Purifying Selection Shapes the Dynamics of P-element Invasion in *Drosophila* Populations

Anna M. Langmüller[1,2,3,4,*], Benjamin C. Haller[3], Viola Nolte[2], Christian Schlötterer[2,*]

[1] Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

[2] Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

[3] Department of Computational Biology, Cornell University, 237 Tower Road, Ithaca NY 14850, USA

[4] Aarhus Institute of Advanced Studies, Aarhus University, Høegh-Guldbergs Gade 6B, 8000 Aarhus C, Denmark

[*] Corresponding authors

E-mail adresses:

Anna M. Langmüller: annamaria.langmueller@gmail.com

Benjamin C. Haller: bhaller@mac.com

Viola Nolte: Viola.Nolte@vetmeduni.ac.at

Christian Schlötterer: Christian.Schloetterer@vetmeduni.ac.at

Abstract

Background

Transposable elements (TEs) are DNA sequences that can move within a host genome. Many new TE insertions have deleterious effects on their host and are therefore removed by purifying selection. The genomic distribution of TEs thus reflects a balance between new insertions and purifying selection. However, the inference of purifying selection against deleterious TE insertions from the patterns observed in natural populations is challenged by the confounding effects of demographic events, such as population bottlenecks and migration.

Results

We used Experimental Evolution to study the role of purifying selection during the invasion of the P-element, a highly invasive TE, in replicated *Drosophila simulans* populations under controlled laboratory conditions. Because the change in P-element copy number over time provides information about the transposition rate and the effect of purifying selection, we repeatedly sequenced the experimental populations to study the P-element invasion dynamics. Based on these empirical data we used Gaussian Process surrogate models to infer parameter values characterizing the observed P-element invasion trajectory. We found that 73% of P-element copies are under purifying selection with a mean selection coefficient of -0.056, highlighting the central role of selection in shaping P-element invasion dynamics.

Conclusion

This study underscores the power of Experimental Evolution as a tool for studying transposable element invasions, and highlights the pivotal role of purifying selection in regulating P-element dynamics.

Keywords

Transposable Elements, P-element, *Drosophila*, Experimental Evolution, purifying selection, individual-based modeling, Gaussian Process

Background

Transposable Elements (TEs) are DNA sequences that can move and amplify within a host genome [1]. The evolutionary fate of TEs depends not only on their ability to replicate within the genome of their host, but also on the harm their activity might cause to their host [2]. Despite the challenges posed by their potentially disruptive behavior, TEs are ubiquitous in the tree of life [3,4], thriving in countless species, and have fascinated evolutionary biologists since their discovery more than seven decades ago [5].

It is widely accepted that some TEs insertions are harmful for their hosts [2,3,6], but the fraction of new TE insertions that are deleterious, and the distribution of their fitness effects, remain open questions [7,8]. Because measuring the effect of individual TE insertions on host fitness is tedious and a large number of independent insertions must be studied to obtain representative patterns, most evidence for purifying selection operating against TEs relies on the low abundance of TE insertions in specific regions of the host genome [9–11]. One limitation of this approach is the assumption of a uniform

insertion probability of TEs across the genome. However, TEs can exhibit strong insertion preferences for specific genomic regions [3,12,13], which makes it difficult to distinguish between insertion bias and purifying selection.

An alternative method for detecting genomic signatures of purifying selection is based on the frequency spectrum of TEs within natural populations [14,15]. As with SNPs, demographic events such as population bottlenecks or admixture events pose a major challenge to the interpretation of TE frequency spectra [6,16]. Furthermore, recent TE activity can cause a shift in the site frequency spectrum that resembles the pattern expected from purifying selection, again making it difficult to distinguish between insertion bias and purifying selection [14,15,17]. Finally, the low population frequencies of TE insertions — caused by recent TE activity [15] and purifying selection — present a significant obstacle because detecting a specific TE insertion and estimating its frequency reliably requires high sequencing coverage.

Experimental Evolution (EE) [18,19], a powerful approach to study evolution under controlled laboratory conditions, is becoming increasingly popular as an alternative method to study TE invasion dynamics [20–22]. By combining EE with whole-genome sequencing of pooled individuals, an approach called Evolve and Resequence [23–25], researchers can minimize confounding factors that complicate the interpretation of polymorphism patterns in natural populations. Moreover, EE enables the replication of the invasion process, allowing stochastic (random) effects to be distinguished from deterministic patterns across multiple evolutionary replicates [18,19,24].

The P-element [26,27] — one of the best-studied TEs in eukaryotes [28] — is a DNA transposon with a remarkable invasion record [29–31]. Likely introduced into *Drosophila simulans* from *D. melanogaster* in a single horizontal transfer event [32], the P-element spread through *D. simulans* populations worldwide in less than a decade [31]. *Drosophila* species curb P-element invasions through the piRNA pathway, a specialized defense mechanism involving small RNAs that target and silence TEs [33,34]. Most piRNAs originate from specific genomic regions known as piRNA clusters, which collectively comprise approximately 3.5% of the *Drosophila* genome [33]. Current studies suggest that the effective silencing of P-elements across the genome relies on the presence of at least one P-element insertion within a piRNA cluster, which serves as a template for the production of P-element-specific piRNAs [35,36].

In this study, we used EE to investigate how purifying selection shapes the invasion dynamics of the P-element in experimental *D. simulans* populations. We followed the increase in P-element copy number in replicated EE studies across time. Using these time-resolved empirical invasion curves together with an individual-based simulation framework matching our experimental setup, we show that only 27% of the P-element copies can be considered as effectively neutral, while the remaining 73% are subject to purifying selection with an average selection coefficient of -0.056. The genome-wide average selection coefficient for P-element copies outside piRNA clusters is -0.041, highlighting the importance of purifying selection in shaping P-element invasion dynamics.

Results

The P-element rapidly invades experimental populations

To explore the influence of purifying selection on TE invasion dynamics, we analyzed the P-element invasion in replicate *D. simulans* populations from two EE experiments conducted under identical environmental conditions (Figure 1). These experimental populations were established from about 200 isofemale lines derived from a natural *D. simulans* population (Tallahassee, Florida). During the collection of the isofemale lines this natural *D. simulans* population was at the onset of a P-element invasion [32], where 25–44% of these isofemale lines were estimated to be P-element carriers [31]. The 1st wave EE experiment was started shortly after the isofemale line collection, thus all evolutionary replicates had only a small number of P-element copies at generation 0. The 2nd wave experiment on the other hand was established about 4 years after the 1st wave experiment from the same isofemale lines. During this period those isofemale lines that already contained a P-element at the time of collection, acquired additional P-element copies [11]. As a consequence, the 2nd wave experiment started from a higher number of initial P-element copies. Because of the small effective population size during the maintenance of the isofemale lines, many of the P-element insertions that occurred in the isofemale lines between the setup of the 1st and 2nd wave experiment were deleterious. Purifying selection could only operate once the flies were maintained in large, outbred populations [11]. To track P-element invasion dynamics, we sequenced the replicate populations using the Pool-Seq approach [37] in both EE experiments and estimated P-element copy numbers per haploid genome using DeviaTE [38].

Our data confirmed that, the average initial P-element copy number in the 2nd wave experiment was significantly higher than in the 1st wave experiment (6.92 vs. 0.86, Figure

2). Despite distinct initial dynamics, both EE experiments reached a similar P-element copy number plateau: approximately 15 copies per haploid genome, after around 20 generations (Figure 2). This suggests that the timing and ultimate copy number plateau are consistent across both waves, indicating that these aspects of the P-element invasion are robust to initial conditions.

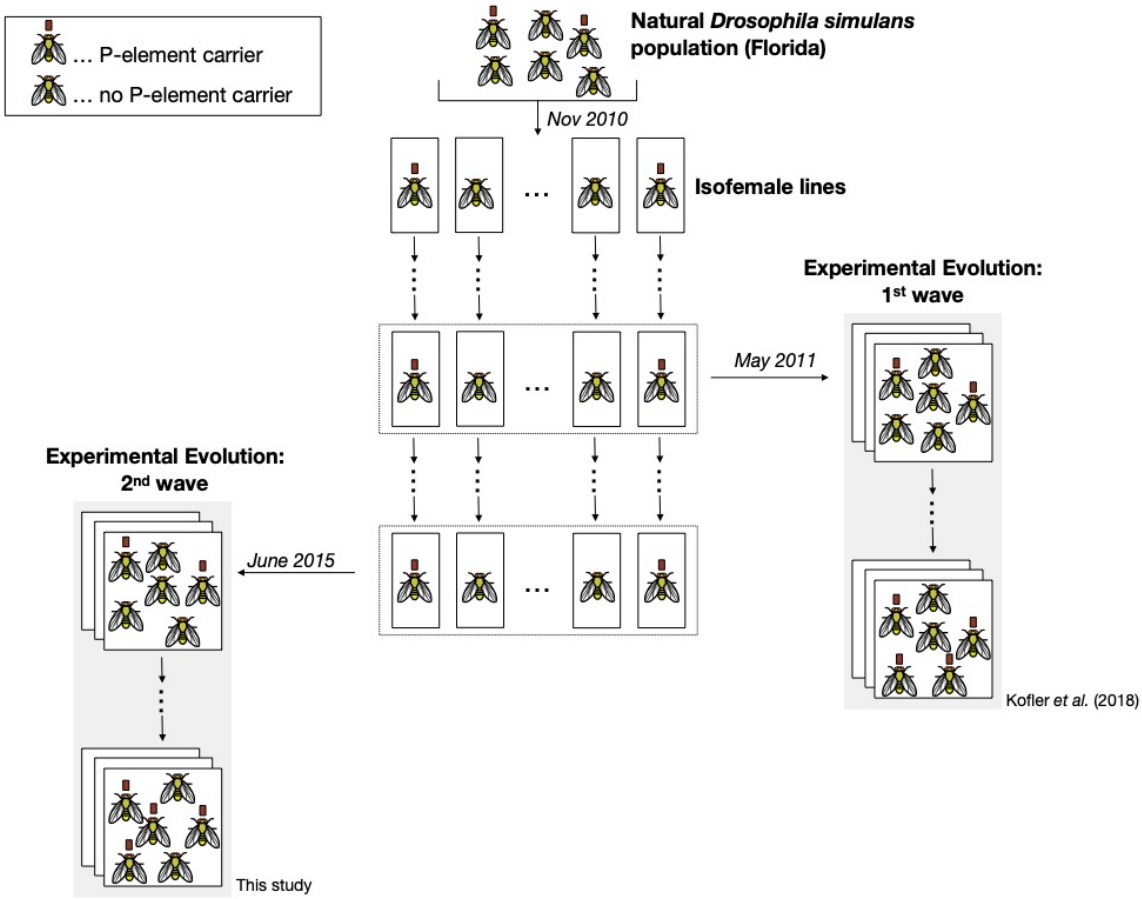


Figure 1. Schematic overview of the experimental design. Isofemale lines were established from a natural population of *Drosophila simulans*, where the P-element has been invading [32]. Isofemale lines carrying an active P-element (indicated by red squares) are expected to accumulate P-element copies until an active defense mechanism is triggered [33,39]. In May 2011, three large replicate populations were established, using 202 isofemale lines. These populations, referred to as the "1st wave" Experimental Evolution (EE) experiment, are maintained with non-overlapping generations under a cycling hot temperature regime and were sequenced every 10th generation to monitor the P-element invasion dynamics [20]. In June 2015, a second EE experiment was initiated using three new replicate populations derived from 191 surviving isofemale lines [11]. These populations, referred to as the "2nd wave" EE experiment, were maintained under the same environmental conditions and were also sequenced over time to track the P-element invasion.

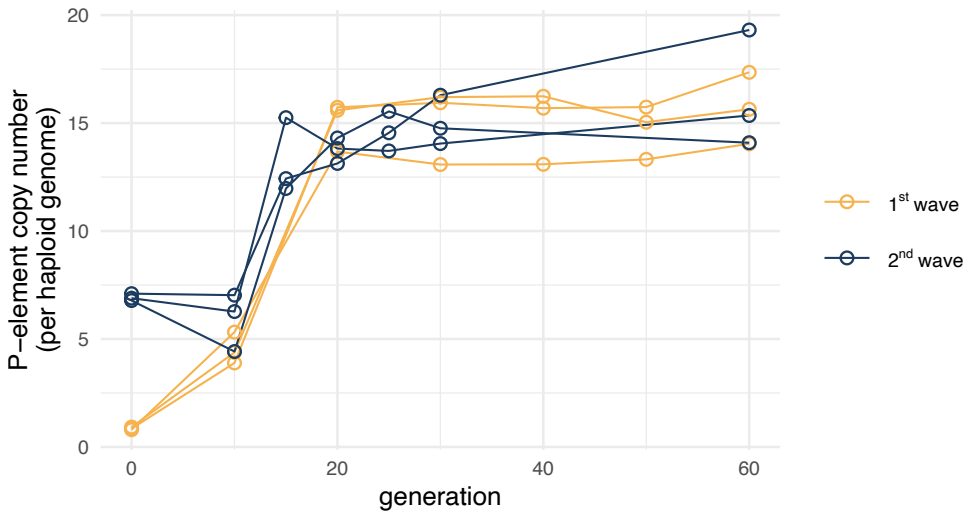


Figure 2. Estimated P-element copy number per haploid genome from the 1st (amber) and 2nd (dark navy blue) wave Experimental Evolution study. Sequenced time points are shown as circles.

Purifying selection shapes P-element invasion dynamics

We estimated similar P-element copy numbers per haploid genome in generations 0 and 10 in the 2nd wave experiment (means: 6.92 vs. 5.91, Figure 2). We previously hypothesized that this pattern arises from the balance between purifying selection, which removes deleterious P-element insertions, and transposition activity, which introduces new P-element copies [11]. Nevertheless, a comprehensive understanding of P-element invasion dynamics necessitates the availability of reliable estimates of the strength of selection operating on new P-element insertions. While genetic editing enables the inference of selection operating on specific TE insertions [40], it is evident that this approach is not feasible to describe the costs associated with insertions throughout the entire genome. Therefore, we used computer simulations to estimate the strength of purifying selection in our EE experiments.

Simulation Framework

To investigate the potential influence of purifying selection on P-element invasion dynamics in the 1st and 2nd wave experiments, we developed an individual-based model for transposon dynamics in SLiM [41]. A comprehensive description of the model can be found in the Materials & Methods section. Figure 3 provides a schematic overview of the individual-based model, and Table 1 gives an overview of all model parameters. In brief, the individual-based model is able to simulate P-element invasions in EE experiments reflecting the empirical setups of the 1st and 2nd wave, respectively. As a host defense mechanism, we implemented a classic "trap model" [35,36,39]: P-elements transpose with a defined probability unless silenced by an insertion in a piRNA cluster [33], which immediately inactivates all P-elements in the genome. Outside piRNA clusters, purifying selection acts on P-elements with a selection coefficient *s* drawn from a beta distribution and a dominance coefficient *h* of 0.5 (co-dominance). By modifying the parameters of the beta distribution, it is possible to create a gradual transition between simulation scenarios where the majority of P-elements are effectively neutral and scenarios where the majority of P-elements are subject to strong purifying selection.

By comparing simulated invasion outcomes to empirical data and quantifying the fit, systematic exploration of this model's parameter space (Table 1) would in principle allow us to determine the level of purifying selection that best fits the observed P-element copy number trajectories from the 1st and 2nd wave experiments. However, exhaustively probing our five-dimensional parameter space in this manner would be very computationally intensive [42], even using an optimized model. To address this challenge, we used Gaussian Process (GP) surrogate modeling [43]. A GP is a statistical model that serves as a surrogate for the individual-based model, allowing rapid

prediction of the individual-based model's behavior based on previously simulated data. The utility of the GP approach lies in its capacity to extrapolate from sparse data, thereby predicting the behavior of a system, such as our individual-based model, at untested parameter combinations. We trained two separate GPs — one for the 1st wave and one for the 2nd wave — using simulation outcomes from the individual-based model as described further in the Materials & Methods section. We then explored the accuracy of our trained GPs with a test dataset of 5,000 additional simulation outcomes spanning a broad range of parameter combinations, for both the 1st and 2nd wave experiments. For this test dataset, the GPs accurately predicted trajectories of P-element copy numbers across generations (Figure 4). Based on these results, we conclude that the GPs provide an efficient and precise alternative to the individual-based model. This allowed us to use the GPs to find the model parameters that best replicate the experimental data, ultimately allowing us to estimate the level of purifying selection for which our individual-based model best fits the empirical time-series data.

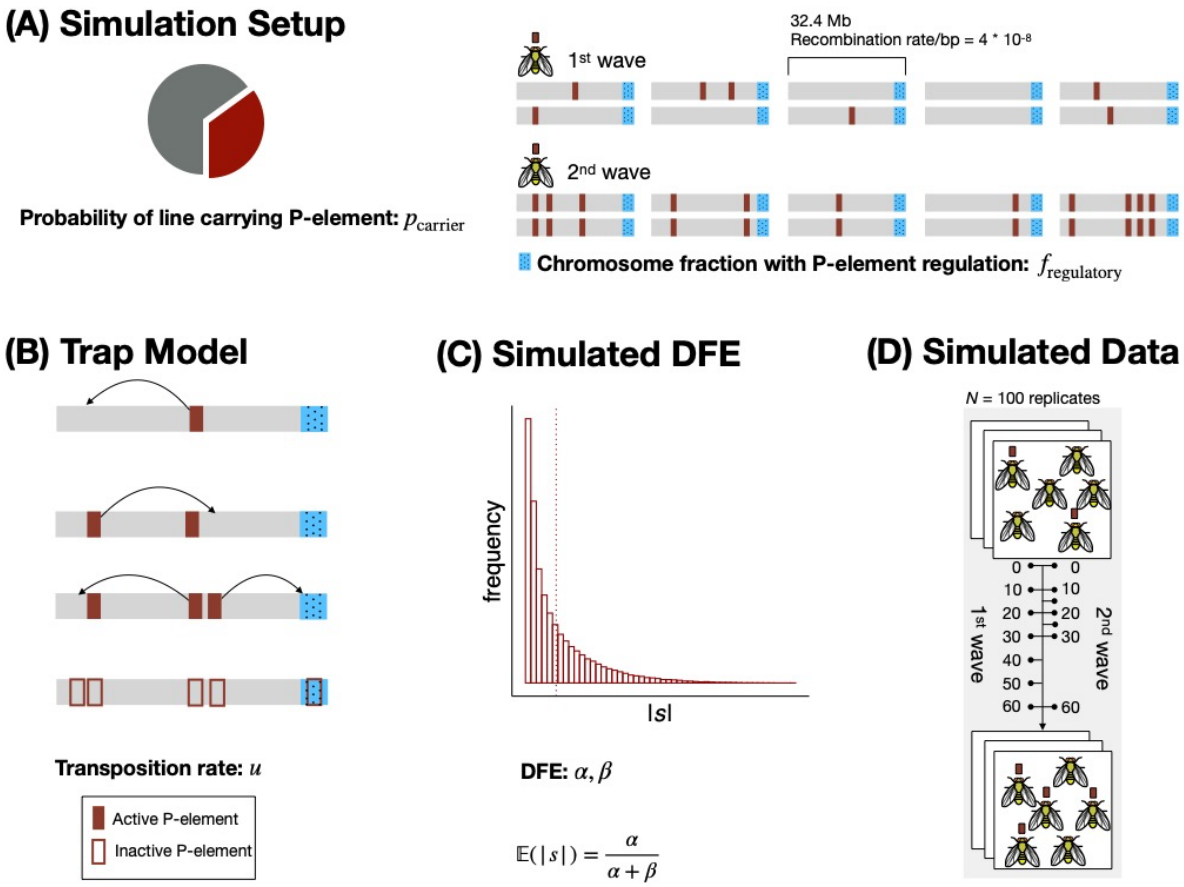


Figure 3. Schematic overview of the individual-based simulation model mimicking the P-element invasion dynamics in our Experimental Evolution experiments. **(A) Simulation Setup:** Ancestral outbred populations are generated by mixing about 200 isofemale lines (five flies each). Each line carries the P-element with probability $p_{carrier}$. We modeled diploid individuals with five chromosomes, each with a fixed length of 32.4 Mb and a recombination rate of 4×10^{-8} per bp per generation. For the 1st wave, P-element insertions are assumed to be heterozygous. For the 2nd wave, since the experiment started with isofemale lines that had been maintained at small populations sizes for 4.5 years [11], the model assumes that all P-element insertions are homozygous due to increased inbreeding and the likely establishment of a defense mechanism. The parameter $f_{regulatory}$ defines the fraction of each chromosome with P-element-regulatory properties (piRNA clusters; blue rectangles). **(B) Trap Model:** The P-element remains active (filled rectangle) unless one of its copies transposes into a piRNA cluster (blue rectangles). The probability of transposition for a single P-element per generation is controlled by the transposition rate u. Once a piRNA cluster acquires a single P-element insertion, all P-elements in the genome are immediately inactivated (unfilled rectangles with red borders). **(C) Simulated Distribution of Fitness Effects (DFE):** The DFE for P-elements is modeled using a beta distribution with parameters α and β. Positive selection is not considered in our model. **(D) Simulated Data:** The simulation output used in the analyses contains the average P-element copy number per haploid genome across 100 simulation runs, taken at the same time points used in the two Experimental Evolution studies.

Table 1. Model parameters of the individual-based P-element invasion model and their considered ranges. The range for $p_{carrier}$ was guided by a previous study estimating that 25–44% of the isofemale lines used in the 1st and 2nd wave experiment carried the P-element [31], while the range for $f_{regulatory}$ was based on the estimate that approximately 3.5% of the *Drosophila* genome consists of piRNA clusters with TE-regulatory properties [33].

Parameter	Description	Range
$p_{carrier}$	The probability that an isofemale line carries the P-element at the beginning of the simulation	[0.15, 0.5]
$f_{regulatory}$	piRNA cluster size: the fraction of the chromosome with P-element-regulatory properties (once at least one P-element insertion is acquired)	[0.01, 0.1]
u	The probability of transposition of a single P-element per generation	[0.15, 0.5]
α	The α parameter of the beta distribution that determines the distribution of fitness effects for P-elements outside piRNA clusters	[0.001, 0.5]
β	The β parameter of the beta distribution that determines the distribution of fitness effects for P-elements outside piRNA clusters	[10, 20]

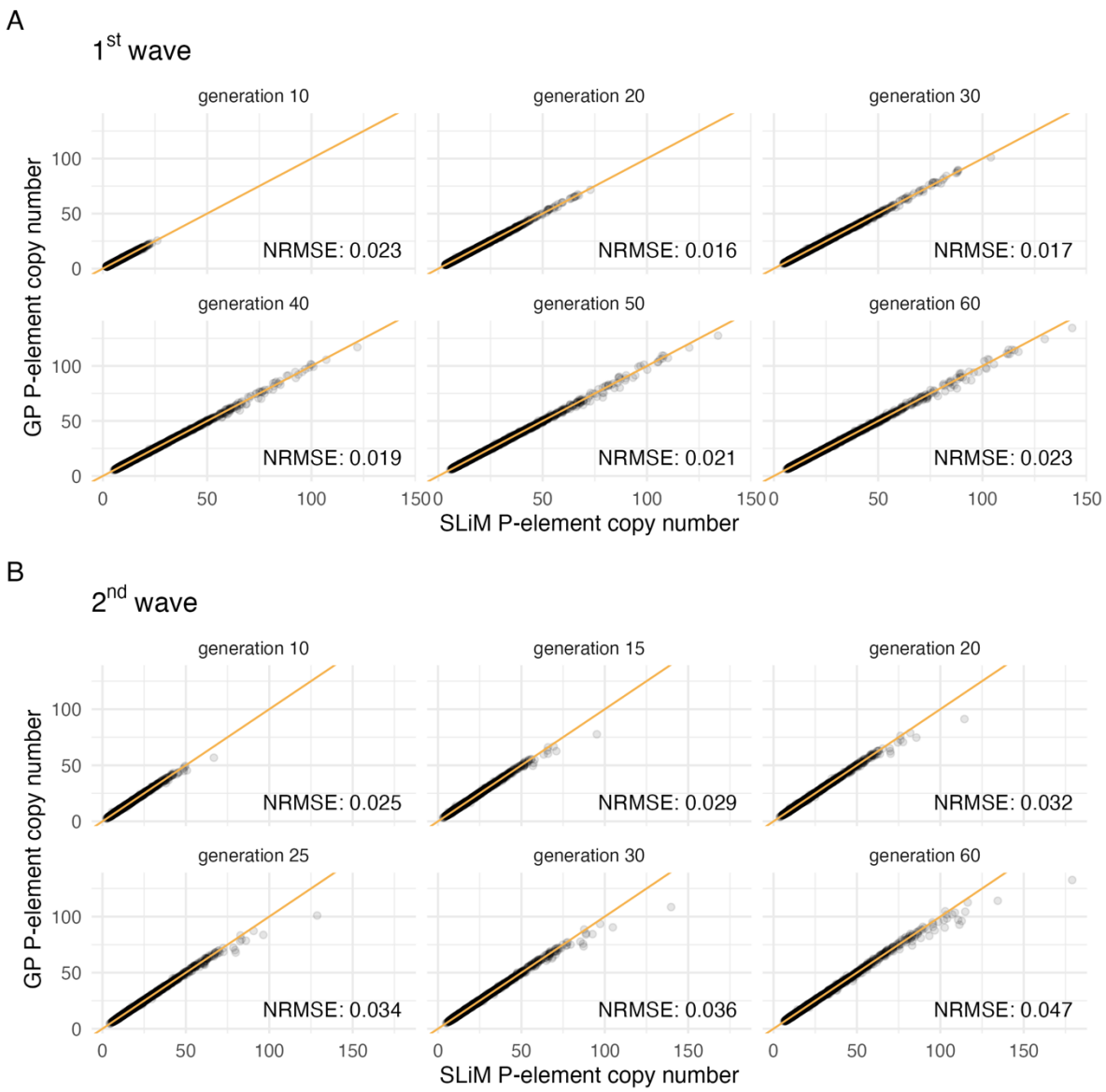


Figure 4. Gaussian Process (GP) performance: A test dataset consisting of 5,000 data points was simulated with the individual-based model (*x* axis) for the (A) 1st wave and (B) 2nd wave and compared to the predictions of the GP (*y* axis). One data point in this test dataset comprises a specific combination of five model parameters (Table 1) and the corresponding predicted P-element copy numbers for six time points. The observed and predicted P-element copy numbers are shown (gray dots) for each of those six time points (the six panels), with amber lines indicating the identity line ($x = y$). Normalized root mean square errors (NRMSE) for the time points are shown at the bottom right of the corresponding panels. The analysis shows that the GP can predict the copy number observed in the individual-based model very accurately. Only at extremely high copy numbers — well beyond the empirical estimates (Figure 2) — can a slight underestimation be observed.

Purifying selection as a central mechanism: model fitting results

After validating the performance of the GPs with the simulated test dataset, we used the GPs to predict P-element invasion dynamics — the P-element copy number per haploid genome for six evolved time points corresponding to the sequenced time points of the empirical EE experiments — across 10^6 model parameter combinations. To ensure that later generations, which tend to have higher copy numbers (Figure 2), did not disproportionately influence the fit, we quantified the fit between the model predictions and the experimental data using the normalized root mean square error (NRMSE; please refer to the M&M section for further details), where a lower NRMSE value indicates a better agreement between the predicted and empirical P-element invasion curves.

By fitting GP predictions from 10^6 parameter combinations to the 1st wave invasion dynamics, we observed a wide range of NRMSE values, spanning from 0.60 to 39.28. The best fit, characterized by the lowest NRMSE, corresponded to a parameter combination with notable purifying selection outside piRNA clusters ($\bar{s} = -0.045$, Figure 5). Although the initial invasion dynamics of the 1st and 2nd wave EE experiments differ considerably, we reasoned that the genome-wide distribution of fitness effects for P-elements should be constant between the two experiments. Thus, we used the parameters estimated from the 1st wave experiment to predict the invasion trajectory of the 2nd wave experiment. Interestingly, this resulted in a remarkably accurate prediction of the invasion dynamics of the 2nd wave experiment, which has been conducted four years later with an 8-fold higher initial P-element copy number (Figure 5). This result highlights the robustness and predictive power of our parameter estimates even for rather distinct invasion trajectories in the first generations.

After confirming that the parameters inferred from the 1st wave also explain the 2nd wave data, we proceeded to jointly infer parameters using data from both experiments. We reasoned that this could further improve parameter estimates because the estimates will be based on twice the amount of data. The overall fit metric was calculated as the sum of the NRMSEs for the 1st and 2nd wave experiments (NRMSE sum). We observed a considerable degree of variation in the NRMSE sums across the evaluated parameter combinations, with values ranging from 1.340 to 80.365 (Figure S1A). As anticipated, the NRMSE values were strongly correlated between the 1st and the 2nd wave experiment (Spearman's $\rho = 0.87$, Figure S1B), consistent with similar evolutionary dynamics driving the P-element invasion in both EE experiments.

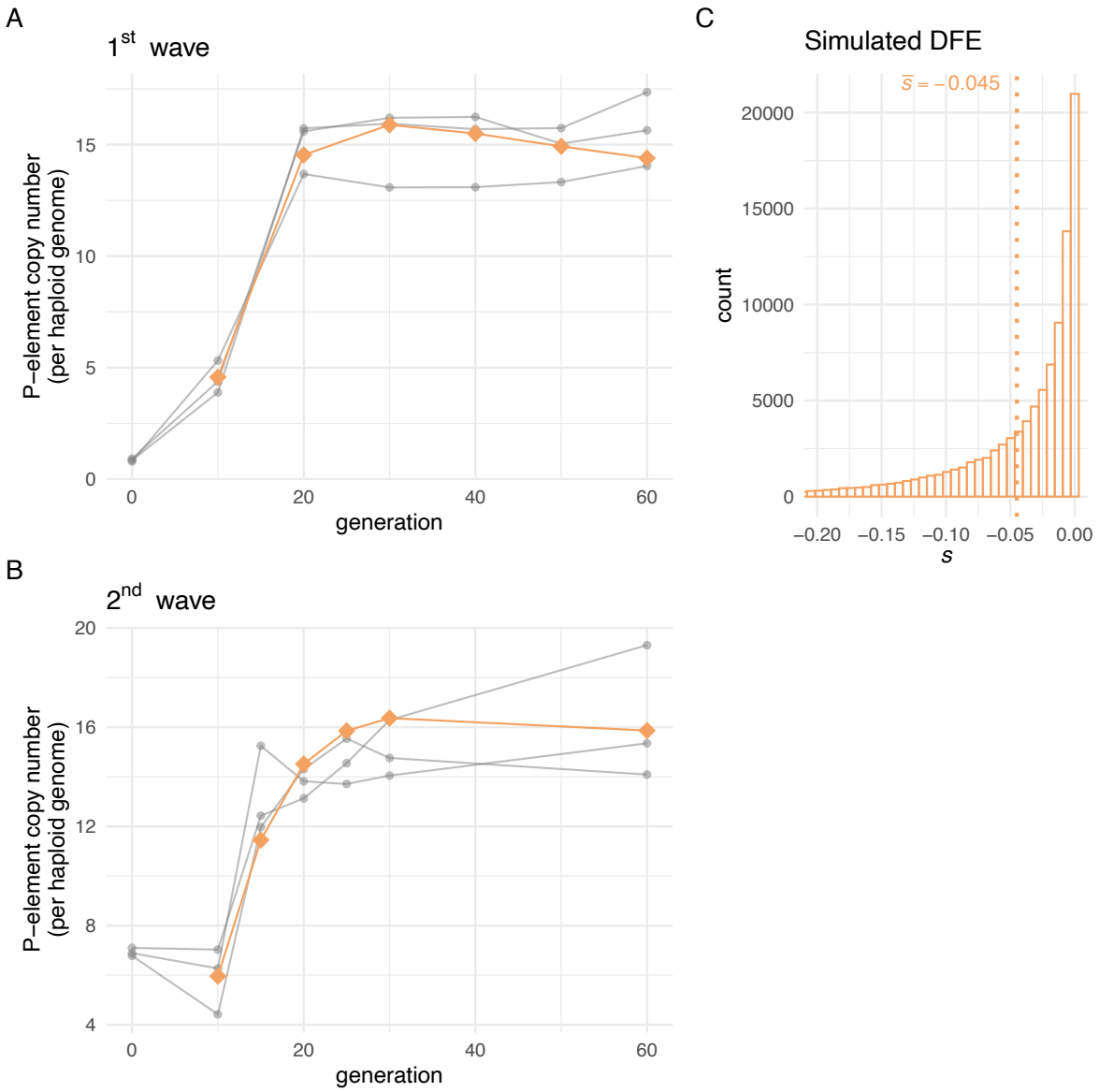
In the best-fitting scenario (Table 2), approximately 20% of the isofemale lines were predicted to carry the P-element, which is slightly below the empirical estimate of 25–44% [31]. Additionally, the estimated piRNA cluster size (fraction of each chromosome with the P-element-regulatory properties) under this scenario is 1.7%, which is approximately half the size as suggested previously [33]. The results of the parameter tuning indicated that a significant amount of purifying selection against single P-element insertions is necessary to account for the observed invasion dynamics. The best-fitting parameter combination, as indicated by the lowest NRMSE sum, Table 2) was obtained for $\alpha = 0.461$ and $\beta = 10.743$, which yielded to an average genome-wide selection coefficient *s* of -0.041 for P-element insertions outside the piRNA clusters. This indicates that selection is effective against P-element insertions in experimental *Drosophila* populations, which have typically effective population sizes $N_e < 300$ [44,45]. Based on the average N_e estimate from the 1st wave experiment [45] and the aforementioned best-

297 fitting parameter combination, our results indicate that only 27% of P-elements are
298 effectively neutral (i.e., $N_e|s| \leq 1$, Table 2) in our experimental populations. Additionally,
299 our findings suggest that strong selection can occur, with the 95th percentile of the
300 selection coefficient ($|s|$) reaching 0.16 (Table 2).
301
302 We also used the mean N_e estimate from the 1st wave experiment [45] to categorize
303 parameter combinations into two distinct scenarios: those exhibiting efficient purifying
304 selection against the average P-element insertion, and effectively neural ones. In the
305 latter case, genetic drift is so strong that the invasion trajectory of an average P-element
306 insertion is indistinguishable from a neutral one, even if the insertion is slightly
307 deleterious. Simulation scenarios classified as effectively neutral had the best fit to the
308 data when they included larger piRNA clusters and higher P-element activity (Table 2).
309
310 Nevertheless, simulations incorporating efficient purifying selection yielded a better
311 agreement with empirical invasion dynamics. These scenarios accurately reproduced
312 both the rapid increase in P-element copy number that occurred during the 1st wave and
313 the transient plateau that was observed between generations 0 and 10 in the 2nd wave. In
314 contrast, neutral scenarios consistently failed to replicate these patterns (Figure 6). This
315 conclusion holds even when potential biases were considered, such as focusing solely
316 on the parameter combination with the lowest NRMSE sum. The qualitative pattern
317 remains consistent when using average parameter values from the 100 combinations
318 with the lowest NRMSE sums (Figure S2). Moreover, our results remain consistent when
319 the N_e estimate was modified, as this parameter serves as a threshold for neutrality in our
320 approach. Doubling the N_e estimate from 221 to 442 did not result in any qualitative

321 alterations to the outcomes, thereby reinforcing the conclusion that purifying selection

322 plays a pivotal role in shaping the observed P-element invasion dynamics (Table S1).

323



324

325 **Figure 5.** Gaussian Process (GP) predictions provide a good fit to the empirical data — model parameters

326 estimated only from 1st wave data: (A) 1st wave experiment (B) 2nd wave experiment. Each grey line

327 represents an empirical evolution replicate, with sequenced time points indicated by dots. GP predictions

328 are indicated by amber diamonds. (C) The simulated distribution of fitness effects (DFE) used for

329 generating predictions in panels (A) and (B). The parameter combination used for the GP prediction is: $p_{carrier}$

330 $= 0.189$, $f_{regulatory} = 0.014$, $u = 0.278$, $\alpha = 0.484$, $\beta = 10.327$. Predictions based on parameters inferred from

331 the 1st wave accurately describe the P-element invasion dynamics in the 2nd wave.

Table 2. Summary statistics describing the 1 ("best fit") versus 100 ("top 100") parameter combination(s) with the lowest NRMSE sum (NRMSE sum = NRMSE 1st wave + NRMSE 2nd wave) from the parameter exploration with the Gaussian Process model, grouped by average expected selection efficacy (N_e = 221 [45]). Values in square brackets represent the range between the 2.5th and 97.5th percentiles of individual metrics. $|\bar{s}|$ = mean selection coefficient outside piRNA clusters ($\mathrm{E}(|s|) = \alpha/(\alpha+\beta)$); 95th percentile $|s|$ = 95th percentile of $|s|$ outside piRNA clusters; Neutral fraction = expected fraction of effectively neutral ($|s| \leq 1/N_e$) P-element insertions outside of piRNA clusters; $|\bar{s}_{\mathrm{effective}}|$ = mean selection coefficient for P-element insertions outside piRNA clusters where selection is strong enough to overcome genetic drift ($|s| > 1/N_e$); NRMSE = Normalized root mean square error; Results are robust to changes in the effective population size estimate (Table S1).

	Best fit		**Top 100**	
	$\frac{\mathrm{E}(\lvert s\rvert)}{2} \leq \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} > \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} \leq \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} > \frac{1}{2N_e}$
Parameter	"neutral"	"selection"	"neutral"	"selection"
p_{carrier}	0.151	0.204	0.180	0.192
			[0.151, 0.247]	[0.158, 0.237]
$f_{\mathrm{regulatory}}$	0.092	0.017	0.094	0.021
			[0.087, 0.099]	[0.015, 0.027]
u	0.343	0.290	0.352	0.288
			[0.317, 0.385]	[0.258, 0.317]
α	0.065	0.461	0.061	0.446
			[0.033, 0.083]	[0.362, 0.500]
β	15.696	10.743	15.297	11.676
			[10.297, 19.509]	[10.216, 14.764]
$\lvert\bar{s}\rvert$	0.004	0.041	0.004	0.037
			[0.003, 0.005]	[0.031, 0.045]
95th percentile $\lvert s\rvert$	0.024	0.160	0.023	0.145
			[0.013, 0.027]	[0.121, 0.170]
Neutral fraction	0.866	0.273	0.873	0.297
			[0.845, 0.919]	[0.243, 0.365]
$\lvert\bar{s}_{\mathrm{effective}}\rvert$	0.029	0.056	0.029	0.052
			[0.025, 0.038]	[0.044, 0.059]
NRMSE 1st + 2nd wave	2.254	1.340	2.286	1.462
			[2.260, 2.305]	[1.361, 1.507]

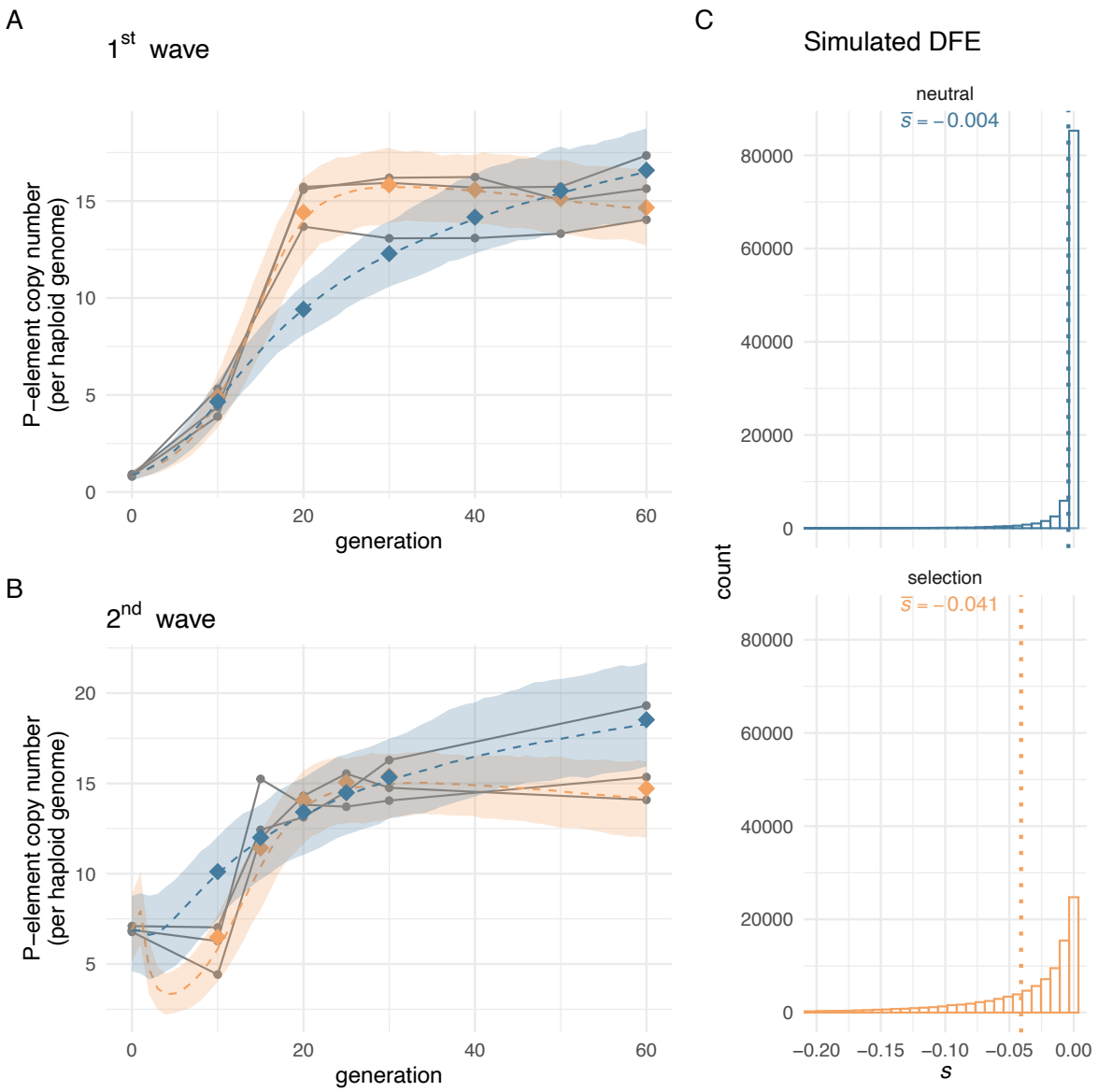


Figure 6. Gaussian Process (GP) predictions provide a good fit to the empirical data — parameters estimated jointly from 1st and 2nd wave data. GP predictions from the parameter combinations with the lowest NRMSE sum when compared against empirical data are shown for scenarios without expected effective purifying selection ($E(|s|/2) \leq 1/(2N_e)$, steel blue) and scenarios with expected effective purifying selection ($E(|s|/2) > 1/(2N_e)$, orange), N_e = 221 [45]. Each grey line represents an empirical evolution replicate, with sequenced time points indicated by dots. GP predictions are marked by diamonds, while the dashed colored lines represent the average P-element copy number per haploid genome across 100 simulation runs with the individual-based model. The colored ribbons show the range between the 2.5th and 97.5th percentiles of P-element copy number trajectories simulated with the individual-based model. (A) 1st wave experiment. (B) 2nd wave experiment. (C) Simulated distribution of fitness effects (DFE) used in (A) and (B). Results are robust to averaging parameter values across 100 combinations with the lowest NRMSE sums (Figure S2).

Discussion

The power of Experimental Evolution to study TE dynamics

To evaluate the role of purifying selection in controlling TEs, we investigated the dynamics of P-element invasions in two *Drosophila* EE experiments. Because low-frequency TE insertions, a hallmark of the early stages of an invasion process, are very difficult to study empirically, we used the P-element copy number as a summary statistic to describe the invasion dynamics. Through careful model analysis, we estimated the distribution of fitness effects of P-element insertions in these experiments.

While beneficial fitness effects have been documented for some TE insertions (e.g., [46–50]), we reasoned that such cases are rare, and thus unlikely to substantially impact the population-level dynamics of a P-element invasion. Furthermore, a previous analysis of the 1st wave experiment found no evidence for positive selection acting on any P-element insertion site in 10 replicate populations over 60 generations [44]. We thus focused our analysis on deleterious effects of P-element insertions. We fitted an individual-based simulation model to data from the two EE studies to infer the distribution of deleterious fitness effects, first from the 1st wave alone, and then from both the 1st and 2nd waves combined. Both approaches yielded similar estimates. Remarkably, parameters inferred only from the 1st wave accurately predicted the invasion dynamics observed in the 2nd wave experiment, conducted four years later under different initial conditions. Overall, we observed that substantial purifying selection outside of piRNA clusters ($\bar{s} = -0.041$, Table 2) was necessary to accurately capture the observed P-element invasion patterns in both experiments. TE invasions are challenging to analyze in natural populations due

379 to their rapid dynamics and confounding demographic factors such as gene flow.
380 Excitingly, this study not only demonstrates that purifying selection is a key evolutionary
381 force shaping P-element invasions, but also highlights the utility of EE, in combination
382 with simulation and modeling-based approaches, for studying TE invasions in the lab.
383

384 **Alternative Mechanisms of TE Regulation**

385 A key feature of our study is the individual-based modeling framework used to simulate
386 P-element invasion dynamics in experimental *Drosophila* populations. As with any
387 model, it is based on a number of simplifying assumptions that inevitably omit some
388 aspects of biological realism, and these assumptions might influence the model
389 outcomes. For example, our individual-based simulations are based on the classic trap
390 model, which posits that a single TE insertion in a piRNA cluster silences all TEs of that
391 type. Given that the distribution of TEs in piRNA clusters differs from predictions under
392 the trap model [51,52] and that the removal of piRNA clusters does not necessarily
393 induce TE activity [53], alternative mechanisms of TE silencing must be considered.
394

395 One potential mechanism could be paramutations, whereby insertions outside of piRNA
396 clusters can be converted into piRNA-producing loci in the presence of maternally
397 deposited piRNAs [54–56]. In a recent study, Scarpa *et al*. (2023) demonstrated with
398 comprehensive computer simulations that this additional layer of defense against TEs
399 can help to resolve some of the discrepancies between observed and predicted TE copy
400 numbers under the classic trap model in natural *Drosophila* populations [54]. Since this
401 study focused on the overall abundance of TE families in *D. melanogaster*, it remains
402 unclear to what extent paramutations are shaping TE invasion processes in natural and

experimental populations over time. Given the excellent fit between the simulations under the trap model and our empirical data, it is possible that paramutations are not essential to explain the observed P-element dynamics in our two EE experiments. Even more importantly, the transient P-element copy number plateau observed between generations 0 and 10 in the 2nd wave experiment, along with the subsequent increase in copy number in later generations, is more readily explained by purifying selection rather than by paramutations. Further work is required to evaluate the extent to which paramutations influence the invasion dynamics of the P-element and other TEs.

Most TE insertions are deleterious

The effective population size is a key factor in determining whether a mutation is "seen" by selection or behaves effectively neutrally [57]. In EE experiments with *Drosophila*, such as those presented in this work, estimates of effective population size typically range from 200 to 400 [44,45]. At this effective population size, purifying selection can only act on mutations with strongly deleterious effects. In natural *Drosophila* populations, the effective population size is several orders of magnitude larger [58,59]. This implies that in natural *Drosophila* populations, only a small fraction of P-elements is likely to be effectively neutral — much fewer than the 27% estimated from our empirical data. This idea aligns well with the observation that only a small fraction of the genome shares P-element insertions between *D. melanogaster* and *D. simulans* [32]. While this is typically attributed to non-random insertions of actively transposing P-elements [12], we propose that purifying selection in large natural populations might drive independent enrichment of P-element insertions at genomic regions subject to relaxed purifying selection.

Conclusions

Our study highlights the crucial role of purifying selection in shaping TE invasion dynamics, specifically in the context of P-element invasions in *D. simulans*. Using an individual-based model, we demonstrated that substantial purifying selection is needed to replicate the observed invasion dynamics in two EE studies. This study not only highlights EE as a valuable tool for understanding TE behavior, but also illustrates how computational modeling can complement experimental work to study TE invasions. Future work could expand this modeling framework to include additional regulatory mechanisms such as paramutations, advancing our understanding of TE behavior in diverse environments.

Methods

Experimental Evolution

The details of the 1st wave EE experiment have been previously published [20,45]. Additionally, Pool-Seq data for generations 0 and 10 of the 2nd wave EE experiment are available in Langmüller *et al*. (2023) [11]. This study adds generation 15–60 of the 2nd wave experiment. This section only provides an overview of the experiments and specific information regarding already published data can be found in those sources. To maintain consistency and facilitate comparison between studies, we used the same replicate identifiers as in the original publications.

Setup & Maintenance of Experimental *D. simulans* Populations

449 For the 1st wave [20], replicate populations were established from 202 isofemale lines
450 collected from a natural *D. simulans* population in Tallahassee, Florida [60], which
451 experienced a P-element invasion at the time point of sampling [32]. For the 2nd wave
452 experiment, three replicate populations were established using the surviving 191 out of
453 202 isofemale lines that were originally collected [11]. All replicate populations were
454 maintained with non-overlapping generations in a cycling hot environment (12 hours of
455 light at 28°C; 12 hours of darkness at 18°C) and at a constant population size (N = 1,000
456 for the 1st wave; N = 1,250 for the 2nd wave) for 60 generations.
457
458 **Genomic Sequencing**
459 All experimental populations were sequenced using the Pool-Seq approach [37] with at
460 least 500 flies per sample. In the 1st wave, populations were sequenced every 10th
461 generation, resulting in 21 samples (7 time points; 3 replicate populations) [20,45]. To
462 more effectively monitor the rapid invasion of the P-element, we increased the number
463 of sequenced time points at the beginning of the 2nd wave experiment. For this wave,
464 populations were sequenced at generation 0, as well as at generations 10, 15, 20, 25, 30,
465 and 60, also resulting in 21 samples. Genomic DNA for all new samples was extracted
466 from half the number of individuals that had contributed to the next generation
467 (approximately 600 flies). To facilitate comparison across studies, we adopt the Fxx
468 nomenclature for sequencing samples, where 'xx' represents the generation in our EE
469 experiment. Paired-end libraries for generation F60 were prepared and sequenced along
470 with those for generation F0 using the NEBNext Ultra II DNA Library Prep Kit (New England
471 Biolabs, Ipswich, MA) with an insert size of 300bp. Paired-end libraries for generations
472 F15, F20, F25, and F30 were processed in the same way as those for generation F10, using

473 a protocol with 10% of the reagents provided in the NEBNext Ultra II FS DNA Library Prep
474 Kit (New England Biolabs, Ipswich, MA). Steps for all library preparations are described in
475 more detail in [11]. We sequenced all libraries using a 2 × 150 bp protocol on a HiSeq X
476 Ten.
477
478 **P-element Copy Number as a Summary Statistic for Invasion Dynamics**
479 Given the rapid spread of the P-element in our experimental populations, most individual
480 P-element insertions are expected to remain at low frequency. This poses a major
481 challenge for accurately estimating empirical P-element insertion frequenies,
482 irrespective of whether pools of individuals or single individuals are being analyzed.
483 Therefore, we used the estimated P-element copy number per haploid genome as a
484 summary statistic for describing the P-element invasion in the experimental populations.
485 We used DeviaTE [38] to estimate P-element copy number per haploid genome in both
486 EE experiments. DeviaTE calculates P-element copy number by comparing the number
487 of reads mapped to the P-element consensus sequence with the number of reads
488 mapped to single copy genes [38,61]. For the 1st wave, P-element copy number estimates
489 were previously published in Kofler *et al.* 2022 and can be found in the supplemental data
490 (replicates 1, 3, and 5 from the hot environment) [45].
491
492 For 2nd the wave, we used ReadTools (v1.5.2.) [62] to demultiplex *(–maximumMismatches*
493 *3 –maximumN 2*) and trim barcoded sequencing files based on sequencing quality (–
494 *mottQualityThreshold 15 –disable5pTrim –minReadLength 25*). Since DeviaTE does not
495 take paired-end read information into account [38], we merged paired-end read files prior
496 mapping. We mapped reads to a reference genome that included the *Drosophila* P-

497 element consensus sequence [63] along with three single copy genes (*rpl32, traffic jam,*
498 *rhino),* using bwa bwasw (v0.7.17; *–M*) [64]. Finally, we estimated P-element copy number
499 per haploid genome with DeviaTE (*–minID 1*) [38].
500
501 Our genomic analysis pipeline follows Kofler *et al*. 2022 with some minor differences
502 (e.g., we used a more recent version of bwasw) [45]. To verify the consistency between
503 pipelines, we re-analyzed three randomly selected samples from Kofler *et al*. (2022) using
504 our updated pipeline. The copy number estimates from the two pipelines differed by less
505 than 5 %, which we considered an acceptable level of consistency (data not shown).
506

507 P-element Invasion Models

508 Individual-based Models

509 We implemented an individual-based P-element invasion model in SLiM (v4.0.1) [41]. Our
510 individual-based model simulates an obligatory outcrossing population with discrete,
511 non-overlapping generations and a constant census size of 1,000 individuals (flies). Each
512 fly has a diploid genome consisting of five chromosomes, each with a length of 32.4 Mb
513 and a recombination rate of 4×10^{-8} per bp per generation, mimicking *Drosophila simulans*
514 [65,66]. At the end of each chromosome, we modeled a piRNA cluster region capable of
515 regulating P-element activity [33,39]. The size of simulated piRNA clusters is regulated by
516 the model parameter $f_{\text{regulatory}}$. The only genomic variants included in the individual-based
517 model are P-element insertion sites (Figure 3A).
518
519 The ancestral population (generation 0) is initialized by mixing 200 inbred fly strains
520 (isofemale lines), using five flies per line. Single lines might be P-element carriers, with a

521 user-specified probability $p_{carrier}$ (Table 1). If a line carries the P-element, the number of
522 initial P-element insertion sites is drawn from a Poisson distribution, where the rate
523 parameter λ is chosen such that, given the value of $p_{carrier}$, the average P-element copy
524 number per haploid genome for the simulated EE experiment fits the empirical data (1st
525 wave: 0.86 P-element copies per haploid genome; 2nd wave: 6.92). All flies from the same
526 isofemale line are isogenic, meaning they share the same set of P-element insertion
527 sites. We note that this assumption likely oversimplifies the genetic structure of
528 isofemale lines. Any residual within-line heterogeneity could introduce additional
529 variance in P-element dynamics across simulation runs. However, we averaged
530 individual-based model results across 100 simulation runs for each parameter
531 combination, which would mitigate the effect of this additional variance, such that our
532 conclusions about the overall invasion dynamics remain robust despite this
533 simplification. For the 1st wave, we modeled heterozygous initial P-element insertions,
534 because the P-element invasion was recent [32]. For the 2nd wave, we modeled
535 homozygous initial P-element insertions, given that the isofemale lines had been
536 maintained at small population sizes for 4.5 years [11], causing increased inbreeding and
537 a likely establishment of defense mechanisms against further P-element proliferation.
538
539 To account for a host defense mechanism against the P-element, we incorporated the
540 trap model into our individual-based model [35,36]. Under the trap model, P-element
541 activity is regulated by insertions into piRNA clusters [33]. In our individual-based model,
542 single P-elements transpose with probability u per generation, unless a fly carries at least
543 one P-element insertion in a piRNA cluster. We assumed that P-element proliferation
544 operates exclusively via a copy-and-paste mechanism, with no excision events. This

545 simplification is based on evidence from previous studies that P-elements can increase
546 their copy number through mechanisms such as sister chromatid-mediated gap repair
547 following excision [67] and transposition into unreplicated DNA during cell division,
548 which facilitates copy number increase without significant loss of existing elements [12].
549 If a single P-element insertion occurs in any piRNA cluster, all P-elements within the
550 genome are rendered inactive immediately and are unable to transpose further (Figure
551 3B).
552
553 We assumed that P-element insertions within piRNA clusters are neutral (selection
554 coefficient $s = 0$). For all other P-element insertions, the absolute value of s is drawn from
555 a beta distribution with shape parameters α and β (Figure 3C). We considered only
556 purifying selection against P-elements ($s <= 0$), because no signature of adaptive
557 selection was detected in 10 replicate populations in the 1st wave experiment [44]. We
558 modeled co-dominant fitness effects ($h = 0.5$), such that individuals with no P-element
559 insertions at a single locus have a fitness of 1, heterozygous individuals have a fitness of
560 $1 + s/2$, and individuals homozygous for the P-element insertion have a fitness of $1 + s$,
561 where s is always $<= 0$. With more than one P-element insertion site, the fitness effects
562 from each site are combined with the others multiplicatively.
563
564 We systematically varied five parameters: the probability that a single line carries the P-
565 element, the fraction of the chromosome capable of triggering a defense against the P-
566 element (piRNA cluster size), the transposition probability, and two parameters governing
567 the distribution of fitness effects (Table 1). For each parameter combination, simulations
568 were repeated 100 times, and the average P-element copy number per haploid genome

was recorded at the same time points used in the empirical EE experiments (Figure 3D). This approach allowed us to directly compare simulation outcomes with observed P-element copy numbers.

Gaussian Process Surrogate Models

Because our individual-based model has five parameters that can be varied (Table 1), exploring the full range of this parameter space with the individual-based model would be a complex and time-intensive task [42]. To streamline this process, we employed statistical emulation, replacing the individual-based model with a surrogate model that can rapidly and accurately predict the individual-based model's behavior [68]. GP surrogate models [43,69] are particularly well-suited for this task because they are non-parametric models that define distributions over functions. This characteristic allows them to efficiently extrapolate between sparsely sampled data points. For a detailed introduction to GPs, we defer to Rasmussen and Williams [43].

We implemented our GPs in Python (v3.10.6) using the GpyTorch (v1.11) [70] and PyTorch (v2.0.1) [71] libraries for efficient GP modeling. We developed two GPs, one for each of the EE experiments (1st and 2nd wave). Specifically, we used a multi-task GP [70,72] to model the relationship between six tasks — the P-element copy number for the evolved generations of the corresponding EE experiment — and five model parameters (Table 1), which serve as predictors across all tasks. Due to time-related dependencies of P-element copy numbers within one experiment, we assumed that tasks are correlated and modeled their relationship using a rank-1 covariance structure. This means that any correlation between single tasks is determined by one single latent factor. The multi-task

GP consists of a constant mean function providing a baseline P-element copy number estimate for each task (constant mean = 0 copies), and a covariance function that accounts for both: input-dependent correlation within one task (k_{input}) and correlations across tasks (k_{task}):

$$k\big((x,t),(x',t')\big) = k_{\text{input}}(x,x') * k_{\text{task}}(t,t')$$

Where $k_{\text{input}}(x,x')$ is a radial basis function kernel modeling the covariance between two input data points *x* and *x'*[43], and $k_{\text{task}}(t,t')$ models the relationship between tasks *t* and *t'*[70,72]. Using a rank-1 task-kernel simplifies the task covariance matrix K_{task} to:

$$K_{\text{task}} = \omega\omega^T + \sigma^2 I$$

Where ω is a vector of task weights and σ^2 represents task-specific, independent noise. During GP training, the task weights ω, k_{input} hyper-parameters, and task-specific noise are learned by maximizing the log marginal likelihood over observed training data.

We trained GPs on a dataset of 1,000 points using Latin Hypercube Sampling (LHS), a method that ensures evenly distributed exploration of the entire input domain (Table 1), where P-element copy number time series were generated with the experiment-specific individual-based model. We used the Adam optimizer from PyTorch with a learning rate of 0.01 [71] to train each GP over 40 rounds (50 iterations per round). After each round, a snapshot of the GP was saved and evaluated against an independent validation dataset consisting of 5,000 LHS points to determine the optimal training duration and avoid overfitting. We selected the GP snapshot with the lowest root mean square error (RMSE) on the validation dataset for further analysis (snapshot 18 for 1st wave, snapshot 19 for 2nd wave).

To evaluate the GP's performance, we predicted P-element copy number time series for a test dataset consisting of 5,000 LHS data points. To make GP model performance comparable between the two EE experiments, we report RMSE values normalized by the average observed P-element copy number per time point. This normalization was necessary because the 2nd wave, which has higher absolute P-element copy numbers in the ancestral generation, would otherwise naturally result in higher RMSEs. The trained GP was only challenged when predicting large P-element copy numbers in the test data.

This issue is likely due to our choice of a constant mean of 0 across all tasks (P-element copy number per generation). When data points are sparse, the GP reverts to the prior mean of 0, which tends to downwardly bias the predictions (an alternative approach would be to set the mean to the average observed value in the training data for each respective generation). Another potential source of error lies in the kernel choice. Simulated P-element invasions often involve rapid changes, and the RBF kernel we used might not always fully capture these rapid dynamics. Alternative kernels, such as the Matérn kernel [43,73], might be more effective at modeling abrupt shifts in behavior, offering a better fit for this type of system. Additionally, the GP's Bayesian nature offers a way to improve its predictive power. In Bayesian frameworks, predictions are not single values but are accompanied by uncertainty estimates. This uncertainty can guide further sampling in regions where the GP's predictions are less confident, a technique known as active learning [73]. Because the trained GP predicted P-element copy numbers within the empirical range with high accuracy, we did not explore alternative mean function, kernel choices, and active learning strategies in this study.

Parameter Tuning with Gaussian Process Surrogate Models

We used our trained GPs to comprehensively explore simulated P-element dynamics across the whole input domain (Table 1). For each EE experiment (1st and 2nd wave), we predicted P-element copy number time series for 10^6 LHS data points. The GPs allowed for rapid predictions: simulating P-element invasion curves for all 10^6 data points took only 2 seconds on a local machine with an i5-12600K CPU and a GeForce RTX4090 GPU. In comparison, running 10^8 individual-based simulation runs for the same type of analysis would take several days on a computing cluster, depending on available resources.

To identify the parameter values that best explain the observed P-element invasions, we assessed the fit between the predicted and observed P-element copy number time series. For each evolved generation, we computed the RMSE between the GP prediction and the three replicate experimental populations. Next, we normalized the RMSE by dividing by the average observed P-element copy number at the corresponding generation. Finally, we summed the normalized RMSE across all six evolved generations to generate a single error metric between the simulated and empirical P-element invasion curves for each EE experiment. The fit metric across both EE experiments — NRMSE sum — is simply the sum of the NRMSEs for the 1st and 2nd wave experiments. A lower NRMSE value indicates a better fit between the predicted and observed P-element invasion dynamics.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Information regarding data availability and processing for the 1st wave can be found in Kofler *et al*. 2018 and Kofler *et al*. 2022. For the 2nd wave, raw reads of generations 0 and 10 have been previously published (Langmüller *et al*. 2023) and are available from the European Nucleotide Archive (ENA) under project accession number PRJEB54573 (samples SAMEA110271575 – SAMEA110271580). Raw reads of the remaining generations of the 2nd wave experiment are available from ENA under project accession number PRJEB83416. Time-resolved P-element copy number estimates in the experimental populations, our individual-based P-element invasion models implemented in SLiM [41], simulated data, trained GPs, as well as a Jupyter notebook demonstrating the usage of the GPs are available on GitHub at https://github.com/AnnaMariaL/EE-P_element.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was mainly supported by the European Research Council (ERC, ArchAdapt), and the Austrian Science Funds (FWF, W1225). This research was in part supported by the National Science Foundation (NSF PHY-174958), the Gordon and Betty Moore Foundation (Grant No. 2919.02), and the Aarhus University Research Foundation (AIAS-AUFF).

Author's contributions

CS designed and supervised the experimental evolution study. AML was responsible for the modeling aspects of the study. VN was responsible for the molecular work. BCH contributed to the design of the individual-based simulation model. AML and CS wrote the initial manuscript draft. All authors revised the manuscript and approved the final version.

Acknowledgements

We thank all members of the Messer lab for feedback and support. Special thanks to Marlies Dolezal for helpful discussions and to all current and former members of the Institute of Population Genetics, who contributed to the maintenance of the experimental populations over many years.

References

1. Charlesworth B, Charlesworth D. The population dynamics of transposable elements. Genetical Research. 1983;42:1–27.

2. Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica. 1999;107:129.

3. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biology. 2018;19:199.

4. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 2007;41:331–68.

5. McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 1950;36:344–55.

6. Bourgeois Y, Boissinot S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. Genes. 2019;10.

7. Arkhipova IR. Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. Mol Biol Evol. 2018;35:1332–7.

8. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42.

9. Bartolomé C, Maside X, Charlesworth B. On the Abundance and Distribution of Transposable Elements in the Genome of *Drosophila melanogaster*. Molecular Biology and Evolution. 2002;19:926–37.

10. Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. Population Genomics of Transposable Elements in *Drosophila melanogaster*. Molecular Biology and Evolution. 2011;28:1633–44.

11. Langmüller AM, Nolte V, Dolezal M, Schlötterer C. The genomic distribution of transposable elements is driven by spatially variable purifying selection. Nucleic Acids Research. 2023;51:9203–13.

12. Spradling AC, Bellen HJ, Hoskins RA. *Drosophila* P elements preferentially transpose to replication origins. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:15948–53.

13. Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. Nature Reviews Genetics. 2017;18:292–308.

14. Mérel V, Boulesteix M, Fablet M, Vieira C. Transposable elements in *Drosophila*. Mobile DNA. 2020;11:23.

15. Horvath R, Menon M, Stitzer M, Ross-Ibarra J. Controlling for Variable Transposition Rate with an Age-Adjusted Site Frequency Spectrum. Genome Biology and Evolution. 2022;14:evac016.

16. González J, Macpherson JM, Messer PW, Petrov DA. Inferring the strength of selection in *Drosophila* under complex demographic models. Mol Biol Evol. 2009;26:513–26.

17. Kofler R, Betancourt AJ, Schlötterer C. Sequencing of Pooled DNA Samples (Pool-Seq) Uncovers Complex Dynamics of Transposable Element Insertions in *Drosophila melanogaster*. PLOS Genetics. 2012;8:e1002487.

18. Garland T, Rose M. Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments. Oakland: University of California Press; 2009.

19. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. Experimental evolution. Trends in ecology & evolution. 2012;27:547–60.

20. Kofler R, Senti KA, Nolte V, Tobler R, Schlötterer C. Molecular dissection of a natural transposable element invasion. Genome Research. 2018;28:824–35.

21. Wang L, Zhang S, Hadjipanteli S, Saiz L, Nguyen L, Silva E, et al. P-element invasion fuels molecular adaptation in laboratory populations of *Drosophila melanogaster*. Evolution. 2023;77:980–94.

22. Bergthorsson U, Sheeba CJ, Konrad A, Belicard T, Beltran T, Katju V, et al. Long-term experimental evolution reveals purifying selection on piRNA-mediated control of transposable element expression. BMC Biology. 2020;18:162.

23. Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in *Drosophila melanogaster*. PLoS Genetics. 2011;7:e1001336.

24. Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity. 2015;114:431–40.

25. Long A, Liti G, Luptak A, Tenaillon O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature Reviews Genetics. 2015;16:567–82.

26. O'Hare K, Rubin GM. Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell. 1983;34:25–35.

27. Kidwell MG, Kidwell JF, Sved JA. Hybrid Dysgenesis in *Drosophila melanogaster*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics. 1977;86:813–33.

28. Kelleher ES. Reexamining the P-element invasion of *Drosophila melanogaster* through the lens of piRNA silencing. Genetics. 2016;203:1513–31.

29. Kidwell MG. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 1983;80:1655 LP – 1659.

30. Anxolabéhère D, Kidwell MG, Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. Molecular biology and evolution. 1988;5:252–69.

31. Hill T, Schlötterer C, Betancourt AJ. Hybrid Dysgenesis in *Drosophila simulans* Associated with a Rapid Invasion of the P-Element. PLOS Genetics. 2016;12:e1005920.

32. Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. The recent invasion of natural *Drosophila simulans* populations by the P-element. Proceedings of the National Academy of Sciences. 2015;112:6659 LP – 6663.

33. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. Cell. 2007;128:1089–103.

34. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in *Drosophila*. Science. 2007;315:1587–90.

35. Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. Genome Biol. 2006;7:R112.

36. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nature Reviews Genetics. 2019;20:89–108.

37. Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nature Reviews Genetics. 2014;15:749–63.

38. Weilguny L, Kofler R. DeviaTE: Assembly-free analysis and visualization of mobile genetic element composition. Molecular ecology resources. 2019;19:1346–54.

39. Kofler R. Dynamics of transposable element invasions with piRNA clusters. Molecular Biology and Evolution. 2019;36:1457–72.

40. Merenciano M, González J. The Interplay Between Developmental Stage and Environment Underlies the Adaptive Effect of a Natural Transposable Element Insertion. Molecular Biology and Evolution. 2023;40:msad044.

41. Haller BC, Messer PW. SLiM 4: Multispecies Eco-Evolutionary Modeling. The American Naturalist. 2023;201:E127–39.

42. Bellman R. Dynamic Programming. Dover Publications; 1957.

43. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. The MIT Press. The MIT Press; 2005.

44. Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, et al. Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLOS Biology. 2019;17:1–31.

45. Kofler R, Nolte V, Schlötterer C. The Transposition Rate Has Little Influence on the Plateauing Level of the P-element. Molecular Biology and Evolution. 2022;39:msac141.

46. Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. Molecular Ecology. 2019;28:1537–49.

47. Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534:102–5.

48. Rech GE, Radío S, Guirao-Rico S, Aguilera L, Horvath V, Green L, et al. Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. Nat Commun. 2022;13:1948.

49. Battlay P, Leblanc PB, Green L, Garud NR, Schmidt JM, Fournier-Level A, et al. Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the *Drosophila* Genetic Reference Panel. G3 Genes|Genomes|Genetics. 2018;8:3489–97.

50. Aminetzach YT, Macpherson JM, Petrov DA. Pesticide Resistance via Transposition-Mediated Adaptive Gene Truncation in *Drosophila*. Science. 2005;309:764–7.

51. Tomar SS, Hua-Van A, Rouzic AL. A population genetics theory for piRNA-regulated transposable elements. Theoretical Population Biology. 2023;150:1–13.

52. Wierzbicki F, Kofler R. The composition of piRNA clusters in *Drosophila melanogaster* deviates from expectations under the trap model. BMC Biology. 2023;21:224.

53. Gebert D, Neubert LK, Lloyd C, Gui J, Lehmann R, Teixeira FK. Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. Molecular Cell. 2021;81:3965-3978.e5.

54. Scarpa A, Kofler R. The impact of paramutations on the invasion dynamics of transposable elements. Genetics. 2023;225:iyad181.

55. de Vanssay A, Bougé A-L, Boivin A, Hermant C, Teysset L, Delmarre V, et al. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. Nature. 2012;490:112–5.

56. Hollick JB. Paramutation and related phenomena in diverse species. Nature Reviews Genetics. 2017;18:5–23.

57. Kimura M. On the probability of fixation of mutant genes in a population. Genetics. 1962;47:713–9.

58. Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, et al. Recurrent Collection of *Drosophila melanogaster* from Wild African Environments and Genomic Insights into Species History. Molecular Biology and Evolution. 2020;37:627–38.

59. Lange JD, Bastide H, Lack JB, Pool JE. A Population Genomic Assessment of Three Decades of Evolution in a Natural *Drosophila* Population. Molecular Biology and Evolution. 2022;39:msab368.

60. Barghi N, Tobler R, Nolte V, Schlötterer C. *Drosophila simulans* : A Species with Improved Resolution in Evolve and Resequence Studies. G3 Genes|Genomes|Genetics. 2017;7:2337–43.

61. Schwarz F, Wierzbicki F, Senti K-A, Kofler R. Tirant Stealthily Invaded Natural *Drosophila melanogaster* Populations during the Last Century. Molecular Biology and Evolution. 2020;

62. Gómez-Sánchez D, Schlötterer C. ReadTools: A universal toolkit for handling sequence data from different sequencing platforms. Molecular Ecology Resources. 2018;18:676–80.

63. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS computational biology. 2005;1:166–75.

64. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.

65. Bosco G, Campbell P, Leiva-Neto JT, Markow TA. Analysis of *Drosophila* Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species. Genetics. 2007;177:1277–90.

66. Howie JM, Mazzucco R, Taus T, Nolte V, Schlötterer C. DNA Motifs Are Not General Predictors of Recombination in Two *Drosophila* Sister Species. Genome Biology and Evolution. 2019;11:1345–57.

67. Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J. High-frequency P element loss in *Drosophila* is homolog dependent. Cell. 1990;62:515–25.

68. Jerome Sacks, William J. Welch, Toby J. Mitchell, Henry P. Wynn. Design and Analysis of Computer Experiments. Statistical Science. 1989;4:409–23.

69. Krige DG. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa. 1951;52:119–39.

70. Gardner JR, Pleiss G, Bindel D, Weinberger KQ, Wilson AG. Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. Advances in Neural Information Processing Systems. 2018;12:7576–86.

71. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2019. p. 8024–35.

72. Bonilla EV, Chai K, Williams C. Multi-task Gaussian Process Prediction. In: Platt J, Koller D, Singer Y, Roweis S, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2007.

73. Nguyen Q. Bayesian Optimization in Action. Shelter Island: Manning Publications; 2023.

Supplementary Information

Supplementary Tables

Table S1. Summary statistics describing the 1 ("best fit") versus 100 ("top 100") parameter combination(s) with the lowest NRMSE sum (NRMSE sum = NRMSE 1st wave + NRMSE 2nd wave) from the parameter exploration with the Gaussian Process model, grouped by average expected selection efficacy (N_e = 442). Values in square brackets represent the range between the 2.5th and 97.5th percentiles of individual metrics. $|\bar{s}|$ = mean selection coefficient outside piRNA clusters ($\mathrm{E}(|s|) = \alpha/(\alpha+\beta)$); 95th percentile $|s|$ = 95th percentile of $|s|$ outside piRNA clusters; Neutral fraction = expected fraction of effectively neutral ($|s| \leq 1/N_e$) P-element insertions outside of piRNA clusters; $|\bar{s}_{\mathrm{effective}}|$ = mean selection coefficient for P-element insertions outside piRNA clusters where selection is strong enough to overcome genetic drift ($|s| > 1/N_e$); NRMSE = Normalized root mean square error

	Best fit		**Top 100**	
	$\frac{\mathrm{E}(\lvert s\rvert)}{2} \leq \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} > \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} \leq \frac{1}{2N_e}$	$\frac{\mathrm{E}(\lvert s\rvert)}{2} > \frac{1}{2N_e}$
Parameter	"neutral"	"selection"	"neutral"	"selection"
p_{carrier}	0.151	0.204	0.274	0.192
			[0.155, 0.476]	[0.158, 0.237]
$f_{\mathrm{regulatory}}$	0.098	0.017	0.098	0.021
			[0.093, 0.100]	[0.015, 0.027]
u	0.343	0.290	0.338	0.288
			[0.314, 0.365]	[0.258, 0.317]
α	0.039	0.461	0.027	0.446
			[0.012, 0.042]	[0.362, 0.500]
β	17.379	10.743	14.701	11.676
			[10.272, 19.687]	[10.216, 14.764]
$\lvert\bar{s}\rvert$	0.002	0.041	0.002	0.037
			[0.001, 0.002]	[0.031, 0.045]
95th percentile $\lvert s\rvert$	0.011	0.159	0.007	0.145
			[0.001, 0.011]	[0.121, 0.171]
Neutral fraction	0.898	0.201	0.924	0.219
			[0.894, 0.964]	[0.174, 0.286]
$\lvert\bar{s}_{\mathrm{effective}}\rvert$	0.021	0.051	0.023	0.047
			[0.019, 0.030]	[0.040, 0.054]
NRMSE 1st + 2nd wave	2.313	1.340	2.371	1.462
			[2.328, 2.387]	[1.361, 1.507]

Supplementary Figures

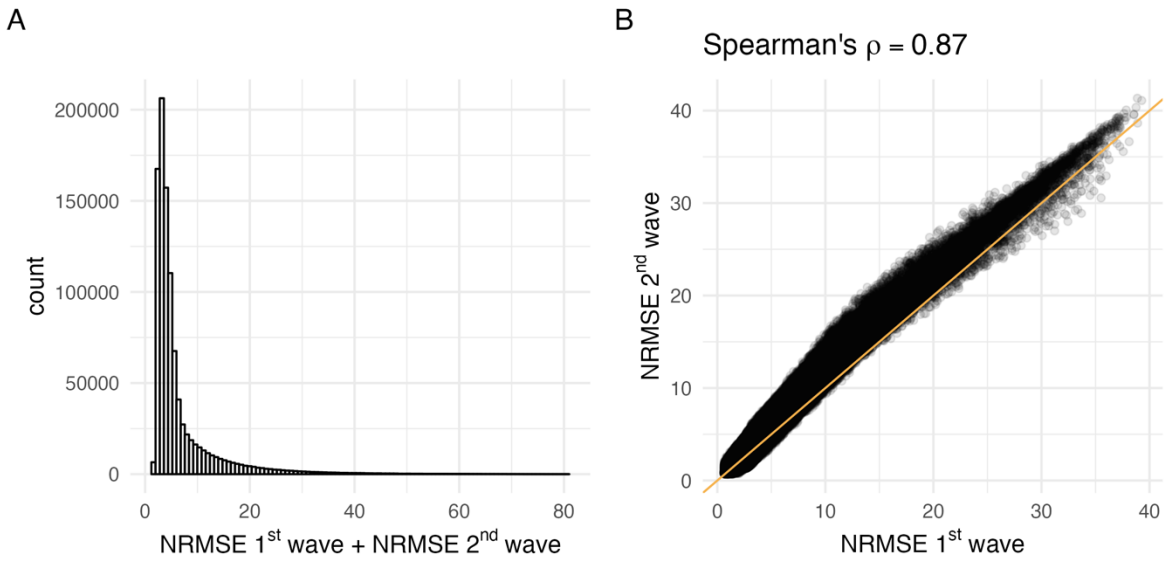


Figure S1. Normalized root mean square error (NRMSE) for GP predictions using 10^6 different parameter combinations. (A) Distribution of NRMSE sums, showing variation in the fits between GP predictions and empirical data. While most values are moderate, a few combinations have very low or very high NRMSE values. (B) Comparison of NRMSE for the 1st wave (*x* axis) versus the 2nd wave (*y* axis). Each point represents the NRMSEs for one of the 10^6 data points. The amber line represents the identity line ($x = y$). Overall, there is strong agreement between the NRMSEs, with a Spearman rank correlation coefficient of 0.87.

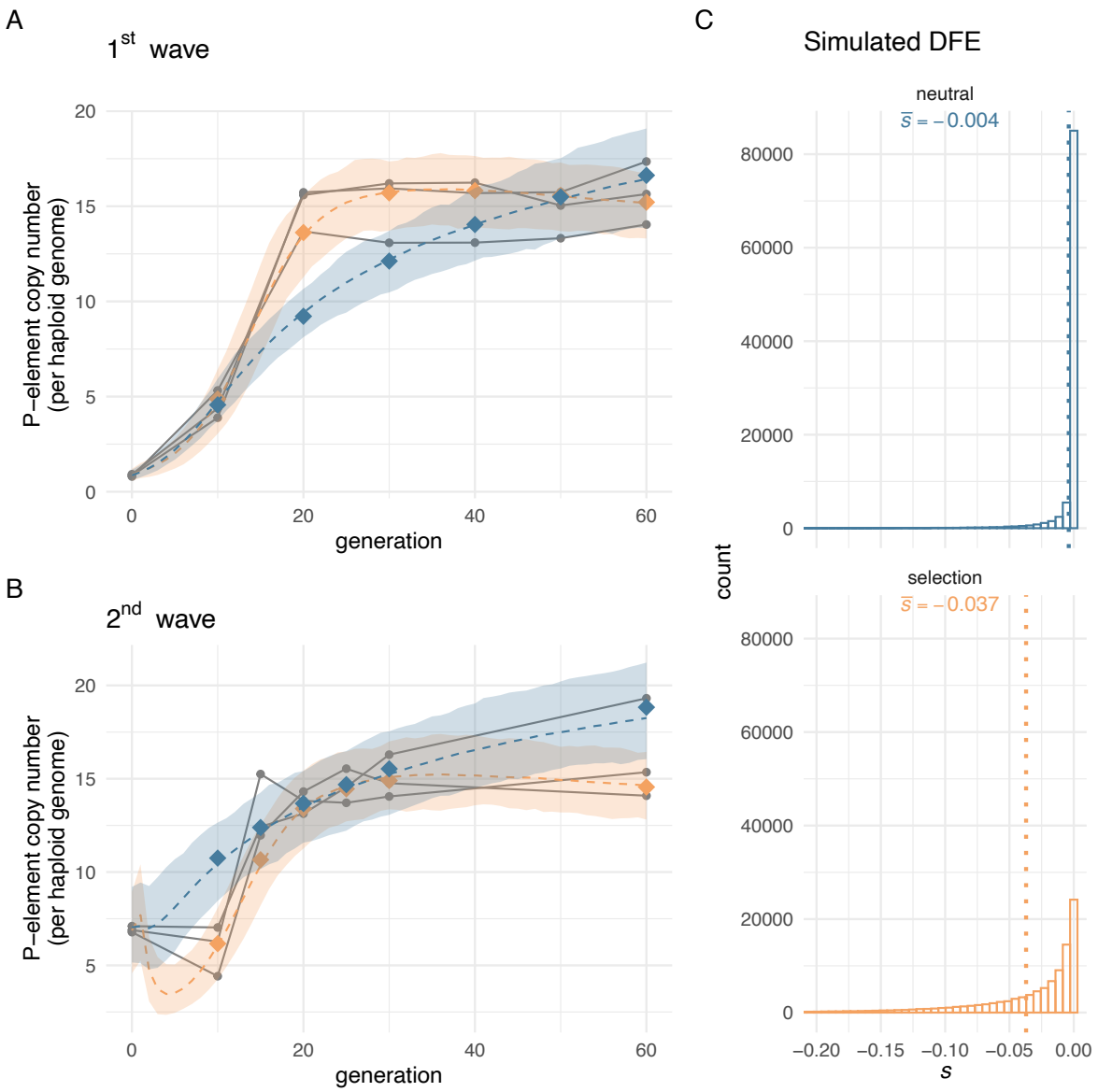


Figure S2. Gaussian Process (GP) predictions provide a good fit to the empirical data — parameters estimated jointly from 1st and 2nd wave data (average parameter values from the 100 combinations with the lowest NRMSE sums). GP predictions from the parameter combinations with the lowest NRMSE sum when compared against empirical data are shown for scenarios without expected effective purifying selection ($E(|s|/2) \leq 1/(2N_e)$, steel blue) and scenarios with expected effective purifying selection ($E(|s|/2) > 1/(2N_e)$, orange), $N_e = 221$ [45]. Each grey line represents an empirical evolution replicate, with sequenced time points indicated by dots. GP predictions are marked by diamonds, while the dashed colored lines represent the average P-element copy number per haploid genome across 100 simulation runs with the individual-based model. The colored ribbons show the range between the 2.5th and 97.5th percentiles of P-element copy number trajectories simulated with the individual-based model. (A) 1st wave experiment. (B) 2nd wave experiment. (C) Simulated distribution of fitness effects (DFE) used in (A) and (B).