Coexistence of piRNA and KZFP Defense Systems: Evolutionary Dynamics of Layered Defense against Transposable Elements

Yusuke Nabeka* and Hideki Innan*

*SOKENDAI, Research Center for Integrative Evolutionary Science, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan

*Corresponding author: SOKENDAI, Research Center for Integrative Evolutionary Science, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan. Email: nabeka_yusuke@soken.ac.jp (Y.N.); innan_hideki@soken.ac.jp (H.I.)

Abstract

Transposable elements (TEs) pose a persistent threat to genome stability, and host organisms have consequently evolved sophisticated defense mechanisms to restrain them. In animals, the most prominent systems include PIWI-interacting RNAs (piRNAs) and Krüppel-associated box zinc finger proteins (KZFPs). Because both systems recognize TEs in a sequence-specific manner and induce epigenetic silencing, they appear functionally redundant at first glance. However, KZFPs are a relatively recent innovation that emerged and diversified in genomes where the piRNA pathway was already established. This raises an important question: under what conditions can a second, seemingly redundant defense system invade and persist? To address this, we constructed a mathematical model integrating the evolutionary dynamics of TEs, piRNAs, and KZFPs. Our approach focuses on a key mechanistic asymmetry between the two defense systems: whereas piRNA-mediated suppression is dependent on TE activity, KZFPs provide constitutive suppression that does not rely on ongoing TE activity. We show that these distinct modes of action generate interactions that extend beyond simple redundancy or additivity. We derive analytical conditions under which KZFPs can invade a pre-existing TE-piRNA equilibrium and characterize the evolutionary logic that enables stable coexistence of these multilayered defense strategies. Together, our results provide a theoretical framework for understanding how complex, layered genome defense systems can evolve and persist.

Keywords: transposable elements; genome defense; piRNA pathway; KRAB zinc finger proteins; population genetics; mathematical models

Introduction

Transposable elements (TEs) are mobile genetic elements capable of moving and replicating within host genomes, posing a significant threat to genomic stability (Kazazian 2004). Uncontrolled proliferation of TEs can lead to gene disruption and chromosomal rearrangements, which, in extreme cases, may result in the extinction of entire host populations. Consequently, hosts have evolved multiple defense mechanisms to suppress TE activity.

Among these suppression mechanisms, epigenetic regulation that silences TE transcription is a prominent strategy. Two well-characterized examples are the PIWI-interacting RNA (piRNA) pathway and Krüppel-associated box (KRAB) zinc finger proteins (KZFPs) (Lawlor and Ellison 2023). piRNAs are small RNAs present in most eukaryotes and are derived from specific genomic loci known as piRNA clusters (Ozata *et al.* 2019). When a TE inserts into a piRNA cluster, the host generates piRNAs complementary to the TE sequence. These piRNAs bind to PIWI clade proteins, targeting TEs through sequence complementarity. The piRNA/PIWI complex recruits histone methyltransferases (HMTs) and DNA methyltransferases (DNMTs), which catalyze histone methylation (H3K9me3) and DNA methylation. These epigenetic modifications induce heterochromatin formation around the TE insertion site, effectively suppressing TE transcription (Brennecke *et al.* 2007; Halic and Moazed 2009; Cosby *et al.* 2019; Iwasaki *et al.* 2025).

In contrast, KZFPs represent a more recent evolutionary innovation, found in coelacanths, lungfish, and tetrapods, originating approximately 420 million years ago (Huntley *et al.* 2006; Tadepally *et al.* 2008; Emerson and Thomas 2009; Iwasaki *et al.* 2025). KZFPs consist of a DNA-binding zinc finger array and a KRAB domain. The zinc fingers recognize specific DNA triplets, enabling sequence-specific binding to TEs, while the KRAB domain recruits TRIM28 (also known as KAP1), DNMTs, and other chromatin modifiers to establish repressive heterochromatin around the targeted TE locus, robustly suppressing its transcription (Emerson and Thomas 2009; Levin and Moran 2011; Ecco *et al.* 2017; Yang *et al.* 2017; Imbeault *et al.* 2017; Cosby *et al.* 2019; Almeida *et al.* 2022; Rosspopoff and Trono 2023; de Tribolet-Hardy *et al.* 2023; Iwasaki *et al.* 2025).

While both piRNAs and KZFPs suppress TE activity through sequence-specific recognition and epigenetic modifications, it remains unclear why KZFPs evolved in the presence of the already effective piRNA pathway. One might hypothesize that KZFPs operate more efficiently than the piRNA pathway; however, this does not fully explain their coexistence, as less effective systems would typically be purged from the genome over evolutionary time. Conversely, if their suppressive effects were merely additive, one might expect hosts to accumulate as many defense mechanisms as possible. The reality—that these two systems coexist, while new, similar systems have not arisen indefinitely—presents an intriguing question.

A clue to resolving this question may lie in the different mechanisms upon which their functions depend. The piRNA system requires TE transposition, as piRNAs are generated from TE insertions within piRNA clusters (Brennecke *et al.* 2007; Kofler 2019; Tomar *et al.* 2023; Srivastav *et al.* 2024). Although it relies on stochastic insertion events, it provides complete suppression once established. In contrast, KZFPs can evolve recognition specificity over longer timescales by accumulating mutations in their DNA-binding zinc-finger arrays (Nardelli *et al.* 1992; Looman *et al.* 2002; Ecco *et al.* 2017; Najafabadi *et al.* 2017; Zhang *et al.* 2022; Wells *et al.* 2023). Once evolved, their persistence is determined by a balance between their own suppressive efficacy

and cost, independent of the TE activity level. In other words, a newly emerged KZFP is less specific at first, and their silencing effect may be modest until the binding array refines. It is these asymmetries that may give rise to interactions between the two systems that go beyond simple exclusion or additivity.

We thus hypothesize that these mechanistic differences in TE dependence and suppressive specificity are the key to the coexistence of the two TE suppression mechanisms: the two distinct TE suppression systems are neither mutually exclusive nor function in a simple additive manner, but rather interact with each other. To test this hypothesis, we construct a mathematical model that describes the dynamical interactions among TEs, piRNAs, and KZFPs, reflecting their different modes of suppression systems. Through the analysis of this model, we elucidate the evolutionary logic of multilayered genome defense strategies by clarifying (1) the conditions under which KZFPs can invade and become established in an existing TE-piRNA equilibrium, (2) the conditions under which both systems can stably coexist, and (3) the consequences of their coexistence for the host's TE copy number and fitness.

Model

Our objective in this study is to explain why seemingly redundant TE suppression mechanisms are maintained within genomes and to determine the evolutionary conditions for their emergence and coexistence. To this end, we have constructed a mathematical model that integrates the evolutionary dynamics of TE copy number, piRNA-producing alleles, and KZFP alleles. This model is based on the "trap model" of piRNA-mediated suppression described by Tomar et al. (2023), which is extended in this work by explicitly incorporating the evolutionary dynamics of KZFP-mediated TE suppression (Tomar *et al.* 2023).

We consider a deterministic model of an infinite, diploid population where genetic drift can be ignored. We also assume that the population is always in Hardy-Weinberg equilibrium. The genome of each individual consists of a pair of chromosomes, each with an infinite number of potential TE insertion sites, a single piRNA cluster locus, and a single KZFP locus. For simplicity, we assume a single TE family whose sequence does not mutate and no linkage disequilibrium between TE copies. Under these assumptions, we construct the system of three coupled ordinary differential equations, describing (i) the change in TE copy number, (ii) piRNA allele frequency and (iii) KZFP allele frequency. We define p and x as the frequency of TE suppression alleles for piRNA and KZFP, respectively. We assume the piRNA locus has two allelic states: a piRNA-producing allele that fully suppresses TEs, and a null allele with no suppressive function. On the other hand, the KZFP loci involves multiple alleles. We assume KZFP allele i with frequency x_i has a suppression efficacy which ranges from 0 to 1, denoted by f_i, and $\bar{f}$ denotes the its average suppressive effect over all individuals in the population. The detailed derivations for (i), (ii), and (iii) are provided as follows.

Change in TE Copy Number

First, we derive an equation for the dynamics of the mean TE copy number, which is governed by the balance between transpositional gain and loss due to selection. We define n as the number of TE copies located outside of the piRNA clusters, and those within the piRNA clusters are ignored because their direct contribution to the total TE load is negligible (the piRNA clusters occupy only $\approx 0.37\%$ of the human genome, (Rosenkranz *et al.* 2022)). Since new transpositions arise from existing TE copies, the potential rate of transposition is proportional to the current copy number n, scaled by the basal transposition rate u, giving a total rate of $u \cdot n$. This transposition is suppressed by both piRNA and KZFP systems and we assume their effects are multiplicative. The fraction of individuals permissive to piRNA-mediated suppression (i.e., homozygous for non-piRNA alleles) is $(1-p)^2$, assuming Hardy-Weinberg equilibrium and dominance of the piRNA-producing allele. The average suppression effect of KZFPs is $\bar{f}$, so the fraction of events escaping KZFP suppression is $(1-\bar{f})$. Then, the effective transposition rate of TE is $u(1-\bar{f})(1-p)^2 n$. The derivations of p and $\bar{f}$ are detailed in the following sections.

For the decrease of n via selection, we assume that each TE copy, regardless of its genomic location, reduces the host fitness by a selection coefficient of s. We assume that the fitness of an individual carrying n TE copies is given by $W(n) = e^{-s \cdot n}$ and the corresponding Malthusian fitness is $m(n) = \log W(n) = -s \cdot n$. Treating the TE copy number (n) as a quantitative trait, the change of n due to selection in continuous time can be approximated by $\Delta n \approx \mathrm{Var}(n)\partial m(n)/\partial n$. Assuming random mating with no linkage disequilibrium, the TE copy number is approximately Poisson-distributed, so $\mathrm{Var}(n) \approx n$. Because $\partial m(n)/\partial n = -s$, this yields a linear reduction $-s \cdot n$, which is a standard approximation in quantitative genetic models of TE copy-number dynamics (Charlesworth and Charlesworth 1983). Combining these gives the full expression of the equation for the change in TE copy number:

$$\frac{dn}{dt} = u(1-\bar{f})(1-p)^2 n - s \cdot n. \tag{1}$$

Change in piRNA Allele Frequency

We consider the dynamics of the piRNA allele frequency, p, which is determined by its generation via TE insertions and its loss due to selection. We assume the piRNA cluster locus has in two allelic states: a non-functional allele (frequency $1-p$) and a functional, piRNA-producing allele (frequency p). We also assume that the functional allele is dominant with respect to its suppressive effect. The change in the frequency of the piRNA-producing allele (p) is given by the difference between its gain from new TE insertions and its loss due to selection, so that $\frac{dp}{dt}$ consists of a "gain" term and a "selection" term.

The gain term is derived from the rate at which non-functional, null alleles are converted into functional, piRNA-producing ones. This process requires two conditions: a TE transposes, and it lands within the piRNA cluster. Here, we define π as the fraction of the genome occupied by piRNA clusters. As described above, the effective transposition rate of TE is $u(1-\bar{f})(1-p)^2 n$. A fraction π of these transposition events will occur within the piRNA cluster, potentially creating a new functional, piRNA-producing allele. Therefore, the rate of these insertion events is $\pi \cdot u(1-\bar{f})(1-p)^2 n$. To translate this rate of allele creation into a rate of frequency change in a diploid population, we must account for the total number of alleles at this locus in the population (two per individual). This introduces a scaling factor of $1/2$. Thus, the rate of increase in the frequency p, or the gain term, is given by $\frac{\pi}{2} u(1-\bar{f})(1-p)^2 n$.

The derivation of the selection term is based on the standard model of selection against deleterious alleles with additive effects. Here, we explicitly assume that the piRNA-producing allele itself contains one deleterious TE insertion within the piRNA cluster, thus incurring a fitness cost equivalent to carrying one

TE copy (with selection coefficient s). This cost is accounted for separately from the main TE load described in the dn/dt equation (1). We define the functional piRNA-producing allele as '+' and the null allele, which has no associated cost, as '−'. Assuming additive fitness effects, the genotype fitnesses are

$$W_{--} = 1, \quad W_{+-} = 1 - s \quad \text{and} \quad W_{++} = 1 - 2s,$$

from which, we have the mean population fitness:

$$\bar{W} = (1-p)^2 + 2p(1-p)(1-s) + p^2(1-2s) = 1 - 2s \cdot p.$$

Following selection, the frequency of the + allele in the next generation, p', is

$$p' = \frac{p^2(1-2s) + p(1-p)(1-s)}{1-2s\cdot p} = \frac{p - s\cdot p - s\cdot p^2}{1-2s\cdot p}.$$

The change in allele frequency over a single generation, Δp, is therefore the difference between the post-selection and pre-selection frequencies:

$$\Delta p = p' - p = \frac{-s\cdot p(1-p)}{1-2s\cdot p}.$$

In continuous time, the rate of change $\frac{dp}{dt}$ is approximated by the change per generation, Δp. Thus, with this selection term, $\frac{dp}{dt}$ becomes

$$\frac{dp}{dt} = \frac{\pi}{2}u(1-\bar{f})(1-p)^2 n - \frac{s\cdot p(1-p)}{1-2s\cdot p}. \tag{2}$$

Change in KZFP Allele Frequency

We derive the dynamics of the KZFP suppressor allele frequency by focusing on its growth rate due to the host's fitness increase. KZFPs are proteins that exert their suppressive effect by binding to TE sequences via a DNA recognition motif. Unlike the piRNA system, this mechanism does not require a TE to insert into a specific locus to trigger a response. Instead, we assume that the suppression efficacy is a function of the sequence similarity between the KZFP's recognition motif and the target TE copy. We therefore consider a model that allows a finite number of KZFP alleles. Allele i has its suppression efficacy, f_i, ranging from 0 to 1, and its frequency, x_i. For simplicity in this study, we do not specifically model of the sequence the KZFP's recognition motif, rather we treat the efficacy f_i itself as a trait value. An allele with zero efficacy ($f_i = 0$) is referred to as a null allele. The change in frequency of allele i over time is determined by natural selection and mutation:

$$\frac{dx_i}{dt} = x_i(\bar{W}_i - \bar{W}) + \sum_{j\neq i}\mu_{ji}\cdot x_j - \sum_{k\neq i}\mu_{ik}\cdot x_i, \tag{3}$$

where $\sum_i x_i = 1$. $\bar{W}_i$ is the mean fitness of individuals carrying allele i, and $\bar{W}(=\sum_j x_j W_j)$ is the population mean fitness. μ_{ji} is the mutation rate from allele j to allele i per generation. Assuming dominance, the suppression efficacy of a diploid individual with genotype (i,j) is given by $\max(f_i, f_j)$. Under Hardy-Weinberg equilibrium, the average suppression efficacy across the population is:

$$\bar{f} = \sum_i\sum_j x_i\cdot x_j \max(f_i, f_j). \tag{4}$$

For analytical tractability, we reduce this finite-allele model to a simple two-allele model. This system consists of only two kinds of alleles: One is a functional suppressor allele that has a suppression efficacy $f_1(>0)$ and a frequency of x_1, and the other is a non-functional null allele, which has no suppressive function $f_0(=0)$ and a frequency of $1-x_1$. We assume that the suppressor allele is dominant over the null allele with respect to its suppressive function. We hereafter ignore mutation because we focus on a very short period immediately after the invasion of KZFP suppressor allele, where the effect of mutation is very small (i.e., $\mu_{ji} \approx 0$). Then, $\frac{dx_1}{dt}$ and $\bar{f}$ under these simplified assumptions can be obtained from Equations (3)-(4):

$$\frac{dx_1}{dt} = x_1(W_1 - \bar{W}) \tag{5}$$

and

$$\begin{aligned}\bar{f} &= x_1^2\cdot f_1 + 2x_1(1-x_1)\cdot f_1 + (1-x_1)^2\cdot 0\\ &= (x_1^2 + 2x_1 - 2x_1^2)f_1\\ &= x_1(2-x_1)f_1.\end{aligned} \tag{6}$$

The growth rate of the suppressor allele, denoted by σ, is given by the benefit of TE suppression minus the maintenance cost. The benefit arises from the reduction in the number of new deleterious TE insertions. This can be quantified by comparing the rate of transposition experienced by individuals with and without the suppressor allele. For individuals without the suppressor allele, the rate of new TE insertions is $n\cdot u(1-p)^2$. For individuals carrying at least one suppressor allele, the transposition rate is reduced by a factor of f_1, resulting in a rate of $n\cdot u(1-f_1)(1-p)^2$. The difference between these two rates represents the number of TE insertions prevented per generation by carrying the suppressor allele:

$$\underbrace{\left[n\cdot u(1-p)^2\right]}_{\text{without suppressor allele}} - \underbrace{\left[n\cdot u(1-f_1)(1-p)^2\right]}_{\text{with suppressor allele}} = n\cdot u\cdot f_1(1-p)^2.$$

Since each TE copy carries a fitness cost of s, the total benefit conferred by the suppressor allele is this reduction in TE copy numbers multiplied by s. We also assume a constant maintenance cost of c for carrying the suppressor allele, which is independent of its suppression efficacy (the null allele is assumed to have no cost). Therefore, the growth rate of suppressor allele, σ, is given by

$$\sigma = s\cdot n\cdot u\cdot f_1(1-p)^2 - c. \tag{7}$$

Under the standard population genetic model for a dominant advantageous allele, the population mean fitness is given by a function of σ:

$$\begin{aligned}\bar{W} &= x_1^2\cdot W_{11} + 2x_1(1-x_1)W_{01} + (1-x_1)^2 W_{00}\\ &= (x_1^2 + 2x_1 - 2x_1^2)(1+\sigma) + (1-x_1)^2\\ &= 1 + (2x_1 - x_1^2)\sigma.\end{aligned} \tag{8}$$

The average fitness of individuals carrying at least one suppressor allele, W_1, is written as

$$W_1 = x_1\cdot W_{11} + (1-x_1)W_{01} = 1 + \sigma. \tag{9}$$

Then, substituting $\bar{W}$ and W_1 into Equation (5), the change in KZFP allele frequency becomes

$$\begin{aligned}\frac{dx_1}{dt} &= x_1(W_1 - \bar{W}) = x_1\{\left[1 + \sigma - 1 + (2x_1 - x_1^2)\sigma\right]\}\\ &= \sigma x_1(1-x_1)^2.\end{aligned} \tag{10}$$

The three differential equations (1), (2) and (10) provide the entire theoretical basis used in the following analyses.

Results

In this article, we analyze the long-term outcomes after a second defense mechanism, the KZFP system, is introduced into a stable TE-piRNA equilibrium, which we refer to as the "pre-invasion" state. Our analysis proceeds as follows: First, we derive the analytical conditions that determine whether a rare KZFP allele can invade in this pre-invasion state. We then characterize the two possible long-term outcomes of a successful invasion: either the KZFP allele spreads to fixation in the population (see below for definition) or it is maintained at a polymorphic equilibrium.

Conditions for KZFP Invasion

This work considers the invasion of a KZFP allele into a stable system composed only of TEs and piRNAs. To do so, we need to first investigate the conditions for the TE-piRNA pre-invasion equilibrium before the KZFP invades. The TE-piRNA pre-invasion equilibrium can be explored by setting the KZFP frequency to zero ($x_1 = 0$) to reduces the three-variable model to a two-variable dynamical system with TE copy number (n) and piRNA allele frequency (p):

$$\frac{dn}{dt} = u(1-p)^2 n - s \cdot n, \tag{11}$$

$$\frac{dp}{dt} = \frac{\pi}{2} u(1-p)^2 n - \frac{s \cdot p(1-p)}{1-2s \cdot p}. \tag{12}$$

A nontrivial equilibrium (n^*, p^*) of this system is given analytically by solving $\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$:

$$p^* = 1 - \sqrt{s/u}, \tag{13}$$

$$n^* = \frac{2s \cdot p^*}{\pi\sqrt{s \cdot u}(1-2s \cdot p^*)}. \tag{14}$$

This equilibrium requires $s < u$ for both TEs and piRNAs to persist. At this equilibrium, p^* is determined solely by the ratio s/u, whereas n^* additionally depends on π.

Given this equilibrium, we are interested in whether the KZFP system can invade into the population. In practice, in our deterministic treatment, we investigate the growth rate of KZFP suppressor allele in frequency when it is very rare (i.e, $x_1 \approx 0$). From Equation (10), $\frac{dx_1}{dt}$ at $x_1 \approx 0$ is approximately given by

$$\left.\frac{dx_1}{dt}\right|_{x_1 \to 0} \approx \sigma^* x_1, \tag{15}$$

where the initial growth rate (σ^*) is

$$\sigma^* = s \cdot n^* \cdot u \cdot f_1 (1-p^*)^2 - c. \tag{16}$$

For a KZFP to successfully invade, the initial growth rate σ^* must be positive, meaning the benefit from TE suppression must exceed its maintenance cost. By substituting the pre-invasion equilibrium values of n^* and p^* by Equations (13) and (14) into Equation (16), we have the invasion condition:

$$\frac{2s^3(1-\sqrt{s/u})f_1}{\pi\sqrt{s \cdot u}[1-2s(1-\sqrt{s/u})]} > c. \tag{17}$$

To characterize the system's representative dynamics, Figure 1A shows the results for four parameter sets, where $u = 0.01$ and $\pi = 0.01$ are fixed and the selection parameter s is varied: $s = 5 \times 10^{-3}, 2.5 \times 10^{-3}, 1 \times 10^{-3}$ and 7.5×10^{-4}. We define the boundary indicated by the red curve as the invasion threshold, which is analytically given by solving $\sigma^* = 0$. Equation (17) indicates that the invasion of KZFP is possible in the region below the invasion threshold in the suppression efficacy (f_1) and maintenance cost (c) parameter plane (logarithmic scale for c). As shown in Figure 2A, increasing s shifts the invasion threshold upward, unless s is unrealistically large. This means that a large s allows KZFPs to invade even when their maintenance cost c is high. This is because TE with a large s imposes a strong deleterious effect on the host fitness so that suppressing TEs is beneficial, thereby making such KZFPs easy to invade.

The effects of the other two parameters, u and π, are also investigated in Figure 2B and C. Figure 2B shows that a large u value allows only KZFP with low cost to invade, except when u is very small. This can be explained as follows: when transposition is highly active (i.e., u is large), the piRNA pathway is already strongly engaged in TE suppression, resulting in a high p^* and low n^* (Equations (13)-(14)). In this state, additional benefit of KZFP invasion is limited. As a result, KZFP invasion is only possible when the cost c is sufficiently low. An exception is when u is very small, as indicated by the purple dashed line in Figure 2B. In such a case, transposition activity is very low and TEs are already rare, making the benefit of KZFP is small. As a result, only KZFPs with very low cost can invade. Likewise, Figure 2C shows that large π values also restrict the invasion to KZFPs with low cost. This is because a large piRNA cluster enhances the efficiency of the piRNA pathway, leading to a reduction in n^* (Equation (14)), which diminishes the benefit of KZFP.

Characterization of Possible Long-Term Outcomes after KZFP Invasion

To understand possible long-term outcomes, we here consider the equilibrium after successful invasion . This equilibrium is referred to as the "post-invasion equilibrium". Let $(n_{\mathrm{eq}}, p_{\mathrm{eq}}, x_{1\mathrm{eq}})$ denote (n, p, x_1) in the post-invasion equilibrium, which can be obtained by solving Equations (1),(2) and (10).

Let us focus on $x_{1\mathrm{eq}}$, the equilibrium frequency of the KZFP allele long after the invasion, which is visualized in the (f_1, c) parameter-space in Figure 1A. It is clearly shown that $x_{1\mathrm{eq}}$ is almost 1 in the majority region in yellow. We define this parametric region as the "KZFP-fixation zone". It should be noted that, in our deterministic model, the KZFP allele cannot mathematically reach fixation in the population; however, we consider it effectively fixed when there is a consistent selective pressure driving its frequency toward 1 across all frequencies.

Figure 1A also shows that $x_{1\mathrm{eq}}$ has an intermediate value in the region in green. We define this region as the "KZFP-polymorphic zone". To explore the boundary between these two zones, we again focus on, σ , the growth rate of the KZFP suppressor allele in frequency, which should be given by a function of x_1. As defined, within the KZFP-fixation zone, the growth rate should be positive for any value of x_1 between 0 and 1, so that the KZFP suppressor allele increases to 1. In contrast, this should not hold in the KZFP-polymorphic zone, resulting in an equilibrium KZFP frequency $x_{1\mathrm{eq}}$ less than 1. Detailed derivation of this boundary is as follows.

To obtain the growth rate of KZFP as a function of x_1, we use a quasi-equilibrium approximation in which it is assumed that the variables n and p immediately reach their quasi-equilibrium value, denoted by $(n_{\mathrm{qs}}, p_{\mathrm{qs}})$. This is justified by a clear time-scale separation between the fast (n, p) dynamics and the slow x_1 dynamics (see Appendix B). This reduces the 3D system to an effective 1D dynamics in x_1. Under this assumption, $\frac{dx_1}{dt}$ is

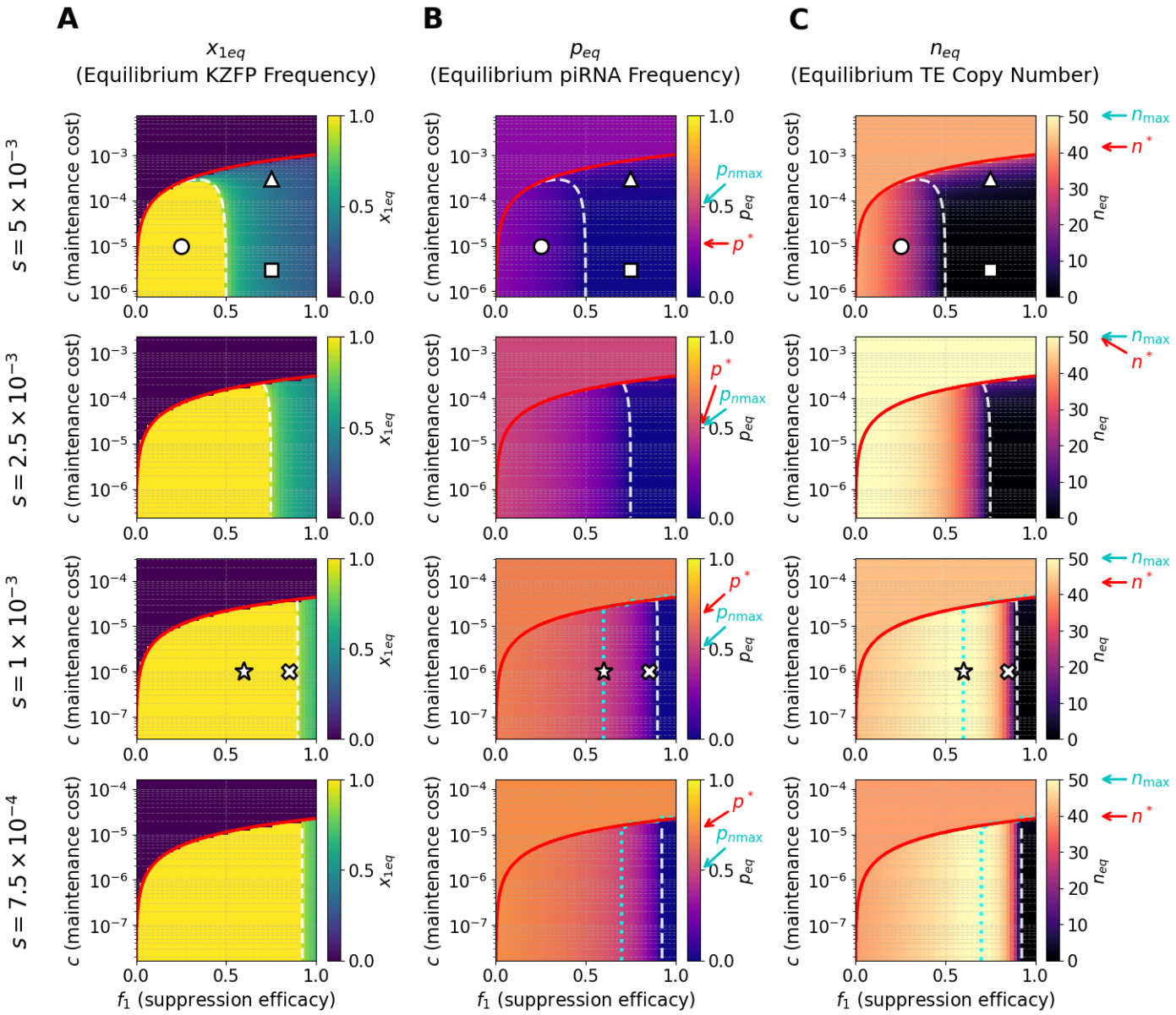

Figure 1 Post-invasion equilibrium states in the (f_1, c) parameter space. (A) Equilibrium frequency of the KZFP allele ($x_{1\mathrm{eq}}$), (B) equilibrium frequency of the piRNA-producing allele (p_{eq}), and (C) equilibrium TE copy number (n_{eq}) after KZFP invasion plotted in the parameter space of suppression efficacy of KZFP (f_1) and maintenance cost of KZFP (c). Four different values of the selection coefficient s are used, while $u = 0.01$ and $\pi = 0.01$ are fixed. The post-invasion equilibrium states are numerically computed by using the `solve_ivp` ODE solver in the `scipy` package (Virtanen *et al.* 2020). The color bars are to present the value of focal quantities in the post-invasion equilibrium. The red curve shows the invasion threshold from Equation (17), below which KZFP invasion is possible. The white dashed curve shows the upper boundary of the KZFP-fixation zone from Equation (27). The cyan dotted line in the lower two panels of (B) and (C) marks (f_1, c) at which the piRNA frequency ($p_{n\max}$) maximizes the equilibrium TE copy number, $n = n_{\max}$ (see Equation (35)). The cyan arrows on the color bars indicate $p_{n\max}$ and $n_{\max}$, together with the pre-invasion equilibria p^* and n^* with red arrows. Symbols mark the parameter sets whose dynamics are shown in Figure 5: circle–Figure 5A ($s = 5 \times 10^{-3}$, $f_1 = 0.25$, $c = 1 \times 10^{-5}$); triangle–Figure 5B ($s = 5 \times 10^{-3}$, $f_1 = 0.75$, $c = 3 \times 10^{-4}$); square–Figure 5C ($s = 5 \times 10^{-3}$, $f_1 = 0.75$, $c = 3 \times 10^{-6}$); cross–Figure 5D ($s = 1 \times 10^{-3}$, $f_1 = 0.85$, $c = 1 \times 10^{-6}$); star–Figure 5E ($s = 1 \times 10^{-3}$, $f_1 = 0.6$, $c = 1 \times 10^{-6}$).

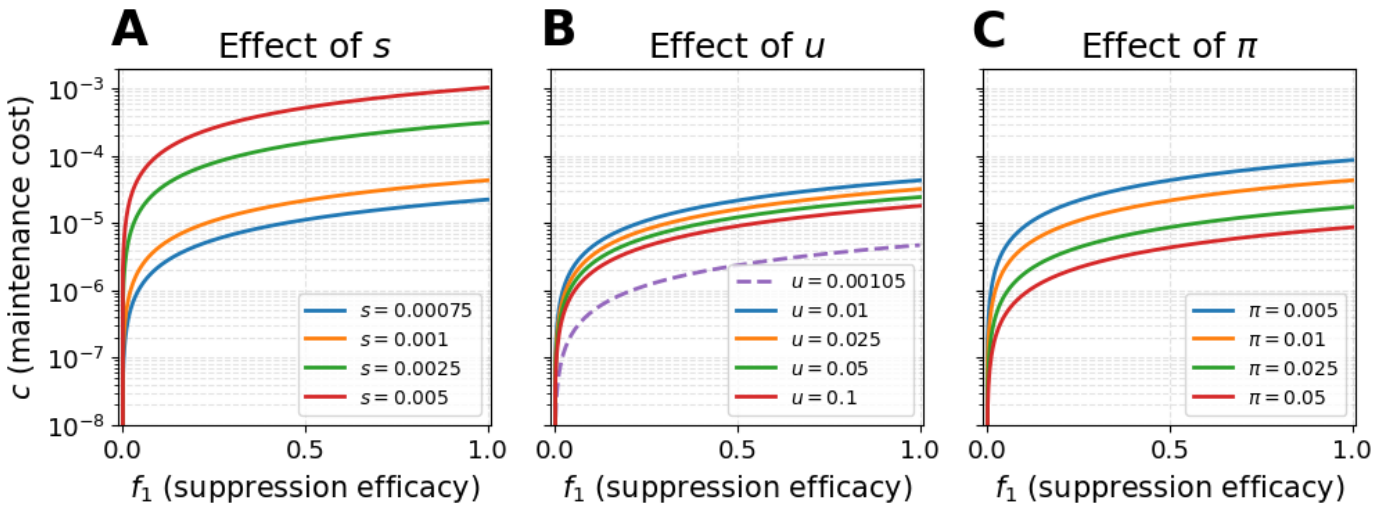

Figure 2 Effect of the parameters on the KZFP invasion boundary. Each curve represents the analytical invasion threshold derived from Equation (17), plotted in the suppression efficacy (f_1) versus maintenance cost (c) parameter space. KZFP invasion is possible in the region below each respective curve. (A) Effect of varying the selection coefficient, s. (B) Effect of varying the basal transposition rate, u. (C) Effect of varying the piRNA cluster fraction, π. Unless otherwise specified in the panel legend, the fixed parameters are $s = 0.001$, $u = 0.01$, and $\pi = 0.01$.

expressed as

$$\frac{dx_1}{dt} \approx \sigma_{qs} x_1 (1 - x_1)^2, \tag{18}$$

where σ_{qs}, the growth rate of KZFP at the quasi-equilibrium, is given by Equation (7):

$$\sigma_{qs} = s \cdot n_{qs} \cdot u \cdot f_1 (1 - p_{qs})^2 - c. \tag{19}$$

The quasi-equilibrium (n_{qs}, p_{qs}) is given by solving $\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$ from Equations (1) and (2) with $\bar{f} = x_1(2 - x_1) f_1$:

$$p_{qs} = 1 - \sqrt{\frac{s}{u(1 - \bar{f})}} \tag{20}$$

and

$$n_{qs} = \frac{2}{\pi} \sqrt{\frac{s}{u(1 - \bar{f})}} \frac{1 - \sqrt{\frac{s}{u(1-\bar{f})}}}{[1 - 2s \cdot (1 - \sqrt{\frac{s}{u(1-\bar{f})}})]}. \tag{21}$$

Here, we define y as

$$y = \sqrt{\frac{s}{u(1 - \bar{f})}} = \sqrt{\frac{s}{u(1 - x_1(2 - x_1) f_1)}} \quad \text{for } s < u(1 - \bar{f}). \tag{22}$$

Note that y is a monotonically increasing function of x_1 in interval $[0, 1]$. As x_1 varies over $[0, 1]$, y ranges over $y_0 \le y \le y_1$, where

$$y_0 = \sqrt{\frac{s}{u}} \quad \text{at } x_1 = 0 \qquad \text{and} \qquad y_1 = \sqrt{\frac{s}{u(1 - f_1)}} \quad \text{at } x_1 = 1. \tag{23}$$

Therefore we should evaluate σ_{qs} only for $y \in [y_0, y_1]$. Substituting y into Equations (20) and (21), we obtain p_{qs} and n_{qs}:

$$p_{qs} = 1 - y \qquad \text{and} \qquad n_{qs} = \frac{2y(1 - y)}{\pi(1 - 2s + 2s \cdot y)}. \tag{24}$$

These equations show that the piRNA frequency and TE copy number at the post-invasion equilibrium in the KZFP-fixation zone are independent of the KZFP cost, c. We then obtain the explicit form of the growth rate of KZFP, σ_{qs}, as a function of y, by substituting (n_{qs}, p_{qs}) into Equation (10),

$$\begin{aligned} \sigma_{qs} &= s \cdot u \cdot n_{qs} (1 - p_{qs})^2 f_1 - c \\ &= \frac{2s \cdot u \cdot f_1}{\pi} \frac{y^3 (1 - y)}{1 - 2s + 2s \cdot y} - c. \end{aligned} \tag{25}$$

This equation is very useful to visualizes how the fate of KZFP (fixed or polymorphic) is determined. Figure 3 plots σ_{qs} with y for two representative values of f_1. σ_{qs} is a function of y with $\sigma_{qs}(y = 0) = \sigma_{qs}(y = 1) = -c$. We can only focus on σ_{qs} within its biologically feasible range, $[y_{lower}, y_{upper}]$, shown by the orange shaded region in Figure 3. The feasible range of y is the overlap of two biologically relevant intervals. The first interval is between the gray dashed lines $0 \le y \le 1$, which show the relevant ranges of TE copy number and piRNA frequency, that is, $n_{qs} \ge 0$ and $0 \le p_{qs} \le 1$ (from Equation (24)). The second is between the blue dotted lines where y must lie in $[y_0, y_1]$ because KZFP frequency, x_1, also needs to range in $[0, 1]$. As y_{lower} is y_0 because $y_0 > 0$, the feasible range of y is given by

$$y_{lower} = y_0 \qquad \text{and} \qquad y_{upper} = \min\left\{1, y_1\right\}. \tag{26}$$

As y increases from y_{lower}, σ_{qs} increases to a single peak and then declines. If σ_{qs} crosses zero inside the feasible range of y, we denote the value of y at this intersection by y^* (i.e., $\sigma_{qs}(y = y^*) = 0$).

Figure 3 shows two plots, representing typical cases in the KZFP-fixation and -polymorphic zones. Figure 3A shows a plot for $f_1 = 0.35$, in which σ_{qs} is positive in the entire feasible range of y, indicating KZFP can "essentially" fix. By contrast, Figure 3B illustrates a case in the KZFP-polymorphic zone ($f_1 = 0.55$). In this plot, σ_{qs} is 0 at y^*, which is inside the feasible range of y. It is indicated that the growth rate, σ_{qs}, is positive for $y < y^*$,

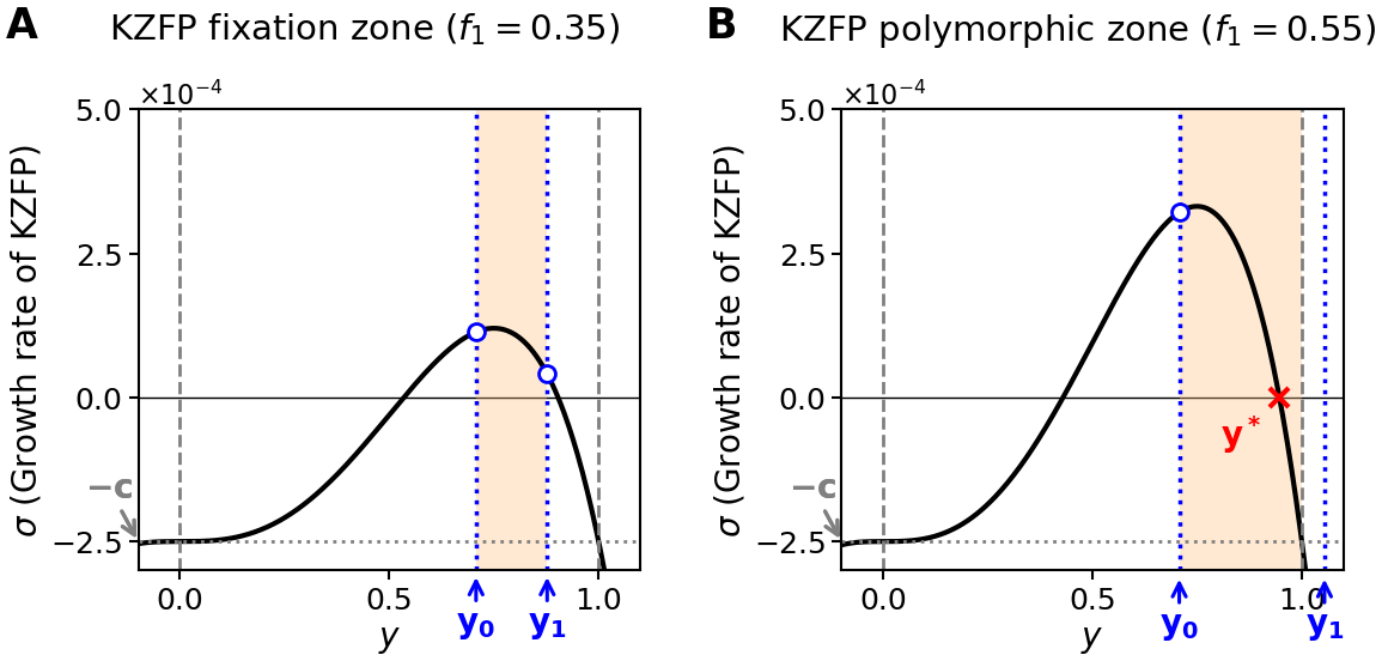

Figure 3 Growth rate of the KZFP allele frequency (σ) as a function of y for (A) $f_1 = 0.35$, representing the KZFP-fixation zone and for (B) $f_1 = 0.55$, representing the KZFP-polymorphic zone. $s = 5 \times 10^{-3}$, $u = 0.01$, $\pi = 0.01$, and $c = 2.5 \times 10^{-4}$ are fixed. The black curve shows σ_{qs} with y. The gray vertical dashed lines at $y = 0$ and $y = 1$ delimit the biologically relevant range for $n_{qs} \geq 0$ and $0 \leq p_{qs} \leq 1$ (Equation (24)). The blue dotted lines at $y_0 = \sqrt{s/u}$ and $y_1 = \sqrt{s/[u(1-f_1)]}$ indicate the biologically relevant range of the KZFP frequency. The orange shaded region indicates the feasible range of y, which is the overlap of these two constraints. The red cross in (B) shows the point where σ_{qs} is 0 at y^* within the feasible range of y. In (A), σ_{qs} is positive for the entire interval and x_1, resulting in a fixation of the KZFP allele. In (B), $\sigma_{qs}(y = y_0) > 0$ but $\sigma_{qs}(y = y_1) < 0$, indicating that the system converges to a stable equilibrium at y^* with $0 < x_{1eq} < 1$.

whereas negative for $y > y^*$. Given this behavior of the growth rate, the KZFP frequency reaches a stable equilibrium with an intermediate value, $0 < x_{1eq} < 1$. In other words, the KZFP suppressor allele is polymorphic.

Obviously, the border between the KZFP-fixation and -polymorphic zones is given when $y_1 = y^*$. The border is obtained by solving

$$\sigma_{qs}(y = y_1) = \frac{2s \cdot u \cdot f_1}{\pi} \frac{y_1^3(1-y_1)}{1 - 2s + 2s \cdot y_1} - c = 0, \qquad (27)$$

which is shown by the white dashed curve in Figure 1.

KZFP-fixation zone

We explore the behavior of the frequency of piRNA and TE copy number in the KZFP-fixation zone. Figure 1B shows the equilibrium frequency of piRNA, p_{eq}, computed by Equation (2). In the KZFP fixation zone, p_{eq} is always lower than p^* (from orange to blue in Figure 1B). This is intuitively easy to understand. At equilibrium, $\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$ hold simultaneously. From $\frac{dn}{dt} = 0$ we obtain

$$u(1-\bar{f})(1-p_{eq})^2 = s \qquad (28)$$

because $u(1-\bar{f})(1-p)^2 n - sn = 0$, and $n_{eq} > 0$ allows cancellation of n, where $\bar{f} = x_{1eq}(2 - x_{1eq})f_1$ is the average suppression efficacy of KZFP at its equilibrium from Equation (6). In the KZFP fixation zone, as $x_{1eq} = 1$ and $\bar{f} = f_1$ hold, by solving Equation (28) for p_{eq}, we obtain

$$p_{eq} = 1 - \sqrt{\frac{s}{u(1-f_1)}}. \qquad (29)$$

Thus, for fixed u and s, KZFP with any $f_1 (> 0)$ necessarily lowers p_{eq} relative to p^*. It is indicated that the two suppression systems cooperate in the fixation zone, with KZFP almost fixed while the frequency of piRNA decreased.

Figure 1C illustrates the equilibrium TE copy number, n_{eq}, computed from Equation (1). In the KZFP-fixation zone, n_{eq} is typically smaller than its pre-invasion level n^*, when selection is strong (Figure 1C, upper two panels). However, this does not necessarily hold when selection is weak as shown in the lower two panels in Figure 1C with moderate KZFP suppression efficacy, f_1, where n_{eq} exceeds n^*, which seems counter-intuitive though. The condition for $n_{eq} > n^*$ could be obtained by expressing n_{eq} as a function of p_{eq}. At equilibrium, $\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$ hold simultaneously, therefore we have $u(1-\bar{f})(1-p_{eq})^2 = s$. Substituting this into $\frac{dp}{dt} = 0$ yields

$$\begin{aligned} \frac{dp}{dt} &= \frac{\pi}{2} u(1-\bar{f})(1-p_{eq})^2 n_{eq} - \frac{s \cdot p_{eq}(1-p_{eq})}{1 - 2s \cdot p_{eq}} \\ &= \frac{\pi}{2} s \cdot n_{eq} - \frac{s \cdot p_{eq}(1-p_{eq})}{1 - 2s \cdot p_{eq}} = 0. \end{aligned} \qquad (30)$$

We then obtain the explicit relationship between n_{eq} and p_{eq} at

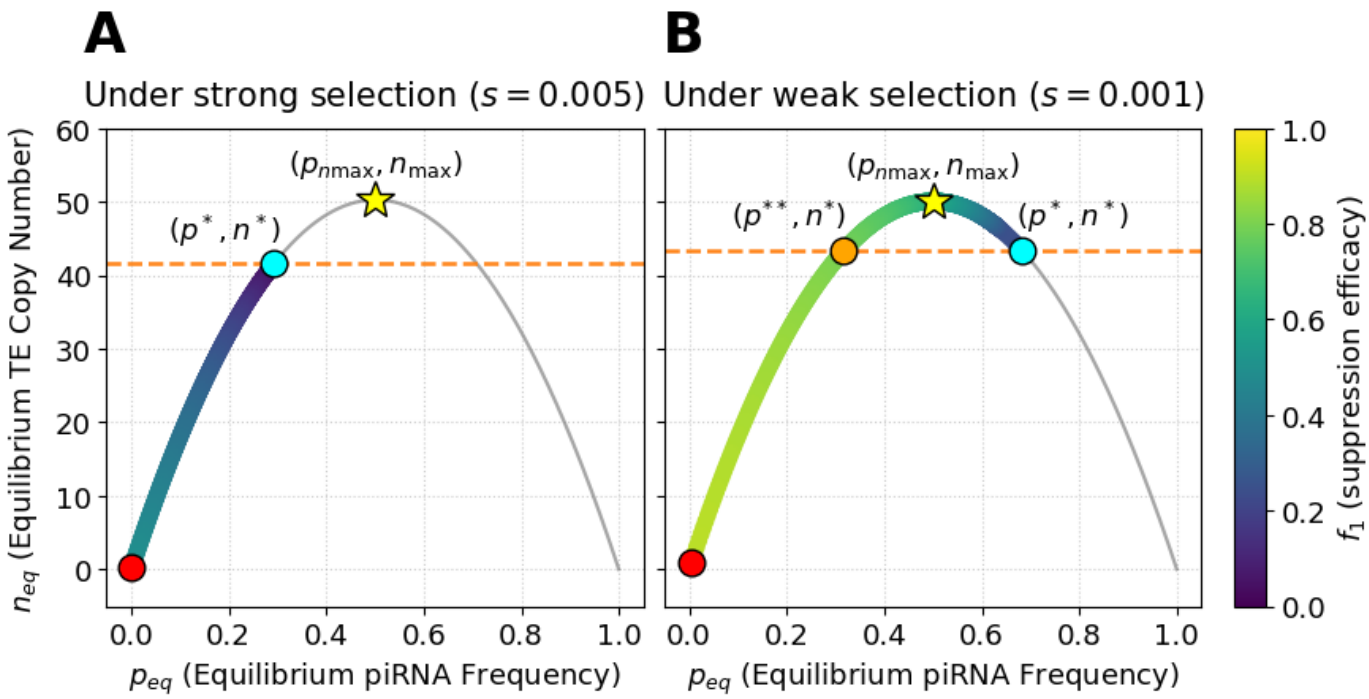

Figure 4 Post-invasion equilibrium relationship between TE copy number and piRNA frequency $(n_{\mathrm{eq}}, p_{\mathrm{eq}})$ as a function of the KZFP suppression efficacy f_1. Fixed parameters: $u = 0.01$, $\pi = 0.01$, and $c = 10^{-5}$. (A) Strong selection ($s = 0.005$). The pre-invasion value of p (p^*) lies to the left of the peak ($p^* < p_{n_{\max}}$), so all post-invasion equilibria have n_{eq} at or below the pre-invasion level n^*. (B) Weak selection ($s = 0.001$). The pre-invasion value of p (p^*) lies to the right of the peak ($p^* > p_{n_{\max}}$). In the range $p^{**} < p_{\mathrm{eq}} < p^*$, n_{eq} exceeds the pre-invasion level n^*. The gray curve shows the analytical equilibrium mapping given by Equation (31). The color bar represents the value of f_1. Cyan circle: pre-invasion equilibrium (n^*, p^*) at $f_1 = 0$. Red circle: post-invasion equilibrium at full suppression efficacy ($f_1 = 1$). Star: local maximum $n_{\max}$ attained at $p_{n_{\max}}$. Orange horizontal line: pre-invasion TE copy number n^*. Orange circle: point (n^*, p^{**}) where $n_{\mathrm{eq}} = n^*$ at $f_1 \approx 0.8$.

equilibrium:

$$n_{\mathrm{eq}} = \frac{2}{\pi} \cdot \frac{p_{\mathrm{eq}}(1 - p_{\mathrm{eq}})}{1 - 2s \cdot p_{\mathrm{eq}}}. \tag{31}$$

Equation (31) provides a convenient tool for assessing under what condition n_{eq} can exceed the pre-invasion level, n^*. Figure 4 plots the relationship between n_{eq} and p_{eq} given by Equation (31) for strong and weak selection cases ($s = 0.005$ for the strong case; Figure 4A, $s = 0.001$ for the weak case; Figure 4B). In both cases, n_{eq} is unimodal with a single peak at $p_{n\max}$, indicated by the star. Here, we denote the value of p_{eq} that maximizes the TE copy number by $p_{n\max}$, where $n_{\max}$ is the value of n_{eq} at the peak (we will derive $p_{n\max}$ and $n_{\max}$ later). p^{**} is the value that produces $n_{\mathrm{eq}} = n^*$ in the left side of the star (orange circle). It is important to note that the feasible range of n_{eq} is determined by the feasible range of p_{eq} in one-to-one correspondence, and that, within the fixation zone, p_{eq} is given by a monotonically increasing function of f_1 (see Equation (29)). This relationship is visualized by showing the corresponding f_1 value by colored points along the curve of Equation (31), from $f_1 = 0$ by cyan circle to the border value of f_1 by red circle (calculated by solving Equation (27)). In the strong-selection case in Figure 4A, $p^* < p_{n\max}$ holds with any f_1 value, so that n_{eq} is maximized at $f_1 = 0$ and never exceeds n^*. By contrast, in the weak case in Figure 4B, as p^* exceeds $p_{n\max}$, the feasible range of p_{eq} contains the star, at which n_{eq} has its maximum $n_{\max}$. This means that, if f_1 is low ($f_1 <\sim 0.8$ so that $p_{\mathrm{eq}} > p^{**}$), n_{eq} exceeds n^*. If f_1 is sufficiently high (i.e., $f_1 >\sim 0.8$), as well as the strong selection case, n_{eq} is smaller then n^*. Thus, our analytical result indicates n_{eq} can exceed n^* under some condition, although it may be somehow counter intuitive.

The condition under which n_{eq} can exceed n^* (i.e., $p^* > p_{n_{\max}}$) can be expressed as a function of the ratio s/u as follows. We first derive the piRNA frequency that maximizes TE copy number, $p_{n\max}$. Differentiation of Equation (31) with respect to p_{eq} gives

$$\frac{dn_{\mathrm{eq}}}{dp_{\mathrm{eq}}} = \frac{2}{\pi} \cdot \frac{1 - 2p_{\mathrm{eq}} + 2s \cdot p_{\mathrm{eq}}^2}{(1 - 2s \cdot p_{\mathrm{eq}})^2}. \tag{32}$$

Because $2/\pi(1 - 2s \cdot p_{\mathrm{eq}})^2 > 0$, the sign of $\frac{dn_{\mathrm{eq}}}{dp_{\mathrm{eq}}}$ is determined solely by the quadratic numerator $1 - 2p_{\mathrm{eq}} + 2s \cdot p_{\mathrm{eq}}^2$. Then, the biologically relevant root of this quadratic, denoted by $p_{n\max}$, is obtained exactly as

$$p_{n\max} = \frac{1 - \sqrt{1 - 2s}}{2s}. \tag{33}$$

Substituting $p_{n\max}$ into Equation (31), we have a local maximum, denoted by $n_{\max}$:

$$n_{\max} \approx \frac{2}{\pi} \cdot \frac{p_{n\max}(1 - p_{n\max})}{1 - 2s \cdot p_{n\max}}. \tag{34}$$

For a small s, $p_{n\max}$ can be approximately given by a simple function of s:

$$p_{n\max} \approx \frac{1}{2} + \frac{s}{4} + O(s^2). \tag{35}$$

For n_{eq} to be greater than n_*, $p^* > p_{n\max}$ has to hold:

$$p^* > p_{n\max} \iff 1 - \sqrt{\frac{s}{u}} > \frac{1}{2} + \frac{s}{4} + O(s^2)$$
$$\implies \frac{s}{u} \lesssim \frac{1}{4} \quad (\text{for small } s), \tag{36}$$

where p^* is from Equation (13). It is indicated that the ratio of s to u is an important factor to determine whether n_{eq} increases or decreases after the invasion of KZFP.

In addition to the equilibrium values of these three quantities, x_1, p and n, we explore how they change over time after the invasion of KZFP. Figure 5 shows the results for five parameter sets at the locations of the open circle, triangle, square, star, and cross in Figure 1. Here we still focus on the KZFP fixation zone, that is, the open circle, star, and cross, in Figures 5A–C. When $p^* < p_{n\max}$ (i.e., strong selection), Figure 5A shows that x_1 rises to fixation while p declines moderately (e.g. from ∼0.3 to ∼0.2), consistent with Equation (29). n also decreases monotonically. When $p^* > p_{n\max}$ (weak selection), as shown in Figures 5B and C, x_1 and p follow similar dynamics to those in Figure 5A, except that the behavior of n is complicated. In Figure 5B with a large f_1, n initially increases above n^* (from $\sim$ 42 to $\sim$ 50) and then declines to $\sim$ 10, whereas with intermediate f_1 (Figure 5C), n increases from $\sim$ 42 to $\sim$ 50 and remains elevated at equilibrium.

KZFP-polymorphic zone

We next consider the KZFP-polymorphic zone where the KZFP suppressor allele does not fix but converges to an intermediate frequency (the green region in Figure 1A). In this zone, both the piRNA frequency, p_{eq}, and the TE copy number, n_{eq}, are generally low, particularly for KZFP with high efficacy (f_1) and low cost (c).

Although of mathematical interest, we are able to obtain $x_{1\mathrm{eq}}$, p_{eq} and n_{eq} as simple approximate formulas under some conditions. When $y_1 > 1$ (i.e. $f_1 > 1 - s/u$), the upper endpoint of y is $y_{upper} = 1$ with $\sigma(1) = -c < 0$ (see Figure 3B). Hence, when KZFP can invade (i.e., $\sigma(y_0) > 0$), an interior equilibrium always exists in interval $(y_0, 1)$. For $c \to 0$ in this regime, the root y^* lies very close to $y = 1$, so the first-order expansion of Equation (25) around $y = 1$ (derivation in Appendix) gives

$$y^* \approx 1 - \frac{\pi c}{2s \cdot u \cdot f_1}, \tag{37}$$

so that the piRNA frequency and the TE copy number at equilibrium are given by

$$p_{\mathrm{eq}} = 1 - y^* \approx \frac{\pi c}{2suf_1} \tag{38}$$

and

$$n_{\mathrm{eq}} = \frac{2y^*(1-y^*)}{\pi(1-2s+2s\cdot y^*)} \approx \frac{c}{suf_1}. \tag{39}$$

Moreover, as $c \to 0$, the KZFP equilibrium frequency is approximated as

$$x_{1\mathrm{eq}} \approx 1 - \sqrt{1 - \frac{1-\frac{s}{u}}{f_1}}. \tag{40}$$

This reveals a key difference from the KZFP-fixation zone: The equilibrium state in the KZFP-fixation zone is independent of the KZFP cost (c), whereas in the KZFP-polymorphic zone, c becomes a critical parameter that directly determines the equilibrium levels of piRNA and TEs.

The evolutionary dynamics of x_1, p and n in the KZFP-polymorphic zone are investigated in Figures 5D and E. With a large c in Figure 5D, x_1 rises and settles at an intermediate level (∼0.4) while p and n drops more strongly than those in the KZFP-fixation zone with a smaller f_1. With a small c in Figure 5E, x_1 overshoots transiently and then converges on ∼0.4 while p and n approach near zero and occasionally exhibit sharp spikes.

This oscillation arises from a delayed negative feedback linking x_1, p and n through the KZFP suppression. As x_1 increases, the average suppression efficacy $\bar{f} = x_1(2 - x_1)f_1$ increases, leading to suppression of TE activity and a decline of n to near zero. When n becomes small, the fitness advantage of carriers of the KZFP suppressor allele over non-carriers vanishes, and the growth rate of KZFP, $\sigma = s \cdot n \cdot u \cdot f_1(1-p)^2 - c$, falls below zero. Consequently, once $\sigma < 0$, x_1 begins to decrease. The resulting weakening of suppression allows n to rebound, which in turn raises σ to positive values and drives x_1 upward again. Because p responds to TE activity with a lag, via the first term in Equation (2), the phase delay between x_1 and p produces an overshoot and a damped oscillation before the system stabilizes. In summary, in the KZFP-polymorphic zone, both systems coexist, but the high efficacy of KZFP drives the piRNA allele to a very low equilibrium frequency, sometimes near elimination.

Discussion

Our study addresses the evolutionary puzzle of how multiple, seemingly redundant defense mechanisms against transposable elements can coexist within a single genome. By developing a three-component dynamical model of TEs, the piRNA system, and newly introduced KZFP, we analytically characterized the conditions governing their interaction. Specifically, we derived the invasion threshold for a KZFP allele into an established TE-piRNA equilibrium (Equation (17)). Following a successful invasion, the KZFP allele can either spread to fixation or be maintained at an intermediate frequency; its fixation is guaranteed when the growth rate remains positive throughout its increase in frequency, with the boundary determined by the condition based on Equation (27).

Our analysis reveals that, in the KZFP-fixation zone, piRNA and KZFP can coexist. In general, KZFP has an intermediate suppression efficacy and spreads to fixation due to the selective advantage by its additional suppressive effect on TEs, whereas the burden on piRNA system is reduced and its allele frequency declines but remains polymorphic. As a result, the cooperation of the two suppression systems effectively reduces TEs. When the suppression efficacy of KZFP is either too weak or too strong, it does not fix and instead maintains a stable polymorphic state, thereby forming a KZFP-polymorphic zone.

However, this general pattern does not always hold. We find that TE copy number can paradoxically increase after KZFP introduction when selection against TEs is weak and/or the basal transposition rate is high. Under these conditions, TE proliferation is intrinsically strong, and the piRNA system plays a major role in TE suppression as its benefit outweighs its cost, whereas KZFP has only intermediate, rather than very strong, suppression efficacy (see above). This situation results in an interesting situation where KZFP partially suppresses TE activity but simultaneously weakens piRNA activity, because piRNA production requires TE insertions into piRNA clusters (i.e., piRNA tends to be more active when TEs are more active). In such cases, the total suppression capacity decreases because the KZFP is not strong enough to fully compensate for the weakened piRNA

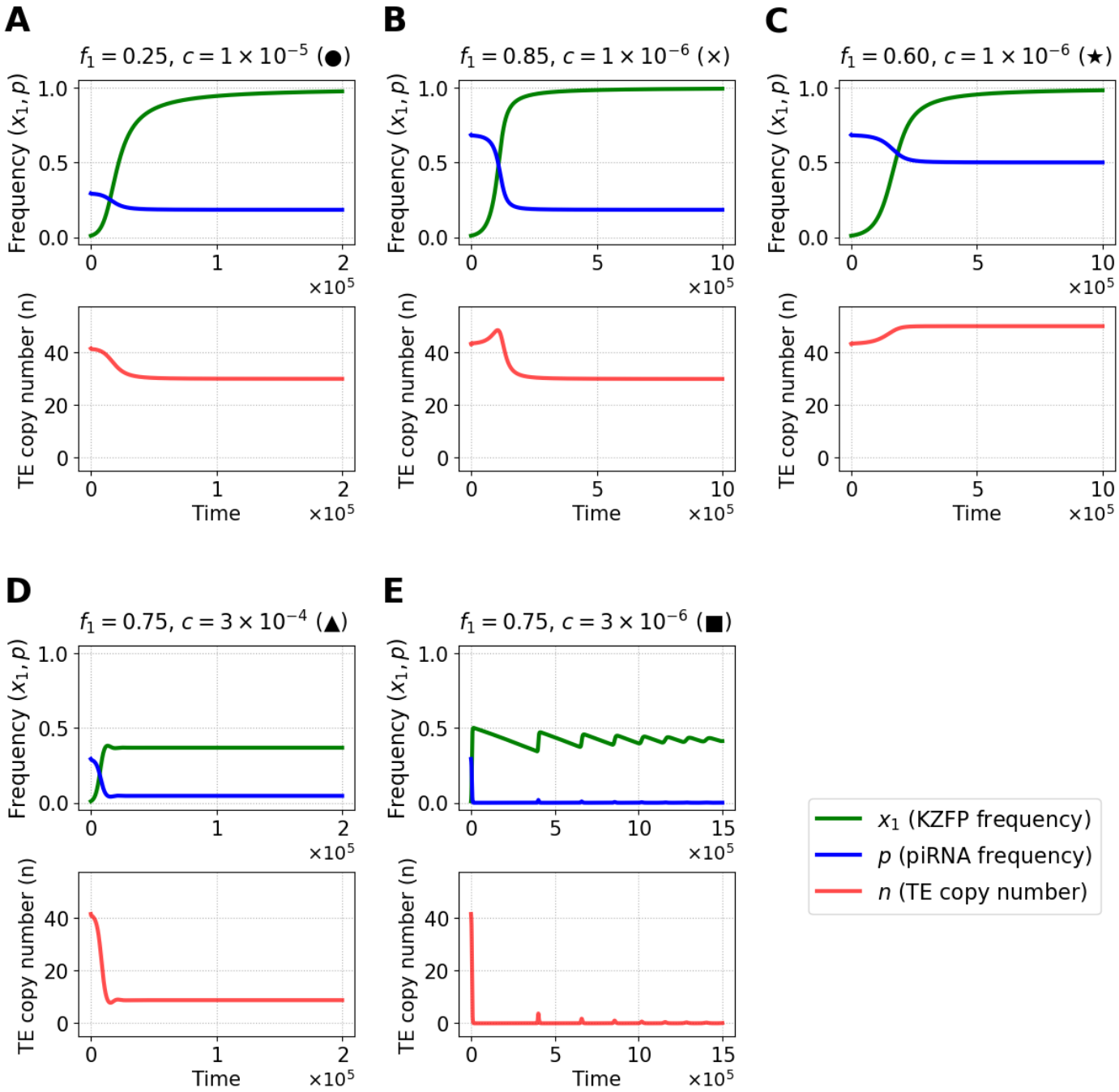

Figure 5 Dynamics of KZFP frequency (x_1), piRNA frequency (p), and TE copy number (n) for five representative parameter sets sampled from Figure 1. In each column, the upper panel shows the time courses of x_1 (green) and p (blue), and the lower panel shows the time course of n (red). All trajectories start from the pre-invasion equilibrium (n^*, p^*) with assuming an initial KZFP frequency of $x_1 = 0.01$. (A) KZFP-fixation zone with strong selection ($s = 5 \times 10^{-3}$, $f_1 = 0.25$, $c = 1 \times 10^{-5}$; circle in Figure 1). (B) KZFP-fixation zone with weak selection and high suppression efficacy ($s = 1 \times 10^{-3}$, $f_1 = 0.85$, $c = 1 \times 10^{-6}$; cross in Figure 1). (C) KZFP-fixation zone with weak selection and moderate suppression efficacy ($s = 1 \times 10^{-3}$, $f_1 = 0.6$, $c = 1 \times 10^{-6}$; star in Figure 1). (D) KZFP-polymorphic zone with high maintenance cost ($s = 5 \times 10^{-3}$, $f_1 = 0.75$, $c = 3 \times 10^{-4}$; triangle in Figure 1). (E) KZFP-polymorphic zone with low maintenance cost ($s = 5 \times 10^{-3}$, $f_1 = 0.75$, $c = 3 \times 10^{-6}$; square in Figure 1).

system. Consequently, a net increase in TE copy number occurs despite the addition of a new defense system. This paradoxical increase in TE number suggests that the allele-level selection does not necessarily maximize the population-level fitness. For carriers of KZFP, the allele confers a fitness advantage for individuals by reducing deleterious TEs ($\sigma > 0$), driving its spread. Yet as the KZFP allele becomes common, the piRNA system decreases in frequency, allowing more copies of TEs can survive, potentially lowering mean population fitness compared with the pre-invasion equilibrium.

In the KZFP-polymorphic zone, piRNA and KZFP cannot coexist stably and KZFP becomes dominant. KZFP are maintained at an intermediate frequency, but their suppression is so strong that TE activity is nearly eliminated. Since the piRNA system depends on active TE transposition, it effectively disappears once TEs are silenced.

When the maintenance cost c is very small, TE copy number n declines to almost zero. Under our deterministic infinite-population model, n never reaches zero exactly (Figure 5E), so that KZFP frequency x_1 settles at an low value and does not disappear. However, this result may not hold in a finite population, where TE is expected to go extinct stochastically due to random genetic drift. Once n reaches zero, the benefit of TE suppression vanishes and the growth rate of KZFP turns negative. In other words, KZFP-mediated suppression is so effective that it eliminates the TE threat entirely, making KZFP itself no longer necessary.

Our analytical result for the invasion threshold of KZFP (Equation (17)) produces a testable prediction: if a genome has a very large piRNA cluster fraction π, KZFP should be less likely to invade and persist, because a large π reduces n^* (Equation (14)) and the first term in σ (Equation (16)). Although a systematic phylogenetic analysis remains to be done, this prediction is qualitatively consistent with two well-studied species: *Drosophila melanogaster*, which lacks KZFP system, has a relatively large piRNA cluster fraction (approximately 3% of its genome (Brennecke *et al.* 2007)), whereas humans, with a large KZFP repertoire, have a substantially smaller cluster fraction (roughly 0.3-0.4% (Rosenkranz *et al.* 2022)). These examples are illustrative rather than definitive, and a formal comparative analysis that regresses KZFP repertoire size against π across taxa—ideally using phylogenetic comparative methods and controlling for genome size, effective population size, and TE landscape—would provide a decisive test.

Our framework extends the established "trap model" of piRNA (Kofler 2019; Tomar *et al.* 2023), in which TE insertions into piRNA clusters generate sequence-specific small RNAs that suppress those transposable elements. In the trap model, the TE copy number and the frequency of piRNA at equilibrium are determined by the balance of transposition and purifying selection (Equations (13) and (14)). This is exactly the same as the pre-invasion equilibrium (n^*, p^*) in our model. Introducing KZFP in the trap model adds a second layer of sequence-specific suppression that directly reduces the "effective" transposition rate from u to $u(1-\bar{f})$ in both the n- and p-equations (1) and (2). As a result, it generally diminishes the influx of new TEs and piRNA alleles, but, when s/u is small, TEs can paradoxically increase instead (see Equation (36)). If KZFP is disabled ($f_1 = 0$), our model reduces to the trap model, and if piRNA is also absent ($f_1 = 0$, $\pi = 0$) it further collapses to the classical transposition-selection balance model (Charlesworth and Charlesworth 1983).

Our analysis is based on a simplified model that assumes a single TE family, but this simplification omits several features of real genomes. In reality, multiple TE families coexist and compete for shared resources such as insertion sites or replication machinery (Abrusén and Krambeck 2006; Venner *et al.* 2009; Lawlor and Ellison 2023). The expansion of one family can displace others and may alter their suppression dynamics in ways that our model does not capture. Additionally, our model does not incorporate the possibility of a continuous influx of novel TE families. Given these factors that are ignored in our modeling framework, some of our results should be interpreted with caution. For example, when KZFP has strong suppressive efficacy and a low maintenance cost, our model predicts near extinction of the resident TE family. The elimination of all TE copies would remove the selective pressure to maintain the costly KZFP and piRNA defense systems, leading to their eventual decay in frequency. However, this scenario is unlikely in real genomes: the loss of both defense systems would leave the genome vulnerable to invasion by new TE families. Future models that incorporate multiple TE families and multiple layers of defense systems will provide a more comprehensive understanding of these dynamics.

A further simplification of our model is the assumption of static sequences for both TEs and the two defense systems, which works for the analysis on short-term frequency dynamics. Over longer evolutionary timescales, however, sequence evolution is inevitable. TEs can accumulate mutations that allow them to escape recognition by piRNAs or KZFPs. In response, host genomes can evolve new defense variants—such as novel piRNA cluster insertions or mutations in the DNA-binding domains of KZFPs—that restore suppression against these escaping TEs. This reciprocal process is a molecular "arms race" between TEs and host defenses. Empirical observations support these recurrent cycles of conflict and adaptation at the sequence level. For example, TE families are often lineage-specific and show rapid turnover across species, meaning that new TE families emerge and expand while older ones become inactive or extinct. Similarly, the KZFP gene family exhibits rapid diversification, with frequent birth of new paralogs and loss of older ones (Thomas and Schneider 2011; Cosby *et al.* 2019; Bruno *et al.* 2019; Kosuge *et al.* 2024). Future models that incorporate sequence evolution would be able to explicitly simulate this arms race, providing deeper insights into the long-term co-evolutionary dance between TEs and their hosts.

Furthermore, our model focuses exclusively on the interaction between only the two key sequence-specific defense systems. In reality, these systems operate within a broader network of genome defense that also includes sequence-independent but context-dependent silencing mechanisms, such as the HUSH complex in mammals, which senses long intronless nascent transcripts and reinforces H3K9me3-dependent repression (Seczynska *et al.* 2022; Nikolopoulos *et al.* 2025; Bloor *et al.* 2025). Incorporating such a mechanism into future models could provide a more comprehensive understanding of the multilayered TE suppression architecture.

Appendices

Appendix A: Jacobian Matrices at Equilibrium

To show whether the pre- or post-invasion equilibrium is stable or not, this appendix provides the Jacobian matrices for the two-variable (TE-piRNA) and three-variable (TE-piRNA-KZFP)

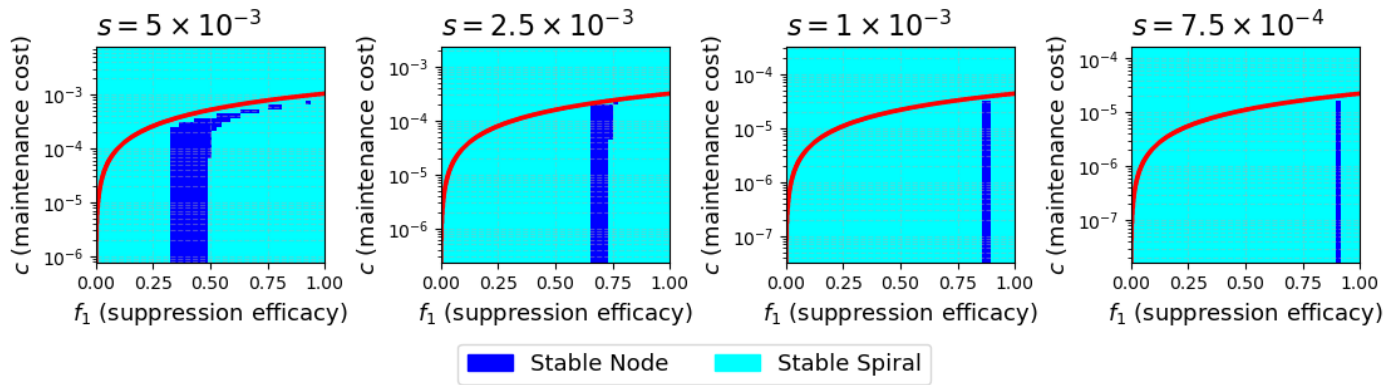

Figure A1 Local stability of the post-invasion equilibrium $(n_{\text{eq}}, p_{\text{eq}}, x_{1\text{eq}})$ in the (f_1, c) plane. Parameter sets where the equilibrium is a stable node are shown in blue, and those where it is a stable spiral are shown in cyan.

systems, evaluated at their respective non-trivial equilibrium points. Substituting the equilibrium conditions to the Jacobian matrices allows for simplification of the matrix elements, clarifying the stability analysis.

Jacobian for the TE-piRNA System ($x_1 = 0$)

The two-variable system is described by Equations (13)-(14) with $\bar{f} = 0$. The non-trivial equilibrium point (n^*, p^*) satisfies the conditions,

$$u(1-p^*)^2 - s = 0 \quad \text{and} \quad \frac{\pi}{2}u(1-p^*)^2 n^* - \frac{s \cdot p^*(1-p^*)}{1-2s \cdot p^*} = 0.$$

Evaluating the Jacobian matrix at the equilibrium point (n^*, p^*), we have

$$J^* = \begin{pmatrix} J_{11}^* & J_{12}^* \\ J_{21}^* & J_{22}^* \end{pmatrix} = \begin{pmatrix} 0 & -2n^* \cdot u(1-p^*) \\ \frac{\pi \cdot s}{2} & -\pi \cdot n^* \cdot u(1-p^*) - \frac{s(1-2p^*+2s \cdot (p^*)^2)}{(1-2s \cdot p^*)^2} \end{pmatrix},$$

where the (1,1) element of the Jacobian, J_{11}^*, is $u(1-p^*)^2 - s = 0$. The stability of the system is determined by the trace ($tr(J^*) = J_{22}^*$) and the determinant ($det(J^*) = -J_{12}^* J_{21}^*$). Since $J_{12}^* < 0$ and $J_{21}^* > 0$, the determinant is always positive. Thus, stability depends solely on the sign of J_{22}^*, requiring $J_{22}^* < 0$ for the equilibrium to be stable.

Jacobian for the TE-piRNA-KZFP System ($x_1 > 0$)

The full three-variable system is described by Equations (1),(2) and (10). Its non-trivial internal equilibrium $(n_{\text{eq}}, p_{\text{eq}}, x_{1\text{eq}})$ satisfies the conditions:

(1) $u(1-\bar{f})(1-p_{\text{eq}})^2 - s = 0$

(2) $\frac{\pi}{2}u(1-\bar{f})(1-p_{\text{eq}})^2 n_{\text{eq}} - \frac{s \cdot p_{\text{eq}}(1-p_{\text{eq}})}{1-2s \cdot p_{\text{eq}}} = 0$

(3) $s \cdot n_{\text{eq}} \cdot u \cdot f_1(1-p_{\text{eq}})^2 - c = 0 \quad (0 < x_{1\text{eq}} < 1)$.

Under these conditions, we have the Jacobian matrix:

$$J = \begin{pmatrix} 0 & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & 0 \end{pmatrix}$$

with

$$\begin{aligned}
J_{12} &= -2n_{\text{eq}} \cdot u(1-\bar{f})(1-p_{\text{eq}}) \\
J_{13} &= -2n_{\text{eq}} \cdot u \cdot f_1(1-x_{1\text{eq}})(1-p_{\text{eq}})^2 \\
J_{21} &= \frac{\pi \cdot s}{2} \\
J_{22} &= -\pi \cdot n_{\text{eq}} \cdot u(1-\bar{f})(1-p_{\text{eq}}) - \frac{s(1-2p_{\text{eq}}+2s \cdot p_{\text{eq}}^2)}{(1-2s \cdot p_{\text{eq}})^2} \\
J_{23} &= -\pi \cdot n_{\text{eq}} \cdot u \cdot f_1(1-x_{1\text{eq}})(1-p_{\text{eq}})^2 \\
J_{31} &= s \cdot u \cdot f_1(1-p_{\text{eq}})^2 \cdot x_{1\text{eq}}(1-x_{1\text{eq}})^2 \\
J_{32} &= -2s \cdot n_{\text{eq}} \cdot u \cdot f_1(1-p_{\text{eq}}) \cdot x_{1\text{eq}}(1-x_{1\text{eq}})^2,
\end{aligned}$$

where $J_{11} = J_{33} = 0$ from the conditions (1) and (3). The presence of zero elements on the main diagonal simplifies the analysis of the characteristic polynomial, $det(J - \lambda I) = 0$, which is used to determine the eigenvalues and thus the stability of the three-variable equilibrium.

Using the eigenvalues we obtained above, we examined whether the post-invasion equilibrium is stable or not in the (f_1, c) plane. Figure A1 demonstrates that, regardless of whether KZFP fixes or remains polymorphic, the system consistently converges to a stable equilibrium. Each point in the (f_1, c) plane is colored based on the eigenvalue structure of the Jacobian matrix J evaluated at equilibrium, confirming that the equilibrium is locally stable across all parameter sets examined. The blue region, located near the boundary between the KZFP-fixation and KZFP-polymorphic zones, corresponds to a stable node, where all variables converge directly to the equilibrium without oscillations. The surrounding cyan region indicates a stable spiral, where convergence occurs with oscillations.

Appendix B: Quasi-equilibrium approximation—time-scale separation and validation

In the main text we treat the TE copy number n and the piRNA allele frequency p as fast variables and the KZFP allele frequency x_1 as a slow variable. Here we quantify the assumption by comparing the relaxation rate of the (n, p) subsystem with the growth rate driving x_1.

For a fixed x_1, let $(n_{\text{qs}}, p_{\text{qs}})$ be the quasi-steady state obtained by solving $\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$ from Equations (1)-(2). Let $J(n, p|x_1)$ be a 2×2 Jacobian of the (n, p) subsystem, and let λ_i denote its eigenvalues at $(n_{\text{qs}}, p_{\text{qs}})$. When the equilibrium

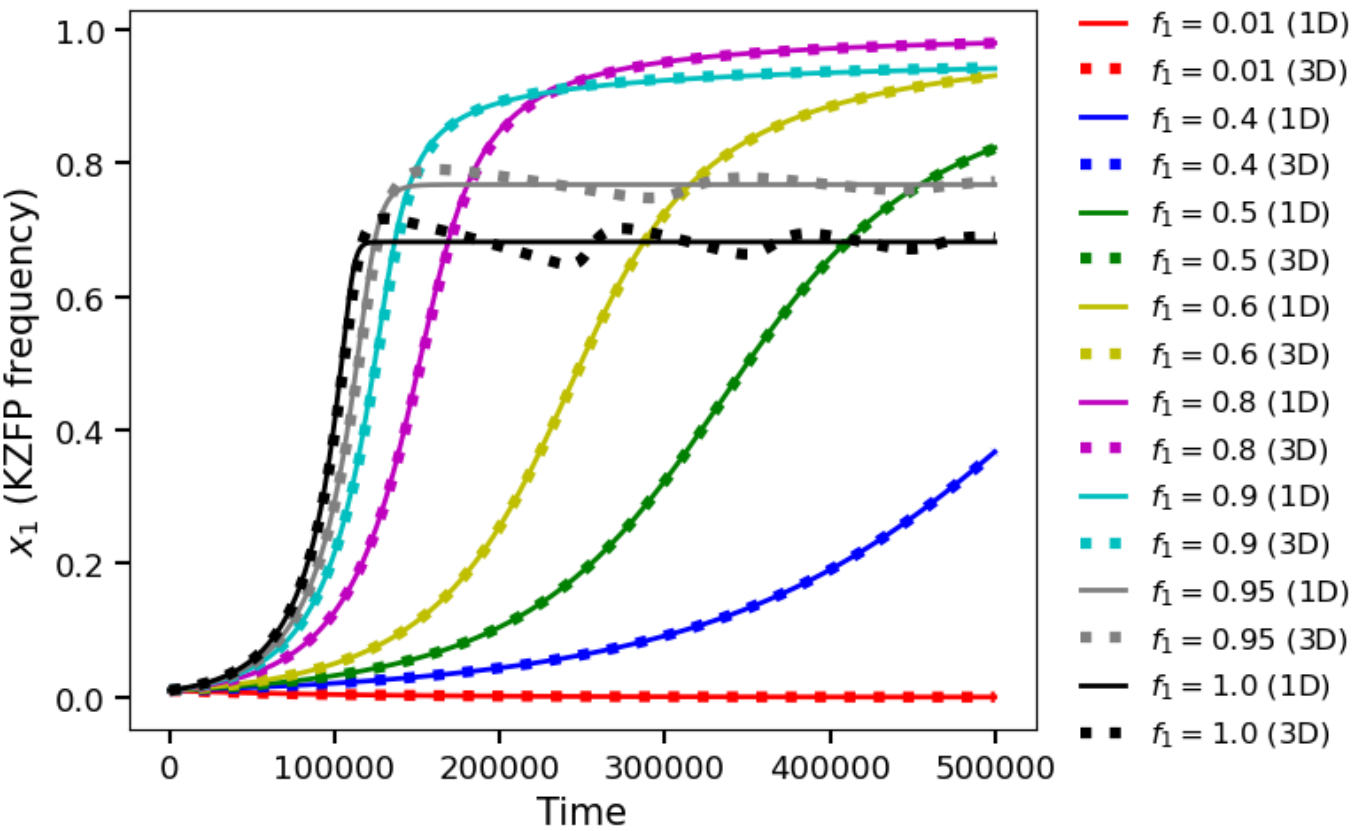

Figure A2 Validation of the quasi-equilibrium approximation. Time courses of KZFP frequency (x_1) from the full three-dimensional (3D) system (solid lines) and the one-dimensional (1D) reduced model (dashed lines) for representative values of suppression efficacy f_1. $u = 0.01$, $s = 0.001$, $\pi = 0.01$, and $c = 1 \times 10^{-5}$ are assumed. The reduced model closely tracks the full dynamics across all tested conditions. Minor deviations appear only at high efficacy ($f_1 \gtrsim 0.898$), where the time separation is less accurate.

is stable, the real part of an eigenvalues, denoted by $\Re\lambda_i$, is negative. Then we define the relaxation rate of fast variables (n, p) as κ and it holds

$$\kappa \equiv \min_i(\Re\lambda_i) < 0.$$

Under quasi-equilibrium approximation, we have the growth rate of KZFP as

$$\sigma = s \cdot u \cdot n_{\mathrm{qs}}(1 - p_{\mathrm{qs}})^2 f_1 - c$$

(cf. Equation (18)).

We quantify the timescale separation by separation index, denoted by ε:

$$\varepsilon \equiv \frac{|\sigma|}{|\kappa|}$$

Heuristically, $\varepsilon \ll 1$ indicates that (n, p) relaxes to quasi-equilibrium much faster than x_1 changes, justifying the quasi-equilibrium approximation at that x_1.

We also dynamically validate the quasi-equilibrium approximation by comparing the time courses of KZFP frequency x_1 from the full three-dimensional (3D) model given by Equations (1), (2), and (10), with those from the one-dimensional (1D) reduced approximation described by Equation (18). Figure A2 shows the temporal dynamics of x_1 for varying suppression efficacy f_1: dashed lines represent the full 3D system, and solid lines represent the 1D approximation. The trajectories from both models closely overlap, demonstrating that the reduced 1D model accurately captures the dynamics of the full system. This agreement holds even outside the fixation zone, where f_1 is large ($f_1 \gtrsim 0.898$ in Figure A2), showing only mild phase lags when κ is small. These lags do not alter our qualitative conclusions and are not responsible for key phenomena such as the paradoxical increase in TE copy number.

Appendix C: Derivation of n_{eq}, p_{eq} and $x_{1\mathrm{eq}}$ for small c in the KZFP-polymorphic zone

We analyze the KZFP-polymorphic zone, where $\sigma(y)$ has a unique root y^* in the interval $(y_0, 1)$. We now assume a small maintenance cost c, so that this root lies close to $y = 1$, and derive approximations for n_{eq}, p_{eq} and $x_{1\mathrm{eq}}$.

We write

$$y = 1 - \varepsilon, \qquad 0 < \varepsilon \ll 1,$$

where ε is a very small positive quantity. Substituting this into Equation (25) and expanding, we obtain

$$\begin{aligned}\frac{y^3(1-y)}{1-2s+2s\cdot y} &= \frac{(1-\varepsilon)^3\varepsilon}{1-2s(1-\varepsilon)} \\ &= \left(\varepsilon - 3\varepsilon^2 + O(\varepsilon^3)\right)\left(1 + 2s\cdot\varepsilon + O(\varepsilon^2)\right) \\ &= \varepsilon + (2s-3)\varepsilon^2 + O(\varepsilon^3). \qquad (41)\end{aligned}$$

Hence

$$\sigma(y = 1 - \varepsilon) = \frac{2s \cdot u \cdot f_1}{\pi}\varepsilon - c + O(\varepsilon^2). \tag{42}$$

Setting $\sigma(y = 1 - \varepsilon^*) = 0$ gives the distance of the root y^* from $y = 1$:

$$\varepsilon^* = 1 - y^* = \frac{\pi \cdot c}{2s \cdot u \cdot f_1} + O(c^2). \tag{43}$$

From Equation (24), the equilibrium values corresponding to y are

$$p_{\text{eq}} = 1 - y, \qquad n_{\text{eq}} = \frac{2y(1-y)}{\pi(1 - 2s + 2s \cdot y)}.$$

Evaluating these at $y = y^* = 1 - \varepsilon^*$, we obtain

$$p_{\text{eq}} = 1 - y^* = \varepsilon^* = \frac{\pi \cdot c}{2s \cdot u \cdot f_1} + O(c^2), \tag{44}$$

and

$$\begin{aligned} n_{\text{eq}} &= \frac{2y^*(1-y^*)}{\pi(1 - 2s + 2s \cdot y^*)} = \frac{2\varepsilon^*}{\pi} + O(\varepsilon^{*2}) \\ &= \frac{c}{s \cdot u \cdot f_1} + O(c^2). \end{aligned} \tag{45}$$

To obtain $x_{1\text{eq}}$ explicitly, we recall from the definition of y (Equation (22)) that

$$\bar{f} = x_1(2 - x_1)f_1 = 1 - \frac{s}{u \cdot y^2}. \tag{46}$$

With $y^* = 1 - \varepsilon^*$, we expand

$$y^{*2} = 1 - 2\varepsilon^* + O(\varepsilon^{*2}) \quad \Rightarrow \quad \frac{1}{y^{*2}} = 1 + 2\varepsilon^* + O(\varepsilon^{*2}),$$

so the right-hand side of Equation (46) becomes

$$1 - \frac{s}{u \cdot y^{*2}} = \left(1 - \frac{s}{u}\right) - \frac{2s}{u}\varepsilon^* + O(\varepsilon^{*2}). \tag{47}$$

Let us define a as

$$a \equiv \frac{1}{f_1}\left[\left(1 - \frac{s}{u}\right) - \frac{2s}{u}\varepsilon^*\right].$$

Then Equation (46) can be written as

$$x_1(2 - x_1) = a + O(\varepsilon^{*2}).$$

Solving $x^2 - 2x + a = 0$ gives the biologically relevant root (since $x_{1\text{eq}} < 1$):

$$x_{1\text{eq}} = 1 - \sqrt{1 - a} + O(\varepsilon^{*2}).$$

Writing $1 - a = A + B$, where A and B are given by

$$A = 1 - \frac{1 - \frac{s}{u}}{f_1} \qquad \text{and} \qquad B = \frac{2s}{u \cdot f_1}\varepsilon^*,$$

and using that ε^* is small, we obtain

$$\sqrt{A + B} = \sqrt{A} + \frac{B}{2\sqrt{A}} + O(B^2).$$

Therefore,

$$\begin{aligned} x_{1\text{eq}} &= 1 - \sqrt{A} - \frac{B}{2\sqrt{A}} + O(B^2) \\ &= 1 - \sqrt{1 - \frac{1 - \frac{s}{u}}{f_1}} - \frac{\frac{2s}{u \cdot f_1}\varepsilon^*}{2\sqrt{1 - \frac{1 - \frac{s}{u}}{f_1}}} + O(\varepsilon^{*2}). \end{aligned} \tag{48}$$

Finally, substituting Equation (43) into Equation (48) gives

$$x_{1\text{eq}} = 1 - \sqrt{1 - \frac{1 - \frac{s}{u}}{f_1}} - \frac{\pi \cdot c}{2u^2 \cdot f_1^2\sqrt{1 - \frac{1 - \frac{s}{u}}{f_1}}} + O(c^2). \tag{49}$$

When c is sufficiently small ($c \to 0$), this reduces to Equation (40).

Appendix D: Post-invasion equilibrium states in other parameter spaces

The overall patterns of the KZFP-fixation and KZFP-polymorphic zones remain robust across different parameter planes. Figure A3 shows the post-invasion equilibrium values of $x_{1\text{eq}}$, p_{eq}, and n_{eq} plotted in the parameter space of basal transposition rate (u) and selection coefficient against TE insertions (s). Figure A4 shows the corresponding equilibrium values in the parameter space of piRNA cluster fraction (π) and selection coefficient against TE insertions (s).

Data Availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article. The Python codes used for numerical analysis and generating the figures are available at https://github.com/YusukeNabeka/TE-coexist.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Abrusén G, Krambeck HJ. 2006. Competition may determine the diversity of transposable elements. Theoretical Population Biology. 70:364–375.

Almeida MV, Vernaz G, Putman AL, Miska EA. 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends in Genetics. 38:529–553.

Bloor S, Wit N, Lehner PJ. 2025. Rna binding by periphilin plays an essential role in initiating silencing by the hush complex. Nucleic Acids Research. 53:gkae1165.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell. 128:1089–1103.

Bruno M, Mahgoub M, Macfarlan TS. 2019. The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. Annual Review of Genetics. 53:393–416.

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

Cosby RL, Chang NC, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. Genes & Development. 33:1098–1116.

de Tribolet-Hardy J, Thorball CW, Forey R, Planet E, Duc J, Coudray A, Khubieh B, Offner S, Pulver C, Fellay J *et al*. 2023. Genetic features and genomic targets of human KRAB-zinc finger proteins. Genome Research. 33:1409–1423.

Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. Development. 144:2719–2729.

Emerson RO, Thomas JH. 2009. Adaptive evolution in zinc finger transcription factors. PLoS Genetics. 5.

Halic M, Moazed D. 2009. Transposon Silencing by piRNAs. Cell. 138:1058–1060.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. Genome Research. 16:669–677.

Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 543:550–554.

Iwasaki YW, Shoji K, Nakagwa S, Miyoshi T, Tomari Y. 2025. Transposon-host arms race: a saga of genome evolution. Trends in Genetics. 41:369–389.

Kazazian HH. 2004. Mobile Elements: Drivers of Genome Evolution. Science. 303:1626–1632.

Kofler R. 2019. Dynamics of Transposable Element Invasions with piRNA Clusters. Molecular Biology and Evolution. 36:1457–1472.

Kosuge M, Ito J, Hamada M. 2024. Landscape of evolutionary arms races between transposable elements and KRAB-ZFP family. Scientific Reports. 14:23358.

Lawlor MA, Ellison CE. 2023. Evolutionary dynamics between transposable elements and their host genomes: mechanisms of suppression and escape. Current Opinion in Genetics & Development. 82:102092.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. Nature Reviews Genetics 2011 12:9. 12:615–627.

Looman C, brink MA, Mark C, Hellman L. 2002. KRAB Zinc Finger Proteins: An Analysis of the Molecular Mechanisms Governing Their Increase in Numbers and Complexity During Evolution. Molecular Biology and Evolution. 19:2118–2130.

Najafabadi HS, Garton M, Weirauch MT, Mnaimneh S, Yang A, Kim PM, Hughes TR. 2017. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. Genome Biology. 18.

Nardelli J, Gibson T, Charnay P. 1992. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic Acids Research. 20:4137–4144.

Nikolopoulos N, Oda Si, Prigozhin DM, Modis Y. 2025. Structure and methyl-lysine binding selectivity of the hush complex subunit mpp8. Journal of Molecular Biology. 437:168890.

Ozata DM, Gainetdinov I, Zoch A, OCarroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs with big functions. Nature Reviews Genetics 2018 20:2. 20:89–108.

Rosenkranz D, Zischler H, Gebert D. 2022. piRNAclusterDB 2.0: update and expansion of the piRNA cluster database. Nucleic Acids Research. 50:D259–D264.

Rosspopoff O, Trono D. 2023. Take a walk on the KRAB side. Trends in Genetics. 39:844–857.

Seczynska M, Bloor S, Cuesta SM, Lehner PJ. 2022. Genome surveillance by hush-mediated silencing of intronless mobile elements. Nature. 601:440–445.

Srivastav SP, Feschotte C, Clark AG. 2024. Rapid evolution of piRNA clusters in the Drosophila melanogaster ovary. Genome Research. 34:711–724.

Tadepally HD, Burger G, Aubry M. 2008. Evolution of C2H2-zinc finger genes and subfamilies in mammals: Species-specific duplication and loss of clusters, genes and effector domains. BMC Evolutionary Biology. 8.

Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. Genome Research. 21:1800–1812.

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

Venner S, Feschotte C, Biémont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. Trends in Genetics. 25:317–323.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J *et al*. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 17:261–272.

Wells JN, Chang NC, McCormick J, Coleman C, Ramos N, Jin B, Feschotte C. 2023. Transposable elements drive the evolution of metazoan zinc finger genes. Genome Research. 33:1325–1340.

Yang P, Wang Y, Macfarlan TS. 2017. The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution.

Zhang Y, He F, Zhang Y, Dai Q, Li Q, Nan J, Miao R, Cheng B. 2022. Exploration of the regulatory relationship between KRAB-Zfp clusters and their target transposable elements via a gene editing strategy at the cluster specific linker-associated sequences by CRISPR-Cas9. Mobile DNA. 13:1–17.

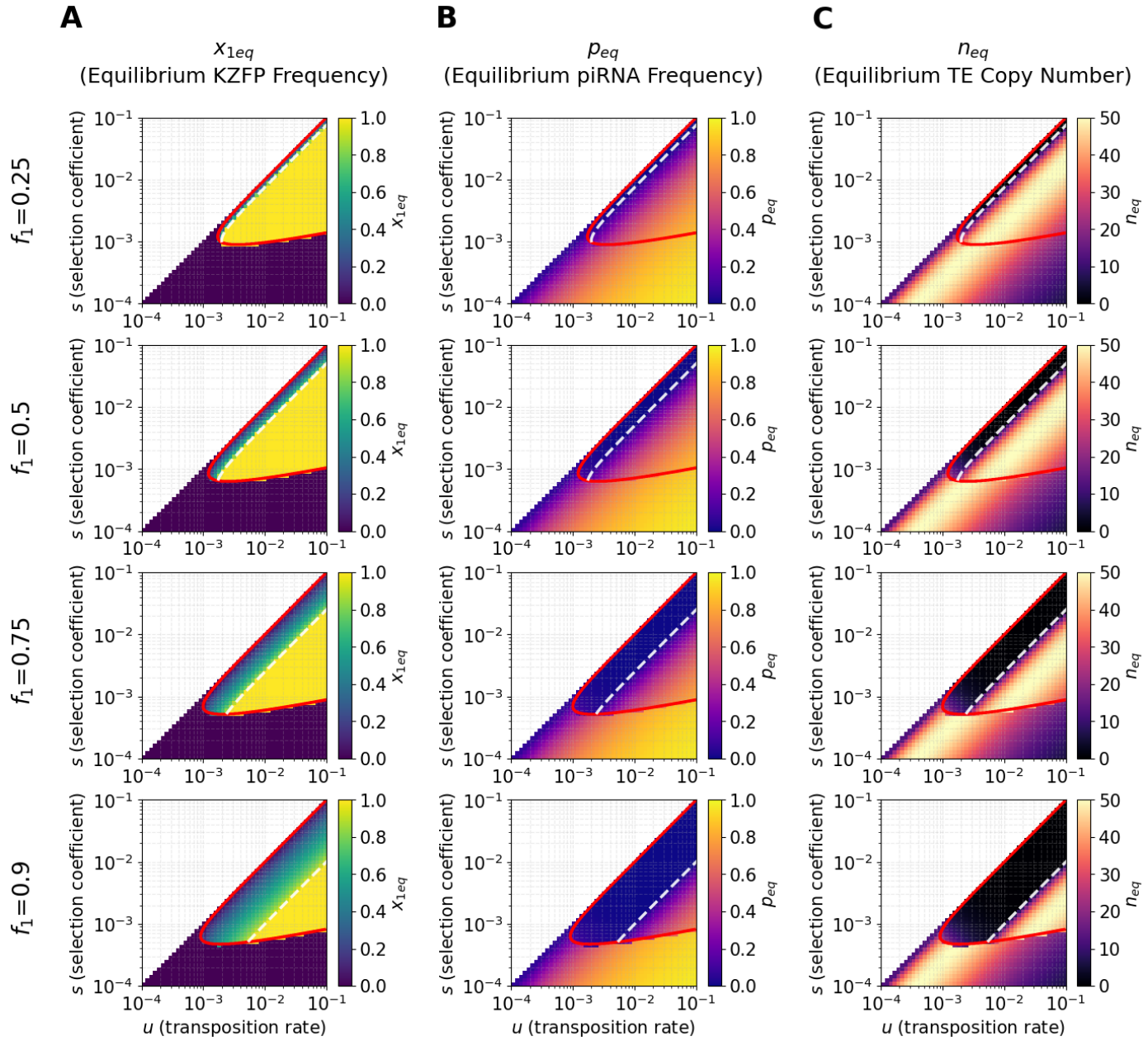

Figure A3 Post-invasion equilibrium states in the (u, s) parameter space. (A) Equilibrium frequency of the KZFP allele ($x_{1\text{eq}}$), (B) equilibrium frequency of the piRNA-producing allele (p_{eq}), and (C) equilibrium TE copy number (n_{eq}) after KZFP invasion plotted in the parameter space of basal transposition rate (u) and selection coefficient against TE insertions (s). Four different values of the suppression efficacy f_1 are used ($f_1 = 0.25, 0.5, 0.75$ and 0.9), while $\pi = 0.01$ and $c = 1 \times 10^{-5}$ are fixed. The post-invasion equilibrium states are numerically computed by using the `solve_ivp` ODE solver in the `scipy` package (Virtanen *et al.* 2020). The color bars are to present the value of focal quantities in the post-invasion equilibrium. The red line shows the analytical invasion boundary from Equation (17), below which KZFP invasion is possible. The white dashed curve shows the upper boundary of the KZFP-fixation zone from Equation (27). The white area in the upper left corresponds to parameter combinations with $s > u$, where the TE-piRNA pre-invasion equilibrium does not exist because $p^* = 1 - \sqrt{s/u} < 0$ and $n^* < 0$ (Equations (13) and (14)).

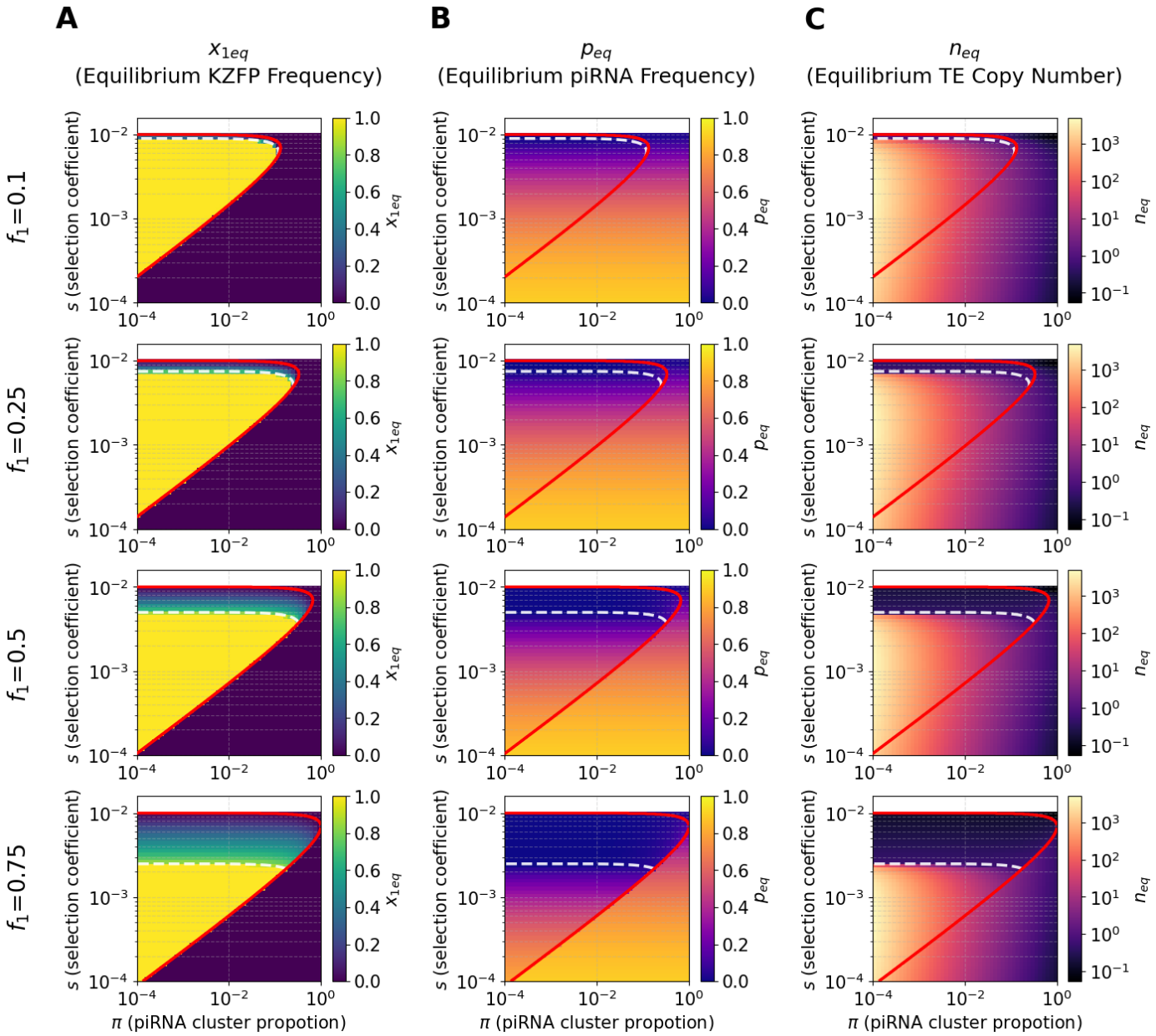

Figure A4 Post-invasion equilibrium states in the (π, s) parameter space. (A) Equilibrium frequency of the KZFP allele ($x_{1\mathrm{eq}}$), (B) equilibrium frequency of the piRNA-producing allele (p_{eq}), and (C) equilibrium TE copy number (n_{eq}) after KZFP invasion plotted in the parameter space of piRNA cluster fraction (π) and selection coefficient against TE insertions (s). Four different values of the suppression efficacy f_1 are used ($f_1 = 0.1, 0.25, 0.5$ and 0.75), while $u = 0.01$ and $c = 1 \times 10^{-5}$ are fixed. The post-invasion equilibrium states are numerically computed by using the `solve_ivp` ODE solver in the `scipy` package (Virtanen *et al.* 2020). The color bars are to present the value of focal quantities in the post-invasion equilibrium. The red curve shows the invasion threshold from Equation (17), below which KZFP invasion is possible. The white dashed curve shows the upper boundary of the KZFP-fixation zone from Equation (27). The white area in the upper corresponds to parameter combinations with $s > u$, where the TE-piRNA pre-invasion equilibrium does not exist because $p^* = 1 - \sqrt{s/u} < 0$ and $n^* < 0$ (Equations (13) and (14)).