

# Coexistence vs collapse in transposon populations

Aria Yom

*Department of Physics, University of California, San Diego.*

Nathan E. Lewis\*

*Departments of Pediatrics and Bioengineering, University of California, San Diego.*

*Center for Molecular Medicine, Complex Carbohydrate Research Center,*

*and Department of Biochemistry and Molecular Biology University of Georgia.*

(Dated: May 21, 2025)

Transposons are small, self-replicating DNA sequences found in every branch of life. Often, one transposon will parasitize another, forming a tiny intracellular ecosystem. In some species these ecosystems thrive, while in others they go extinct, yet little is known about when or why this occurs. Here, we present a stochastic model for these ecosystems and discover a transition from stable coexistence to population collapse when the propensity for a transposon to replicate comes to exceed that of its parasites. Our model also predicts that replication rates should be low in equilibrium, which appears to be true of many transposons in nature.

## I. INTRODUCTION

Transposons are tiny intragenomic parasites found in every branch of life [1–3]. In humans, they comprise over 40% of our DNA and can contribute to a number of diseases, such as hemophilia and cancer [4–6]. Unlike most pathogens, transposons are rarely exchanged between hosts, passing instead from host to offspring like any typical gene. The uniqueness, ubiquity, and simplicity of the transposon makes it an ideal candidate for biophysical modeling.

Two major fissures quarter the landscape of transposon species: the Class I / Class II divide and the distinction between autonomous and non-autonomous elements. Class I elements replicate through an RNA template, while Class II elements replicate entirely as DNA. Autonomous elements encode all the proteins necessary for their replication, while non-autonomous elements must steal these resources from their autonomous counterparts in order to reproduce. It is common for transposons of both classes to come in autonomous/non-autonomous pairs [7–9].

The non-autonomous element parasitizes the autonomous element by consuming its reproductive machinery. The autonomous element in turn parasitizes the host cell. Researchers have thus advocated an “ecological” view of elements cohabiting the same host species [10, 11]. In particular, it has been suggested that the autonomous/non-autonomous element interaction could stabilize transposon populations, much like the predator/prey interaction stabilizes populations in the Lotka-Volterra model [10].

Unfortunately, early models found that when autonomous and non-autonomous elements cohabited a host species, one or both of them went extinct [12–14]. Later models for Class II ecosystems found stable coex-

istence only with the introduction of an ad hoc nonlinear fitness function to prevent populations from diverging [15, 16], a precondition unlikely to play a role in stabilizing most natural populations (see Appendix A). Of note, however, Ref. [17] found stable populations in a simple linear model for Class I element ecosystems. Thus, past models disagree over the fate of transposon ecosystems.

In nature, some transposons persist stably over long timescales, while others flare up sporadically or go extinct [18, 19]. These two behaviors mirror the two dynamics described in past models. One may therefore wonder, what exactly determines the stability or instability of transposon ecosystems?

Here, we probe this question by way of a general model for interacting transposon strains in sexually reproducing hosts. We derive our model from first principles based on the life cycle of a simple Class II element known as a helitron. In this case, we find that ecosystems achieve a stable equilibrium only when the non-autonomous elements’ ability to obtain reproductive resources exceeds that of the autonomous elements. This criterion divides transposon populations into two phases, a coexisting phase and a collapsing phase. We also show that transposition rates and autonomous/non-autonomous element ratios should be low in equilibrium, which appears to be true of many transposons in nature.

## II. MODEL DERIVATION

For concreteness, we base our model on the life cycle of a simple type of transposon known as a helitron, which uses a single transposase protein to replicate itself (Figure 1, reviewed in Refs. [2, 20]). Upon binding to the helitron, the transposase copies the helitron DNA sequence and inserts this copy elsewhere into the genome, potentially harming the host. Autonomous helitrons produce transposase, while non-autonomous helitrons must steal it from their autonomous counterparts. Since this process requires only one transposase, the mean trans-

\* natelewis@uga.edu

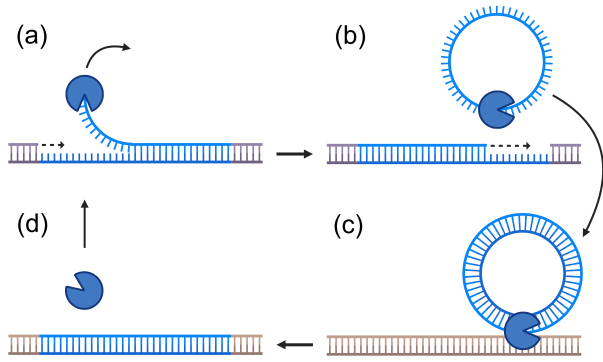


FIG. 1: Helitron replication mechanism. (a) The transposase binds to one end of the helitron and begins peeling one strand. (b) The transposase excises and circularizes the single strand of helitron DNA. The cell repairs the missing material. (c) The helitron is converted into double stranded form and moved to a new location in the genome. (d) The helitron is inserted into a new genomic location. Note that some details of this process remain unknown. Please see Refs. [21–23] for more details.

position rate should scale linearly with the transposase concentration, at least up to a point.

There are four processes for which any model of transposon populations must account:

1. Sex
2. Transposition
3. Transpositional toxicity
4. Transposon loss

In the sections that follow, we derive the effects of each of these four processes on the distribution of transposons in our host species. We will take the host population to be infinite until Section IIID, wherein we analyze finite size effects. For simplicity, we assume that mating is random and indiscriminate among members of the host population. We also assume that mating occurs much more frequently than the per-element rate of transposition, which appears to be true for many real-life transposons in equilibrium [24–29]. The effects of toxicity and loss are also assumed to be small, as is typically the case in nature [30–33].

We define a *strain*  $i$  to be a population of transposons with a defined set of parameters  $\alpha_i$ ,  $\omega_i$ , and  $r_i$  that characterize its propensity to replicate, to produce transposase, and to be lost from the genome, respectively. We analyze ecosystems of many interacting strains, but we do not consider events wherein one strain mutates into another.

We shall conclude this section by deriving an equation for the complete dynamics of transposon populations in

our model. Let us therefore begin by computing the effect of each of these subprocesses on the mean transposon count.

### A. Sex

Let  $\phi_i$  be a random variable denoting the number of elements of strain  $i$  in a randomly chosen host. So long as the population is in linkage equilibrium, which occurs whenever mating is random, frequent, and indiscriminate among hosts,  $\phi_i$  will be Poissonian [14]. It follows that we may compute the entire evolution of the transposon distribution simply by accounting for changes in the mean  $\lambda_i = \langle \phi_i \rangle$ . Random mating has no effect on  $\lambda_i$ , so let us move on to processes which do.

### B. Transposition

We assume that each strain  $i$  in this ecosystem replicates at a rate proportional to  $\alpha_i$  and produces transposase at a rate  $\omega_i$ . The sum  $\sum_j \omega_j \phi_j$  quantifies the total amount of transposase being produced within the cell, and acts as a background field stimulating the replication of transposons. For simplicity, and because the half-life of a transposase is typically much smaller than the replicative timescale [34–36], we imagine that the transposase production rate and the cellular transposase concentration are proportional. The probability of replication for a strain  $i$  will thus be given by  $\frac{\delta p_i}{\delta t} = \alpha_i \phi_i \sum_j \omega_j \phi_j R(\sum_k \alpha_k \phi_k)$ , where we have introduced the functional response  $R$  to account for the reduction in transposition that occurs when elements must compete for a limited amount of transposase.

In our model, we take the functional response  $R$  to be Holling’s type II,  $R(x) = \frac{a}{b+cx}$ , although the precise form of this function is not important. What matters is that it decays like  $\frac{1}{x}$ , which should be true of any model, since at some point the replication rate must be bounded by the availability of free transposase or another resource. Without loss of generality, we may absorb the constants  $a, b, c$  into  $\alpha_i$  and  $\omega_i$ , to get  $R(x) = \frac{1}{1+x}$ . Let us now proceed to calculate the change in the mean transposon count  $\lambda_i$ :

$$\left(\frac{d\lambda_i}{dt}\right)_{\text{trans}} = \alpha_i \left\langle \frac{\phi_i \sum_j \omega_j \phi_j}{1 + \sum_k \alpha_k \phi_k} \right\rangle \quad (1)$$

$$= \alpha_i \lambda_i \left\langle \frac{\omega_i}{1 + \sum_k \alpha_k \phi_k + \alpha_i} + \sum_j \frac{\omega_j \lambda_j}{1 + \sum_k \alpha_k \phi_k + \alpha_i + \alpha_j} \right\rangle \quad (2)$$

$$= \frac{\alpha_i \lambda_i (\omega_i + \sum_j \omega_j \lambda_j)}{1 + \sum_k \alpha_k \lambda_k} \left(1 + O\left(\frac{\max_l \alpha_l}{\sum_m \alpha_m \lambda_m}\right)\right) \quad (3)$$

$$\sim \frac{\alpha_i \lambda_i (\omega_i + \Omega)}{1 + A} \quad (4)$$

where we have defined  $A = \sum_k \alpha_k \lambda_k$ ,  $\Omega = \sum_k \omega_k \lambda_k$ , and assumed  $\alpha_j \ll \sum_k \alpha_k \phi_k \forall j$ , which is true when the total transposon activity eclipses the activity of any single element in the cell. The brackets  $\langle \rangle$  denote the Poisson expectation value, and we have employed the formula  $\langle \phi_i f(\phi_i) \rangle = \lambda_i \langle f(\phi_i + 1) \rangle$  to obtain Equation (2). Note that the approximation (4) is valid because  $\frac{1}{1+A} \sim \frac{1}{A}$  for large  $A$ , and because we expect the mean number of helitrons per host to be large [37–40]. We shall use this approximation with abandon throughout our analysis.

### C. Transpositional toxicity

In the process of replication, a transposon may kill its host with some probability  $q$ . The probability of replication for a single strain is  $\frac{\delta p_i}{\delta t} = \frac{\alpha_i \phi_i \sum_j \omega_j \phi_j}{\sum_k \alpha_k \phi_k}$ ; thus, the probability that the host is killed in this scenario is  $\frac{\delta p_{\text{death}}}{\delta t} = q \frac{\sum_i \alpha_i \phi_i \sum_j \omega_j \phi_j}{\sum_k \alpha_k \phi_k}$ .

The result of this small change in the host population will be a slight alteration to the  $\phi_i$  distribution. Let  $N(\vec{\phi})$  denote the number of hosts carrying  $\phi_i$  transposons of each strain  $i$ . The net effect of a change in  $N(\vec{\phi})$  on the mean transposon count  $\lambda_i$  will be:

$$\delta \lambda_i = \delta \frac{\sum_{\vec{\phi}} \phi_i N(\vec{\phi})}{\sum_{\vec{\phi}} N(\vec{\phi})} \quad (5)$$

$$= \frac{\sum_{\vec{\phi}} \phi_i \delta N}{\sum_{\vec{\phi}} N} - \frac{\sum_{\vec{\phi}} \phi_i N}{\sum_{\vec{\phi}} N} \frac{\sum_{\vec{\phi}} \delta N}{\sum_{\vec{\phi}} N} \quad (6)$$

$$= \left\langle (\phi_i - \lambda_i) \frac{\delta N}{N} \right\rangle \quad (7)$$

Thus, the change in  $\lambda_i$  will be:

$$\left(\frac{d\lambda_i}{dt}\right)_{\text{tox}} = -q \left\langle \frac{(\phi_i - \lambda_i) \sum_j \omega_j \phi_j \sum_k \alpha_k \phi_k}{1 + \sum_l \alpha_l \phi_l} \right\rangle \quad (8)$$

$$\sim -\frac{q \lambda_i (\alpha_i \Omega + \omega_i A)}{1 + A} \quad (9)$$

One may also wonder whether the helitron places some fitness burden on its host even when it is not replicating. In that case, we would take the toxicity to be proportional to the number of helitrons and get  $\frac{d\lambda_i}{dt} \propto \lambda_i$ . As we shall see, such a term can simply be absorbed into the term for transposon loss.

### D. Transposon loss

Finally, we consider the effects of genetic drift and transposon excision. Since each element has some rate  $r_i$  of decaying or being lost in each generation, the effect on  $\lambda_i$  will be:

$$\left(\frac{d\lambda_i}{dt}\right)_{\text{loss}} = -r_i \langle \phi_i \rangle = -r_i \lambda_i \quad (10)$$

### E. Total mean change in transposon count

Having considered in detail the individual effects of each of these processes, we may now obtain the total mean change in transposon count:

$$\boxed{\frac{d \log \lambda_i}{dt} = \frac{\alpha_i (\omega_i + \Omega) - q \omega_i A}{1 + A} - r_i} \quad (11)$$

where we have replaced  $\sim$  with  $=$  since this will serve as our model for the rest of this paper. This equation is valid so long as  $\alpha_i \ll \sum_j \alpha_j \lambda_j$  and  $q$  and  $r_i$  are small.

## III. RESULTS

### A. Stable ecosystems consist of one autonomous and one non-autonomous strain

When do transposon strains coexist in equilibrium in our model? So long as we are not interested in timescales, we may reparameterize our model such that  $dt = (1 + A)d\tau$ . Equation (11) then simplifies to a generalized Lotka-Volterra model:

$$\frac{d \log \lambda_i}{d\tau} = \alpha_i \omega_i - r_i + \sum_j (\alpha_i \omega_j - q \omega_i \alpha_j - r_i \alpha_j) \lambda_j \quad (12)$$

If all strains are non-autonomous ( $\omega_i = 0$ ), then there will be no transposase production, and all elements will go extinct. However, if  $\omega_i > 0$  for all strains  $i$ , then since  $q$  and  $r_i$  are small,  $\frac{d \log \lambda_i}{d\tau} > \alpha_i \omega_i$ , and  $\lambda_i$  will diverge. As  $\lambda_i$  grows, the toxicity of the transposons will eventually lead to the collapse of the host population. Therefore, stable populations must consist of at least one autonomous and one non-autonomous strain.

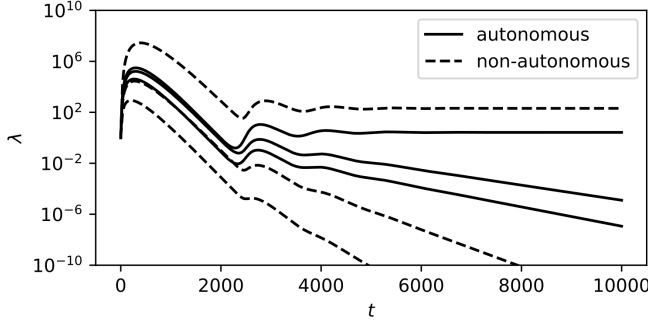


FIG. 2: Simulation of a stable system of 3 autonomous and 3 non-autonomous strains, each starting at  $\lambda_i(0) = 1$ , with  $t$  measured in arbitrary units.  $\alpha_i, \omega_i, q$ , and  $r_i$  were randomly chosen for each strain. In this instance,  $\alpha_i = \{.75, .36, .50, .51, .58, .56\}$ ,  $\omega_i = \{0, 0, 0, .61, .77, .42\}$ ,  $q = .00076$ ,  $r_i = .0099$ . Note that of the six initial strains, only two persist to  $t = \infty$ , one autonomous and one non-autonomous.

We now proceed to show that, except on a measure-zero subset of our parameter space, stable populations will always consist of precisely one autonomous and one non-autonomous strain. Let us consider the long-time behavior of orbits by defining the long-time averages  $\bar{A} = \lim_{T \rightarrow \infty} \frac{\int_0^T d\tau A}{T}$  and  $\bar{\Omega} = \lim_{T \rightarrow \infty} \frac{\int_0^T d\tau \Omega}{T}$ . For any stable strain,  $\frac{d \log \lambda_i}{d\tau}$  will average to zero over the long term. Thus:

$$\alpha_i \omega_i - r_i + \alpha_i \bar{\Omega} - (q \omega_i + r_i) \bar{A} = 0 \quad (13)$$

For these linear equations to yield a solution  $(\bar{A}, \bar{\Omega})$ , the  $3d$  vectors  $(\alpha_i \omega_i - r_i, \alpha_i, -q \omega_i - r_i)$  must all lie on the same  $2d$  plane. This is not possible without fine-tuning for  $n > 2$  strains. It follows that stable ecosystems should consist of exactly one autonomous and one non-autonomous strain. This can be observed in simulations of our model (Figure 2), wherein less fit strains are whittled away until only one autonomous and one non-autonomous strain remain.

### B. Coexistence collapses when $\alpha_a > \alpha_n$

Having shown that stable orbits consist of one autonomous and one non-autonomous strain, we now restrict our attention to the following two-strain model:

$$\frac{d \log \lambda_a}{d\tau} = \alpha_a - r_a + (1 - q - r_a) \alpha_a \lambda_a - (q + r_a) \alpha_n \lambda_n \quad (14)$$

$$\frac{d \log \lambda_n}{d\tau} = -r_n + (\alpha_n - r_n \alpha_n) \lambda_a - r_n \alpha_n \lambda_n \quad (15)$$

where  $a$  and  $n$  denote autonomy and non-autonomy, and where we have taken  $\omega_a = 1$ , which can be done without

loss of generality by rescaling  $r_a, r_n, \tau$ . Note that this operation makes all of our parameters dimensionless.

The fixed point of these equations occurs at:

$$\lambda_a^0 = \frac{r_n(q + \alpha_a)}{\alpha_n(r_a + q) - \alpha_a r_n} \quad (16)$$

$$\sim \frac{r_n \alpha_a}{\alpha_n(r_a + q) - \alpha_a r_n} \quad (17)$$

$$\lambda_n^0 = \frac{1}{\alpha_n} \frac{\alpha_a \alpha_n - r_a \alpha_n + r_n \alpha_a (1 - q - \alpha_a)}{\alpha_n(r_a + q) - \alpha_a r_n} \quad (18)$$

$$\sim \frac{\alpha_a}{\alpha_n(r_a + q) - \alpha_a r_n} \quad (19)$$

where we have again employed the approximation  $q, r_a, r_n \ll 1$ .

Since  $\lambda_a, \lambda_n > 0$ , this fixed point only exists when  $\frac{q+r_a}{\alpha_a} > \frac{r_n}{\alpha_n}$ . The existence of a fixed point is our first requirement for coexistence. We must now assess the stability of this fixed point. It is easy to linearize our equations about their fixed point and find the eigenvalues,  $\alpha_a - \alpha_n \pm \sqrt{(\alpha_a - \alpha_n)^2 - 4 \frac{\alpha_a \alpha_n^2}{r_n} \left( \frac{q+r_a}{\alpha_a} - \frac{r_n}{\alpha_n} \right)}$ . The fixed point is stable when the real parts of these eigenvalues are all negative, which occurs if and only if  $\alpha_n > \alpha_a$ . To summarize:

Orbits are stable iff  $\frac{q+r_a}{\alpha_a} > \frac{r_n}{\alpha_n}$  and  $\alpha_n > \alpha_a$

We illustrate the transition from stable coexistence to diverging/collapsing orbits in Figure 3. It is worth noting that because autonomous elements have more potential points of failure and are more problematic for the cell, we should expect  $r_a > r_n$ . We may also imagine that the deleteriousness of the transposon is more significant than its decay rate, leading to  $q > r_a, r_n$ . Either of these conditions will imply that whenever  $\alpha_n > \alpha_a$ , the model automatically satisfies  $\frac{q+r_a}{\alpha_a} > \frac{r_n}{\alpha_n}$ . Thus for real life transposons,  $\alpha_a = \alpha_n$  is probably the only relevant phase boundary.

### C. Low transposition rates and few autonomous elements in equilibrium

Taking the ratio of  $\lambda_a$  and  $\lambda_n$ , we can see that autonomous elements come to be greatly outnumbered by their non-autonomous partners in equilibrium:

$$\frac{\lambda_a^0}{\lambda_n^0} \sim r_n \quad (20)$$

As a result, the per-element transposition rate becomes low near the fixed point:

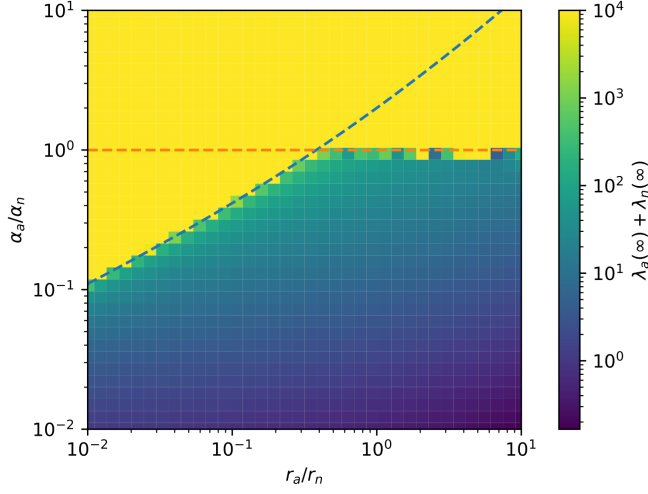


FIG. 3: Total transposon counts after long simulations. At each point,  $\lambda_a + \lambda_n$  was evaluated at  $t = 10000$ , by which time the system had usually equilibrated. Parameters satisfied  $\alpha_a \alpha_n = 1$ ,  $r_a r_n = 10^{-4}$ , and  $q = 10^{-2}$ . The dotted curves denote the phase boundaries  $\alpha_a = \alpha_n$  and  $\frac{q+r_a}{\alpha_a} = \frac{r_n}{\alpha_n}$ .

$$\left( \frac{d \log(\lambda_a + \lambda_n)}{dt} \right)_{\text{trans}} = \frac{\lambda_a}{\lambda_a + \lambda_n} \frac{\alpha_a + A}{1 + A} \sim r_n \quad (21)$$

Intriguingly, both of these results that naturally fall out of our model appear to be the norm for helitrons and other transposons in nature [2, 8, 24–29, 37, 39, 41–44].

#### D. Finite population effects

What happens when the number  $N$  of hosts is finite? In this case, the randomness of events such as mating cannot be neglected, and the evolution of  $\lambda_i$  becomes stochastic. Our system may therefore be described by a multidimensional Fokker-Planck equation whose drift term comes from Equation (11) and whose diffusion term we may compute from the variances of each of the four subprocesses in our model.

Near the fixed point, the only non-negligible contribution to the variance comes from random mating, which contributes a factor of  $\sigma_i^2 = \frac{2\lambda_i}{N}$  per generation. As our populations approach equilibrium, the  $\lambda_i$  distribution becomes approximately normal, centered about  $\lambda_i^0$ , with variance  $\Sigma_{ij}$ . We may solve for  $\Sigma_{ij}$  numerically using the Lyapunov equation:

$$J\Sigma + \Sigma J^T + Q = 0 \quad (22)$$

where  $J_{ij} = (\alpha_i \omega_j - (q\omega_i + r_i)\alpha_j) \frac{\lambda_i^0}{1+A^0}$ ,  $Q_{ij} = \frac{2\gamma\lambda_i^0}{N} \delta_{ij}$ , and  $\gamma$  denotes the number for generations per unit time, which we may take to be  $\gamma = 1$  without loss of generality.

Figure 4 illustrates a sample trajectory from our model. We can see that even for relatively small populations ( $N \sim 10^4$ ), orbits converge to the expected equilibria, and our estimates for the fluctuation sizes are accurate as well.

## IV. DISCUSSION

### A. Stability in Class I and instability in Class II models

We are now in a position to understand why previous Class II models tended towards instability. In each of these models [12–16], the replication rates were taken to be proportional to the transposase concentration, placing them into the same framework analyzed herein. However, in each of these models, the parameters  $\alpha_a$  and  $\alpha_n$  were taken to be equal for simplicity. Unfortunately, by taking  $\alpha_a = \alpha_n$ , past researchers placed their populations precisely on the phase boundary between coexistence and collapse. Hence, the instability.

What can we say about the stable equilibrium found in Ref. [17] for long and short interspersed nuclear elements (LINEs and SINEs), a common pair of autonomous and non-autonomous Class I elements? To simplify their model to its mean-field essence, their equations were roughly of the form:

$$\frac{d\lambda_a}{dt} = \alpha_a \Omega - r_a \quad (23)$$

$$\frac{d\lambda_n}{dt} = \alpha_n \Omega \lambda_a - r_n \quad (24)$$

These equations differ from ours in two important ways: (a) they have no functional response, and (b) the autonomous transposition rate is not proportional to  $\lambda_a$ . It is the latter difference that accounts for the stability of these equations, since the analogous transposition parameters,  $\alpha_a^{\text{eff}} = \frac{\alpha_a}{\lambda_a}$  and  $\alpha_n^{\text{eff}} = \alpha_n$ , do not have a constant ratio. Therefore, if populations begin to diverge, they will eventually reach a regime in which  $\alpha_n^{\text{eff}} > \alpha_a^{\text{eff}}$ , and the system will self-stabilize.

The essential difference between this model and ours arises from the fact that in the LINE replication process, the transposase is imagined to bind immediately to the LINE transcript that generates it, rather than binding to a random element in the DNA [45, 46]. This is known as “*cis*-preference” and occurs in some other Class I elements as well, but does not occur in Class II elements [47, 48]. It follows that *cis*-binding Class I ecosystems should be stable, while Class II ecosystems may be stable or unstable.

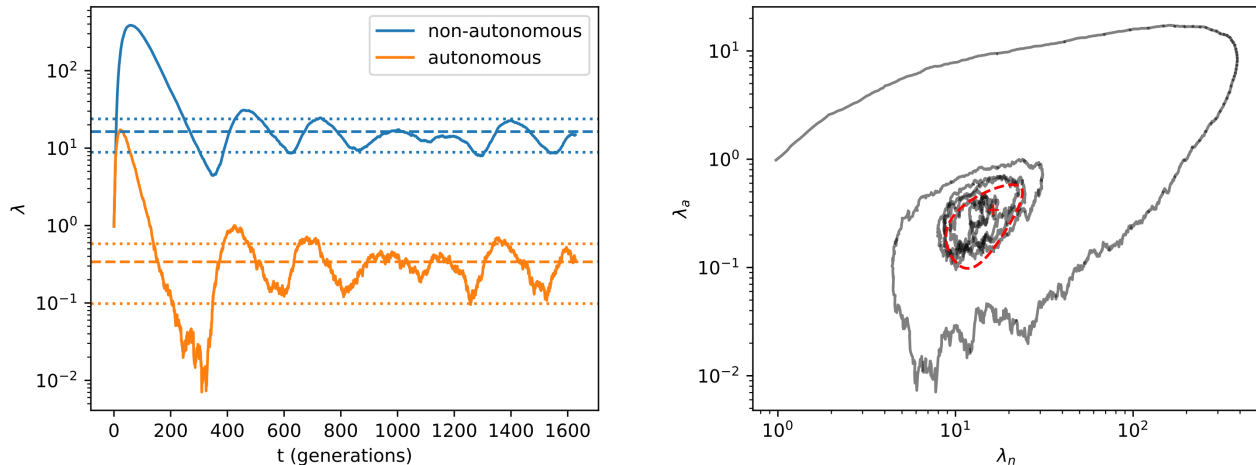


FIG. 4: Simulation of a finite population of autonomous and non-autonomous elements. Parameters used:  $\lambda_a(0) = \lambda_n(0) = 1$ ,  $\alpha_a = 1$ ,  $\alpha_n = 2$ ,  $q = r_a = r_n = 10^{-2}$ ,  $N = 4096$ . (left) Dynamics of  $\lambda_a$  and  $\lambda_n$ . The infinite population equilibrium values  $\lambda_a^0$  and  $\lambda_n^0$  are indicated by dashed lines, while the dotted lines denote fluctuations of size  $\sqrt{4\Sigma_{aa}}$  and  $\sqrt{4\Sigma_{nn}}$ . (right) Orbit of  $\lambda_a$  and  $\lambda_n$  with fixed point and  $4\Sigma$  ellipse.

## B. Predictions of the model

Aside from the coexistence or collapse of helitron populations, our model also makes the following two quantitative predictions:

- (i) That non-autonomous elements should outnumber autonomous ones, at least in equilibrium.
- (ii) That per-element transposition rates should be low in stable populations.

Our first prediction, although counterintuitive, has been confirmed numerous times [2, 37, 39]. Heuristic arguments for the fitness advantage of non-autonomous elements over autonomous ones have been given by many previous authors (see [10] for example). If our results are correct, then we have derived the first formula to quantify this preponderance.

Our second prediction, that per-element transposition rates should be low, has been verified in helitrons and many other transposons [24–29]. One explanation for this phenomenon could be that hosts have evolved to suppress transposons, but if this is the case, then why haven't hosts been able to suppress viruses to a similar extent? The threat posed by viruses is far greater, yet their replication rates remain many orders of magnitude higher. In our model, we show instead that transposons can be kept in check by other transposons rather than by the host.

Do these results hold for transposons other than helitrons? Let us consider the aforementioned case of LINEs and SINEs, which together occupy one third of the human genome [4]. Since active SINEs outnumber LINEs at least ten to one, they clearly satisfy (i) [42, 44].

And since both elements transpose on the order of once per hundred generations, amounting some  $10^{-4}$  replications per active element per generation, they also satisfy (ii) [41, 43]. Therefore, these results may be much more universal than one may expect from this simple model.

## C. Concluding remarks

Can this model explain the survival or collapse of real-life transposon populations? This question is difficult to answer, as the relevant parameters for natural populations have not been measured. However, since  $\alpha_i$  is essentially a measure of transposase affinity, one could perform an experiment to test our  $\alpha_n > \alpha_a$  criterion in any number of transposon species and compare the results to natural populations.

In this article, we derived some simple equations for the stability and properties of transposon ecosystems which appear to explain the behavior of diverse transposon species. In particular, we have shown how subtle parameter changes can cause transposon populations to diverge dramatically, which could shed light on why similar species can have such vastly different genome sizes [49]. Transposons are an ancient and ubiquitous part of life on Earth, and are the most abundant elements of our DNA. These results shed light both on the evolution of Eukaryotes and on the dynamics of modern genomes.

All code used to generate our plots is freely available at [github.com/moyja/transposon\\_paper](https://github.com/moyja/transposon_paper)

## ACKNOWLEDGMENTS

This work was supported in part by funding from NIGMS (R35 GM119850).

### Appendix A: On superlinear toxicities

Consider a single transposon which replicates at a rate  $\beta$  and which causes the host to perish at a rate  $\gamma_1$ . In this case, the mean rate of change in transposon count will be  $\frac{d\lambda}{dt} = (\beta - \gamma_1)\lambda$ . Since such a model admits no nonzero equilibrium, we may be inclined to introduce a nonlinear term to the toxicity, resulting in the following equation:

$$\frac{d\lambda}{dt} = (\beta - \gamma_1)\lambda - \gamma_2\lambda^2 \quad (\text{A1})$$

This simple approach was advocated in the earliest models for transposon population dynamics to ensure the existence of a stable equilibrium population, and has continued to be used in more recent studies as well [15, 16, 50, 51]. However, in order to obtain such a term, there must be some mechanism by which the transposon toxicity compounds more than linearly in  $\lambda$ . Such a mechanism can either be transposition-dependent or transposition-independent.

The former case includes effects like insertional mutagenesis. Conceivably perhaps, the transposition of two elements in close succession could kill the host with high probability, while a single transposition would not. But as we mentioned earlier, the most abundant transposons in the human genome replicate just once per thousand years [41, 43]. It is therefore difficult to imagine how any such interaction could occur.

The latter case includes effects like ectopic recombination, which was once the most popular hypothesis to explain this phenomenon [52, 53]. In this case it is imagined that the presence of transposons in the genome may disrupt the DNA recombination machinery and lead to fatal errors in crossing over. This view is articulated in Ref. [13]:

*“A balance between transposition and selection can be achieved, but only if the logarithm of the fitness of an individual with  $i$  copies of a transposable element declines more than linearly with  $i$ . This selection could arise through ectopic recombination with the transposable elements acting as scattered sites of homology in the*

*chromosomes. The frequencies of such events will increase with the square of transposable element copy number. One interesting aspect is that the selection arises purely because these sequences are interspersed repetitive, and not because of any transposition process that they are undergoing.”*

There is ample evidence that transposons do indeed cause ectopic recombination, but this only implies a nonzero contribution to  $\gamma_1$  [54, 55]. To gain insight into the contribution of ectopic exchange or of any other transposition-independent mechanism to  $\gamma_2$ , let us consider the case of transposition-incompetent elements.

Though our genomes are saturated with transposons, the vast majority of them are no longer capable of transposition, having lost or damaged sequences critical to their replication [11, 18, 38, 56–60]. If these “dead” transposons posed a meaningful threat to host fitness, then we would expect to see them quickly removed from the population. Their “live” cousins may also be deleterious, but at least they have the capacity to replicate to offset this shortcoming, and they should therefore be fitter and more abundant than their transposition-incompetent counterparts.

To make this argument more mathematically precise, let us decompose our simple model into “live” and “dead” populations:

$$\frac{d\lambda_{\text{live}}}{dt} = (\beta - \gamma_1)\lambda_{\text{live}} - \gamma_2\lambda_{\text{live}}(\lambda_{\text{live}} + \lambda_{\text{dead}}) \quad (\text{A2})$$

$$\frac{d\lambda_{\text{dead}}}{dt} = -\gamma_1\lambda_{\text{dead}} - \gamma_2\lambda_{\text{dead}}(\lambda_{\text{live}} + \lambda_{\text{dead}}) \quad (\text{A3})$$

If  $\lambda_{\text{live}} \ll \lambda_{\text{dead}}$ , we may neglect its quadratic contribution to the  $\gamma_2$  term, leaving us again with a linear differential equation with no equilibrium. Furthermore, since  $\frac{d\lambda_{\text{dead}}}{dt} < \frac{d\lambda_{\text{live}}}{dt}$ , the dead transposons will be whittled out of the population. Thus, the transposition-independent hypothesis is not consistent with the observed preponderance of transposition-incompetent elements.

Finally in all of these models, we would expect transposons to evolve to ever-increasing replication rates, which would appear to be possible by comparison with viral replication rates, but in nature we tend to see very low transposon replication rates [24–29]. On all of these grounds, we must reject the notion that nonlinear toxicities are relevant to stabilizing transposon populations.

---

[1] P. Siguier, E. Gournbeyre, and M. Chandler, Bacterial insertion sequences: their genomic impact and diversity, *FEMS Microbiology Reviews* **38**, 865 (2014), <https://academic.oup.com/femsre/article-pdf/38/5/865/18142951/38-5-865.pdf>.

[2] J. Thomas and E. J. Pritham, Helitrons, the eukaryotic rolling-circle transposable elements, *Microbiology Spectrum* **3**, 10.1128/microbiolspec.mdna3 (2015), <https://journals.asm.org/doi/pdf/10.1128/microbiolspec.mdna3-0049-2014>.

- [3] J. Wells and C. Feschotte, A field guide to eukaryotic transposable elements, *Annu Rev Genet* **54**, 539 (2020), pMID: 32955944.
- [4] E. Lander, L. Linton, B. Birren, C. Nusbaum, M. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczkzy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. Waterston, R. Wilson, L. Hillier, J. McPherson, M. Marra, E. Mardis, L. Fulton, A. Chinwalla, K. Pepin, W. Gish, S. Chissoe, M. Wendl, K. Delehaunty, T. Miner, A. Delehaunty, J. Kramer, L. Cook, R. Fulton, D. Johnson, P. Minx, S. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. Gibbs, D. Muzny, S. Scherer, J. Bouck, E. Sodergren, K. Worley, C. Rives, J. Gorrell, M. Metzker, S. Naylor, R. Kucherlapati, D. Nelson, G. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. Davis, N. Federspiel, A. Abola, M. Proctor, R. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. Cox, M. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. Evans, M. Athanasiou, R. Schultz, B. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. McComb, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. Brown, C. Burge, L. Cerutti, H. Chen, D. Church, M. Clamp, R. Copley, T. Doerks, S. Eddy, E. Eichler, T. Furey, J. Galagan, J. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. Johnson, T. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. Kent, P. Kitts, E. Koonin, I. Korf, D. Kulp, D. Lancet, T. Lowe, A. McLysaght, T. Mikkelsen, J. Moran, N. Mulder, V. Pollara, C. Ponting, G. Schuler, J. Schultz, G. Slater, A. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. Wolf, K. Wolfe, S. Yang, R. Yeh, F. Collins, M. Guyer, J. Peterson, A. Felsenfeld, K. Wetterstrand, A. Patrinos, M. Morgan, P. de Jong, J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. Chen, and J. Szustakowski, Initial sequencing and analysis of the human genome, *Nature* **409(6822)**, 860 (2001), pMID: 11237011.
- [5] J. Kazazian, HH, C. Wong, H. Youssoufian, A. Scott, D. Phillips, and S. Antonarakis, Haemophilia a resulting from de novo insertion of Ll sequences represents a novel mechanism for mutation in man, *Nature* **332(6160)**, 164 (1988), pMID: 2831458.
- [6] R. K. Pradhan and W. Ramakrishna, Transposons: Unexpected players in cancer, *Gene* **808**, 145975 (2022).
- [7] E. Prak and H. J. Kazazian, Mobile elements and the human genome, *Nat Rev Genet* **1(2)**, 134 (2000), pMID: 11253653.
- [8] E. Havecker, X. Gao, and D. Voytas, The diversity of ltr retrotransposons, *Genome Biol* **5(6)**, 225 (2004), pMID: 15186483.
- [9] Venkatesh and B. Nandini, Miniature inverted-repeat transposable elements (mites), derived insertional polymorphism as a tool of marker systems for molecular plant breeding, *Mol Biol Rep* **47(4)**, 3155 (2020), pMID: 32162128.
- [10] T. Leonardo and S. Nuzhdin, Intracellular battlegrounds: conflict and cooperation between transposable elements, *Genet Res* **80(3)**, 155 (2002), pMID: 12688654.
- [11] J. Brookfield, The ecology of the genome - mobile dna elements and their hosts, *Nat Rev Genet* **6(2)**, 128 (2005), pMID: 15640810.
- [12] N. Kaplan, T. Darden, and C. Langley, Evolution and extinction of transposable elements in mendelian populations, *Genetics* **109(2)**, 459 (1985), pMID: 2982700.
- [13] J. Brookfield, Models of repression of transposition in p-m hybrid dysgenesis by p cytotype and by zygotically encoded repressor proteins, *Genetics* **128(2)**, 471 (1991), pMID: 1649073.
- [14] J. Brookfield, Models of the spread of non-autonomous selfish transposable elements when transposition and fitness are coupled, *Genetics Research* **67(3)**, 199 (1996).
- [15] A. Le, Rouzic and P. Capy, Population genetics models of competition between transposable element subfamilies, *Genetics* **174(2)**, 785 (2006), pMID: 16888345.
- [16] A. Le, Rouzic, T. Boutin, and P. Capy, Long-term evolution of transposable elements, *Proc Natl Acad Sci U S A* **104(49)**, 19375 (2007), pMID: 200.
- [17] C. Xue and N. Goldenfeld, Stochastic predator-prey dynamics of transposons in the human genome, *Phys Rev Lett* **117(20)**, 208101 (2016), pMID: 27886494.
- [18] C. Feschotte, N. Jiang, and S. Wessler, Plant transposable elements: where genetics meets genomics, *Nat Rev Genet* **3(5)**, 329 (2002), pMID: 11988759.
- [19] U. Oggenfuss, T. Badet, T. Wicker, F. E. Hartmann, N. K. Singh, L. Abraham, P. Karisto, T. Vonlanthen, C. Mundt, B. A. McDonald, and D. Croll, A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen, *eLife* **10**, e69249 (2021).
- [20] D. Barro-Trastoy and C. Köhler, Helitrons: genomic parasites that generate developmental novelties, *Trends Genet* **40(5)**, 437 (2024), pMID: 38429198.
- [21] I. Grabundzija, S. Messing, J. Thomas, R. Cosby, I. Bilic, C. Miskey, A. Gogol-Döring, V. Kapitonov, T. Diem, A. Dalda, J. Jurka, E. Pritham, F. Dyda, Z. Izsvák, and Z. Ivics, A helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes, *Nat Commun* **7**, 10716 (2016), pMID: 26931494.
- [22] I. Grabundzija, A. Hickman, and F. Dyda, Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition, *Nat Commun* **9(1)**, 1278 (2018), pMID: 29599430.
- [23] P. Chakrabarty, R. Sen, and S. Sengupta, From parasites to partners: exploring the intricacies of host-transposon



- dynamics and coevolution, *Funct Integr Genomics* **23**(3), 278 (2023), PMID: 37610667.
- [24] M. Shen, E. Raleigh, and N. Kleckner, Physical analysis of tn10- and is10-promoted transpositions and rearrangements, *Genetics* **116**(3), 359 (1987), PMID: 3038673.
- [25] K. Harada, K. Yukuhiro, and T. Mukai, Transposition rates of movable genetic elements in drosophila melanogaster, *Proc Natl Acad Sci U S A* **87**(8), 3248 (1990), PMID: 2158108.
- [26] S. Nuzhdin and T. Mackay, Direct determination of retrotransposon transposition rates in drosophila melanogaster, *Genet Res* **63**(2), 139 (1994), PMID: 8026740.
- [27] D. Suh, E. Choi, T. Yamazaki, and K. Harada, Studies on the transposition rates of mobile genetic elements in a natural population of drosophila melanogaster, *Mol Biol Evol* **12**(5), 748 (1995), PMID: 7476122.
- [28] X. Maside, S. Assimacopoulos, and B. Charlesworth, Rates of movement of transposable elements on the second chromosome of drosophila melanogaster, *Genet Res* **75**(3), 275 (2000), PMID: 10893864.
- [29] A. Sousa, C. Bourgard, L. Wahl, and I. Gordo, Rates of transposition in escherichia coli, *Biol Lett* **9**(6), 20130838 (2013), PMID: 24307531.
- [30] W. F. Eanes, C. Wesley, J. Hey, D. Houle, and J. W. Ajioka, The fitness consequences of p element insertion in drosophila melanogaster, *Genetics Research* **52**, 17–26 (1988).
- [31] B. Charlesworth and C. Langley, The population genetics of drosophila transposable elements, *Annu Rev Genet* **23**, 251 (1989), PMID: 2559652.
- [32] D. Houle and S. V. Nuzhdin, Mutation accumulation and the effect of copia insertions in drosophila melanogaster, *Genetics Research* **83**, 7 (2004).
- [33] E. Pasyukova, S. Nuzhdin, T. Morozova, and T. Mackay, Accumulation of transposable elements in the genome of drosophila melanogaster is associated with a decrease in fitness, *Journal of Heredity* **95**, 284 (2004).
- [34] M. Steiniger, C. D. Adams, J. F. Marko, and W. S. Reznikoff, Defining characteristics of tn 5 transposase non-specific dna binding, *Nucleic Acids Research* **34**, 2820 (2006).
- [35] S. Bire, D. Ley, S. Casteret, N. Mermoud, Y. Bigot, and F. Rouleux-Bonnin, Optimization of the piggybac transposon using mrna and insulators: toward a more reliable gene delivery system, *PLoS One* **8**(12), e82559 (2013), PMID: 24312663.
- [36] I. Querques, A. Mades, C. Zuliani, C. Miskey, M. Alb, E. Grueso, M. Machwirth, T. Rausch, H. Einsele, Z. Ivics, M. Hudecek, and O. Barabas, A highly soluble sleeping beauty transposase improves control of gene insertion, *Nat Biotechnol* **37**(12), 1502 (2019), PMID: 31685959.
- [37] L. Yang and J. L. Bennetzen, Structure-based discovery and description of plant and animal  $\pi$ helitron*si*/i $\pi$ , *Proceedings of the National Academy of Sciences* **106**, 12832 (2009), <https://www.pnas.org/doi/pdf/10.1073/pnas.0905563106>.
- [38] L. Yang and J. L. Bennetzen, Distribution, diversity, evolution, and survival of  $\pi$ helitron*si*/i $\pi$  in the maize genome, *Proceedings of the National Academy of Sciences* **106**, 19922 (2009), <https://www.pnas.org/doi/pdf/10.1073/pnas.0908008106>.
- [39] C. Du, N. Fefelova, J. Caronna, L. He, and H. K. Dooner, The polychromatic  $\pi$ helitron*si*/i $\pi$  landscape of the maize genome, *Proceedings of the National Academy of Sciences* **106**, 19916 (2009), <https://www.pnas.org/doi/pdf/10.1073/pnas.0904742106>.
- [40] W. Xiong, L. He, J. Lai, H. K. Dooner, and C. Du, Helitronscanner uncovers a large overlooked cache of  $\pi$ helitron*si*/i $\pi$  transposons in many plant genomes, *Proceedings of the National Academy of Sciences* **111**, 10263 (2014), <https://www.pnas.org/doi/pdf/10.1073/pnas.1410068111>.
- [41] R. Cordaux, D. J. Hedges, S. W. Herke, and M. A. Batzer, Estimating the retrotransposition rate of human alu elements, *Gene* **373**, 134 (2006).
- [42] E. Bennett, H. Keller, R. Mills, S. Schmidt, J. Moran, O. Weichenrieder, and S. Devine, Active alu retrotransposons in the human genome, *Genome Res* **18**(12), 1875 (2008), PMID: 18836035.
- [43] J. Xing, Y. Zhang, K. Han, A. Salem, S. Sen, C. Huff, Q. Zhou, E. Kirkness, S. Levy, M. Batzer, and L. Jorde, Mobile elements create structural variation: analysis of a complete human genome, *Genome Res* **19**(9), 1516 (2009), PMID: 19439515.
- [44] C. Beck, J. Garcia-Perez, R. Badge, and J. Moran, Line-1 elements in structural variation and disease, *Annu Rev Genomics Hum Genet* **12**, 187 (2011), PMID: 21801021.
- [45] C. Esnault, J. Maestre, and T. Heidmann, Human line retrotransposons generate processed pseudogenes, *Nat Genet* **24**(4), 363 (2000), PMID: 10742098.
- [46] W. Wei, N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, and J. V. Moran, Human l1 retrotransposition: cis preference versus trans complementation, *Molecular and Cellular Biology* **21**, 1429 (2001), PMID: 11158327, <https://doi.org/10.1128/MCB.21.4.1429-1439.2001>.
- [47] E. Bolton, C. Coombes, Y. Eby, M. Cardell, and J. Boeke, Identification and characterization of critical cis-acting sequences within the yeast ty1 retrotransposon, *RNA* **11**(3), 308 (2005), PMID: 15661848.
- [48] M. J. Curcio, S. Lutz, and P. Lesage, The ty1 ltr-retrotransposon of budding yeast,  $\pi$ saccharomyces cerevisiae*si*/i $\pi$ , *Microbiology Spectrum* **3**, 10.1128/microbiolspec.mdna3 (2015), <https://journals.asm.org/doi/pdf/10.1128/microbiolspec.mdna3-0053-2014>.
- [49] T. R. GREGORY, The c-value enigma in plants and animals: A review of parallels and an appeal for partnership, *Annals of Botany* **95**, 133 (2005).
- [50] J. Brookfield, Interspersed repetitive dna sequences are unlikely to be parasitic, *J Theor Biol* **94**(2), 281 (1982), PMID: 7078209.
- [51] B. Charlesworth and D. Charlesworth, The population dynamics of transposable elements, *Genetics Research* **42**(1), 1–27 (1983).
- [52] E. Montgomery, B. Charlesworth, and C. Langley, A test for the role of natural selection in the stabilization of transposable element copy number in a population of drosophila melanogaster, *Genet Res* **49**(1), 31 (1987), PMID: 3032743.
- [53] C. Langley, E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth, On the role of unequal exchange in the containment of transposable element copy number, *Genet Res* **52**(3), 223 (1988), PMID: 2854088.
- [54] A. Delprat, B. Negre, M. Puig, and A. Ruiz, The transposon galileo generates natural chromosomal inversions in drosophila by ectopic recombination, *PLoS One* **4**(11),

- e7883 (2009), pMID: 19936241.
- [55] M. Santana, J. Silva, E. Mizubuti, E. Araújo, B. Condon, B. Turgeon, and M. Queiroz, Characterization and potential evolutionary impact of transposable elements in the genome of *cochliobolus heterostrophus*, *BMC Genomics* **15**(1), 536 (2014), pMID: 24973942.
  - [56] C. F. Voliva, C. L. Jahn, M. B. Comer, I. Hutchison, Clyde A., and M. H. Edgell, The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end, *Nucleic Acids Research* **11**, 8847 (1983), <https://academic.oup.com/nar/article-pdf/11/24/8847/7048814/11-24-8847.pdf>.
  - [57] S. C. Hardies, S. L. Martin, C. F. Voliva, r. Hutchison, C A, and M. H. Edgell, An analysis of replacement and synonymous changes in the rodent L1 repeat family., *Molecular Biology and Evolution* **3**, 109 (1986), <https://academic.oup.com/mbe/article-pdf/3/2/109/11167308/2hard.pdf>.
  - [58] D. Sassaman, B. Dombroski, J. Moran, M. Kimberland, T. Naas, R. DeBerardinis, A. Gabriel, G. Swergold, and H. J. Kazazian, Many human l1 elements are capable of retrotransposition, *Nat Genet* **16**(1), 37 (1997), pMID: 9140393.
  - [59] E. Ostertag and H. J. Kazazian, Twin priming: a proposed mechanism for the creation of inversions in l1 retrotransposition, *Genome Res* **11**(12), 2059 (2001), pMID: 11731496.
  - [60] B. Brouha, J. Schustak, R. Badge, S. Lutz-Prigge, A. Farley, J. Moran, and H. J. Kazazian, Hot l1s account for the bulk of retrotransposition in the human population, *Proc Natl Acad Sci U S A* **100**(9), 5280 (2003), pMID: 12682288.