OXFORD GENETICS

The impact of paramutations on the invasion dynamics of transposable elements

Almorò Scarpa,^{1,2} Robert Kofler (D^{1,*}

¹Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, Wien 1210, Austria ²Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria

*Corresponding author: Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, Wien 1210, Austria. Email: rokofler@gmail.com

According to the prevailing view, the trap model, the activity of invading transposable elements (TEs) is greatly reduced when a TE copy jumps into a piRNA cluster, which triggers the emergence of piRNAs that silence the TE. One crucial component in the host defence are paramutations. Mediated by maternally deposited piRNAs, paramutations convert TE insertions into piRNA producing loci, thereby transforming selfish TEs into agents of the host defence. Despite this significant effect, the impact of paramutations on the dynamics of TE invasions remains unknown. To address this issue, we performed extensive forward simulations of TE invasions with piRNA clusters and paramutations. We found that paramutations significantly affect TE dynamics, by accelerating the silencing of TE invasions, reducing the number of insertions accumulating during the invasions and mitigating the fitness cost of TEs. We also demonstrate that piRNA production induced by paramutations, an epigenetically inherited trait, may be positively selected. Finally, we show that paramutations may account for three important open problems with the trap model. Firstly, paramutated TE insertions may compensate for the insufficient number of insertions in piRNA clusters observed in previous studies. Secondly, paramutations may explain the discrepancy between the observed and the expected abundance of different TE families in *Drosophila melanogaster*. Thirdly, piRNA clusters may be crucial to trigger the host defence, but paramutations render the clusters dispensable once the defence has been established. This could account for the lack of TE activation when three major piRNA clusters were deleted in a previous study.

Keywords: paramutations; transposable elements; simulations; piRNAs; population genetics

Introduction

Transposable elements (TEs) are short stretches of DNA that selfishly multiply within genomes (Doolittle and Sapienza 1980; Orgel and Crick 1980). They are highly successful, having invaded virtually all eukaryotic species investigated so far (Wicker et al. 2007). The effects of TE insertions on host fitness are still debated. While some TE insertions may confer a selective advantage (González et al. 2008; Casacuberta and González 2013), it is likely that the majority of TEs are either neutral or deleterious to the host (Mackay et al. 1991; Pasyukova et al. 2004; Arkhipova 2018). It is feasible that TEs may even threaten the persistence of species, as theoretical and experimental work suggests that TE invasions could lead to the extinction of populations (Kofler 2019; Selvaraju et al. 2022; Wang et al. 2023).

To combat the spread of TEs, host organisms have evolved sophisticated defence mechanisms, frequently involving small RNAs (Sarkies et al. 2015). In mammals and invertebrates, the host defence relies on piwi-interacting RNAs (piRNAs), small RNAs ranging in size from 23 to 29nt (Brennecke et al. 2007; Gunawardane et al. 2007). piRNAs bound to PIWI clade proteins mediate the transcriptional and post-transcriptional silencing of TEs (Brennecke et al. 2007; Gunawardane et al. 2007; Sienski et al. 2012; Le Thomas et al. 2013). These piRNAs are largely derived from discrete genomic loci, the piRNA clusters (Brennecke et al. 2007). In Drosophila melanogaster, for example, piRNA clusters account for about 3% of the genome (Brennecke et al. 2007). According to the prevailing view, the trap model, the activity of an invading TE is greatly reduced when a copy of the invading TE jumps into a piRNA cluster, which triggers the emergence of piRNAs complementary to the invading TE (Bergman et al. 2006; Malone et al. 2009; Zanni et al. 2013; Goriaux et al. 2014; Yamanaka et al. 2014; Ozata et al. 2018). Several lines of evidence support this model. The insertion of artificial sequences into piRNA clusters leads to the emergence of piRNAs complementary to the inserted sequence (Kawaoka et al. 2012; Muerdter et al. 2012; Luo et al. 2022). A single insertion in a piRNA cluster, such as X-TAS or 42AB, is sufficient to silence a reporter (Josse et al. 2007; Luo et al. 2022). Finally, piRNA clusters are largely composed of many different TEs, thought to represent an "immunological memory" of past TE invasions (Brennecke et al. 2007; Zanni et al. 2013; Wierzbicki et al. 2023; Wierzbicki and Kofler 2023).

However, recently some doubts about the trap model have emerged. First, TE invasions in experimental populations showed that the number of cluster insertions in later generations, where the host defence has been largely established, is lower than expected under the trap model (Kofler et al. 2018, 2022; Selvaraju et al. 2022). In agreement with this, an investigation of the composition of piRNA clusters in long-read assemblies of different D. *melanogaster* strains showed that many TE families have fewer insertions than expected (Wierzbicki and Kofler 2023). Second, theoretical work shows that the number of insertions accumulating

Downloaded from https://academic.oup.com/genetics/article/225/4/iyad181/7306651 by Veterinarmedizinische Universitat Wien user on 21 August 2024

Received on 14 July 2023; accepted on 28 September 2023

[©] The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

during an invasion largely depends on the size of piRNA clusters (Kelleher et al. 2018; Kofler 2019; Tomar et al. 2022). If the size of the piRNA clusters is known (as a percentage of the genome), quantitative predictions can be made about the expected frequency of the different TE families. However, the abundance of most families in D. melanogaster is much lower than expected under the trap model (Kofler 2019). Third, the deletion of three major piRNA clusters in D. melanogaster did not lead to an activation of TEs (Gebert et al. 2021). The authors argue that dispersed piRNA-producing TE insertions outside of piRNA clusters could compensate for the loss of piRNA clusters. Such dispersed piRNA-producing loci were observed for most TE families in D. melanogaster (Shpiz et al. 2014; Wierzbicki and Kofler 2023). By mediating the installation of distinct chromatin marks, maternally deposited piRNAs are thought to direct the conversion of TE insertions into piRNA producing loci (de Vanssay et al. 2012; Le Thomas et al. 2014). This conversion of TE insertions resembles paramutations (de Vanssay et al. 2012; Hollick 2017). Paramutations were first discovered in maize and generally describe heritable changes in the behavior of alleles due to trans-acting (epigenetic) interactions (Brink 1958; Hollick 2017). According to a strict definition the term paramutation refers to inheritable changes among alleles. However, in Drosophila it was realized that paramutation-like behavior may also occur between distinct sites that could even be located on different chromosomes (e.g. P-1152 on the X-chromosome and BX2 on the 2nd chromosome de Vanssay et al. 2012; Hermant et al. 2015). In agreement with previous works (de Vanssay et al. 2012; Hermant et al. 2015), we thus use the term paramutation to refer to the conversion of a TE insertion into a piRNA producing locus, irrespective of whether the converted TE is at the same or a distant locus. We note that we are therefore deviating from the strict definition of paramutation that is solely referring to allelic interactions.

However, apart from generating dispersed piRNA-producing TE insertions, maternally deposited piRNAs may also direct the formation of piRNA clusters in the next generation (de Vanssay et al. 2012). Paramutations could thus be necessary for the long-term maintenance of the location of piRNA clusters across generations (although some genomic motifs may also be involved Baumgartner et al. 2022). Since paramutations depend on maternally deposited piRNAs, they can only explain the maintenance of TE silencing, but they cannot account for the emergence of novel piRNAs. Paramutations may thus solely act as an extension to existing models of the de novo silencing of TE invasions and not as an alternative hypothesis. One event that may trigger the emergence of such novel piRNAs is an insertion into a piRNA cluster. Therefore, piRNA clusters may be important to trigger the emergence of the first piRNAs complementary to an invading TE, but they may be dispensable once a host defence is established (Chen and Aravin 2021). An alternative hypothesis suggests that siRNAs could trigger the emergence of the first piRNAs complementary to an invading TE (Luo et al. 2022). Mediated by Dicer-2, dsRNA composed of sense and antisense transcripts of a TE may be cleaved into siRNAs that could, similarly to paramutations, turn TE-rich regions into piRNA-producing loci. Interestingly, loci that generate such siRNAs (e.g. hairpins) may not necessarily themselves be converted into piRNA-producing loci (Luo et al. 2022). The sites that could trigger siRNA production are thus distinct from the sites that may produce piRNAs.

Previous theoretical works have studied the dynamics of TE invasions with piRNA clusters (i.e. under the trap model) (Li et al. 2009; Kelleher et al. 2018; Kofler 2019, 2020; Tomar et al. 2022). Some frequent assumptions of these theoretical studies of the trap model are: TEs spread in populations with a constant transposition rate, insertion sites are random, piRNA clusters occupy a certain fraction of the genome and a single insertion in a piRNA cluster silences a TE. These theoretical works provided important insights. For example, they showed that, due to recombination among segregating TE insertions, on average, about four cluster insertions are necessary to stop a TE invasion, TE invasions with piRNA clusters have three distinct phases, TE insertions in piRNA clusters may be positively selected, and piRNA clusters may prevent the extinction of populations, provided that the clusters have a minimum size (Lu and Clark 2009; Kelleher et al. 2018; Kofler 2019, 2020; Tomar et al. 2022). Although paramutations are a crucial component of the host defence, the impact of paramutations on the dynamics of TE invasions is unclear. Investigating the effect of paramutations is especially important since it has been speculated that they may account for some of the important open problems with the trap model, such as the insufficient number of TE insertions found in piRNA clusters (see above; Kofler et al. 2022; Wierzbicki and Kofler 2023).

To address this issue, we performed extensive computer simulations of TE invasions with piRNA clusters and paramutations. We also explore TE invasions under a model where siRNAs initiate the silencing of an invading TE. Using our novel simulator, InvadeGO, we show that paramutations accelerate the silencing of TE invasions, limit the amount of TEs accumulating during invasions and reduce the fitness burden of TEs. We also found that piRNA production induced by paramutations, an epigenetically inherited trait, may be positively selected. Finally, we show that paramutations can account for several of the important open issues with the trap model. First, paramutations reduce the number of cluster insertions necessary to stop a TE invasion and, therefore, may account for the low number of cluster insertions found in experimental populations. Second, since paramutations reduce the amount of TEs accumulating during TE invasion, they may explain why the TE abundance in D. melanogaster is lower than expected under the trap model. A model with 10% paramutable loci and 3% piRNA clusters roughly captures the abundance of the TE families in D. melanogaster, even in the absence of negative selection against TEs. Finally, we find that the redundancy in piRNA-producing loci generated by paramutations renders piRNA clusters dispensable once a host defence is established, explaining why Gebert et al. (2021) were able to delete three major clusters without effect on TE activity.

Materials and methods Simulation software

To simulate TE invasions with paramutations, we implemented a novel simulation software in the Go programming language: InvadeGO (v0.2.3). Several algorithms implemented in InvadeGO are inspired by our previous simulation tool, Invade, which is implemented in Java (Kofler 2019). InvadeGO performs individualbased forward simulations of TE invasions in populations of diploid organism using discrete (nonoverlapping) generations. A TE insertion is modeled as position (integer) in the half-open interval [0, q), where q is the genome size. The TE insertions in a haploid genome are thus modeled as a list of integers (i.e. the genomic insertion sites). A diploid individual carries two separate lists of insertion sites. Each chromosome occupies a unique nonoverlapping territory in the genomic interval [0, g), such that each TE insertion is part of exactly one chromosome. The piRNA clusters occupy sub-regions of the chromosomes, such that a TE insertion may be part of either none or one piRNA cluster. Recurrent sites, such as paramutable loci and the siRNA-trigger-sites are modeled using the modulo operation. For example, with x%10 = =3 the sites 3, 13, 23...would be paramutable loci (except for sites in piRNA clusters). Maternally deposited piRNAs are modeled as a binary trait, where each individual may have (m > 0) or not-have (m = 0) maternal piRNAs. Each individual has a fitness w, which solely depends on the number of TE insertions w = 1 - xn, where x is the negative effect of a single TE insertion and *n* is the number of TE insertions per diploid individual. The fitness determines the mating probability (i.e. fecundity selection). We simulated two sexes, where "males" may only mate with "females." Each parent generates a single gamete that is passed to the offspring. To create a gamete, first recombination and random assortment among chromosomes are simulated and then novel transposition events are introduced into the recombined gamete. We assumed that TEs multiply with a given transposition rate *u*, which is the probability that a TE insertion will generate a novel insertion in the next generation. A transposition rate of zero (u = 0) was used for individuals carrying a piRNA producing locus (either a cluster insertion or a paramutated TE). To avoid excessive computation times, we calculated, for each gamete, the number of novel insertion sites based on a Poisson distributed random variable with $\lambda = u * n/2$. Novel insertion sites were randomly chosen in the genomic interval [0, g]. If a site was already occupied, the novel insertion was ignored. The sex was randomly assigned to each offspring. Only females transmitted the maternal piRNA status to the offspring.

InvadeGO allows providing a wide range of different parameters, such as the number of chromosomes, the size of the chromosomes, the size of the piRNA clusters, the distribution of the paramutable loci and of the siRNA-trigger-sites, the recombination rate, the transposition rate, the population size, the number of generations, the number of TE insertions in the base population, the negative effect of TEs and a flag indicating whether or not cluster insertions are selectively neutral. For the base population, it is also feasible to provide a file that contains for each individual the positions of the TE insertions, sex, and the maternal piRNA status.

To test whether piRNA clusters are dispensable after the piRNA-based host defence was established, we developed a novel branch of InvadeGO (remove cluster; v.br.rc-0.2.3) that allows removing a given number of piRNA clusters at the specified time.

InvadeGO was thoroughly tested with unit-tests (all tests can be run with the command go test ./ \cdots). Furthermore, we validated the correct implementation of all relevant population genetic forces, such as drift, recombination, selection and transposition (Supplementary Figs. S1–S7).

Simulations and data analysis

To minimize the parameter space, we performed simulations with default conditions and varied the parameter of interest. For our default conditions, we used a genome consisting of 5 chromosomes each having a size of 10 Mbp (-genome MB:10,10,10,10,10,10), a recombination rate of 4cM/Mbp (-rr 4,4,4,4), 10% paramutable loci (-paramutation 10:1), piRNA clusters with a size of 300kbp at the beginning of each chromosome (-cluster kb:300,300,300, 300,300; i.e. 3% of the genome), a transposition rate of 0.1 (-u 0.1), population size of 1,000 (-N 1000) and a base population with 100 randomly distributed TE insertions (-basepop 100). Initially, we simulated neutral TE insertions (-x 0) but later explored more complex scenarios with negatively selected TEs. To cover the parameter space in two dimensions (Fig. 3c,d), we used Python scripts that launched about 10,000 simulations with

randomly chosen parameter combinations. All data analysis was done in R (R Core Team 2012) and data were visualized with ggplot2 (Wickham 2016).

Results

Model assumptions and implementation

In this work, we explore the influence of paramutations on the dynamics of TE invasions. We rely on the following definitions and model assumptions: A 'paramutation' refers to the conversion of a regular TE insertion—outside of piRNA clusters—into a piRNA-producing locus (Fig. 1a: scenario 7). Paramutations are mediated by maternally deposited piRNAs. A 'paramutable locus' is a genomic region (outside of piRNA clusters) where a TE insertion could be paramutated (Fig. 1a: scenarios 3, 4, 7, 8; violet areas). Previous works have shown that not all TE insertions are paramutated (de Vanssay et al. 2012; Shpiz et al. 2014; Wierzbicki and Kofler 2023). We therefore assumed a variable fraction of paramutable loci. It is unclear what determines whether or not an insertion may be paramutated but it is likely that the genomic neighborhood plays a role. For example, a TE insertion into a repetitive region may be more readily paramutated than an insertion in a gene. We thus assumed that the genomic site determines whether or not a locus is paramutable. As a consequence, all individuals in the population have the same set of paramutable loci. A 'paramutable TE' is a TE insertion in a paramutable locus which does not produce piRNAs, i.e. the TE has not (yet) been paramutated (Fig. 1a: scenario 3; TE on the right). A 'paramutated TE' is a TE insertion in a paramutable locus that produces piRNAs, i.e. the TE has been paramutated (Fig. 1a: scenario 7; TE on the right). Since paramutations require both maternally deposited piRNAs and a TE insertion in a paramutable locus, a TE silenced in the mother may be reactivated in the offspring if the maternally deposited piRNAs but not the paramutable TE (nor a cluster insertion) is inherited (Fig. 1a: scenario 5). This is in agreement with previous work which has shown that the maternally deposited component without chromosomal component is not sufficient to maintain the silencing of a reporter construct (Josse et al. 2007; Luo et al. 2022). As a consequence, we assume that a TE is active in all individuals not having at least one piRNA-producing locus, i.e. either an insertion into a piRNA cluster or a paramutated TE (Fig. 1a: scenarios 1,3,5). One important consideration is that paramutations may maintain the silencing of an invading TE, but since they depend on maternally deposited piRNAs, they cannot explain the origin of the first piRNAs against an invading TE. Initially, we explored a scenario where a TE insertion in a piRNA cluster triggers the emergence of the first piRNAs against an invading TE (Fig. 1a: scenarios 2,4). Later, we consider the case that siRNAs might activate the piRNA-based host defence (Luo et al. 2022).

To explore the influence of paramutations on TE invasions, we performed individual-based forward simulations. With our novel simulator InvadeGO, developed in the Go programming language, we simulated TE invasions in panmictic populations of diploid organisms with discrete generations. Since paramutations depend on maternally transmitted piRNAs, we simulated two sexes (males and females) in contrast to our previous work where only hermaphrodites were simulated (Kofler 2019, 2020). The presence/absence of maternally deposited piRNAs was simulated as a binary trait. We thoroughly validated the novel simulator by comparing the observed results against theoretical expectations for all relevant population genetic forces, such as transposition, drift, recombination and selection (Supplementary Figs. S1–S4).



Fig. 1. An overview of our simulations. a) Overview of model assumptions and the notation used in this work. b) Overview of the simulated genomes. We simulated five chromosomes, each with a piRNA cluster at the end. In the simulations with paramutations, the paramutable loci were evenly distributed in the genome. c) Abundance of TE insertions during TE invasions with different fractions of paramutable loci (para; top panel). Note that paramutations substantially reduce the number of TE insertions accumulating during a TE invasion. Colors indicate the three distinct phases of TE invasions (see text).

Additionally, we validated the correct implementation of the genomic landscape, piRNA clusters and paramutations (Supplementary Figs. S5–S7). All simulations and analyses conducted in this work have been carefully documented with RMarkdown and deposited, together with the resulting figures, at GitHub (https://github.com/Almo96/Paramutations'TEs; see the md-files in the subfolders).

In this work, we simulated 5 chromosome arms with a length of 10Mb, a uniform recombination rate of 4cM/Mbp and piRNA clusters at the beginning of each chromosome (Fig. 1b). We assumed that the paramutable loci are regularly distributed in the genome, e.g. with 10% paramutable loci a TE insertion at each 10th locus could be paramutated (Fig. 1b).

Unless mentioned otherwise, we used the following default parameters: a transposition rate of u = 0.1, a population size of N = 1,000, piRNA clusters with a size of 300 kb (i.e. 3% of the genome) and 10% paramutable sites. Initially, we explored neutral TE insertions (x = 0.0) but later simulated more complex scenarios with negatively selected TEs. We triggered the TE invasions by distributing 100 insertions randomly in a population of 1000 individuals (population frequency $f = 1/(2 \cdot 1,000)$) to avoid the early stochastic stage of TE invasions, where TEs are frequently lost by genetic drift (Le Rouzic and Capy 2005; Kofler 2019).

We first asked whether TE invasions can be stopped under a trap model with paramutations. We define an invasion to be "stopped" (inactive phase) once a piRNA producing locus (either a cluster insertion or paramutated TE) becomes fixed in the population, since the TE can no longer be active in any single individual at any generation following fixation. All TE insertions start at a low population frequency (f = 1/2N) but due to genetic drift the population frequency of all TE insertions (insertions in piRNA clusters, paramutable loci and the rest of the genome) increases gradually until eventually most insertions reach fixation (Supplementary Fig. S8). Irrespective of the number of paramutable loci, 92–100% of the invasions were stopped after 5,000 generations and all after 10,000 generations (Supplementary Table S1; Fig. 1c).

Next, we asked whether paramutations have an influence on the dynamics of TE invasions. To test this, we simulated TE invasions in populations with varying fractions of paramutable loci (0%, 1%, 10% and 100%; (Fig. 1c top panel). Note that throughout the manuscript the scenario without paramutable loci (0%) corresponds to the classic trap model, which holds that TE invasions are solely controlled by TE insertions in piRNA clusters. Interestingly, we observed that paramutations exert a profound effect on TE invasions (Fig. 1c), highlighting the importance of investigating the impact of paramutations on the dynamics of TE invasions.

Effect of paramutations on TE invasions

To quantify the effect of paramutations on TE invasions, we relied on the three phases of TE invasions identified in a previous work (Kofler 2019). In our previous work, we only considered insertions in piRNA clusters to define the three phases (no paramutations were considered). For this work, we extended the definition of the phases to incorporate paramutated loci. Thus, we define the three phases as follows; In the first phase, the rapid invasion phase, the TE may spread in the population largely uninhibited by the host defence (Fig. 1c green). In the second phase, the shotgun phase, the spread of the TE is controlled by segregating piRNA producing loci (i.e. either a cluster insertions or paramutated loci). The onset of this phase is defined as the time when at least 99% of the individuals carry at least one piRNA producing locus (Fig. 1c yellow). In the final phase, the inactive phase, the TE is inactivated by a fixed piRNA producing locus, where fixation of any of the many segregating piRNA producing loci is sufficient to irrevocably inactivate the TE. With neutral TE insertions (as simulated initially) piRNA producing loci may get fixed due to genetic drift (Supplementary Fig. S8). The onset of the phase is the time when the first piRNA producing locus reaches fixation (Fig. 1c red).

Based on these definitions, we found that paramutations significantly reduce the length of the rapid invasion phase (Fig. 2a; Kruskal–Wallis rank sum test $\chi^2 = 274.0$, $p = 2.2 \cdot 10^{-16}$). Therefore, paramutations enable hosts to control TE invasions more rapidly. Paramutations only had a moderate effect on the length of the shotgun phase (Fig. 2a; Kruskal–Wallis rank sum test $\chi^2 = 16.7$, $p = 8.2 \cdot 10^{-4}$). Next, we investigated the influence of paramutations on the abundance of TEs. Given that the time required to establish host control differs among replicate populations, we measured the TE abundance not at a fixed time point but rather at the onset of the shotgun and inactive phase. This corresponds to the moment when a TE invasion is initially brought under host control, either through segregating or fixed piRNA-producing TE insertions. We found that the abundance of the TEs dramatically decreased with increasing numbers of



Fig. 2. Effect of paramutations on different key properties of TE invasions. Neutral TEs were simulated (x = 0). a) Length of the phase b) TE abundance per diploid individual at the start of the phase c) Cluster insertions per diploid individual at the start of the phase d) Fraction of individuals with at least one cluster insertion at the start of the phase e) Paramutable TEs per diploid individual at the start of the phase f) Fraction of individuals with at least one paramutated TE at the start of the phase. g) Effect of the fraction of paramutable loci (p) on the average number of TE insertions per diploid individual (at generation 5,000). h) Effect of paramutable loci on the number of cluster insertions (at generation 5,000).



Fig. 3. Paramutations reduce the fitness burden generated by TE invasions. a) Average fitness of individuals during TE invasions with different fractions of paramutable loci (top panel). The lowest fitness during the invasion (minimum fitness) is indicated. b) Paramutations reduce the fitness cost of TE invasions (fitness cost = 1 - minimum fitness). The significance of the fitness reduction was estimated with Wilcoxon rank sum tests (***p < 0.001). c) Paramutations reduce the fitness cost of TE invasions, for many different negative effects of TEs. Each dot represents the result of a simulated TE invasion with a randomly chosen negative effect of TEs (y-axis) and a fraction of paramutable loci (x-axis). Colors indicate the minimum fitness during the invasion. Simulations which lost all TEs are shown in gray. Negative selection is not effective when N · x < 1 (dashed line). Additional data for simulations without paramutable loci (classic trap model) are shown in a separate left panel. d) Paramutations reduce the fitness cost of TE invasions for different sizes of piRNA clusters. Each dot represents the results the result of a different size of piRNA cluster (y-axis) and the fraction of paramutable loci (x-axis) were randomly chosen. Extinct populations (minimum fitness <0.1) are shown in white.

paramutable loci (Fig. 2b; shotgun phase, Kruskal–Wallis rank sum test $\chi^2 = 350.1$, $p = 2.2 \cdot 10^{-16}$, inactive phase, Kruskal–Wallis rank sum test $\chi^2 = 328.5$, $p = 2.2 \cdot 10^{-16}$). The accelerated silencing of TEs in simulations with paramutations is likely responsible for this decreased number of TEs accumulating during the invasions.

The average number of piRNA cluster insertions per diploid individual also decreased with the fraction of paramutable loci (Fig. 2c; shotgun phase Kruskal–Wallis rank sum test $\chi^2 = 368.5$, $p = 2.2 \cdot 10^{-16}$, Kruskal–Wallis rank sum test $\chi^2 = 301.6$, $p = 2.2 \cdot 10^{-16}$). Accordingly, the fraction of individuals with at least one cluster insertion also decreased with the fraction of paramutable loci (Fig. 2d; shotgun phase Kruskal–Wallis rank sum test $\chi^2 = 365.6$, $p = 2.2 \cdot 10^{-16}$, inactive phase Kruskal–Wallis rank sum test $\chi^2 = 281.0$, $p = 2.2 \cdot 10^{-16}$). On the other hand, the number of paramutable TEs and the fraction of individuals with a paramutated TE increased with the abundance of paramutable loci (Fig. 2e,f; number of paramutated TEs: shotgun phase, Kruskal–Wallis rank sum test $\chi^2 = 380.0$, $p = 2.2 \cdot 10^{-16}$, inactive phase, Kruskal–Wallis rank sum test $\chi^2 = 280.97$, $p = 2.2 \cdot 10^{-16}$, fraction of individuals with paramutated TE: shotgun phase Kruskal–Wallis rank sum test $\chi^2 = 373.9$, $p = 2.2 \cdot 10^{-16}$, inactive phase Kruskal–Wallis rank sum test $\chi^2 = 311.9$, $p = 2.2 \cdot 10^{-16}$). Our data show that paramutated TEs may, to some extent, compensate for cluster insertions, even though cluster insertions are necessary to initiate piRNA production. To more carefully investigate the effect of paramutations on the abundance of TEs, we performed simulations where we varied the fraction of paramutable loci in steps of 5% (Fig. 2g,h). Both the abundance of TEs and cluster insertions declined rapidly with increasing numbers of paramutable loci until about 10–20% of the loci were paramutable loci only had a minor effect on the abundance of TEs, suggesting that around 10–20% paramutable loci are sufficient to markedly reduce the number of TEs accumulating during TE invasion.

Next, we explored the interaction between the size of piRNA clusters (in % of the genome) and the fraction of paramutable loci (Supplementary Fig. S9). In the absence of paramutable loci (i.e. trap model), increasing the size of piRNA clusters from 1% to 50% had little effect on the abundance of cluster insertions. In agreement with our previous work, about four cluster

insertions were necessary to control TE invasion under the trap model, regardless of the cluster size (Supplementary Fig. S9a; Kofler 2019). Notably, four cluster insertions are necessary to stop a TE invasion in a population even when we assume that a single insertion is sufficient to stop a TE in any individual. This can be explained by the fact that most TE insertions in piRNA clusters will be segregating in the population, and that recombination among these segregating cluster insertions will lead to a heterogeneous distribution of cluster insertions in the next generation, where some individuals will carry many cluster insertions and some solely a few or even none (Kofler 2019). The TE will be active in the individuals without cluster insertions and thus the average number of cluster insertions in the population will increase. Only when individuals carry around four cluster insertions, do most individuals in the population end up with at least a single insertion. However, when 10% paramutable loci were simulated, the size of piRNA clusters had a substantial effect: the number of cluster insertions required to silence an invasion increased from 0.3 with a cluster size of 1% to 4 when clusters account for 50% of the genome. In contrast, the number of paramutated TE insertions decreased from about 4 (cluster size 1%) to about 0.34 (cluster size 50%; Supplementary Fig. S9b). Therefore, paramutations are most important when the size of piRNA clusters is small. This further supports our finding that paramutated TEs may, to some extent, compensate for cluster insertions. Previous works established that the size of piRNA cluster is a major factor determining the abundance of TEs accumulating during an invasion (Kelleher et al. 2018; Kofler 2019; Tomar et al. 2022). As paramutations increase the proportion of the genome that produces piRNAs, it is perhaps not surprising that paramutable loci reduce the abundance of TEs accumulating during invasions. It is however unclear to which extent paramutable loci can substitute for piRNA clusters. We were thus interested in the amount of TEs accumulating during invasions with shifting proportions of piRNA clusters and paramutable loci, such that sum of these two regions was kept constant (%-cluster + %paramutable loci = constant). We performed additional simulations with varying proportions of piRNA cluster and paramutable loci where the total proportion of piRNA producing loci remained constant at 3% (e.g. 3%-cluster/0%-para, 2%-cluster/1%-para, 1%-cluster/2%-para, etc). The fewest TEs accumulated when all piRNA producing loci were piRNA cluster (i.e. 3%-cluster/0%-para; Supplementary Fig. S10). This is expected as solely piRNA clusters may trigger as well as maintain piRNA production whereas insertions in paramutable loci may solely maintain piRNA production. Interestingly the amount of TEs accumulating during invasions increased only slowly with the paramutable-loci until about 0.5%-cluster/ 2.5%-para. However with even lower proportions of piRNA cluster a steep increase in the amount of TEs accumulating during invasions was noticed (e.g. 0.05%-cluster/2.95%-para).

In summary, we showed that paramutations speed up the silencing of TE invasions and thus reduce the amount of TEs accumulating during TE invasions. Additionally, paramutated TEs may compensate for cluster insertions. When paramutations are considered, the expected number of TE insertions in piRNA clusters is lower than under the classical trap model (without paramutations).

Paramutations reduce the fitness burden of TEs

Motivated by our finding that paramutations reduce the number of TEs accumulating during an invasion, we next asked whether paramutations may also reduce the fitness burden generated by TE invasions. To explore this question, we introduced deleterious effects of TE insertions into the simulations (x > 0). Since negative effects of TEs have a substantial impact on the dynamics of TE invasions (under some parameters a host defence, including paramutations, is not even necessary to control TE invasions Charlesworth and Charlesworth 1983: Kofler 2019) it is not obvious whether paramutations will also reduce the number of accumulated TEs in simulations with deleterious TEs. To avoid an unlikely equilibrium state (TSC-balance (Kofler 2019)) where an invading TE is never silenced in all individuals in a population, we assumed that TE insertions in piRNA cluster are neutral (i.e. $x_{cluster} = 0$). We simulated a linear fitness cost of TE insertions $w = 1 - x \cdot n$ where w is the fitness of an individual, *n* the number of TE insertions per diploid individual (outside of piRNA clusters) and x the negative effect of a TE. The invasion dynamics of TEs with deleterious effects differ substantially from the invasion of neutral TEs (Supplementary Fig. S11). For neutral TEs, copy numbers rapidly increase in early generations. In later generations, piRNA producing loci, which slow down the invasion, start to emerge, such that the TE abundance gradually reaches a stable plateau (Supplementary Fig. S11; Kofler 2019). For deleterious TEs, copy numbers also rapidly increase at early generations. At later generations, when the piRNA-based host defence emerges, the TE abundance declines as deleterious TEs are purged from populations by negative selection (Supplementary Fig. S11; Kelleher et al. 2018; Kofler 2019, 2020). This dynamic is reflected in the average fitness during the invasions of deleterious TEs. Initially, we observe a rapid fitness decline due to the accumulation of deleterious TE insertions (Fig. 3a). At later generations, when the host defence is established, the average fitness rapidly recovers as deleterious insertions are purged from the population by negative selection. Therefore, TE invasions likely reduce host fitness solely temporarily (Fig. 3a; Kelleher et al. 2018; Kofler 2020). Here, we refer to the lowest fitness of individuals during a TE invasion as the "minimum fitness" (Fig. 3a black arrow). The minimum fitness thus represents the maximal fitness burden generated by TE invasions. The minimum fitness may be a crucial factor determining the persistence of host populations as it has been argued that populations could go extinct if this temporary fitness reduction is too severe (Kofler 2020). We found that paramutations have a marked impact on host fitness during TE invasions, where the fitness reduction during invasions is less severe in the presence of paramutable loci (Fig. 3a). Paramutations significantly reduce the fitness cost of TE invasions (fitness cost = 1 - minimum fitness; Fig. 3b). The fitness cost of TE invasions is also reduced by paramutations when assuming that fitness declines multiplicatively with TE copy numbers $(w = (1 - x)^n$; Supplementary Fig. S13).

So far, the simulations were performed with a deleterious TE effect of x = 0.03. However, the fitness effects of TEs remain an important open question (Arkhipova 2018) and the deleterious effect of TEs may vary among TE families, where for example long TE families may be more deleterious than short ones (Petrov et al. 2003). We thus aimed to systematically explore the impact of deleterious TE effects in combination with different fractions of paramutable loci on host fitness. To cover the parameter space, we performed 10,000 simulations. For each simulation, we randomly picked a negative effect of TEs (between x = 0.0001 and x = 0.5) and a fraction of paramutable loci (between para = 0% and para = 50%). We performed additional simulations without any paramutable loci to more clearly show the expectations for the classic trap model (para = 0%; Fig. 3c; left panel). After 5,000 generations, we recorded the minimum fitness for each simulation. In agreement with previous work, we found that all TE insertions quickly got purged from populations when TEs are highly deleterious (here

x > 0.1; Fig. 3c; Kofler 2019, 2020). We found that over most of the explored parameter space, TE invasions led to a substantial fitness burden and that paramutations considerably reduced this fitness burden. The beneficial effect of paramutations was most pronounced when selection against TEs was effective (i.e. above the drift barrier where $N \cdot x > 1$; Fig. 3c; dashed line). Interestingly, about 10% paramutable loci are sufficient to markedly reduce the fitness burden generated by TE invasions (Fig. 3c).

It is feasible that the beneficial effect of paramutations depends on the size of piRNA clusters. Therefore, we again used massive simulations to explore the effect of the size of piRNA clusters (0.1-4%) in combination with different fractions of paramutable loci (0-50%) on host fitness (Fig. 3d). We performed additional simulations (1,000) to more clearly show the expectations under the classic trap model (0% paramutable loci; Fig. 3d). We found that paramutations significantly reduced the fitness cost of TEs for all explored piRNA cluster sizes (Fig. 3d; Supplementary Fig. S12). Again 10% of paramutable loci are sufficient to dramatically reduce the fitness burden of TEs.

In summary, we found that paramutations substantially reduce the fitness burden caused by TE invasions, where about 10% paramutable loci are sufficient to largely achieve the beneficial effect of paramutations.

Epigenetically inherited piRNA production may be positively selected

Mediated by maternally deposited piRNAs, paramutations drive the conversion of TE insertions into piRNA producing loci. Once the first piRNAs against an invading TE have emerged, increasing numbers of TE insertions may be converted into paramutated TEs during TE invasions. We speculated that the piRNA-based silencing of TEs could spread in populations by a chain-reaction where paramutated TEs produce piRNAs that drive further paramutations in the next generation. Hence, we wondered whether paramutation dependent piRNA production (PDPP) could spread in populations in a non-Mendelian manner, similar to a gene drive. We aimed to construct a simple scenario that allows testing this hypothesis. We simulated a genome with five chromosomes (10 Mbp) and a fixed TE insertion at a single paramutable locus (at the beginning of the first chromosome; Fig. 4a). Consequently, each individual at any generation will carry a paramutable TE. The availability of paramutable TEs will, therefore, not be a limiting factor during the simulations. We used a transposition rate of u = 0.0, and a recombination rate of 4cM/Mbp but did not simulate piRNA clusters, Finally, irrespective of the sex, we initiated half of the individuals in the base population with maternally deposited piRNAs and the other half with no piRNAs (Fig. 4a). If our hypothesis is correct and PDPP is spreading in a population similar to a gene drive, PDPP should become fixed in the majority of the simulated populations (if PDPP becomes fixed, all individuals in a population will produce piRNAs). On the other hand, if PDPP is solely subject to genetic drift, the fraction of populations where PDPP becomes fixed or lost should be similar. Our simulations show that the fraction of populations where PDPP was lost/fixed is roughly similar, demonstrating that PDPP is not per se spreading in populations (Fig. 4b). This can be explained by the fact that only half of the individuals can propagate PDPP (females). Although males can inherit maternally deposited piRNAs, they cannot transmit them to the next generation. PDPP may only spread in populations like a gene drive if male gametes would also transmit piRNAs that could trigger paramutations in the next generation.

Next, we wondered whether PDPP, an epigenetically inherited trait, could be positively selected. Individuals with an active

host defence (e.g. PDPP) may accumulate fewer deleterious TE insertions than individuals without a host defence. These fitness differences may result in positive selection of PDPP. To test this idea, we simulated the same scenario as described before (fixed TE at paramutable locus; half of the individuals have piRNAs), with the sole difference being that we introduced deleterious effects of TEs (x > 0) and nonzero transposition rates (u > 0). In these scenarios, PDPP becomes fixed in most simulations, suggesting that PDPP is positively selected (Fig. 4b). Our finding that PDPP may be positively selected is robust to different recombination rates and fitness functions (linear + multiplicative; Supplementary Fig. S14).

In summary, we showed that an epigenetically inherited trait, paramutation dependent piRNA production (PDPP), may be positively selected as it reduces the fitness burden imposed by deleterious TE insertions. As PDPP is an epigenetically inherited trait, positive selection of PDPP may not reflect allele frequency changes at any genomic locus.

Paramutations may account for three important open problems with the trap model

Finally, we asked if paramutations could account for three important open problems with the trap model. First, several works showed that the number of TE insertions in piRNA clusters is lower than the four insertions per diploid individual expected under the trap model (Kofler 2019). Studies that monitored P-element invasions in experimental populations, using Illumina short reads, found that P-element copy numbers stabilized after about 20-80 generations due to the emergence of piRNAs, but the number of P-element insertions in piRNA clusters at this stable plateau was lower than expected under the trap model (Kofler et al. 2018, 2022). A study that monitored P-element copy numbers during invasions, with long-read data, also found that the number of cluster insertions were far lower than expected (Selvaraju et al. 2022). Furthermore, a recent study by Wierzbicki and Kofler (2023) investigated the composition of piRNA clusters in different D. melanogaster strains and found that many TE families have fewer than the four insertions in piRNA clusters expected under the simple trap model (Kofler 2019). Here, we demonstrated that the expected number of TE insertions in piRNA clusters decreases with increasing numbers of paramutable loci (Fig. 2; Supplementary Fig. S9). Therefore, paramutated TEs could compensate for cluster insertions and paramutations may thus account for the insufficient number of cluster insertions observed in previous works (Kofler et al. 2018, 2022; Selvaraju et al. 2022; Wierzbicki and Kofler 2023).

Second, a recent work found that deletion of three major piRNA clusters in D. melanogaster had no effect on the activity of the TEs (Gebert et al. 2021). The authors suggest that dispersed piRNA producing loci outside of piRNA clusters (i.e. likely paramutated TEs) may be responsible for producing the piRNAs that silence the TEs (Gebert et al. 2021). One vital question that remains unanswered in this work is how such dispersed piRNA producing loci arise in the first place. The generation of dispersed piRNA producing loci may require maternally deposited piRNAs and it is not yet clear how the first piRNAs complementary to an invading TE may emerge. In a commentary to this paper, Chen and Aravin (2021) proposed that piRNA clusters could be crucial to trigger the initial host response but may become dispensable later on. Here, we explored this idea with a quantitative model. In particular, we test the hypothesis that paramutations may render piRNA clusters dispensable after a piRNA-based host defence emerged. Insertions in piRNA clusters may trigger the production of the



Fig. 4. An epigenetically regulated trait—i.e. paramutation dependent piRNA production (PDPP)—may be positively selected. a) An overview of the simulated base population. We simulated a fixed TE insertion at a single paramutable locus (circle). Half of the individuals were assigned maternally inherited piRNAs, while the other half were not assigned any. b) Fractions of populations where PDPP is fixed or lost after 5,000 generations. If PDPP is fixed, all individuals in a population will produce piRNAs. We simulated different negative effects of TEs (x) and different transposition rates (u). For neutral TEs (x = 0), we observed roughly equal ratios of populations where PDPP was fixed or lost, demonstrating that PDPP is not *per se* spreading in populations. However, when negative effects of TEs where introduced x > 0 the fraction of populations where PDPP was fixed increased, suggesting that the epigenetically inherited trait, PDPP, is positively selected. The significance was estimated with Chi-squared tests (N.S. not significant, * p < 0.05, *** p < 0.001).

very first piRNAs, that may drive the conversion of TE insertions into paramutated TEs. Once such paramutated TEs have emerged, piRNA clusters may be removed with little impact on the activity of the TEs.

To test this hypothesis, we generated a novel branch of InvadeGO ("remove-cluster") that allows deleting a given number of piRNA clusters at a specified time (Fig. 5a). We again simulated populations of diploid organisms with 5 chromosomes and a piRNA cluster at the end of each chromosome (Fig. 5a). After 3,000 generations, when the host defence has been largely established in all replicates, we removed between 0 and 5 out of 5 clusters from each individual in the population. We performed the simulations in a scenario with 0% (classic trap model) and 10% paramutable loci. Without paramutable loci, TE copy numbers dramatically increased in many replicates after the clusters were removed (Fig. 5b; Wilcoxon rank sum test with average TE abundance after 5,000 generations, clu -4 vs clu -0 W = 1388, $p < 2.2 \cdot 10^{-16}$). However, when 10% paramutable loci were introduced, no significant increase in TE copy numbers was observed even in the scenario where 4 out of the 5 clusters were removed (Fig. 5b; Wilcoxon rank sum test with average TE abundance after 5,000 generations, clu - 4 vs clu - 0 W = 4739, p = 0.52).

As expected, in simulations without paramutable loci, the TE copy numbers increased exponentially in 100% of the replicates when all five piRNA clusters were removed (as the entire host defence was removed after 3,000 generations; Fig. 5b). By contrast, in simulations with paramutable loci, TE copy numbers increased exponentially only in 16% of the replicates when all five clusters were removed (paramutated TEs compensate for the loss of cluster insertions; Fig. 5b). Our simulations thus show that piRNA

clusters could be important to trigger the host defence, but paramutated TEs may render piRNA clusters dispensable once the piRNA-based host defence emerged.

Finally, we tested whether paramutations can account for the discrepancy in the observed and the expected abundance of the different TE families in D. melanogaster (Kofler 2019). Under the trap model, the size of piRNA clusters is a major factor determining the number of TE insertions accumulating during a TE invasion (Kofler 2019). Apart from negative selection against TEs, all other investigated factors, such as transposition rate, genome size and recombination rate, only have a minor influence on the abundance of TEs (Kofler 2019). If the size of piRNA clusters is known, quantitative predictions can be made about the expected abundance of TEs. In D. melanogaster where clusters account for about 3% of the genome, each TE family should have between 67.5–165 insertions per haploid genome (Fig. 6a). However, the actually observed abundance of most TE families is, with around 10 to 40 insertions per haploid genome, far lower than expected. It has been speculated that negative selection against TEs may account for this discrepancy (Kofler 2019). Here, we explore an alternative hypothesis. We suggest that paramutations could also account for this discrepancy. To test this, we simulated TE invasions in populations of diploid organisms with piRNA clusters accounting for 3% of the genome, similar to dual-strand clusters in the germline of D. melanogaster (Brennecke et al. 2007). Additionally, we either simulated 0% (classic trap model) or 10% paramutable loci. Finally, we compared the expected TE abundance (simulations) to the observed abundance of the different TE families in D. melanogaster (considering solely germline TEs; data from Kofler et al. (2015b)). As described previously, for the



Fig. 5. Paramutations render piRNA clusters largely dispensable once a piRNA-based host defence has been established. a) An overview of the simulated scenario. After 3,000 generations, some piRNA clusters were removed from every individual in the population. b) Paramutations stabilize TE copy numbers when piRNA clusters are removed. Results are shown for different numbers of deleted clusters (top panel) and for simulations with 0% and 10% paramutable loci (right panel). Dashed lines indicate the time when the piRNA clusters were removed. Note that in the scenario with paramutations, the TE copy numbers remain largely stable after the clusters are removed (except for some simulations where all five clusters are removed). With paramutations, piRNA clusters are thus important to trigger the host defence but later become dispensable.

classical trap model, the observed TE abundance is much lower than expected (0% paramutable loci; expected: 67.5–165 insertions per haploid genome; 4 out of 98 TE families fit in this range; Fig. 6a; Kofler 2019). However, if 10% paramutable loci are used, the expected TE abundance fits much better with the observed abundance of the TE families (expected: 18.1–42.9 insertions per haploid genome; 29 out of 98 fit in this range; Fig. 6b). This finding is robust to different assumptions about the transposition rate and the recombination rate (Supplementary Fig. S15).

Therefore, a model with 10% paramutable loci and with piRNA clusters accounting for 3% of the genome, roughly captures the observed abundance of the different TE families in *D. melanogaster* (although some families with unusually high copy numbers can be observed; see discussion).

In summary, we have shown that paramutations may account for three important open problems with the trap model. First, paramutations may compensate for cluster insertions, which could explain why the observed number of TE insertions in piRNA clusters is lower than expected under the classic trap model. Second, we showed that paramutations may render piRNA cluster dispensable once the host defence has been established, which could explain why deletion of three major clusters did not lead to an activation of TEs (Gebert et al. 2021). Third, paramutations will reduce the amount of TE insertions accumulating during invasions, which could explain why the abundance of many TE families in *D. melanogaster* is lower than expected under the classic trap model. A model with piRNA clusters (3% of the genome) and paramutations (10% of loci) largely predicts the abundance of TE families in *D. melanogaster*, even in the absence of negative selection against TEs.

Impact of paramutations if siRNAs trigger the host defence

In agreement with the prevailing view, the trap model, we have assumed that a TE insertion into a piRNA cluster triggers the emergence of the first piRNAs required to initiate paramutations (Bergman et al. 2006; Malone et al. 2009; Zanni et al. 2013; Goriaux et al. 2014; Yamanaka et al. 2014; Ozata et al. 2018). However, it was recently suggested that siRNAs, generated from Dicer-2 mediated cleavage of dsRNA, may also initiate the emergence of piRNA producing loci (Luo et al. 2022). Such dsRNA may be formed from sense and antisense transcripts of TEs (van Rij and Berezikov 2009; Luo et al. 2022). It is not obvious how to best capture this scenario with a simple model. Here, we opted to build the model based on the rate-limiting factor for dsRNA formation, the production of antisense RNA. Active TE families usually



Fig. 6. The trap model with paramutations predicts the observed abundance of TE families in *D. melanogaster* more accurately than a model without paramutations. a) Abundance of TEs during the simulated invasions with (10%) and without (0%) paramutable loci. In both scenarios, piRNA clusters account for 3% of the genome. b) Observed and expected abundance of TE families in *D. melanogaster*. Bar plots show the abundance of each TE family, where the colors of bars indicate the average population frequency (blue =0.1, red =1.0). The gray-shaded region highlights the expected abundance of the TE families based on the simulations. Note that the observed abundance of TEs is best captured by a model with paramutations. Data are from Kofler et al. (2015b).

express ample sense RNAs such that sufficient amounts of the enzymes and other components necessary for transposition are generated. Antisense transcripts are usually more rare and are, therefore, most likely to be the bottleneck to any dsRNA formation. For example, we found that during an invasion of the P-element (a widely studied TE family) in experimental populations, about 7% of the P-element transcripts are antisense (Selvaraju et al. 2022). As proposed by Luo et al. (2022), such antisense transcripts could result when a TE is inserted into a transcribed region (e.g. UTR or intron) in the opposite orientation to the gene. We model this notion by assuming that only TE insertions in some genomic regions can lead to the formation of antisense transcripts. We thus distributed siRNA-trigger-sites in the genome. We assumed that a TE insertion in a siRNA-trigger-site will lead to the production of antisense mRNA, that leads to the generation of siRNAs, which in turn drives the conversion of a TE insertion at a paramutable locus into a piRNA producing locus. Since TE insertions into many genes could likely trigger the production of antisense mRNA, we assumed that trigger-sites are more evenly distributed in the genome than piRNA clusters. In contrast to insertions in piRNA clusters, we assumed that insertions in siRNA-trigger-sites cannot themselves produce piRNAs (nor be converted into piRNA producing loci). This is in agreement with Luo et al. (2022), who showed that a siRNA-producing hairpin, which activates piRNA production in trans, is not by itself producing piRNAs. The first piRNAs complementary to an invading TE will therefore emerge in

individuals having (i) at least one insertion in a siRNA-trigger-site and (ii) at least one insertion in a paramutable locus.

Under our siRNA-model, between 90 and 100% of the invasions were stopped after 5,000 generations, and all invasions were stopped after 10,000 generations (Supplementary Table S2). The dynamics of TE invasions under this model are similar to the trap model. After an initial increase, TE copy numbers reach a stable plateau in all replicates (Supplementary Fig. S16). Again similarly to the trap model, three phases of TE invasions can be distinguished. First, TE copy numbers rapidly increase in the population (rapid invasion phase), next the invasion is silenced by segregating paramutated TEs (shotgun phase), and finally, the invasion is silenced by a fixed paramutated TE (Supplementary Fig. S16). Interestingly, once the proportion of trigger sites exceeds a certain minimum ($\geq 1\%$), the amount of trigger sites has little impact on the abundance of TEs accumulating during an invasion (Supplementary Fig. S16). Paramutations also reduce the fitness burden generated by TE invasions under this model (Supplementary Fig. S17). Finally, we found that the observed abundance of different TE families in D. melanogaster can also be predicted under the siRNA model (assuming 10% paramutable loci and 3-30% trigger sites; Supplementary Fig. S18).

To summarize, we found that the siRNA model, which assumes that insertions at some sites may trigger paramutations, is a viable alternative to the trap model. All TE invasions were stopped even in the absence of piRNA clusters and negative selection against TEs. piRNA clusters are, therefore, dispensable. Finally, paramutations again reduce the fitness burden of TEs, and the abundance of TE families in *D. melanogaster* can also be roughly predicted by a siRNA model.

Discussion

Based on computer simulations with our novel tool InvadeGO, we show that paramutations are an important factor governing the dynamics of TE invasions. Paramutations limit the number of TE insertions accumulating during invasions, reduce the fitness burden of TEs, render piRNA clusters dispensable, and may account for the discrepancy in the observed and the expected TE abundance in *D. melanogaster*. Moreover, paramutation dependent piRNA production (PDPP), an epigenetically inherited trait, may be positively selected.

Our conclusions are based on several assumptions: (i) transposons are spreading in populations at a given rate (u), (ii) insertion sites of TEs are randomly distributed in the genome, (iii) piRNAs silence TEs, (iv) maternally deposited piRNAs induce paramutations, (v) some, but not all, TE insertions outside of piRNA clusters may be paramutated, and (vi) an insertion into a piRNA cluster (or a siRNA-trigger-locus) may initiate the production of the first piRNAs complementary to an invading TE. All the aforementioned assumptions are well supported by the literature. First, the multiplication of TEs in genomes and populations has been reported by many studies (Orgel and Crick 1980; Kidwell 1983; Anxolabéhère et al. 1985; Wicker et al. 2007; Biémont 2010; Kofler et al. 2018). Second, although some TEs, such as the P-element, have an insertion bias (Kofler et al. 2015a; Sultana et al. 2017), little insertion bias could be found for most of the investigated TE families in D. melanogaster (Jakšić et al. 2017). Third, there is little doubt that piRNAs repress the activity of TEs (Brennecke et al. 2007; Gunawardane et al. 2007; Sienski et al. 2012; Le Thomas et al. 2013; Ozata et al. 2018). We assumed that the piRNA-based host defence complete silences a TE ($u_{piRNA} = 0$) although it cannot be ruled out that some TEs continue to have some residual activity even when silenced by the piRNA-pathway. It is however unlikely that a residual activity will have any impact on our conclusions about the effect of paramutations on TE invasions as such a residual activity will mostly have an effect only after the host defence was established in all individuals of a population.

Fourth, our assumption that maternally deposited piRNAs may induce paramutations is also well supported by multiple works (de Vanssay et al. 2012; Le Thomas et al. 2014; Hermant et al. 2015; Casier et al. 2019). Fifth, several previous studies have shown that some, but not all, of the TE insertions outside of piRNA clusters may be converted into piRNA producing loci (Shpiz et al. 2014; Wierzbicki and Kofler 2023). It is an important open question which genomic, epigenetic, and cellular factors determine whether or not a TE insertions may be paramutated. The TE family is likely one important factor since the fraction of paramutated TEs varies among the families (Wierzbicki and Kofler 2023). The genomic context is likely also important. For example, Shpiz et al. (2014) suggested that paramutations may happen next to transcribed genes, and de Vanssay et al. (2012) suggested that a tandem arrangement of several TE insertions may be necessary for paramutation. It is conceivable that the composition of maternally deposited piRNAs and even the environment are also important factors influencing whether a locus may be paramutated (e.g. temperature Casier et al. 2019). A related open questions is whether paramutated TEs repress the activity of a TE to the same extent as insertions in piRNA clusters. It is

also feasible that not all paramutated TEs have the same effect on TE activity. Possibly the most important question is which events trigger the emergence of the very first piRNAs complementary to an invading TE. Here, we assumed that either an insertion in a piRNA cluster or siRNA-induced paramutations may initiate piRNA production. Several prior studies have shown that insertion of an artificial sequence into a piRNA cluster will lead to the generation of piRNAs complementary to this artificial sequence (Kawaoka et al. 2012; Muerdter et al. 2012; Luo et al. 2022). For example, Luo et al. (2022) found that insertion of a GFP-reporter in a major piRNA cluster of D. melanogaster, 42AB, leads to the generation of piRNAs complementary to GFP. It is, therefore, reasonable to assume that an insertion into piRNA cluster can trigger the de novo formation of piRNAs. It was recently suggested that siRNAs could also initiate the formation of novel piRNA-producing loci and thus trigger the production of piRNAs (Luo et al. 2022). Such siRNAs may, for example, result from cleavage of dsRNA consisting of sense and antisense transcripts of TEs. We modeled this mechanism by assuming that the expression of antisense transcripts will be the bottleneck in the formation of dsRNA and that only TE insertions at specific genomic sites will yield antisense transcripts (i.e. siRNA-trigger-sites). For example, antisense transcripts of TEs may result from TE insertions in transcribed regions in the opposite direction to the gene (van Rij and Berezikov 2009; Luo et al. 2022). However, there are still many unknowns about this mechanism: Are insertions in transcribed regions the main source of antisense transcripts of TEs? Which cellular processes are involved in the siRNA-mediated initiation of piRNA production? Thus, it is not clear how well our siRNA model captures the process of siRNA-mediated initiation of piRNA production. In any case, our simulations show that the siRNA-model as implemented in this work is a viable alternative to the trap model, as TE copy numbers stabilize after the initial rapid invasion in our siRNA-model. We show that PDPP, an epigenetically inherited trait, may be

positively selected. Individuals that produce piRNAs will accumulate fewer deleterious TE insertions than individuals without piRNAs, and therefore PDPP may be beneficial. This raises the question as to whether PDPP leaves footprints in populations that could be detected. As PDPP is an epigenetically inherited trait, it will not necessarily leave footprints in the genome (Oleksyk et al. 2010). This is illustrated by our simulations where selection on PDPP was observed, although the only TE insertion that could produce piRNAs was fixed in the population (Fig. 4). It is, however, conceivable that some genomic loci, like the cluster insertion that triggered PDPP, could hitchhike with PDPP. Such an indirect increase in the frequency of some alleles may be detectable at the genomic level. Directly determining whether PDPP is positively selected is currently challenging because it requires estimating whether piRNA production is spreading more rapidly in populations than would be expected under a null model (neutral expectations). Apart from the technical difficulty of measuring piRNA production in many individuals during TE invasions, inferring the expectations under the null model is also challenging, especially given that it is not yet entirely clear which events trigger the emergence of the piRNAs complementary to an invading TE.

We demonstrated that a model with paramutations captures the abundance of most TE families in *D. melanogaster* more accurately than a model without paramutations (Fig. 6). Paramutations could thus explain why the abundance of many TE families in *D. melanogaster* is lower than expected under the classic trap model. An alternative, not mutually exclusive hypothesis is that negative selection reduces the amount of TEs accumulating during an invasion, and thus is responsible for the discrepancy between the observed and expected TE abundance. It is likely that negative selection is an important force that could shape the abundance, distribution and population frequency of TEs (Charlesworth and Langley 1989; Nuzhdin 1999; Petrov et al. 2003; Langmüller et al. 2023). However, we find that some TE families, like roo and jockey, have much higher copy numbers than expected under the model with paramutations. It is unclear why some TE families accumulate such extremely high copy numbers. One possible solution comes from recent works which monitored P-element (a DNA transposon) invasions in experimental populations. Usually, the P-element is rapidly and consistently silenced in all experimental populations by the emergence of a piRNA-based host defence (Kofler et al. 2022; Selvaraju et al. 2022). However, Selvaraju et al. (2022) found that in one replicate population, the host-defence could not be established, even after many generations, which also led to the accumulation of unusually high P-element copy numbers. Since this only happened in a single replicate, the failure to establish the host defence is likely a stochastic event. A delay or a failure in the establishment of host control over an invading TE could therefore be a possible explanation for the highly abundant TE families in D. melanogaster. An alternative explanation for the families with unusually high copy numbers could be an insertion bias of the TEs (Sultana et al. 2017). For example, an insertion bias into promoter regions may result in a low probability of jumping into piRNA clusters. TEs with such a bias could then accumulate large numbers of insertions before a sufficient number of cluster insertions appeared.

Paramutations limit the number of TE insertions accumulating during an invasion and thus reduce the fitness burden of TEs. Paramutations may therefore be beneficial to organisms. In agreement with this, we found that PDPP may be positively selected. This raises the question as to whether paramutations evolved as a distinct mechanism to robustly control TEs. An alternative explanation could be that paramutations are just a byproduct of the mechanisms responsible for maintaining the location of piRNA clusters across generations. The position of piRNA clusters in the next generation is, at least partially, determined by maternally deposited piRNAs (Le Thomas et al. 2014). Mediated by maternally deposited piRNAs, silencing chromatin marks necessary for the formation of piRNA clusters are installed (Le Thomas et al. 2014). While these piRNAs will be important to maintain the location of piRNA clusters across generations, they may also, as a side-effect, induce paramutations at dispersed TE insertions outside of piRNA clusters. A related question concerns the influence of paramutations on the evolution of piRNA clusters. If paramutated TEs can compensate for insertions within piRNA clusters, it is possible that selective pressures preserving the composition of piRNA clusters, could be diminished. Consequently, piRNA clusters may undergo accelerated evolution, as indeed found by several studies (Gebert et al. 2021; Srivastav et al. 2023; Wierzbicki et al. 2023).

It has been shown that only some, e.g. around 5 to 10%, of the TE insertions, may be paramutated into piRNA-producing loci (de Vanssay et al. 2012; Mohn et al. 2014; Wierzbicki and Kofler 2023). Why are not more or fewer insertions paramutated? It is feasible that paramutations are costly to organisms. Paramutations of TEs may, for example, reduce the expression level of neighboring genes (Shpiz et al. 2014; Miller et al. 2023). Hence, paramutations may interfere with gene expression, and selection may thus act to reduce the frequency of paramutations. On the other hand, too few paramutations could also be deleterious to hosts. We showed that paramutations reduce

the burden of TEs accumulating during invasions. Since individuals with many paramutations will accumulate fewer TEs, selection may act to increase the frequency of paramutations. However, if paramutations are primarily responsible for maintaining the location of piRNA clusters across generations (see above), then selection may act to increase the frequency of paramutations indirectly. If the frequency of paramutations decreases, the stable inheritance of piRNA clusters across generations may be jeopardized, potentially leading to the reactivation of certain TEs. As this TE activity may be deleterious, selection may act to maintain stable inheritance of piRNA clusters and thus, indirectly, also maintain high frequencies of paramutated loci. Overall, it is conceivable that the frequency of paramutations is in a balance between two opposing evolutionary forces acting to decrease and increase the number of paramutated TEs. It is interesting to note that in D. melanogaster, the fraction of paramutated TEs is approximately 5–10%, which coincides with the range where our simulations suggest that the marginal benefit of paramutations is the highest (i.e. a higher frequency of paramutated TEs would only have a minor effect on the fitness burden caused by TEs; Fig. 3c,d; Wierzbicki and Kofler 2023).

In summary, we showed that paramutations have a substantial impact on the dynamics of TE invasions and may thus be a crucial component of the host defence against selfish DNA.

Data availability

InvadeGO is available under the GNU general public license, version 3 at GitHub (https://github.com/RobertKofler/invadego). The simulations and the data analysis were documented with RMarkdown (Xie et al. 2018) and have been made publicly available, together with the resulting figures, at GitHub https://github. com/Almo96/Paramutations'TEs.

Supplementary material are available at GENETICS online.

Acknowledgments

ChatGPT, a language model developed by OpenAI, and Google Translate contributed to polishing the english of some difficult sections. We thank Filip Wierzbicki and Matthew Beaumont for comments and all members of the Institute of Population Genetics for feedback and support.

Funding

This work was supported by an Austrian Science Fund (FWF) grant (P35093) to RK.

Conflicts of interest

The author(s) declare no conflict of interest.

Author contributions

RK and AS conceived the work. RK designed and implemented InvadeGO. AS performed simulations and analyzed the data. RK and AS wrote the manuscript.

Literature cited

- Anxolabéhère D, Nouaud D, Périquet G, Tchen P. 1985. P-element distribution in Eurasian populations of *Drosophila melanogaster*: a genetic and molecular analysis. Proc Natl Acad Sci USA. 82(16): 5418–5422. doi:10.1073/pnas.82.16.5418
- Arkhipova IR. 2018. Neutral theory, transposable elements, an eukaryotic genome evolution. Mol Biol Evol. 35(6):1332–1337. doi: 10.1093/molbev/msy083
- Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J. 2022. The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. Elife. 11:e80067. doi:10.7554/eLife.80067
- Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol. 7(11):R112. doi:10.1186/gb-2006-7-11-r112
- Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics. 186(4):1085–1093. doi:10.1534/genetics.110.124180
- 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(6): 1089–1103. doi:10.1016/j.cell.2007.01.043
- Brink RA. 1958. Paramutation at the R locus in maize. 23Cold Spring Harbor Laboratory PressCold Spring Harbor symposia on quantitative biology, p. 379–391.
- Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. Mol Ecol. 22(6):1503–1517. doi:10.1111/mec.12170
- Casier K, Delmarre V, Gueguen N, Hermant C, Viodé E, Vaury C, Ronsseray S, Brasset E, Teysset L, Boivin A. 2019. Environmentally-induced epigenetic conversion of a piRNA cluster. eLife. 8:e39842. doi:10.7554/eLife.39842
- Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet Res. 42(1):1–27. doi:10.1017/ S0016672300021455
- Charlesworth B, Langley CH. 1989. The population genetics of Drosophila transposable elements. Annu Rev Genet. 23:251–87. doi:10.1146/genet.1989.23.issue-1
- Chen P, Aravin AA. 2021. Transposon-taming piRNAs in the germline: where do they come from? Mol Cell. 81(19):3884–3885. doi: 10.1016/j.molcel.2021.09.017
- 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(7418):112–115. doi:10.1038/nature11416
- Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature. 284(5757):601–603. doi:10. 1038/284601a0
- Gebert D, Neubert LK, Lloyd C, Gui J, Lehmann R, Teixeira FK. 2021. Large Drosophila germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. Mol Cell. 81: 3965–3978.doi:10.1016/j.molcel.2021.07.011
- González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol. 6(10):e251. doi:10.1371/ journal.pbio.0060251
- Goriaux C, Théron E, Brasset E, Vaury C. 2014. History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. Front Genet. 5:257. doi:10.3389/fgene. 2014.00257
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated

mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 315(5818):1587–1590. doi:10.1126/science. 1140494

- 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. 201(4):1381–1396. doi:10.1534/genetics.115.180307
- Hollick JB. 2017. Paramutation and related phenomena in diverse species. Nat Rev Genet. 18(1):5–23. doi:10.1038/nrg.2016.115
- Jakšić AM, Kofler R, Schlötterer C. 2017. Regulation of transposable elements: interplay between TE-encoded regulatory sequences and host-specific trans-acting factors in *Drosophila melanogaster*. Mol Ecol. 28(19):5149–5159. doi:10.1111/mec.14259
- Josse T, Teysset L, Todeschini A-L, Sidor CM, Anxolabéhère D, Ronsseray S. 2007. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLoS Genet. 3(9):1633–1643. doi:10.1371/journal.pgen. 0030158
- Kawaoka S, Mitsutake H, Kiuchi T, Kobayashi M, Yoshikawa M, Suzuki Y, Sugano S, Shimada T, Kobayashi J, Tomari Y, et al. 2012. A role for transcription from a piRNA cluster in de novo piRNA production. RNA. 18(2):265–273. doi:10.1261/rna.029777.111
- Kelleher ES, Azevedo RBR, Zheng Y. 2018. The evolution of small-RNA-mediated silencing of an invading transposable element. Genome Biol Evol. 10(11):3038–3057. doi:10.1093/gbe/evy218
- Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. Proc Natl Acad Sci USA. 80(6): 1655–1659. doi:10.1073/pnas.80.6.1655
- Kofler R. 2019. Dynamics of transposable element invasions with piRNA clusters. Mol Biol Evol. 36(7):1457–1472. doi:10.1093/ molbev/msz079
- Kofler R. 2020. piRNA clusters need a minimum size to control transposable element invasions. Genome Biol Evol. 12(5):736–749. doi: 10.1093/gbe/evaa064
- Kofler R, Hill T, Nolte V, Betancourt A, Schlötterer C. 2015a. The recent invasion of natural Drosophila simulans populations by the P-element. PNAS. 112(21):6659–6663. doi:10.1073/pnas. 1500758112
- Kofler R, Nolte V, Schlötterer C. 2015b. Tempo and mode of transposable element activity in Drosophila. PLoS Genet. 11(7):e1005406. doi:10.1371/journal.pgen.1005406
- Kofler R, Nolte V, Schlötterer C. 2022. The transposition rate has little influence on the plateauing level of the P-element. Mol Biol Evol. 39(7):msac141.doi:10.1093/molbev/msac141
- Kofler R, Senti K-A, Nolte V, Tobler R, Schlötterer C. 2018. Molecular dissection of a natural transposable element invasion. Genome Res. 28(2):824–835. doi:10.1101/gr.228627.117
- Langmüller AM, Nolte V, Dolezal M, Schlötterer C. 2023. The genomic distribution of transposable elements is driven by spatially variable purifying selection. Nucleic Acids Res. 51(17):9203–9213. doi:10.1093/nar/gkad635
- Le Rouzic A, Capy P. 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics. 169(2): 1033–1043. doi:10.1534/genetics.104.031211
- Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 27(4):390–399. doi:10. 1101/gad.209841.112
- Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YCA, 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(15):1667–1680. doi:10.1101/ gad.245514.114

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/ bioinformatics/btp352
- Lu J, Clark AG. 2009. Population dynamics of PIWI-RNAs (piRNAs) and their targets in *Drosophila*. Genome Res. 20:212–227. doi:10. 1101/gr.095406.109
- Luo Y, He P, Kanrar N, Toth KF, Aravin AA. 2022. Maternally inherited siRNAs initiate piRNA cluster formation. bioRxiv.
- Mackay TF, Lyman RF, Jackson MS. 1991. Effects of P-element insertions on quantitative traits in Drosophila melanogaster. Genetics. 130:315–332. doi:10.1093/genetics/130.2.315
- 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(3):522–535. doi:10.1016/j.cell.2009.03.040
- Miller DE, Dorador AP, Van Vaerenberghe K, Li A, Grantham EK, Cerbin S, Cummings C, Barragan M, Egidy RR, Scott AR. et al. 2023. Off-target piRNA gene silencing in *Drosophila melanogaster* rescued by a transposable element insertion. PLoS Genet. 19(2): e1010598. doi:10.1371/journal.pgen.1010598
- Mohn F, Sienski G, Handler D, Brennecke J. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. Cell. 157(6): 1364–1379. doi:10.1016/j.cell.2014.04.031
- Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, Aravin AA. 2012. Production of artificial piRNAs in flies and mice. RNA. 18(1):42–52. doi:10.1261/ma.029769.111
- Nuzhdin SV. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica. 107(1-3):129–137. doi:10.1023/A:1003957323876
- Oleksyk TK, Smith MW, O'Brien SJ. 2010. Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B Biol Sci. 365(1537):185–205. doi:10.1098/rstb.2009.0219
- Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. Nature. 284(5757):604–607. doi:10.1038/284604a0
- Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2018. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 20(2):89–108. doi:10.1038/s41576-018-0073-3
- Pasyukova E, Nuzhdin S, Morozova T, Mackay T. 2004. Accumulation of TEs in the genome of *D. melanogaster* is associated with a decrease in fitness. J Hered. 95(4):284–290. doi:10.1093/jhered/esh050
- Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. Mol Biol Evol. 20(6):880–892. doi:10. 1093/molbev/msg102
- R Core Team 2012. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org. ISBN 3-900051-07-0.
- Sarkies P, Selkirk ME, Jones JT, Blok V, Boothby T, Goldstein B, Hanelt B, Ardila-Garcia A, Fast NM, Schiffer PM, et al. 2015. Ancient and

novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. PLoS Biol. 13(2): e1002061.doi: 10.1371/journal.pbio.1002061

- Selvaraju D, Wierzbicki F, Kofler R. 2022. P-element invasions in Drosophila erecta shed light on the establishment of host control over a transposable element. bioRxiv.
- Shpiz S, Ryazansky S, Olovnikov I, Abramov Y, Kalmykova A. 2014. Euchromatic transposon insertions trigger production of novel Pi-and endo-siRNAs at the target sites in the germline. PLoS Genet. 10(2):e1004138. doi:10.1371/journal.pgen.1004138
- Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 151(5):964–980. doi:10.1016/j.cell. 2012.10.040
- Srivastav S, Feschotte C, Clark AG. 2023. Rapid evolution of piRNA clusters in the drosophila melanogaster ovary. bioRxiv pages 2023–05.
- Sultana T, Zamborlini A, Cristofari G, Lesage P. 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet. 18(5):292–308. doi:10.1038/nrg.2017.7
- Tomar SS, Hua-Van A, Rouzic AL. 2022. A population genetics theory for piRNA-regulated transposable elements. bioRxiv.
- van Rij RP, Berezikov E. 2009. Small RNAs and the control of transposons and viruses in *Drosophila*. Trends Microbiol. 17(4):163–171. doi:10.1016/j.tim.2009.01.003
- Wang L, Zhang S, Hadjipanteli S, Saiz L, Nguyen L, Silva E, Kelleher E. 2023. P-element invasion fuels molecular adaptation in laboratory populations of Drosophila melanogaster. Evolution. 77(4): 980–994. doi: 10.1093/evolut/qpad017
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O. et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8(12):973–982. doi:10.1038/nrg2165
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. ISBN 978-3-319-24277-4.
- Wierzbicki F, Kofler R. 2023. The composition of piRNA clusters in Drosophila melanogaster deviates from expectations under the trap model. bioRxiv.
- Wierzbicki F, Kofler R, Signor S. 2023. Evolutionary dynamics of piRNA clusters in Drosophila. Mol Ecol. 32:1306–1322. doi:10. 1111/mec.v32.6
- Xie Y, Allaire JJ, Grolemund G. 2018. R Markdown: The Definitive Guide. Boca Raton (FL): Chapman and Hall/CRC.
- Yamanaka S, Siomi MC, Siomi H. 2014. piRNA clusters and open chromatin structure. Mob DNA. 5(1):22. doi:10.1186/1759-8753-5-22
- Zanni V, Eymery A, Coiffet M, Zytnicki M, Luyten I, Quesneville H, Vaury C, Jensen S. 2013. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci. 110(49):19842--19847. doi:10.1073/pnas.1313677110

Editor: D. Barbash