Evolution, 2023, **XX(XX)**, 1–13
https://doi.org/10.1093/evolut/qpad128
Advance access publication 31 July 2023
Original Article

Critical role of insertion preference for invasion trajectory of transposons

Manisha Munasinghe[1], Nathan Springer[1], Yaniv Brandvain[1,2]

[1]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, United States
[2]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN, United States

Corresponding author: Department of Plant and Microbial Biology, University of Minnesota, 140 Gortner Laboratory, 1479 Gortner Avenue, St. Paul, MN 55108, United States. Email: mmunasin@umn.edu

Abstract

It is unclear how mobile DNA sequences (transposable elements, hereafter TEs) invade eukaryotic genomes and reach stable copy numbers, as transposition can decrease host fitness. This challenge is particularly stark early in the invasion of a TE family at which point hosts may lack the specialized machinery to repress the spread of these TEs. One possibility (in addition to the evolution of host regulation of TEs) is that TE families may evolve to preferentially insert into chromosomal regions that are less likely to impact host fitness. This may allow the mean TE copy number to grow while minimizing the risk for host population extinction. To test this, we constructed simulations to explore how the transposition probability and insertion preference of a TE family influence the evolution of mean TE copy number and host population size, allowing for extinction. We find that the effect of a TE family's insertion preference depends on a host's ability to regulate this TE family. Without host repression, a neutral insertion preference increases the frequency of and decreases the time to population extinction. With host repression, a preference for neutral insertions minimizes the cumulative deleterious load, increases population fitness, and, ultimately, avoids triggering an extinction vortex.

Keywords: transposable elements, invasion trajectory, population extinction

Introduction

Transposable elements (TEs) are mobile repetitive DNA sequences that actively increase their copy number through propagating themselves within genomes. A preeminent example of selfish DNA, TEs have been highly successful at invading eukaryotic genomes (Wicker et al., 2007). TEs likely invade naïve populations via horizontal transfer, where the TE moves into the germline of the recipient population and then spreads throughout the genome as well as the population via vertical transmission (Le Rouzic & Capy, 2005). TEs employ either a copy-and-paste or cut-and-paste mechanism to insert into novel positions and increase their mean copy number within a population. Class I elements, or retrotransposons, use an RNA intermediate that is reverse-transcribed and integrated into a new position in the genome, while Class II elements, or DNA transposons, move via a DNA intermediate (Craig et al., 2015; Feschotte & Pritham, 2007). Insertions that can transpose on their own, as they encode the proteins necessary for transposition, are considered autonomous elements, while nonautonomous elements lack these sequences and consequently rely on autonomous TEs of the same type in order to transpose (Feschotte et al., 2002; Wessler, 2006). Classes consist of both autonomous and nonautonomous elements and can be further divided into subclasses, superfamilies, and families depending on their ancestral origins, sequence similarity, and insertional preferences highlighting the genetic and mechanistic diversity of TEs (Arkhipova, 2017; Kapitonov & Jurka, 2008; Seberg & Petersen, 2009).

TE abundance varies greatly between species and is correlated with genome size (Kidwell, 2002; Wells & Feschotte, 2020). The proportion of the genome occupied by TEs ranges from around 10% in *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), 20% in *Drosophila melanogaster* (Quesneville et al., 2005), and 85% in *Zea mays* ssp. *mays* (Schnable et al., 2009). It is unclear what factors determine not only how much of the genome is occupied by TEs but also how that proportion is distributed between the distinct TE families. Some TE families contribute relatively little to this overall proportion, with only a handful to tens of copies present in a genome, while others contain tens of thousands of copies (Baucom et al., 2009; Diez et al., 2014; Stitzer et al., 2021; Sutton et al., 1984). High TE abundance is surprising given the expectation that novel TE insertions are likely to be deleterious, as they may insert into functional genes, alter heterochromatin formation and gene expression patterns, and induce large structural changes via ectopic recombination (Adrion et al., 2017; Hedges & Deininger, 2007; Hollister & Gaut, 2009; Lee & Langley, 2012).

Species have evolved diverse mechanisms of TE regulation, including chromatin modification, DNA methylation, and posttranscriptional modification followed by degradation (Almeida et al., 2022; Borges & Martienssen, 2015; Cosby et al., 2021; Czech et al., 2018; Yoder et al., 1997). In *Drosophila*, distinct heterochromatic loci, often called piRNA clusters, generate primary antisense PIWI-interacting RNAs (piRNAs) that match actively transposing TEs that happen

Received April 12, 2022; revisions received May 1, 2023; accepted July 24, 2023
Associate Editor: Alexander Suh; Handling Editor: Tracey Chapman

to land within these clusters, resulting in repression of these very same TEs (Brennecke et al., 2007; Halic & Moazed, 2009). Reactivation of transposition can occur if the TE copy located within the piRNA cluster degrades, often leading to cyclical bursts of transposition and repression. In *Zea mays*, a single dominant locus, *Mu killer* (*Muk*), can act to silence the autonomous elements necessary for transposition of the *Mutator* family of TEs (Slotkin et al., 2003), and this silencing persists in plants that did not inherit *Muk*, suggesting epigenetic memory that continues to silence TE transposition (Slotkin et al., 2005). In mammalian genomes, KRAB-ZFPs are thought to silence particular TE families by recognizing and binding TE-specific DNA sequences triggering the formation of repressive chromatin (Imbeault et al., 2017; Yang et al., 2017). TEs themselves can even exhibit self-repression where families auto-regulate transposition (albeit under very specific conditions) (Charlesworth & Langley, 1986; Lohe & Hartl, 1996). Ultimately, transposition regulation differs not only between species but also between TE families creating distinct interactions that manage TE copy number evolution.

In the 1980s, Charlesworth and colleagues developed a series of theoretical population genetic models to explore how the mean copy number per individual for a given TE family could change over time given their potentially deleterious effects. These studies demonstrated that the mean copy number could increase and eventually stabilize at an equilibrium point depending on the extent of transposition and excision, the effective population size of the host population, and the strength of selection against high TE copy numbers in individuals (Charlesworth & Charlesworth, 1983; Charlesworth & Langley, 1986; see Charlesworth et al., 1994 for a summary of the effects of these forces). An equilibrium is reached when either all occupiable sites are filled with TEs or the transposition rate is scaled down such that the number of new TEs each generation matches the number lost to drift, selection, and excision. Recent stochastic simulations have shown that copy number evolution is strongly influenced by recombination rate (Dolgin & Charlesworth, 2006, 2008), the introduction of new TE families via horizontal transfer (Groth & Blumenstiel, 2017), the extent of synergistic epistasis (Charlesworth & Charlesworth, 1983; Choi & Lee, 2020), and the specific architecture underlying transposition regulation (Kelleher et al., 2018; Kofler, 2019, 2020; Lu & Clark, 2010). Many of these models make several notable assumptions including: a constant or inverse relationship between transposition and copy number, which acts as a built in mechanism of TE regulation, a fixed or increasing fitness effect for new insertions, and a fixed or infinite population size. These assumptions are quite standard, and there have been efforts to explore the consequences of relaxing some of these assumptions (see Charlesworth, 1991; Kofler, 2019).

One aspect of TE biology that has largely been overlooked by many population genetic models is the tendency of TE families to preferentially insert into specific DNA sequences or features. Insertion preference is, however, an increasingly important aspect of TE biology as TEs exhibit specific insertion preferences in not only a family manner but also a host dependent manner. The P element in *D. melanogaster* shows strong insertional preference for GC-rich regions near gene promoters, while retrotransposons in several species target sequences upstream or downstream of tRNA genes (Liao et al., 2000; Blanc & Adams, 2004; Spradling et al., 2011; Asif-Laidin et al., 2020). In Drosophila, telomeres are composed of arrays of three specialized non-long terminal repeat (non-LTR) retrotransposons (HeT-A, TART, and TAHRE), and, similarly, the Ty5 retrotransposon in Saccharomyces *cerevisiae* preferentially inserts into heterochromatin found at the telomere (Abad et al., 2004; Boeke & Devine, 1998; Casacuberta 2017; Gao et al., 2008; Novikova, 2009; Pardue & DeBaryshe, 2008). In several different grass species, TEs belonging to a Ty3-derived retrotransposon family are found to exclusively localize at centromeric regions and demonstrate the involvement of TEs in the evolution of centromeres (Langdon et al., 2000). By preferentially inserting into non-genic regions, TE families avoid the deleterious effects associated with inserting into and potentially disrupting a gene.

Insertion preference may not only be the result of structural differences between TEs but also an evolved trait that impacts the expected selective effect of new insertions. The genetic load of TE families that preferentially insert into heterochromatic or intronic regions is expected to be less than those that insert into functional or genic regions. Insertion preference therefore represents not just the nucleotide sequence or feature a TE family inserts into but also the underlying distribution of fitness effects for each new insertion. Charlesworth (1991) developed a deterministic model consisting of two classes of TE insertions, selected against or neutral, and found it difficult to obtain combinations of parameters for transposition, excision, and selection against insertions that matched TE copy numbers observed in *Drosophila* (Charlesworth, 1991). However, this work did not consider the stochastic nature of TE replication nor did this work include the possibility that the genetic load imposed by TEs could decrease population growth rates, ultimately leading to population extinction. Previous stochastic models have assumed either a fixed selective effect or allowed for variable selective effects with a fixed proportion of neutral sites (Kelleher et al., 2018; Kofler, 2019; Lu & Clark, 2010), but, to date, no model has explicitly tested how variation in insertion preference influences TE copy number evolution.

Unconstrained TE transposition would inevitably damage the host genome, as all occupiable sites become filled with TEs disrupting genes and plunging host fitness, but we know comparatively little about how variation in TE biology impacts this process. In the absence of mechanisms that manage or eliminate transposition, a clear expectation is that highly replicative TE families should most rapidly expand in copy number and drive populations extinct. Conversely, TE families that evolve a preference for neutral insertion sites should increase in copy number without growing the deleterious genetic load, potentially allowing the host population to survive for longer. However, since nearly all population genetic models assume either a fixed or infinite host population size, we do not have actual confirmation of these hypotheses or an understanding of how these facets of TE biology influence mean copy number and population size over time.

Here, we use a non-Wright–Fisher framework in SLiM 3 to explore how transposition probability and insertion preference influence the evolution of mean TE copy number and host population size (Haller & Messer, 2019). We consider a naïve diploid population that gains a single copy of a TE in the genome of a single individual (analogous to horizontal transfer). This TE belongs to a unique family with an assigned transposition probability and range of fitness effects for novel insertions that represent insertion preference. TEs transpose and increase their mean copy number in the population over time. We first

Table 1. Summary of models constructed.

Model	Number of chromosomes	Recombination rate	Excision	Nonautonomous elements	Transposition regulation
1	Five	1×10^{-5}	No	No	No
2	Five	1×10^{-5}	No	No	Yes
3	One	1×10^{-8}	No	No	No
4	One	1×10^{-8}	Yes	No	No
5	One	1×10^{-8}	No	Yes	No
6	One	1×10^{-8}	Yes	Yes	No

consider a model with no mechanism for silencing TE proliferation and then contrast those results to a model with a generalizable form of host repression where transposition scales inversely with copy number. If the TE family is not initially lost due to either genetic drift or selection, we can track its spread through the population by measuring the mean copy number and population frequency of the TE family. We allow population size to fluctuate depending on the fitness of individuals, such that populations can go extinct. Consequently, we can relate changes in the mean copy number to the mean fitness of the population to explore how key aspects of TE biology influence the invasion trajectories of TE families and under what conditions populations survive the invasion.

Methods

Model setup

We model a diploid, hermaphroditic population that reproduces sexually in SLiM 3 (v3.6). Simulations begin with a single neutral TE in the genome of a single individual. This TE belongs to a unique family with a specified transposition probability (*teJumpP*) and preference for either neutral or deleterious insertions (*neutP*). We then track the mean per-individual TE copy number, host population size, and mean fitness over time. We extend recipe 14.12 (Modeling transposable elements) in the SLiM manual (Haller & Messer, 2016) by employing a non-Wright–Fisher (nonWF) framework (to allow changes in population size) and incorporating insertion preference (to allow TE insertions to have variable fitness effects). Full details of our model can be found in the supplement (Supplementary Appendix 1).

Modeling genome architecture

We consider a five chromosome model ($L = 5 \times 10^5$ occupiable sites divided equally) with a high uniform recombination rate ($r = 1 \times 10^{-5}$). It is worth clarifying that we do not model the actual nucleotide sequence of TEs or the host genome. Each position in the genome represents an occupiable site with an associated fitness effect if occupied by a TE. New insertions do not change the length of the genome, and we do not track the connections between each new insertion (i.e., which TE a novel TE derived from). As previous population genetic theory has demonstrated the influence of recombination rate on TE accumulation patterns as a result of Hill–Robertson effects and Muller's ratchet (Dolgin & Charlesworth, 2008; Hill & Robertson, 1966; Langley et al., 1988; Muller, 1964), we did consider a genomic architecture consisting of a single chromosome with a low uniform recombination rate, but we relegate commentary and analysis of these models to the supplement (Supplementary Appendix 1).

Modeling TE insertion preference

Insertion preference in our model is not a specific sequence or feature that a TE inserts into but instead an underlying distribution of fitness effects for novel insertions. It adjusts the probability that a novel insertion will be neutral (with probability *neutP*) versus deleterious (with probability 1 – *neutP*). The probability that a selected site is mildly deleterious ($s = -0.005$), modestly deleterious ($s = -0.05$), and massively deleterious ($s = -0.5$) are equal (each is set to [1 – *neutP*]/3). Each position assumes a dominance coefficient of $h = 0.5$ such that the full fitness effect is only realized in individuals homozygous for the TE insertion at that site. Individual final fitness is then calculated multiplicatively across all loci. Neutral and selected sites are placed uniformly across the genome, and their position is held constant across all simulations for that parameter combination.

Modeling TE transposition

The number of novel insertions for a single individual is then drawn from a Poisson distribution dependent on both *teJumpP*—the probability that a single TE copies and inserts itself elsewhere in the genome—and the number of autonomous elements present in the genome. The site for each new insertion is randomly chosen, and the fitness effect of the novel insertion is then determined based on the assigned fitness consequence if a TE inserts into that site (see above). In models without host repression (Table 1), *teJumpP* is fixed for the entirety of a simulation run meaning we do not rescale or limit transposition. We use a simple, but generalizable, model of host repression such that transposition scales inversely with the copy number (see Equation 6b in Charlesworth & Charlesworth 1983). Restated here, $teJumpP = teJumpP_0/(1 + kn)$, where n is the copy number in the individual and k is a scalar constant ($k = 0.05$). This allows us to test effects of transposition regulation on our results with losing generalizability by mimicking the biology underlying any specific form of host repression.

Modeling TE biology

Outside of transposition and insertion preference, we consider two other notable features of TE biology. We consider both the inclusion of nonautonomous elements, which contribute to the total copy number and affect fitness but do not contribute to the expected number of new TE insertions, and the random excision of elements, which not only reduces TE copy number but also increases the chance that the TE family could be lost from the population entirely. These features did not meaningfully influence our results, and we limit discussion of these models to the supplement (Supplementary Appendix 1).

Life cycle, population growth, and selection

We initialize the population with K = 1,000 individuals, with K acting as the hard carrying capacity for the population. The generational life cycle in non-Wright–Fisher models in SLiM starts with the creation of offspring. Each generation, we generate a population that is twice the size of the previous generation ($N_{t+1} = 2N_t$) selecting all parents independently and at random. After removing all individuals from the parental generation, we employ viability selection—where an individual's survival probability is simply its expected fitness—assuming additivity within and multiplicative fitness across loci (see above). If N > K, we randomly cull excess offspring to generate a population of size K (or less). Transposition then occurs in the remaining individuals who survived viability selection. We then loop back to the start of the generational cycle with the surviving offspring forming the new parental pool.

Model outcomes

For each distinct parameter combination, we track the mean fitness and size of the population as well as the mean copy number of the TE family and mean frequency of a TE in the family stratified by their fitness effects over time. We consider three endpoints for a replicate simulation run of a given parameter set.

Outcome 1, TE loss

Loss of the TE family from the host population occurs if the mean copy number of autonomous TEs is zero. No further transposition can occur, and we simply track which generation the TE family was lost in.

Outcome 2, population extinction

If the population size hits zero, we record that replicate as resulting in a population extinction event and output the relevant trajectories (population fitness, size, mean TE copy number, and mean TE frequency over time). Population extinction occurs when no individuals in the population survive after viability selection.

Outcome 3, dual survival

If neither of these options occurs, then both the TE family and the host population have survived to the final generation (capped at 50,000). We output the relevant trajectories (same as though outputted in population extinction) for that simulation run.

Models, parameters, and replicates

We consider six distinct models that varied either genome architecture or specific aspects of TE biology detailed above (Table 1). When describing our results, we use the term TE family to refer to a specific parameter combination of transposition probability (*teJumpP*) and insertion preference (*neutP*). We explore the following sets of values for our parameters: *teJumpP* = [1×10^{-4}, 2.5×10^{-4}, 5×10^{-4}, 7.5×10^{-4}, 1×10^{-3}, 2.5×10^{-3}, 5×10^{-3}, 7.5×10^{-3}, 1×10^{-2}, 2.5×10^{-2}, 5×10^{-2}, 7.5×10^{-2}, 1×10^{-1}] (13 values) and *neutP* = [0.010, 0.025, 0.050, 0.075, 0.10, 0.25, 0.50, 0.75, 0.90, 0.925, 0.950, 0.975, 0.99] (13 values) for each model, resulting in a total of 169 distinct parameter combinations per model type.

Each distinct parameter combination is used for a specific simulation run of a model. We initialize the host genome, TE family, and population and consider three endpoints (TE loss, population extinction, or dual survival) for a replicate of a given parameter set. When one of these outcomes occurs, we record the relevant trajectories and end state before looping back to the initialized state. The simulation run is finally complete (i.e., we no longer loop back to the beginning point) if we record either a combined total of 100 population extinction and dual survival events or 1×10^6 TE loss events.

For a subset of the parameter space, quite high TE copy numbers were achieved. This resulted in dramatically increased run times, we therefore reduced the number of technical replicates in this portion of parameter space (down to either 10 population extinction and dual survival events or 1×10^5 TE loss events) and marked them in all figures (Supplementary Table S1). It was similarly computationally unfeasible to obtain results for high transposition probabilities and neutral insertion preferences for our model with transposition regulation (although we hypothesize from the handful of successfully completed replicates that these parameter combinations allow for a very large number of TEs). We relegate full visualizations of the explored parameter space to the supplement. Scripts for all models can be found on GitHub—see Data availability statement.

Results

A set of models were developed to assess the outcome of introducing a TE, which can replicate freely, into a naive population (example showing our five chromosome models in Figure 1). There are three classes of outcomes for this TE introduction: TE loss, population extinction, and dual survival. We explored the proportion of these three classes of outcomes across the parameter space for transposition probabilities and insertion preferences, contrasting our results with and without host repression. Parameter combinations for which computational limitations limited the number of possible replicates are marked in all figures.

Before proceeding, we invite the reader to predict which outcomes may prevail across the parameter space. We had several initial expectations about the likely outcomes. One expectation is that a preference for neutral insertion sites will protect the host population by minimizing the cumulative deleterious load of the TE family, allowing dual survival instead of population extinction. Another is that the highest observed TE copy numbers will occur under the highest transposition probabilities. Combined, we predicted that a TE family with a high transposition probability and strong preference for neutral insertion sites may be the most effective at reaching high copy numbers without rendering the host population extinct.

TE loss is the predominant outcome

Across parameter space, TE loss is the most common outcome. Within all replicates for a given parameter combination across all models, at least 82% of the simulations result in TE loss (Supplementary Table S2). Approximately, 95% of all TE loss outcomes occur within 50 generations, which we expect is primarily due to genetic drift as the initial TE is neutral and begins with frequency 1/2,000 (0.0005). TE loss is likely the result of both genetic drift and selection. When the recombination rate is low, background selection against deleterious TEs causes the removal of linked neutral TEs. While we do observe parameter combinations exclusively resulting in TE loss in our single chromosome model (Supplementary Figure S1), we do not observe any regions of exclusive TE loss in our five chromosome models with

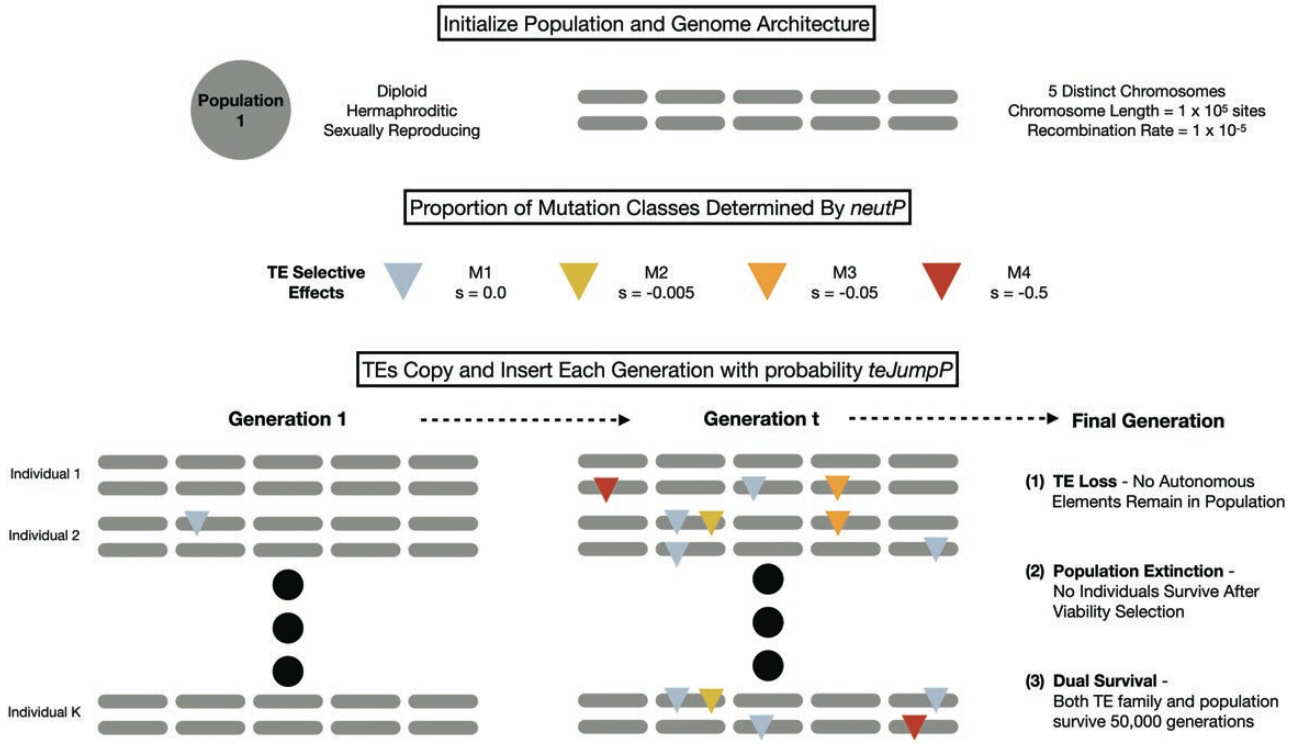

Figure 1. Visual summary of five chromosome model. A visual representation of the five chromosome model design detailed that highlights the population initialization and the role of the two key parameters, *teJumpP* and *neutP*.

higher recombination, as recombination allows neutral TEs to recombine away from deleterious TEs (Figure 2A). Both excision and the inclusion of nonautonomous elements increase the proportion of TE loss outcomes and expand this region of exclusive TE loss (Supplementary Figure S1D and F).

After the introduction of a single TE copy into a naïve population, the expected proportion of populations in which the TE is lost (i.e., no copies of the TE family remain) as the transposition rate tends to 0 are either $1 - 2u$ in an infinite population (where u represents the transposition probability) or $1 - 1/2N$ in a finite population (Kaplan et al., 1985; Le Rouzic & Capy, 2005). These relations can be easily extended to obtain the chance of TE establishment (defined in our simulations as the combined proportion of dual survival and population extinction outcomes). While these predications are inexact for our model, they provide useful references for the expected proportion of TE loss versus TE establishment. We find that, with low transposition (which essentially precludes population extinction), the proportion of TE establishment outcomes is approximately 1/2N (Figure 2E—facet 1.0e-04). Under higher transposition probabilities, we find the 1/2N approximation more appropriate when the neutral insertion preference is low while the $2u$ approximation is more appropriate under high neutral insertion preferences (Figure 2E, Supplementary Figure S2).

TE family dynamics influence trajectories of extinction and survival

TE introductions that do not result in TE loss can be divided into population extinction or dual survival outcomes. The proportion of these outcomes was visualized across the parameter space of TE transposition probability and insertion preference (Figure 2C and D, Supplementary Figure S3). We remind the reader that the proportion of outcomes that are observed in our models depends on the maximum number of generations simulated (50,000). Most TE loss events occur after very few generations, while extinction events occur over a wider range of times. Fitness trajectories for dual survival outcomes (which by definition end at 50,000 generations) indicate that many likely would have gone extinct, especially without host repression, if we had extended the simulations for more generations (Supplementary Figure S4). While this highlights the dependency of the proportion of population extinction and dual survival events on the generational time limit (i.e., with sufficient time most dual survival outcomes would turn into extinctions) and the impact of computational limitations in shaping our results, it also highlights that in some regions of parameter space populations invaded by TEs can survive for a long time.

Models without host repression demonstrated a substantially greater number of population extinction events relative to dual survival. This on its own is not necessarily surprising; however, the first transition from dual survival to population extinction occurs both as we increase the transposition probability and when the preference for neutral insertion sites is high (Figure 2B). This suggests that in the absence of host repression, a preference for neutral insertion sites may not be beneficial.

High neutral insertion preference accelerates population extinction in the absence of host repression

Without host repression, high neutral insertion preferences not only increased the probability of population extinction but also resulted in more rapid population extinctions (Figure 3A, Supplementary Figure S5). If we fixed the transposition

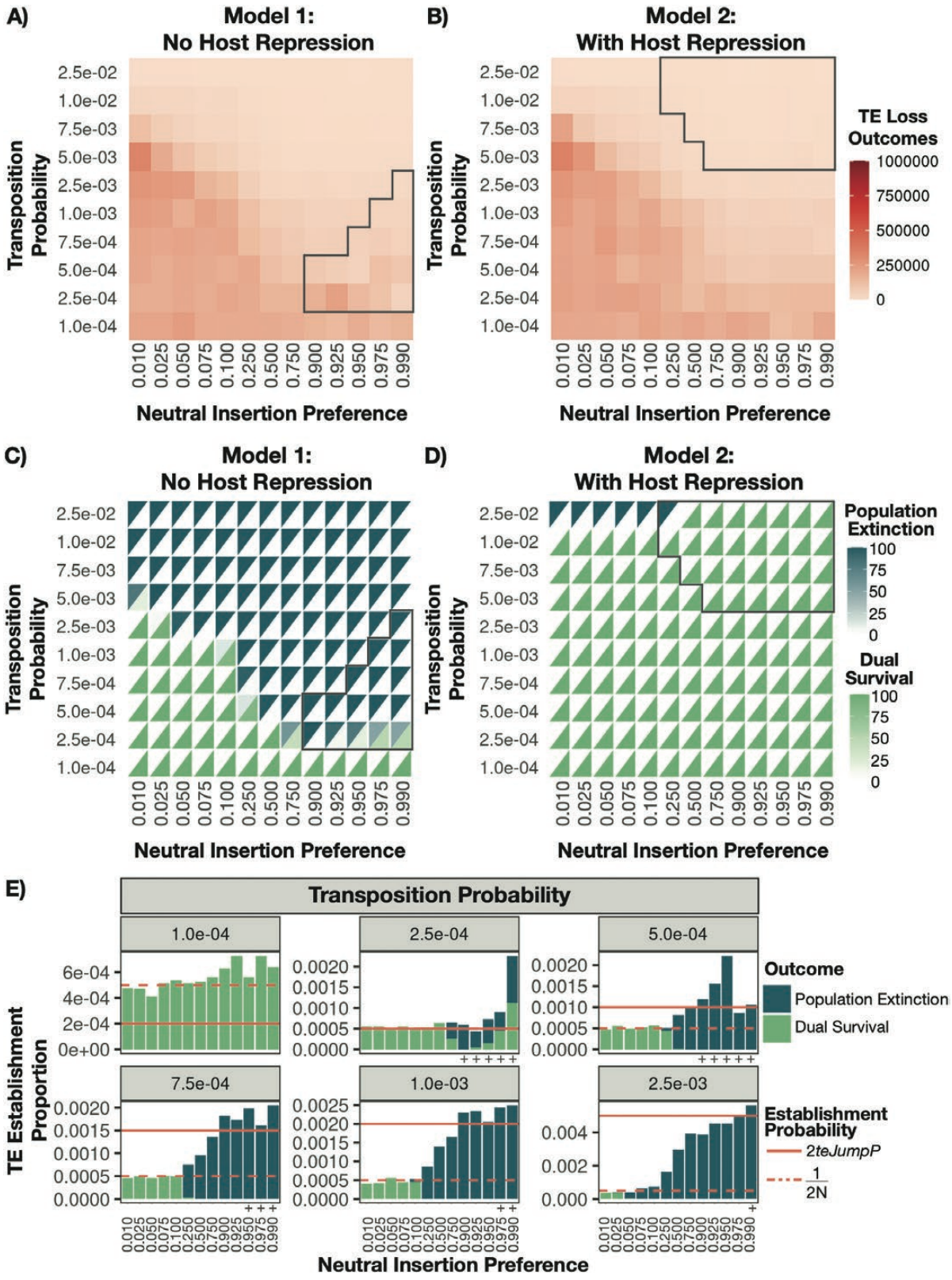

Figure 2. Outcome across our parameter space. The left column visualizes outcomes for our five chromosome model without repression, and the right column shows outcomes for our five chromosome model with repression. The top row (A and B) shows heatmaps colored with the number of transposable element (TE) introductions that resulted in loss of the TE family. The bottom row (C and D) shows heatmaps colored with both the number of TE introductions that resulted in population extinction (upper triangle) or dual survival (lower triangle). For each heatmap, rows represent the transposition probability with increasing probabilities as you move upwards, and columns represent the neutral insertion preference with increasing preferences as you move to the right. The final row (E) highlights the proportion of non-TE loss (dual survival in green, population extinction in blue) for Model 1 for a subset of transposition probabilities. Lines indicate two different establishment probabilities. Cells indicated with gray markers (gray outline in A–D, gray cross in E) represent parameter combinations that, due to computational constraints, had reduced limits detailed in the Methods.

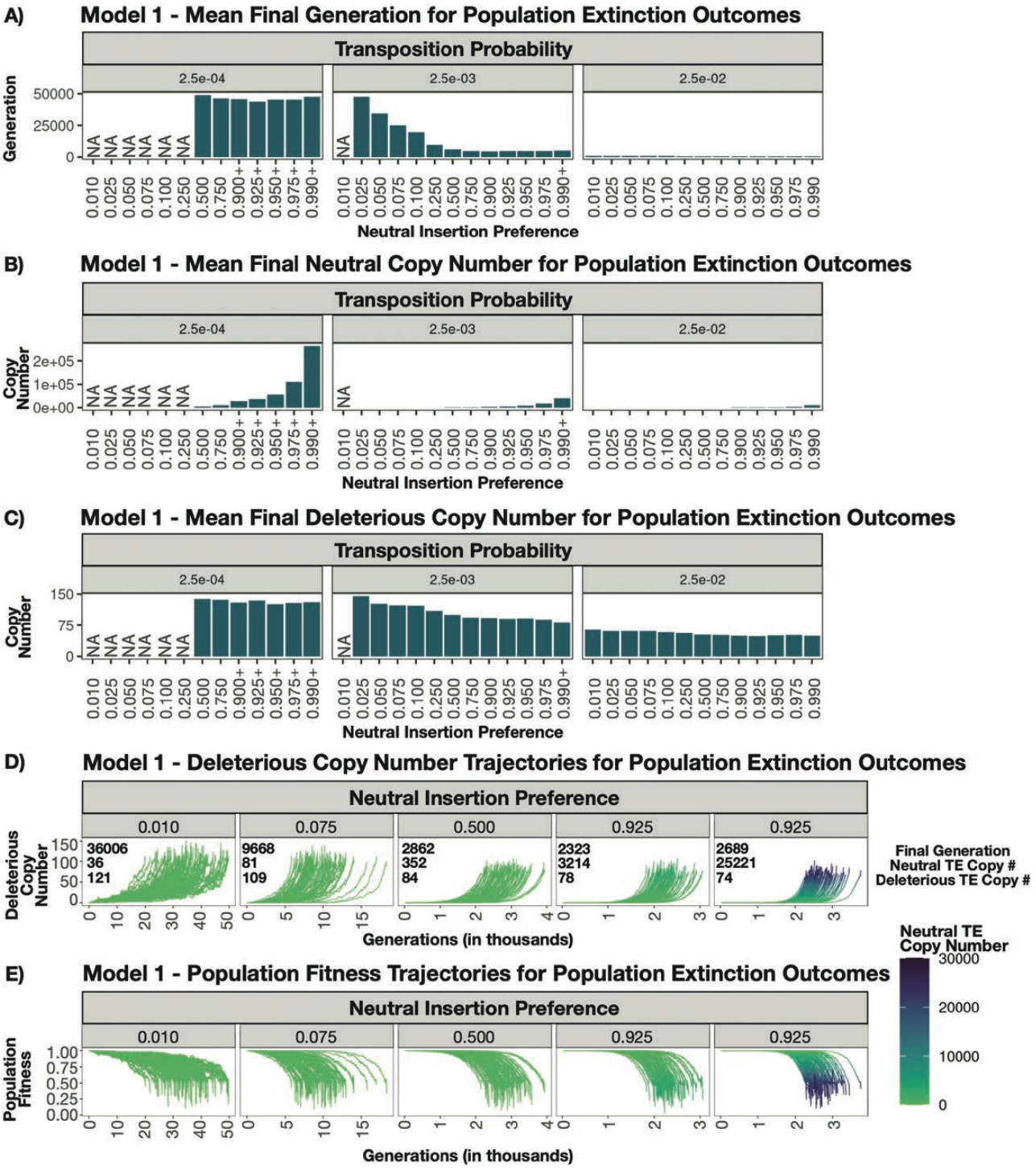

Figure 3. Key dynamics of population extinction outcomes in the five chromosome model. We compared population extinction outcomes across our parameter space by looking at the mean (A) generation with which the population went extinct, (B) neutral transposable element (TE) copy number in this final generation, and (C) deleterious TE copy number in the final generation across replicates for low, medium, and high transposition probabilities (*teJumpP* = 2.5×10^{-4}, 2.5×10^{-3}, and 2.5×10^{-2} respectively). NAs indicate combinations where no population extinction outcomes were observed. The x-axis shows increasing neutral insertion preferences. Gray crosses in A–C indicate parameter combinations with scaled down replicate counts. We then fix the transposition probability at *teJumpP* = 5.0×10^{-3} and visualize the change in (D) the deleterious TE copy number and (E) population mean fitness over time for a subset of neutral probabilities for observed extinction outcomes across our model without host repression (Model 1). We facet these graphs by insertion preference. Note that, the x-axis is not fixed across these plots.

probability, we found that, as we increased the preference for neutral insertions, there was an exponential decline in the time to population extinction which eventually stabilized for our highest insertion preferences. Higher transposition probabilities actually had lower neutral copy numbers (e.g., *teJumpP* = 2.5×10^{-2}, *neutP* = 0.9 ~ 908 neutral copies) than more moderate ones (e.g., *teJumpP* = 2.5×10^{-3}, *neutP* = 0.9 ~ 3,924 neutral copies) (Figure 3B, Supplementary Figure S6),

but the deleterious copy number was more similar (e.g., ~49 to ~92) (Figure 3C, Supplementary Figure S7), reflecting the maximal genetic load met before extinction.

We find that the copy number of deleterious TEs stays low until a critical threshold of TEs is reached. Populations then enter an extinction vortex, where the deleterious copy number grows exponentially causing a matched exponential decline in population fitness (Figure 3D and E). The TE copy number required to trigger this process depends on both the transposition probability and insertion preference. High transposition probabilities require fewer present TEs to generate high numbers of novel insertions, which explains the lower copy numbers observed in populations immediately before extinction under high transposition (Figure 3B and C). Higher preferences for neutral insertions require more TE copies since a smaller portion of novel insertions in the next generation are expected to be deleterious. Since selection acts on deleterious TE insertions, a strong preference for neutral TE insertions allows populations to reach this critical threshold more rapidly as neutral TEs accumulate and provide a source for higher numbers of insertions in the next generation. This explains the tendency to observe shorter times to extinction under high neutral insertion preferences as well.

Contrary to the expectation that a higher preference for neutral insertion sites protects hosts by reducing the deleterious load incurred by insertions, we find that, in the absence of host repression, preferentially inserting into neutral sites increases the risk of population extinction, eliminating both host and TE. While this result may be counterintuitive at first, it makes sense given the design of our model and is consistent with results from Charlesworth (1991). For a fixed transposition probability, the number of expected insertions for a given TE copy number is the same across insertion preferences. It is the proportion of neutral to deleterious TE insertions that varies with insertion preferences. Higher neutral insertion preferences mean the expected proportion of novel insertions that are expected to be neutral is greater, but it does not fully eliminate the chance of observing deleterious insertions. When the number of novel deleterious insertions each generation is small, selection can purge them from the population; however, if the number of deleterious insertions increases each generation at a rate faster than selection can remove them, the population will enter into an extinction vortex as the deleterious copy number grows exponentially. A preference for neutral insertion sites allows TE families to more rapidly reach high copy numbers guaranteeing an increasing number of deleterious TE insertions in subsequent generations, which ultimately results in population extinction.

Dual survival can result in variable final TE copy number and allele frequency

Dual survival rarely occurs in our models without host repression, and it is often limited to the lowest transposition probabilities. With host repression, dual survival becomes common across our parameter space except for when the transposition probability is high and the neutral insertion preference is low (Supplementary Figure S2). Higher TE copy numbers are generally more achievable in our model without host repression (Figure 4A); however, they almost always result in population extinction. Without host repression (Model 1), the median copy number across all population extinction outcomes is 587.0 with a standard deviation of 9,907.0, while dual survival outcomes have a median copy number of 9.2 with a standard deviation of 6,244.5 (Supplementary Figure S8). Similarly, with host repression (Model 2), the median copy number across all population extinction outcomes is 118.0 with a standard deviation of 3,055.7, while dual survival outcomes have a median copy number of 135.1 with a standard deviation of 3,438.9. This is driven by the fact that, under host repression, transposition is regulated before the extinction vortex can be triggered, resulting in higher copy number dual survival outcomes. Preferentially inserting into neutral sites is associated with higher final copy numbers across all models (Supplementary Figure S9). In contrast to the case without host repression (above), an increased neutral insertion preference does not necessarily drive populations extinct when transposition is regulated. When both hosts and transposons survive, the final allele frequency distribution is distinctly bimodal (Figure 4B). Once a TE fixes at a single locus (i.e., the right tail of this bimodal distribution), the TE family cannot be lost from the genome without population extinction.

Transposition regulation facilitates dual survival by reducing the genetic load

Regulation of TE transposition is ubiquitous across all species; however, the underlying mechanism employed within species often varies. We use a generalizable, but simple, model to assess the impact transposition regulation has on our results. Because we do not allow for excision in our model of TE regulation (Model 2), only selection and drift decrease the number of neutral TEs. TE copy number initially increases linearly (Supplementary Figure S10) until the copy number is sufficiently high to limit transposition at which point the change in copy number approaches 0, which suggests an copy number equilibrium point is reached (Supplementary Figure S11). So long as this equilibrium occurs before host population extinction, deleterious mutations can be removed by selection as quickly as they are introduced, resulting in a stable number of deleterious mutations and a stable population fitness (Figure 4C and D, Supplementary Figures S12 and S13). If we fix the transposition probability, we find that the ratio of neutral to deleterious TEs increases (Figure 4C). This results in more fit populations housing larger TE copy numbers (Figure 4D). Here, the expected benefits of preferentially inserting into neutral sites are realized. We could not obtain enough replicates to explore what was occurring in regions of high transposition and high neutral insertion preferences, but we expect, based on preliminary explorations, that dual survival occurs with a very large number of persisting TEs.

Discussion

We initially hypothesized that a preference for neutral insertion sites would allow a TE family to increase in copy number while also minimizing the cumulative deleterious load of the TE family. Consequently, higher copy numbers could be achieved without reducing population fitness or contributing to extinction. Surprisingly, we found that a preference for neutral insertion site was only advantageous in the presence of host repression. Without transposition regulation, preferentially inserting into neutral sites allows the TE copy number to grow rapidly until a critical copy number threshold is reached. This threshold is dependent on the parameter combination and represents the point with which the number of deleterious insertions each generation increases faster

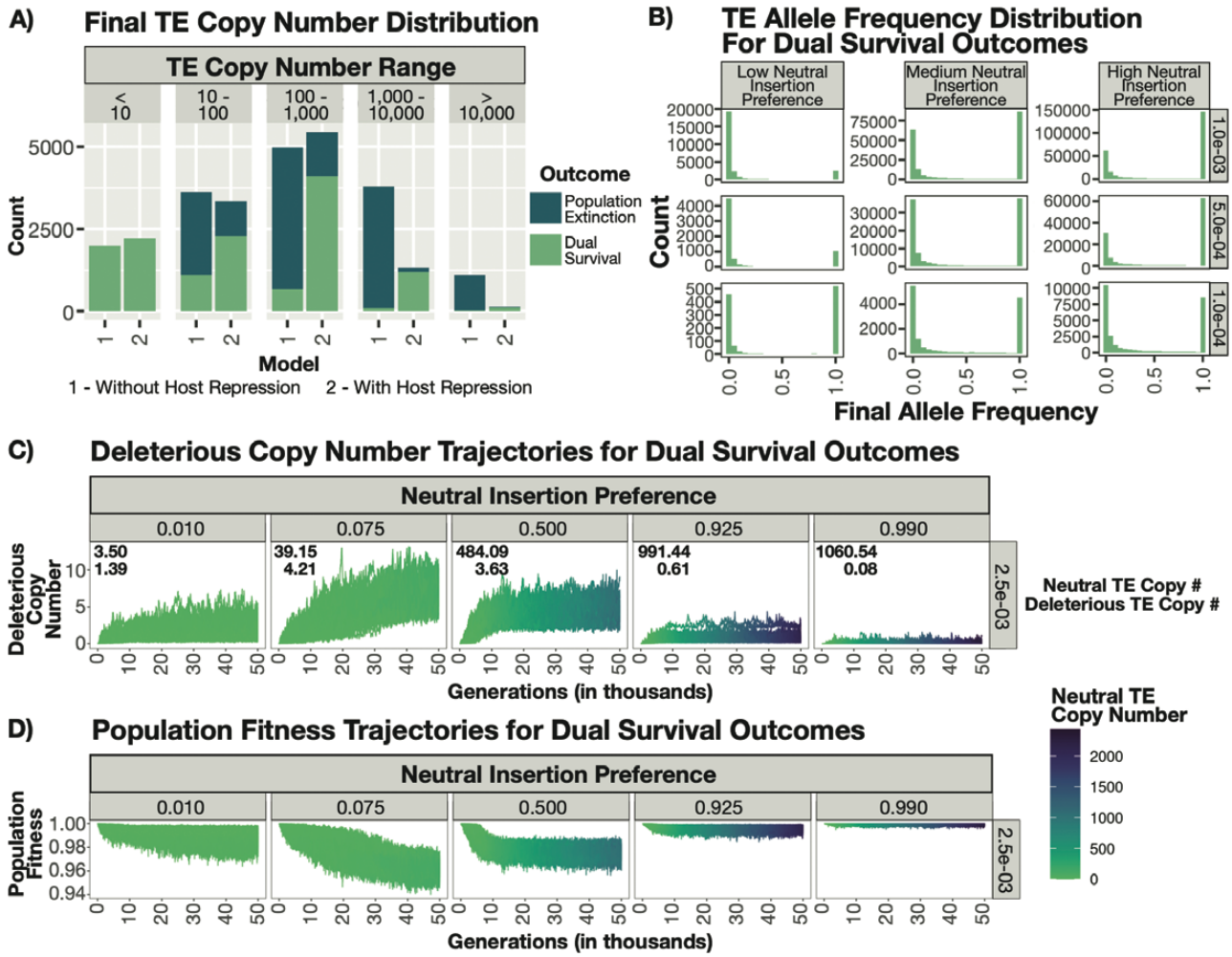

Figure 4. End states observed across our five chromosome models. (A) Shows the distribution of the mean final transposable element (TE) Copy Number of the host population. We bin the final mean copy number for a given replicate into one of four ranges (facet). The x-axis shows results for our models without and with host repression, and the y-axis shows the number of outcomes across our parameter space (colored by whether those outcomes resulted in population extinction [blue] or dual survival [green]). (B) Shows the distribution of final allele frequencies for every specific TE (i.e., for a site in our genome occupied by a TE how many individuals also carry a TE at that site) combined across our models without and with host repression. We bin the neutral insertion preference into either low (0.01–0.075), moderate (0.1–0.9), or high (0.925–0.99) and once again subset three transposition probabilities that span the range of dual survival outcomes. The x-axis of each plot shows the allele frequency, and the y-axis shows how many TEs had that frequency in the final generation. We then fix the transposition probability at *teJumpP* = 2.5×10^{-3} and visualize the change in (D) the deleterious TE copy number and (E) population mean fitness over time for a subset of neutral probabilities for observed dual survival outcomes across our model with host repression (Model 2). We facet these graphs by insertion preference.

than selection can remove them. At this point, the population has entered into an extinction vortex. We observe these population extinction events more often and more rapidly under high transposition probabilities and neutral insertion site preferences. Importantly, our results hold even after including random excision of TEs and nonautonomous elements. Host repression ultimately limits transposition before this critical threshold is reached, which results in dual survival across most of our parameter space.

Heterogeneity in the selective effect of TE insertions has previously been shown to result in high copy number equilibriums that exceed naturally observed copy number (Charlesworth, 1991). Our results build upon this work and demonstrate that, while preferentially inserting into neutral sites does allow the copy number to grow larger, more rapidly, this can put the population at risk if transposition is not restricted. With host repression, high copy numbers can not only be achieved (which is beneficial from the TE's perspective) but also population fitness is comparatively higher as the ratio of neutral to deleterious TE insertions is stronger (which is beneficial to the host population). Dual survival, in the absence of host repression, is characterized by very low copy numbers as it only occurs under the lowest transposition probabilities. While host repression leads to comparatively lower copy numbers for a given parameter combination, it keeps the copy number below the critical threshold resulting in dual survival. Higher copy numbers (>1,000) with host repression tend to occur for greater transposition probabilities (>2.5×10^{-3}) with additional increases in copy number as you increase the preference for neutral insertion sites (Supplementary Figure S10). Studies in *D. melanogaster* have characterized the transposition probability across several TE families and suggest it may range anywhere from 1×10^{-5} to 1×10^{-3} under normal conditions and up to 1×10^{-1} under stressed or dysgenic conditions. Our results align well with previous theory and experimental work; however, it does not necessarily explain why

some families have only a handful of copies in the genome and others have tens of thousands of copies.

Because TEs are characterized as selfish genetic elements seeking to proliferate regardless of their effects on the host (Dawkins, 1976; Orgel & Crick, 1980; Werren et al., 1988), we might expect high copies of transposable elements in most families, as found by previous theory (Charlesworth, 1991). We find that dual survival outcomes with higher copy numbers (>10,000) occur rarely across our parameter space and models. We only observe it in our model with host repression for the highest transposition probabilities. Interestingly, we observe lower copy numbers across our extant populations, which aligns with empirical observations. For example, Stitzer et al. (2021) found that 95% of TE families in the maize genome had less than 10 copies, while only 1.2% of present families exhibited copy numbers greater than 100. Similarly, only four TE families in the *Arabidopsis thaliana* genome have more than 1,000 copies (Ahmed et al., 2011; Quesneville, 2020). These observations lead Stritt et al. (2021) to challenge the idea of TEs as "invasive" genetic elements, suggesting instead that TEs may have evolved strategies to persist at low copy numbers (Stritt et al., 2021). Perhaps, more TE families have evolved moderate to lower transposition probabilities. While this limits their copy number, it does protect the TE family from potentially triggering an extinction vortex if transposition regulation was in any way compromised.

Our model of transposition regulation is built on the first model of transposition regulation (Charlesworth & Charlesworth, 1983), in which transposition scales inversely with copy number. This general model of TE regulation does not directly correspond to any of the diverse known mechanisms for transposition regulation in nature. To date, the best mechanistic models of host repression are built from knowledge of the piRNA pathway studied in *Drosophila* (Groth & Blumenstiel, 2017; Kelleher et al., 2018; Kofler, 2019, 2020; Lu & Clark, 2010). These models clearly demonstrate that host repression facilitates TE copy number expansion by not only limiting transposition but also by mitigating the effects of deleterious TE insertions. Our results generally align with these studies and demonstrate that both TE and host population persistence occur more successfully in the presence of transposition regulation. However, specific results of any such mechanistic model likely depend on the mechanism, and we encourage additional modeling of alternative forms of transposition regulation. Variation in the mechanism underlying host repression may influence which TE families reach high copy number in extant populations and, consequently, merits thorough evaluation.

While modeling known biological mechanisms grounds theory in natural reality, it can limit our understanding of alternative or historical phenomena, and it is therefore important to understand what happens in the absence of transposition regulation. For example, hybridization between populations has been shown to compromise evolved TE regulation mechanisms triggering widespread TE proliferation resulting in a "genomic shock" (McClintock, 1984). Reduced hybrid fitness caused by uncontrollable TE proliferation could act as an isolating mechanism and is an interesting future direction, as most models focus on TE proliferation within a single population. Additionally, while nearly all modern populations exhibit host repression, it is unclear whether it has always been this way. Ancestral TE invasions may have occurred before mechanisms for regulating TE transposition were established. Our work consequently contributes and raises additional questions about these ancestral TE invasions and historical population extinction events.

Our models consider two distinct host genetic architectures (five chromosomes with high recombination or one chromosome with low recombination), key biological features of TE families (random excision of elements, nonautonomous elements, both, or neither), and the presence or absence of host repression. We find that low recombination results in higher proportions of TE loss, especially when the preference for neutral insertion sites is low, in part due to the effects of Muller's ratchet. Random excision and the inclusion of nonautonomous elements do not impact the general trends of our results, but they do make it more difficult for the TE family to initially invade which increases the amount of TE loss events observed. Host repression is the most impactful feature, ultimately resulting in dual survival across most of our parameter space. Missing from our considered genomic architecture is the possibility that a TE insertion could facilitate adaptation (Li et al., 2018). The adaptive potential of TEs has been demonstrated and incorporating it in to future models is warranted, as adaptive insertions could stave off population extinction (Studer et al., 2011). Additionally, we did not consider the nesting feature of certain TE families, where TEs insert into other TE sequences (Jedlicka et al., 2019; SanMiguel et al., 1996, 1998). Nested TE insertions are expected to be neutral, as they are not inserting into host sequence. While we did not explicitly model this, one could imagine that TE families that prefer nested insertions may be shifting their underlying distribution of fitness effects to be more neutral.

Our results highlight the unique role of insertion preference of TE families. The benefit of preferentially inserting into neutral sites is only realized in the presence of host repression. Consequently, a feature that we expected to allow the TE copy number to increase while minimizing the deleterious load, ultimately only provides a benefit if that TE family can no longer transpose which caps its copy number evolution. In the absence of transposition regulation, a high neutral insertion preference is not capable of stabilizing TE copy numbers and actively expedites population extinction. Our simplistic model of host repression suggests that a thorough exploration of the diverse mechanisms underlying host repression likely influences which specific TE families proliferate within genomes. Overall, our results raise interesting evolutionary questions about historical extinction events and the establishment of host repression. What timeframe must host repression "activate" to avoid population extinction, how did host repression evolve in ancestral populations, and were their indirect selective forces wiping out historical TE populations due to extinction events? Population size fluctuations and extinction risk have been mostly ignored from population genetic models of TE copy number evolution. A relevant exception is Kofler (2019) that not only found high transposition probabilities resulted in extinction but also that successful TE invasions occurred in a very narrow portion of their parameter space (Kofler, 2019). Future simulation frameworks which incorporate additional aspects of TE biology and diverse mechanisms of host repression will be crucial in understanding how TE copy numbers increase without resulting in population extinction.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

The scripts for all simulations, job submissions, and figure visualizations can be found on GitHub: https://github.com/mam737/InsertionPreference_TEs. Simulation results can be found at Dryad: doi:10.5061/dryad.k0p2ngfdj.

Author contributions

M.M., Y.B., and N.S. conceived the study and designed the theoretical framework. M.M. incorporated the framework into SLiM and wrote all associated scripts. M.M. analyzed and visualized the results. M.M. wrote the first draft and all authors contributed to the writing of the manuscript. N.S. and Y.B. supervised the project.

Conflict of interest: The author have no conflict of interests to declare.

Funding

This material is based upon work supported by the National Science Foundation Postdoctoral Research Fellowship in Biology under Grant No. #2010908.

Acknowledgments

We would like to thank Candice Hirsch, Jeff Ross-Ibarra, Emily Josephs, Michelle Stitzer, Arvid Ågren, Andrew Read, Erika Magnusson, and four anonymous reviewers for their useful comments, insights, and suggestions regarding model development and the interpretation of results.

References

Abad, J. P., De Pablos, B., Osoegawa, K., De Jong, P. J., Martín-Gallardo, A., & Villasante, A. (2004). TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila telomeres*. *Molecular Biology and Evolution*, 21(9), 1620–1624. https://doi.org/10.1093/molbev/msh180

Adrion, J. R., Song, M. J., Schrider, D. R., Hahn, M. W., & Schaack, S. (2017). Genome-wide estimates of transposable element insertion and deletion rates in *Drosophila melanogaster*. *Genome Biology and Evolution*, 9(5), 1329–1340. https://doi.org/10.1093/gbe/evx050

Ahmed, I., Sarazin, A., Bowler, C., Colot, V., & Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Research*, 39(16), 6919–6931. https://doi.org/10.1093/nar/gkr324

Almeida, M. V., Vernaz, G., Putman, A. L. K., & Miska, E. A. (2022). Taming transposable elements in vertebrates: From epigenetic silencing to domestication. *Trends in Genetics*, 38, 529–553.

Arkhipova, I. R. (2017). Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mobile DNA*, 8, 19. https://doi.org/10.1186/s13100-017-0103-2

Asif-Laidin, A., Conesa, C., Bonnet, A., Grison, C., Adhya, I., Menouni, R., Fayol, H., Palmic, N., Acker, J., & Lesage, P. (2020). A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration upstream of tRNA genes. *EMBO Journal*, 39(17), e104337.

Baucom, R. S., Estill, J. C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J. -M., Westerman, R. P., SanMiguel, P. J., & Bennetzen, J. L. (2009). Exceptional diversity, non-random distribution, and rapid evolution of Retroelements in the B73 maize genome. *PLoS Genetics*, 5(11), e1000732. https://doi.org/10.1371/journal.pgen.1000732

Blanc, V. M., & Adams, J. (2004). Ty1 insertions in intergenic regions of the genome of *Saccharomyces cerevisiae* transcribed by RNA polymerase III have no detectable selective effect. *FEMS Yeast Research*, 4(4–5), 487–491. https://doi.org/10.1016/S1567-1356(03)00199-5

Boeke, J. D., & Devine, S. E. (1998). Yeast retrotransposons: Finding a nice quiet neighborhood. *Cell*, 93(7), 1087–1089. https://doi.org/10.1016/s0092-8674(00)81450-6

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, 16(12), 727–741. https://doi.org/10.1038/nrm4085

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in drosophila. *Cell*, 128(6), 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043

Casacuberta, E. (2017). Drosophila: Retrotransposons making up telomeres. *Viruses*, 9(7), 192. https://doi.org/10.3390/v9070192

Charlesworth, B. (1991). Transposable elements in natural populations with a mixture of selected and neutral insertion sites. *Genetics Research*, 57, 127–134.

Charlesworth, B., & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetics Research*, 42, 1–27.

Charlesworth, B., & Langley, C. H. (1986). The evolution of self-regulated transposition of transposable elements. *Genetics*, 112(2), 359–383. https://doi.org/10.1093/genetics/112.2.359

Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371(6494), 215–220. https://doi.org/10.1038/371215a0

Choi, J. Y., & Lee, Y. C. G. (2020). Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genetics*, 16(7), e1008872. https://doi.org/10.1371/journal.pgen.1008872.

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., & Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, 371(6531), eabc6405. https://doi.org/10.1126/science.abc6405.

Craig, N., Chandler, M., Gellert, M., Lambowitz, A., Rice, P. A., & Sandmeyer, S. B. (2015). *Mobile DNA III* (3rd ed.). American Society for Microbiology (ASM).

Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., & Hannon, G. J. (2018). piRNA-guided genome defense: From biogenesis to silencing. *Annual Review of Genetics*, 52, 131–157. https://doi.org/10.1146/annurev-genet-120417-031441

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Diez, C. M., Meca, E., Tenaillon, M. I., & Gaut, B. S. (2014). Three groups of transposable elements with contrasting copy number dynamics and host responses in the maize (*Zea mays* ssp. *mays*) genome. *PLoS Genetics*, 10(4), e1004298. https://doi.org/10.1371/journal.pgen.1004298

Dolgin, E. S., & Charlesworth, B. (2006). The fate of transposable elements in asexual populations. *Genetics*, 174(2), 817–827. https://doi.org/10.1534/genetics.106.060434

Dolgin, E. S., & Charlesworth, B. (2008). The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*, 178(4), 2169–2177. https://doi.org/10.1534/genetics.107.082743

Feschotte, C., Jiang, N., & Wessler, S. R. (2002). Plant transposable elements: Where genetics meets genomics. *Nature Reviews Genetics*, 3(5), 329–341. https://doi.org/10.1038/nrg793

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, 41, 331–368. https://doi.org/10.1146/annurev.genet.40.110405.090448

Gao, X., Hou, Y., Ebina, H., Levin, H. L., & Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin.

Genome Research, 18(3), 359–369. https://doi.org/10.1101/gr.7146408
Groth, S. B., & Blumenstiel, J. P. (2017). Horizontal transfer can drive a greater transposable element load in large populations. *Journal of Heredity*, 108(1), 36–44. https://doi.org/10.1093/jhered/esw050
Halic, M., & Moazed, D. (2009). Transposon silencing by piRNAs. *Cell*, 138(6), 1058–1060. https://doi.org/10.1016/j.cell.2009.08.030
Haller, B. C., & Messer, P. W. (2016). SLiM: An evolutionary simulation framework. http://benhaller.com/slim/SLiM_Manual.pdf
Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, 36(3), 632–637. https://doi.org/10.1093/molbev/msy228
Hedges, D. J., & Deininger, P. L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation Research*, 616(1–2), 46–59. https://doi.org/10.1016/j.mrfmmm.2006.11.021
Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, 8(3), 269–294.
Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, 19(8), 1419–1428. https://doi.org/10.1101/gr.091678.109
Imbeault, M., Helleboid, P. -Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, 543(7646), 550–554. https://doi.org/10.1038/nature21683
Jedlicka, P., Lexa, M., Vanat, I., Hobza, R., & Kejnovsky, E. (2019). Nested plant LTR retrotransposons target specific regions of other elements, while all LTR retrotransposons often target palindromes and nucleosome-occupied regions: In silico study. *Mobile DNA*, 10, 50. https://doi.org/10.1186/s13100-019-0186-z
Kapitonov, V. V., & Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics*, 9(5), 411–412; author reply 414. https://doi.org/10.1038/nrg2165-c1
Kaplan, N., Darden, T., & Langley, C. H. (1985). Evolution and extinction of transposable elements in Mendelian populations. *Genetics*, 109(2), 459–480. https://doi.org/10.1093/genetics/109.2.459
Kelleher, E. S., Azevedo, R. B. R., & Zheng, Y. (2018). The evolution of small-RNA-mediated silencing of an invading transposable element. *Genome Biology and Evolution*, 10(11), 3038–3057. https://doi.org/10.1093/gbe/evy218
Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115(1), 49–63. https://doi.org/10.1023/a:1016072014259
Kofler, R. (2019). Dynamics of transposable element invasions with piRNA clusters. *Molecular Biology and Evolution*, 36(7), 1457–1472. https://doi.org/10.1093/molbev/msz079
Kofler, R. (2020). piRNA Clusters need a minimum size to control transposable element invasions. *Genome Biology and Evolution*, 12(5), 736–749. https://doi.org/10.1093/gbe/evaa064
Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J. W., Thomas, H., Jones, R. N., & Jenkins, G. (2000). Retrotransposon evolution in diverse plant genomes. *Genetics*, 156(1), 313–325. https://doi.org/10.1093/genetics/156.1.313
Langley, C. H., Montgomery, E., Hudson, R., Kaplan, N., & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research*, 52(3), 223–235. https://doi.org/10.1017/s0016672300027695
Le Rouzic, A., & Capy, P. (2005). The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. *Genetics*, 169(2), 1033–1043. https://doi.org/10.1534/genetics.104.031211
Lee, Y. C. G., & Langley, C. H. (2012). Long-term and short-term evolutionary impacts of transposable elements on Drosophila. *Genetics*, 192(4), 1411–1432. https://doi.org/10.1534/genetics.112.145714
Li, Z. -W., Hou, X. -H., Chen, J. -F., Xu, Y. -C., Wu, Q., González, J., & Guo, Y. -L. (2018). Transposable elements contribute to the adaptation of Arabidopsis thaliana. *Genome Biology and Evolution*, 10(8), 2140–2150. https://doi.org/10.1093/gbe/evy171
Liao, G., Rehm, E. J., & Rubin, G. M. (2000). Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 97, 3347–3351.
Lohe, A. R., & Hartl, D. L. (1996). Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. *Molecular Biology and Evolution*, 13(4), 549–555. https://doi.org/10.1093/oxfordjournals.molbev.a025615
Lu, J., & Clark, A. G. (2010). Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Research*, 20(2), 212–227. https://doi.org/10.1101/gr.095406.109
McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, 226(4676), 792–801. https://doi.org/10.1126/science.15739260
Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1, 2–9.
Novikova, O. (2009). Chromodomains and LTR retrotransposons in plants. *Communicative and Integrative Biology*, 2(2), 158–162. https://doi.org/10.4161/cib.7702
Orgel, L. E., & Crick, F. H. (1980). Selfish DNA: The ultimate parasite. *Nature*, 284(5757), 604–607. https://doi.org/10.1038/284604a0
Pardue, M. -L., & DeBaryshe, P. G. (2008). Drosophila telomeres: A variation on the telomerase theme. *Fly (Austin)*, 2(3), 101–110. https://doi.org/10.4161/fly.6393
Quesneville, H. (2020). Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mobile DNA*, 11, 28. https://doi.org/10.1186/s13100-020-00223-x
Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Computational Biology*, 1(2), e22. https://doi.org/10.1371/journal.pcbi.0010022
SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y., & Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nature Genetics*, 20(1), 43–45. https://doi.org/10.1038/1695
SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., & Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274(5288), 765–768. https://doi.org/10.1126/science.274.5288.765
Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., … Wilson, R. K. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science*, 326(5956), 1112–1115. https://doi.org/10.1126/science.1178534
Seberg, O., & Petersen, G. (2009). A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature Reviews Genetics*, 10(4), 276–276. https://doi.org/10.1038/nrg2165-c3
Slotkin, R. K., Freeling, M., & Lisch, D. (2003). Mu killer causes the heritable inactivation of the mutator family of transposable elements in *Zea mays*. *Genetics*, 165(2), 781–797. https://doi.org/10.1093/genetics/165.2.781
Slotkin, R. K., Freeling, M., & Lisch, D. (2005). Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature Genetics*, 37(6), 641–644. https://doi.org/10.1038/ng1576
Spradling, A. C., Bellen, H. J., & Hoskins, R. A. (2011). Drosophila P elements preferentially transpose to replication origins. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15948–15953.
Stitzer, M. C., Anderson, S. N., Springer, N. M., & Ross-Ibarra, J. (2021). The genomic ecosystem of transposable elements in maize. *PLoS Genetics*, 17(10), e1009768. https://doi.org/10.1371/journal.pgen.1009768
Stritt, C., Thieme, M., & Roulin, A. C. (2021). Rare transposable elements challenge the prevailing view of transposition dynamics in plants. *American Journal of Botany*, 108(8), 1310–1314. https://doi.org/10.1002/ajb2.1709

Studer, A., Zhao, Q., Ross-Ibarra, J., & Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature Genetics*, 43(11), 1160–1163. https://doi.org/10.1038/ng.942

Sutton, W. D., Gerlach, W. L., Peacock, W. J., & Schwartz, D. (1984). Molecular analysis of ds controlling element mutations at the adh1 locus of maize. *Science*, 223(4642), 1265–1268. https://doi.org/10.1126/science.223.4642.1265

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796–815.

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, 54, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

Werren, J. H., Nur, U., & Wu, C. I. (1988). Selfish genetic elements. *Trends in Ecology and Evolution*, 3(11), 297–302. https://doi.org/10.1016/0169-5347(88)90105-X

Wessler, S. R. (2006). Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences*, 103, 17600–17601.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982. https://doi.org/10.1038/nrg2165

Yang, P., Wang, Y., & Macfarlan, T. S. (2017). The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends in Genetics*, 33(11), 871–881. https://doi.org/10.1016/j.tig.2017.08.006

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, 13(8), 335–340. https://doi.org/10.1016/s0168-9525(97)01181-5