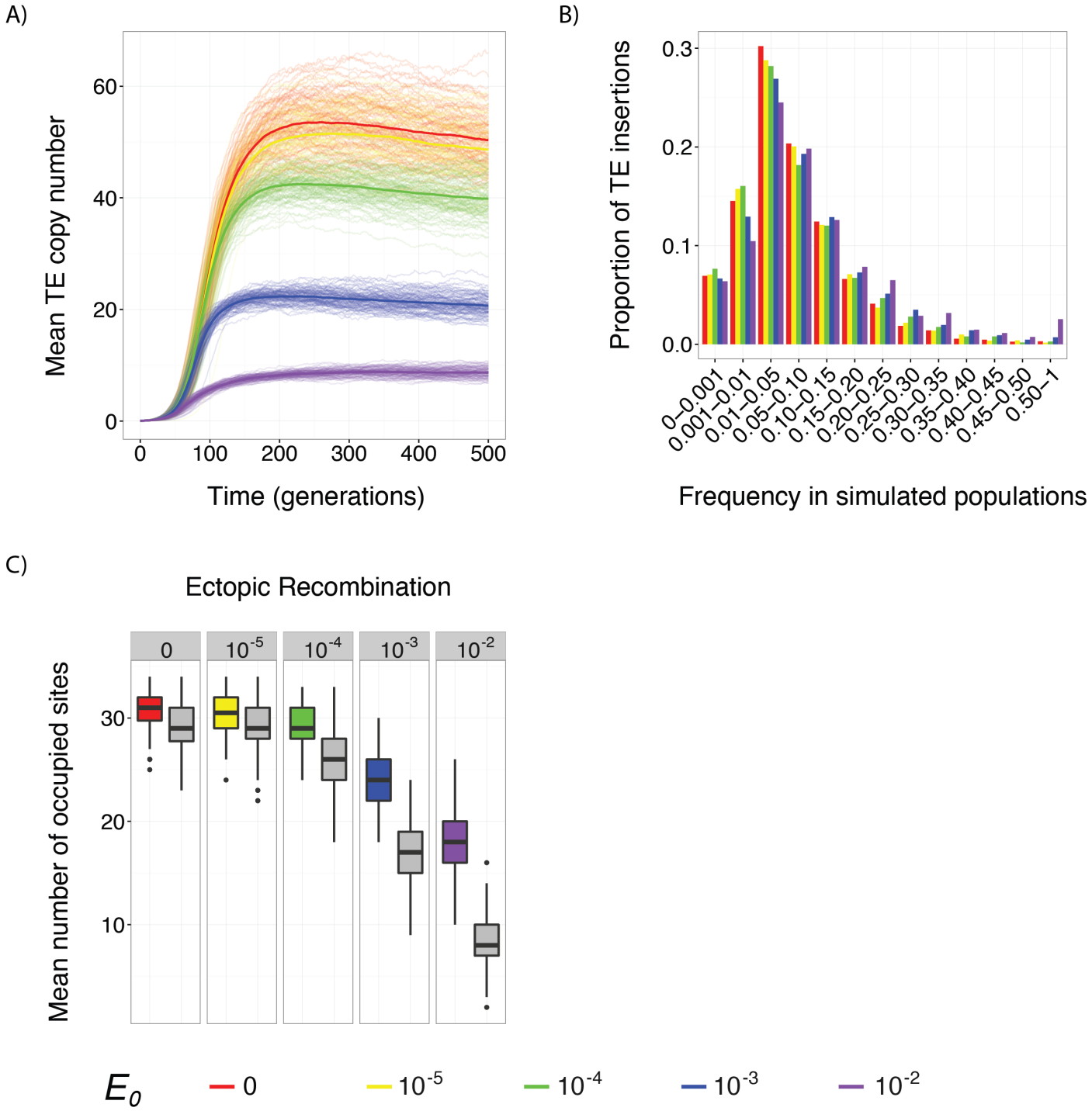

Figure 5 Positive selection on repressor alleles is enhanced by increasing costs of ectopic recombination. Evolution of mean diploid TE copy number (n) over 500 generations of simulated evolution. The thick line represents the average response across all replicate simulated populations. B and C) The site frequency spectra of occupied small RNA producing sites (B) and the average number of occupied small RNA and pseudo-small RNA sites (C) after 500 generations of simulated evolution. Site frequency spectra (B) were calculated by pooling the data from all replicate simulated populations. Site frequency bins include sites whose frequencies are higher than the lower bound, and less than or equal to the upper bound. A site is considered occupied (C) if it carries a TE insertion in at least one individual in a population. Population sizes were $N = 2000$ for all simulations, and individuals experienced no dysgenic sterility ($D = 0$). Colors compare simulations with different values of the baseline ectopic recombination rate (E_0).

the difference between the average number of occupied small RNA sites and pseudo-small RNA sites was increasingly pronounced at higher E_0 values, indicating that selection increasingly retained occupied small RNA sites in simulated populations (Figure 5C). Therefore, although ectopic recombination can stabilize TE copy numbers even in the absence of host repression (Charlesworth and Charlesworth 1983; Dolgin and Charlesworth 2008, Figure 4B), it also strengthens selection for repression when the repressive mechanism (*i.e.*, small RNA silencing) reduces the ectopic recombination rate.

Discussion

We harnessed the deep body of literature on *P*-element transposition and fitness effects (reviewed in Kelleher 2016), as well as piRNA-mediated silencing of TEs (reviewed in Senti and Brennecke 2010), in *D. melanogaster* to construct a forward simulation model of TE invasion and the evolution of host repression that is grounded in known biology. The validity of our model is supported by the degree to which our findings mirror empirical observations of *P*-element invasion in *D. melanogaster*. *P*-elements transitioned from rare to ubiquitous in natural population in < 20 years (Anxolabéhère *et al.* 1988), a rapid timescale that is matched by our simulated populations, in which invasion occurred in < 100 generations (Figure 1A). Similarly, average *P*-element copy numbers in individual genomes in our simulated populations were in the range $\sim 5 - 60$, which matches estimates from wild-derived genomes (Ronsseray *et al.* 1989; Srivastav and Kelleher 2017). Finally, repression in many natural populations evolved in > 30 years (Kidwell 1983), which is echoed by our observation that the vast majority of populations evolve repression in < 200 generations.

Although our model is, by design, informed by the specific biology of *P*-elements in *D. melanogaster*, our findings have general implications for TE invasion and the evolution of TE repression. Horizontal transfer of TEs and post-invasion TE proliferation are ubiquitous across the tree of life (reviewed in Schaack *et al.* 2010). Similarly, small RNA-mediated silencing is a ubiquitous mechanism for TE control, employed by both prokaryotes and eukaryotes (reviewed in Blumenstiel 2011). Our results therefore provide fundamental insights about the trajectory of TE invasion into new genomes, and the regulatory response of genomes to new invaders.

The abundance of small RNA producing sites is a major determinant of post-invasion dynamics

In the *D. melanogaster* genome we used as a model for our simulations ~3% of sites are thought to produce regulatory small RNAs when occupied by a TE (Brennecke *et al.* 2007). A relatively large mutational target for the acquisition small RNA-mediated silencing via transpositional insertion is also observed in other taxa. Although the power to identify loci that give rise to TE-regulating small RNAs can be limited by the quality of the genome assembly as well as its repeat content, large numbers of repeat rich small RNA producing sites are also implicated in TE regulation in mouse (Aravin *et al.* 2006), *Caenorhabditis* (Ni *et al.* 2014) and even *Arabidopsis*(reviewed in Lisch and Bennetzen 2011).

An appealing adaptive explanation for the abundance of small RNA sites is that the large mutation target facilitates the establishment of silencing when new TE families invade (Kelleher 2016). However, our findings do not support this model, as repression evolved within ~200 generations regardless of the number of small RNA sites in the theoretical genome (Figure 2A). Rather, we observed that the mutational target size has a dramatic impact on the maximum copy number the TE achieves after invasion, with genomes containing fewer small RNA sites accumulating many more TE copies (Figure 2C). This accumulation of TE copies can be attributed to longer waiting times for the occurrence of the first repressor allele, during which time TEs enjoy unrestricted transposition (Figure 2A, B).

Our observations suggest an unexpected benefit to having more small RNA sites, namely, that the larger mutational target buffers genomes against post-invasion proliferation. They further provide a potential explanation for rapid, TE-mediated expansions in genome size, which have been observed in certain vertebrate and plant lineages (Sun and Mueller 2014; Lee and Kim 2014). If the relative fraction of the genome that corresponds to small RNA producing sites decreases when genome-size expands, a type of snowballing could occur, in which large genomes experience increasingly dramatic TE proliferations and corresponding decreases in the relative abundance of small RNA producing sites.

Weak selection for repression of transposition?

Charlesworth and Langley (1986) suggested that selection for transpositional repression was likely to be weak in sexually reproducing organisms, because recombination would rapidly separate repressor alleles from chromosomes that they have protected from incurring deleterious mutations. Furthermore, they proposed that the adaptive evolution of transpositional repression relied on a sufficiently high frequency of dominant deleterious insertions and a sufficiently large effective population size. Consistent with this prediction, previous simulations of small RNA-mediated retrotransposon silencing, which assumed dominant deleterious effects of new insertions and large effective population sizes, observed selection for repressor alleles (Lu and Clark 2010). It is counter-intuitive, therefore, that we also observed positive selection for repressor alleles in relatively small populations of $N = 2000$ individuals, with entirely recessive effects of TE insertions on fitness.

The apparent contradiction between our findings and the predictions of Charlesworth and Langley (1986) are at least partially explained by differences in the transposition rates we considered. The empirically estimated transposition rate for an unmarked *P*-element of 0.1 (new insertions/element/genome) (Eggleston *et al.* 1988) is multiple orders of magnitude higher than "typical" estimates of transposition rate in *Drosophila* ($\sim 10^{-4} - 10^{-5}$; Nuzhdin 1999). The high transposition rates of *P*-elements may not be exceptional, however, as other transposition rate estimates were obtained from TE families we now understand to be regulated by the piRNA pathway (Nuzhdin 1999; Brennecke *et al.* 2007). Higher transposition rates cause greater fitness benefits of repression, because they increase the likelihood of a deleterious insertion and therefore the relative advantage of an un-inserted chromosome.

The fitness model that we employed in this study, which was based on the transpositional bias of *P*-elements into functional sites (Bellen *et al.* 2004, 2011; Spradling *et al.* 2011), and the estimated fitness effects of individual *P*-element insertions (Cooley *et al.* 1988; Mackay *et al.* 1992), likely also contributes to the strength of selection we observe. Although they act recessively, the empirically estimated fitness effects of *P*-element insertions are quite deleterious, with an average fitness cost of 12.2%, and 3.7% of sites acting as recessive lethals. Similar to

high transposition rates, these dramatic fitness consequences enhance the benefits of repressing transposition. Transposable elements exhibit a diversity of preferences for particular insertion sites, and the degree to which these preferences influence selection for transpositional repression remains unknown.

Repression of dysgenic sterility and ectopic recombination enhances adaptive evolution of small RNA-mediated TE silencing

In addition to deleterious insertions, we also revealed dysgenic sterility and ectopic recombination can exert positive selection on repressor alleles in simulated populations (Figure 4C, 5B). Our results differ from those of Lee and Langley (2012), who previously considered selection on small RNA independent repressors of dysgenic sterility. They observed that strong selection on the repressor allele was too ephemeral to produce significant changes its population frequency, owing to the rapid loss of non-repressive maternal genotypes from the population. However, they also assumed that the presence of *P*-elements in the maternal genome was sufficient to establish small RNA silencing, rather than explicitly modeling the production of regulatory small RNAs from insertions at particular sites as we do here. Our observations therefore suggest that dysgenic sterility can be an important agent of selection for TE regulation, at least during TE invasion. Dysgenic sterility may further be a common cost of TEs that selects for repression in their hosts. In *Drosophila*, multiple historic TE invasions are associated with dysgenic sterility syndromes (Bingham *et al.* 1982; Bucheton *et al.* 1984; Blackman *et al.* 1987; Evgen'ev *et al.* 1997; Hill *et al.* 2016), and at least three of these syndromes have been attributed to the absence of maternally deposited piRNAs (Blumenstiel and Hartl 2005; Brennecke *et al.* 2008; Chambeyron *et al.* 2008; Rozhkov *et al.* 2010). More broadly, mutations that impact TE regulation are associated with fertility defects in diverse eukaryotes (reviewed in Castañeda *et al.* 2011; Mani and Juliano 2013).

In forward simulations of TE dynamics, costs of ectopic recombination are described by a fitness model that assumes negative epistasis among TE insertions (Dolgin and Charlesworth 2006, 2008; Lu and Clark 2010; Lee and Langley 2012). Negative epistasis appears to be required to stabilize genomic TE copy numbers in the absence of repression (Charlesworth and Charlesworth 1983), a result that is supported by our observation that when repression is not permitted to evolve ($R = 0$), only models that include ectopic recombination exhibit stabilization of TE copy numbers in simulated populations (Figure 3). However, small RNA mediated silencing in particular may reduce the fitness costs of ectopic recombination by establishing a heterochromatic environment at TE loci that suppresses recombination (reviewed in Blumenstiel 2011). Positive selection on small RNA mediated repressors is therefore expected to be stronger in the presence of ectopic recombination. Our model supports this prediction, revealing enhanced positive selection on small-RNA mediated repressors at higher ectopic recombination rates (Figure 5B, C). Because TE copy numbers remained relatively low in the simulations we presented here (generally < 60 copies), the adaptive dynamics we observed could be greatly enhanced in large, repeat-rich genomes, where individual elements can be represented by thousands of copies (*e.g.*, Naito *et al.* 2006). Suppression of ectopic recombination may therefore represent a major benefit of small RNA mediated repressor alleles.

In contrast to the benefits of repressing transposition and potentially ectopic recombination, heterochromatin formation at TE loci (*i.e.*, transcriptional silencing) can also have a deleterious effect if it interferes with host gene expression. Such deleterious effects have been observed in both *Arabidopsis* (Hollister and Gaut 2009) and *Drosophila* (Lee 2015). Exploring trade-offs between the costs and benefits of repression, and how such trade-offs influence transcriptional versus post-transcriptional regulation of TEs, would make an interesting direction of future study.

Acknowledgements

R. Meisel provided useful comments on the manuscript. We used the Maxwell cluster from the Center of Advanced Computing and Data Systems (CACDS) at the University of Houston. CACDS staff provided technical support. The National Science Foundation (grant DEB-1457800 awarded to E.S.K. and grant DEB-1354952 awarded to R.B.R.A.) funded this work.

Literature Cited

Anxolabéhère, D., M. G. Kidwell, and G. Periquet, 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.

Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl, 2006 A novel class of small RNAs bind to MILI protein in mouse testes. Nature **442**: 203–207.

Bellen, H. J., R. W. Levis, Y. He, J. W. Carlson, M. Evans-Holm, E. Bae, J. Kim, A. Metaxakis, C. Savakis, K. L. Schulze, R. A. Hoskins, and A. C. Spradling, 2011 The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. Genetics **188**: 731–743.

Bellen, H. J., R. W. Levis, G. Liao, Y. He, J. W. Carlson, G. Tsang, M. Evans-Holm, P. R. Hiesinger, K. L. Schulze, G. M. Rubin, R. A. Hoskins, and A. C. Spradling, 2004 The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. Genetics **167**: 761–781.

Bingham, P. M., M. G. Kidwell, and G. M. Rubin, 1982 The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell **29**: 995–1004.

Blackman, R. K., R. Grimaila, M. M. Koehler, and W. M. Gelbart, 1987 Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. Cell **49**: 497–505.

Blumenstiel, J. P., 2011 Evolutionary dynamics of transposable elements in a small RNA world. Tr. Genet. **27**: 23–31.

Blumenstiel, J. P. and D. L. Hartl, 2005 Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. Proc. Natl. Acad. Sci. U. S. A. **102**: 15965–15970.

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

Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark, and G. J. Hannon, 2008 An epigenetic role for maternally inherited piRNAs in transposon silencing. Science **322**: 1387–1392.

Bucheton, A., R. Paro, H. M. Sang, A. Pelisson, and D. J. Finnegan, 1984 The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. Cell **38**: 153–163.

Castañeda, J., P. Genzor, and A. Bortvin, 2011 piRNAs, transposon silencing, and germline genome integrity. Mutat. Res. **714**: 95–104.

Chambeyron, S., A. Popkova, G. Payen-Groschêne, C. Brun, D. Laouini, A. Pelisson, and A. Bucheton, 2008 piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. Proc. Natl. Acad. Sci. U. S. A. **105**: 14964–14969.

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

Charlesworth, B. and C. H. Langley, 1986 The evolution of self-regulated transposition of transposable elements. Genetics **112**: 359–383.

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

Cooley, L., C. Berg, and A. Spradling, 1988 Controlling P element insertional mutagenesis. Tr. Genet. **4**: 254–258.

Dolgin, E. S. and B. Charlesworth, 2006 The fate of transposable elements in asexual populations. Genetics **174**: 817–827.

Dolgin, E. S. and B. Charlesworth, 2008 The effects of recombination rate on the distribution and abundance of transposable elements. Genetics **178**: 2169–2177.

Dupuy, A. J., S. Fritz, and D. A. Largaespada, 2001 Transposition and gene disruption in the male germline of the mouse. Genesis **30**: 82–88.

Eggleston, W. B., D. M. Johnson-Schlitz, and W. R. Engels, 1988 P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. Nature **331**: 368–370.

Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston, and J. Sved, 1990 High-frequency P element loss in Drosophila is homolog dependent. Cell **62**: 515–525.

Evgen'ev, M. B., H. Zelentsova, N. Shostak, M. Kozitsina, V. Barskyi, D. H. Lankenau, and V. G. Corces, 1997 Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. Proc. Natl. Acad. Sci. U. S. A. **94**: 196–201.

Fay, J. C. and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155**: 1405–1413.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M. C. Siomi, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. Science **315**: 1587–1590.

Haynes, K. A., A. A. Caudy, L. Collins, and S. C. R. Elgin, 2006 Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. Curr. Biol. **16**: 2222–2227.

Hedges, D. J. and P. L. Deininger, 2007 Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. **616**: 46–59.

Hill, T., C. Schlötterer, and A. J. Betancourt, 2016 Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the *P*-element. PLoS Genet. **12**: e1005920.

Hollister, J. D. and B. S. Gaut, 2009 Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. **19**: 1419–1428.

Huang, C. R. L., K. H. Burns, and J. D. Boeke, 2012 Active transposition in genomes. Annu. Rev. Genet. **46**: 651–675.

Josefsson, C., B. Dilkes, and L. Comai, 2006 Parent-dependent loss of gene silencing during interspecies hybridization. Curr. Biol. **16**: 1322–1328.

Kelleher, E. S., 2016 Reexamining the *P*-element invasion of *Drosophila melanogaster* through the lens of piRNA silencing. Genetics **203**: 1513–1531.

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

Kidwell, M. G., 1983 Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. **80**: 1655–1659.

Klenov, M. S., S. A. Lavrov, A. D. Stolyarenko, S. S. Ryazansky, A. A. Aravin, T. Tuschl, and V. A. Gvozdev, 2007 Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucl. Acids Res. **35**: 5430–5438.

Kocur, G. J., E. A. Drier, and M. J. Simmons, 1986 Sterility and hypermutability in the P-M system of hybrid dysgenesis in *Drosophila melanogaster*. Genetics **114**: 1147–1163.

Kunarso, G., N.-Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.-S. Chan, H.-H. Ng, and G. Bourque, 2010 Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. **42**: 631–634.

Langley, C. H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52**: 223–235.

Lee, S.-I. and N.-S. Kim, 2014 Transposable elements and genome size variations in plants. Genomics Inform. **12**: 87–97.

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

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

Lisch, D. and J. L. Bennetzen, 2011 Transposable element origins of epigenetic gene regulation. Curr. Opin. Plant Biol. **14**: 156–161.

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

Lukic, S. and K. Chen, 2011 Human piRNAs are under selection in Africans and repress transposable elements. Mol. Biol. Evol. **28**: 3061–3067.

Lyman, R. F., F. Lawrence, S. V. Nuzhdin, and T. F. Mackay, 1996 Effects of single *P*-element insertions on bristle number and viability in *Drosophila melanogaster*. Genetics **143**: 277–292.

Lynch, V. J., R. D. Leclerc, G. May, and G. P. Wagner, 2011 Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. **43**: 1154–1159.

Mackay, T. F., R. F. Lyman, and M. S. Jackson, 1992 Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics **130**: 315–332.

Mani, S. R. and C. E. Juliano, 2013 Untangling the web: the diverse functions of the PIWI/piRNA pathway. Mol. Reprod. Devel. **80**: 632–664.

Marin, L., M. Lehmann, D. Nouaud, H. Izaabel, D. Anxolabéhère, and S. Ronsseray, 2000 P-Element repression in Drosophila

melanogaster by a naturally occurring defective telomeric P copy. Genetics **155**: 1841–1854.

Morgan, T. H., 1914 No crossing over in the male of *Drosophila* of genes in the second and third pairs of chromosomes. Biol. Bull. **26**: 195–204.

Naito, K., E. Cho, G. Yang, M. A. Campbell, K. Yano, Y. Okumoto, T. Tanisaka, and S. R. Wessler, 2006 Dramatic amplification of a rice transposable element during recent domestication. Proc. Natl. Acad. Sci. U. S. A. **103**: 17620–17625.

Ni, J. Z., E. Chen, and S. G. Gu, 2014 Complex coding of endogenous siRNA, transcriptional silencing and H3K9 methylation on native targets of germline nuclear RNAi in *C. elegans*. BMC Genomics **15**: 1157.

Nielsen, R., 2005 Molecular signatures of natural selection. Annual Review of Genetics **39**: 197–218.

Nuzhdin, S. V., 1999 Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica **107**: 129–137.

Orsi, G. A., E. F. Joyce, P. Couble, K. S. McKim, and B. Loppin, 2010 *Drosophila* I-R hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. J. Cell Sci. **123**: 3515–3524.

R Core Team, 2016 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ronsseray, S., M. Lehmann, and D. Anxolabéhère, 1989 Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations. Chromosoma **98**: 207–214.

Ronsseray, S., M. Lehmann, and D. Anxolabéhère, 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., M. Lehmann, D. Nouaud, and D. Anxolabéhère, 1996 The regulatory properties of autonomous subtelomeric *P* elements are sensitive to a *Suppressor of variegation* in *Drosophila melanogaster*. Genetics **143**: 1663–1674.

Ronsseray, S., L. Marin, M. Lehmann, and D. Anxolabéhère, 1998 Repression of hybrid dysgenesis in *Drosophila melanogaster* by combinations of telomeric *P*-element reporters and naturally occurring *P* elements. Genetics **149**: 1857–1866.

Rozhkov, N. V., A. A. Aravin, E. S. Zelentsova, N. G. Schostak, R. Sachidanandam, W. R. McCombie, G. J. Hannon, and M. B. Evgen'ev, 2010 Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. RNA **16**: 1634–1645.

Rubin, G. M. and A. C. Spradling, 1982 Genetic transformation of *Drosophila* with transposable element vectors. Science **218**: 348–353.

Schaack, S., C. Gilbert, and C. Feschotte, 2010 Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Tr. Ecol. Evol. **25**: 537–546.

Senti, K.-A. and J. Brennecke, 2010 The piRNA pathway: a fly's perspective on the guardian of the genome. Tr. Genet. **26**: 499–509.

Sentmanat, M. F. and S. C. R. Elgin, 2012 Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. Proc. Natl. Acad. Sci. U. S. A. **109**: 14104–14109.

Sienski, G., D. Dönertas, and J. Brennecke, 2012 Transcriptional silencing of transposons by Piwi and Maelstrom and its impact on chromatin state and gene expression. Cell **151**: 964–980.

Silva-Sousa, R., E. López-Panadès, and E. Casacuberta, 2012 *Drosophila* telomeres: an example of co-evolution with transposable elements. Genome Dyn. **7**: 46–67.

Simmons, M. J., M. W. Meeks, E. Jessen, J. R. Becker, J. T. Buschette, and M. W. Thorp, 2014 Genetic interactions between *P* elements involved in piRNA-mediated repression of hybrid dysgenesis in *Drosophila melanogaster*. G3 **4**: 1417–1427.

Simmons, M. J., J. B. Niemi, D.-F. Ryzek, C. Lamour, J. W. Goodman, W. Kraszkiewicz, and R. Wolff, 2007 Cytotype regulation by telomeric *P* elements in *Drosophila melanogaster*: interactions with *P* elements from M′ strains. Genetics **176**: 1957–1966.

Simmons, M. J., L. M. Ragatz, I. R. Sinclair, M. W. Thorp, J. T. Buschette, and C. D. Grimes, 2012 Maternal enhancement of cytotype regulation in *Drosophila melanogaster* by genetic interactions between telomeric *P* elements and non-telomeric transgenic *P* elements. Genet. Res. **94**: 339–351.

Spradling, A. C., H. J. Bellen, and R. A. Hoskins, 2011 *Drosophila P* elements preferentially transpose to replication origins. Proc. Natl. Acad. Sci. U. S. A. **108**: 15948–15953.

Spradling, A. C., D. Stern, A. Beaton, E. J. Rhem, T. Laverty, N. Mozden, S. Misra, and G. M. Rubin, 1999 The Berkeley Drosophila Genome Project gene disruption project: Single *P*-element insertions mutating 25% of vital Drosophila genes. Genetics **153**: 135–177.

Srivastav, S. P. and E. S. Kelleher, 2017 Paternal induction of hybrid dysgenesis in *Drosophila melanogaster* is weakly correlated with both *P*-element and *hobo* element dosage. G3 **7**: g3.117.040634.

Staveley, B. E., T. R. Heslip, R. B. Hodgetts, and J. B. Bell, 1995 Protected *P*-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in *Drosophila melanogaster*. Genetics **139**: 1321–1329.

Stuart, J. R., K. J. Haley, D. Swedzinski, S. Lockner, P. E. Kocian, P. J. Merriman, and M. J. Simmons, 2002 Telomeric *P* elements associated with cytotype regulation of the *P* transposon family in *Drosophila melanogaster*. Genetics **162**: 1641–1654.

Sun, C. and R. L. Mueller, 2014 Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. Genome Biol. Evol. **6**: 1818–1829.

Wickham, H., 2009 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yannopoulos, G., N. Stamatis, M. Monastirioti, P. Hatzopoulos, and C. Louis, 1987 hobo is responsible for the induction of hybrid dysgenesis by strains of Drosophila melanogaster bearing the male recombination factor *23.5MRF*. Cell **49**: 487–495.

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