DYNAMICS OF SELF-ELIMINATING GENE DRIVE MECHANISMS

A Dissertation

by

JOSEF ZAPLETAL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mark A. Lawley
Co-Chair of Committee,	Zach N. Adelman
Committee Members,	Andrew S. Natsios
	Natarajan Gautam
Head of Department,	Mark A. Lawley

August 2019

Major Subject: Industrial Engineering

ABSTRACT

Vector-borne diseases continue to be endemic in many parts of the world with dengue, malaria, chikungunya, and yellow fever affecting millions of people every year. Recent advances in genetic engineering, such as CRISPR, have allowed for faster and cheaper DNA modification in organisms with potential to suppress the ability to transmit or carry these pathogens. Additionally, gene drive mechanisms that increase the inheritance rates of transgenic DNA have been proposed, which enable the release of very few transgenic organisms to be capable of transforming entire wild populations. The results of such actions could be irreversible with long-term consequences unknown.

Methods to remove transgene DNA have been explored in crops systems and human gene therapy applications. However, such DNA self-elimination mechanisms have not yet been considered to control highly active gene drive transgenes. Here we explore the coupling of three potential gene drive mechanisms (CRISPR, MEDEA, and underdominance) and a proposed self-eliminating mechanism with system dynamics modeling. Our results identify effective parameter spaces for the complete removal of transgenic DNA and restoration of wild-type alleles for all three gene drive mechanisms. Combining gene drive approached with a self-elimination mechanism could allow testing the effects of transgenic populations on the environment, preventing the long-term persistence of the transgene in nature.

To my family, friends, and mentors.

ACKNOWLEDGEMENTS

I feel extremely fortunate for having crossed paths with some of the most inspirational and selfless mentors who have been critical in my successes. I would like to express my deepest appreciation to my committee chair, Dr. Mark Lawley, and my co-chair, Dr. Zach Adelman, for giving me the opportunity to work under their support and guidance. I am grateful for your commitment to continuously challenging me to grow as a researcher and as a person. I would also like to show appreciation to my committee members, Professor Andrew Natsios and Dr. Natarajan Gautam, for their input and support.

I am equally grateful to Dr. Madhav Erraguntla and Jose Vazquez for their perspective and the creative discussions. Thank you for helping me develop into the person I've become.

Lastly, I would also like to thank my family for their mental support. To my parents, thank you for believing in me and pushing me to keep improving. To Sarah, I am grateful for the sacrifices she has made and for her continued patience with me through this process.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Mark Lawley (chair) of the Department of Industrial and Systems Engineering, Dr. Zach Adelman (co-chair) of the Department of Entomology, Dr. Natarajan Gautam of the Department of Industrial and Systems Engineering and Professor Andrew Natsios of the George Bush School of Government. This dissertation is an extension of previous work conducted by former student Neda Najmitabrizi (M.S. 2017).

Funding Sources

Graduate study was supported by the Scowcroft Institute of International Affairs at the George Bush School of Government at Texas A&M University.

This work was also made possible in part by the Defense Health Program (DHP) under Small Business Innovative Research (SBIR) Contract No. W81XWH-15-C-0158 (IPAS) and by the Western Gulf Center of Excellence for Vector-Borne Diseases Grant No. 1U01CK000512-01.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the awarding offices.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Infectious diseases continue to plague human and domestic animal populations across the world. Vector-borne diseases, such as those spread by mosquitoes, fleas, and ticks, account for 17% of all infectious diseases and are endemic through the world[1–5]. Many mosquito-borne diseases such as malaria (spread by female Anopheles mosquitoes), dengue, chikungunya, and Zika virus (spread by female Aedes mosquitoes), impact nearly half of the human population[6–9]. In 2017, malaria alone caused an estimated 219 million infections resulting in 435,000 deaths[10]. Furthermore, increasing temperatures has expanded the suitable ranges for many vector species, exposing more people to the threat of the transmission of these diseases[11–15].

Advances in genome editing, have allowed for an insertion or removal of particular genes to influence the expression of various traits within an organism. In particular, the discovery and advance of clustered regularly interspace palindromic repeats (CRISPR) and the CRISPR-associated protein (Cas9) system has allowed for the modification of precise segments of DNA to be conducted cheaper and more rapidly than previously possible[16–20].

The genetic modification of an organism has been proposed as a potential solution to stop or limit vector-borne diseases in endemic areas by impacting the organism's ability to carry and transmit disease[21–23]. Studies targeting mosquito species to prevent the transmission of malaria and other mosquito-borne pathogens, such yellow fever and

dengue have shown promising results[24–26]. Other approaches have focused on reducing the population size of vector species through the release of sterile individuals[27] or genetically engineered individuals that have reduced reproductive or survival rates[28,29]. While both techniques are successful in reducing the number of individuals, the effects only persist for a few generations before the natural population is restored[30–33]. Hence, these techniques must be applied multiple times to increase their effectiveness over time.

Other approaches focused on the propagation of transgenes throughout a population have also been developed. Gene drives (GD) are selfish inheritance mechanisms that allowed for the deviation from Mendelian inheritance rates by increasing the likelihood that certain traits are passed from the parents to progeny. As a result, these mechanisms ensure that a transgene is able to persist[34]. In heterozygous individuals, homing drive constructs, such as CRISPR/Cas9-based gene drives, work by creating a double stranded break (DSB) at the same location on the opposing chromosome and through homology directed repair (HDR), copying the transgenic material into the break location, using the transgenic chromosome as a template[35–38]. This effectively creates transgenic homozygous individuals from the heterozygous gametes inherited from the parents. Since HDR is not ensured following a DSB, repair by NHEJ can result in resistant alleles, impairing the effectiveness of the GD[38–40]. Other GD mechanisms such as maternal-effect dominant embryonic arrest (MEDEA)[41,42] and maternal-effect lethal underdominance (UD)[43] utilize a maternally-linked toxin to ensure only offspring inheriting the transgenic DNA will be viable.

Gene drive mechanisms are designed to spread and persist in a population. While gene drives are currently strictly limited to confined laboratory settings, several concerns must be addressed before a release of organisms containing gene drive constructs into the environment can occur. Due to the permanent nature of gene drives, an environmental release of gene drive carrying organisms may be irreversible with the consequences unknown. The regulation of such biotechnologies must work to address the adverse risks a release of these organisms would have on the environment; however, without environmental releases, the full effects may not be observed.

Several studies have been conducted on removal of transgenic DNA sequences from the pollen and seed of transgenic plants to prevent the spread of transgenic DNA as a result of cross-fertilization of closely related wild or nontransgenic plants[44]. Similarly, applications in human gene therapy represent an emerging new gene delivery technology with the ability facilitate scarless transgenic excision, reducing chances of deleterious reintegration[45,46].

Excision of DNA segments has been demonstrated through the use of transposable elements, recombinases, and single-strand annealing (SSA) following two nuclease-induced DSB. Excision facilitated by transposases such as piggyBac and p-elements enable the complete removal of DNA segments located on the transposon between two inverted terminal repeats (ITR)[45–50]. Through recombination, such as the Cre/loxP and FLP/FRT systems used in recombinase-mediated cassette exchange (RMCE), the recombinase binds and removes the segment of DNA between the two target sites leaving behind a single target site[51–55]. Lastly, homing endonucleases can be utilized to cause a

DSB at two target sites flanked by direct repeats (DR), excising the DNA between cuts, with single-strand annealing (SSA) rejoining the DNA[56,57]. While these studies have demonstrated the removal of transgenic material, these have not been studied in combination with GD systems.

The regulation of biotechnologies such as gene drives must strike a balance between the benefits and hazards of utilizing this emerging technology. While the use of gene drives may enable the suppression of disease or conservation of endangered species, there is uncertainty pertaining to how transgenic organisms will behave in the wild and what impacts these organisms may have on the health of the environment [58]. Furthermore, since gene drives are designed to persist in the environment, the abilities to contain gene drive carrying organisms and stop or reverse the spread of the gene drive must be considered prior to any release into the environment[58–60].

By introducing a genetic construct with both GD and SEM functionality, we explored the possible impacts of the CRISPR, MEDEA, and UD gene drive mechanisms on the wild-type population when paired with an SEM construct. We explore effective elimination rates that allow for the transgenic DNA to drive into the population with a subsequent restoration of the wild-type population within a short time frame. Our findings show that by using an SEM to remove an introduced gene drive construct, the pervasiveness of the gene drive can be limited and the wild-type population can be restored without depending on the failure of the drive construct.

Literature Review

Gene Drives

The concept of a transgene "driving" throughout a population was first proposed by Curtis in 1968[34], where the use of chromosomal translocations would allow for the spread of wanted traits (such as the inability to transmit disease in mosquitoes) and upon reaching fixation in the population, would protect the area from immigrants. Drive mechanisms utilizing homing endonuclease genes (HEGs) propose the use of drive constructs inducing a DSB on the same location on the opposing chromosome and copying the transgenic material[39,61]. With the discovery of CRISPR, CRISPR-based gene drives have been achieved in fruit flies[62] and mosquitoes [35,36]. However, the CRISPR-based gene drive is limited by the formation of resistance alleles if successful HDR does not occur[63]. To reduce the rate of resistance allele formation, Noble et al. proposed a multiple gRNA gene drive, allowing for multiple chances for a successful HDR to occur, which ensured a stronger and more permanent drive of the transgene. A different approach demonstrated by Kyrou et al.[64] aimed to reduce the presence of resistance alleles by targeting the *doublesex* (dsx) gene in *Anopheles gambiae* mosquitoes. As a result, the formation of resistance alleles and homologous transgene expression disrupted the development of females, but not males. Since there was no disruption in males, the drive spread through the population, biasing against females, ultimately decreasing egg production and eliminating the entire laboratory cage population. With high rates of inducing a DSB and HDR, both approaches are capable of achieve gene drive with the introduction of very few transgenic individuals.

Other gene drive mechanisms spread by causing mortality in the offspring through parentally-linked toxins. The MEDEA construct utilizes a maternally-linked toxin, where the drive elements are spread by causing mortality to offspring of MEDEA-carrying females that do not inherit the transgene (antidote)[65,66]. Applications and modeling for *Drosophila* indicated an initial release of 25% of homozygous MEDEA-bearing males was sufficient to induce a drive of the transgene within 20 generations [42]. The spread of MEDEA gene drives by migration between populations is further explored by Marshall et al. [41], indicating the MEDEA gene drive may be locally contained when a very high fitness cost (50%) is applied to MEDEA-bearing individuals, with 1% of the population migrating every generation.

Toxin-antidote gene drives, such as those utilized by maternal-effect lethal underdominance (UD), use a combination of two toxin-antidote pairs, where the toxin is located on one allele with the antidote located on the another[43]. Both toxins are maternally-linked, with progeny requiring the antidote in order to achieve zygotic rescue. Two-population modeling presented by Akbari et al.[43] indicate this gene drive is containable within a single population when migration rates are below 4% per generation. If migration rates are increased above this threshold, the drive is lost from the both populations since the migration of wild-type individuals into the transgenic release population results in the frequency of transgenic individuals falling below the threshold required for a successful drive[43].

Gene drives such as CRISPR, MEDEA, and UD allow for transgenes to "drive" into the natural populations with the release of transgenic individuals. While the thresholds

for achieving gene drive in a wild-type population vary across mechanisms, the introduction of an adequate number of gene drive organisms could transform entire species if not contained.

Self-Eliminating Mechanisms

The growing public concern over the use of genetic modification has prompted the evaluation of methods to effectively remove transgenic segments of DNA from the genome of genetically modified organisms. Deletion of transgenic sequences has been demonstrated through transposition, recombination, and single-strand annealing.

Transposons or transposable elements are segments of DNA that are capable of moving between locations on a strand of DNA, facilitated by transposases binding to inverted terminal repeats (ITR), leaving a scarless excision site. P elements in *Drosophila,* have been shown to be high efficient and precise in site specific insertion and excision[50,67]. The piggyBac transposable element, isolated from the cabbage looper moth, has been demonstrated in a variety of agricultural[47–49] and human stem cell applications[46,68–70], to show that piggyBac transposons can also facilitate precision insertions and excisions. Furthermore, excision competent/integration defective piggyBac transposases have been developed to facilitate reversible transgenesis without harmful transposon reintegration and have the potential to improve the safety and utility of delivery methods for therapeutic genes[45].

Recombination allows for the removal of DNA segments located between two recombinase binding sites. Unlike transposition, which results in a scarless excision site,

recombination leaves behind a single target. Recombination has been allowed for targeted genome manipulation in *Aedes aegypti* through recombinase mediated cassette exchange (RMCE) utilizing Cre and FLP recombinases[51,71] with uses in vector control and pathogen transmission.

Several applications of recombinases in plants have been evaluated to limit the expression of traits to necessary parts of the plant to create hybrid transgenic plants[44,54]. Using Cre/lox mediated recombination, marker genes and redundant transgenes were removed from pest resistant (*Bacillus thuringiensis*) rice lines[72] and limited the expression of the pest resistant protein to the green tissues[73]. The efficiency of marker genes excision using Cre/lox mediated recombination has also been explored in barley showing cold temperatures were most efficient for increasing recombination frequency[52]. The use of the Cre recombination was also highly efficient method for marker gene excision in *Arabidopsis*[53,74,75]. Similarly, the combination of the Cre/lox and FLP/FRT systems was capable of producing non-transgenic pollen and seeds from transgenic tobacco plants[55] and aspen[76] at increased efficiencies. In the same way, the use of Bxb1 recombinase in switchgrass has demonstrated the excision of transgenes[77]. By limiting the expression of the transgene in the pollen of reproducing plants, the gene dispersal and subsequent cross pollination with closely related species can be reduced. Furthermore, restricting the transgene to non-consumable parts of the plant, the safety concerns regarding genetically modified food may be mitigated and increase consumer confidence[44].

The excision of transgenes homing endonucleases (HEs) occurs when two DNA cuts flanked by direct repeats are repaired through single-strand annealing (SSA). Guided

by the direct repeats, SSA reconnects the ends of the DNA, resulting in the loss of the genetic material between the two cuts. This approach of genetic excision has been demonstrated in yeast[78,79] and *Aedes aegypti*[56,57], showing that the removal of transgenic DNA was achievable through SSA.

Through these self-eliminating mechanisms, specific segments of transgenic material can be removed from the DNA of an organism. Additionally, the use of these mechanisms can allow for the creation of hybrid organisms, only expressing the transgene in certain locations.

Modeling

The models for the evaluation of gene drive mechanisms can be characterized into two main groups: small-scale stochastic models and large-scale deterministic models. One of the earliest deterministic gene drive models was demonstrated by Austin Burt in 2003, modeling a large, randomly mating adult population[39]. To find equilibrium points within the HEG drive construct, Burt evaluated the impacts of transmission-ratio distortion (TRD) to identify feasible fitness costs necessary to produce a viable drive into the wild-type population[39]. This approach has formed the basis for many frameworks modeling the dynamics of various gene drive constructs, including MEDEA[42], X-shredders[40], and CRISPR/Cas9 systems[80,81].

Expansion of these models for CRISPR/Cas9-mediated gene drive reversal strategies through the use of synthetic resistance alleles, reversal drives, and immunizing reversal drives discuss the importance of fitness costs and timing of countermeasures for

stopping the HEG drives[82]. Multiplexing gRNAs for the CRISPR/Cas9-based homing systems, Noble et al. developed deterministic models for the evolution of a large population where a drive construct was introduced[63]. The theoretical results demonstrated the use of five gRNAs for the homing drive would be capable of reducing the rate of resistance allele formation, resulting in the fixation of the transgenic drive construct within the population[63]. Marshall et al. further modified the model structure by dividing the mosquito life cycle into four life stages (egg, larva, pupa, and adult), modeled both male and female adults, and utilized a discrete daily time step[83]. The model utilized density-independent mortality rates for all life stages, with an additional density-dependent mortality rate applied to the larval stage, limiting the growth of the population. The results of this modeling effort show the population suppression is expected to be short-lived, despite high rates of homing in the germline[83].

Due to the computationally intensive nature of stochastic models, environmental releases of gene drive organisms have relied heavily on the use of deterministic models to capture the magnitude of individuals involved. In this way, stochastic modeling has been utilized to support results from deterministic modeling and exploring other possible outcomes of smaller scale laboratory experiments. Using both deterministic and stochastic two-population models, Marshall and Hay consider the dynamics in both small and infinitely large, randomly mating populations for toxin-antidote and killer-rescue gene drive systems[41]. According to this modeling framework, Akbari et al. explore single- and two-locus configurations of maternal-effect lethal underdominance to analyze the spread of the transgene into neighboring populations[43]. To support experimental results,

stochastic models were developed and run multiple times for small populations with a maximum of 650 *Anopheles gambiae* mosquitoes, seeding 25% of the starting population with *doublesex* gene drive carrying males[64]. In this model, random mating was assumed. To limit the model population size, a sample of 650 eggs were randomly selected at each generational step in the model[64].

The type and structure of the model that is used in modeling can determine the level of detail that is captured. Since deterministic assume homogeneity throughout a population or subgroup of a population, the level of detail that is capture is limited to a high level. Stochastic models, such as agent-based models, allow for individual behaviors and traits. However, because deterministic models do not look at the behaviors of individuals, large-scale models are possible with significantly less computational resources than stochastic models. Since deterministic models follow series of differential equations, the mathematical representation of the population may result in extreme cases that would be likely removed in stochastic models due to randomness.

In this study, we utilize a deterministic model to capture the dynamics of an infinitely large population that consists of juvenile and adult populations with equiprobable mating rates between all genotypes to capture the dynamics of SEM-GDs when released into a wild-type population.

CHAPTER II

CRISPR-SEM

In heterozygous individuals, an active CRISPR gene drive introduces a double-stranded DNA break at the homologous position on the wild-type chromosome at probability *q*, with homology-dependent repair resulting in copying the transgenic material from the transgenic chromosome at probability *p*. Thus, heterozygous individuals with one transgenic allele and one wild-type allele are converted into transgenic homozygous individuals at probability *qp*. Based on previous work on highly active CRISPR/Cas9-based gene drives [63], we set *q* and *p* accordingly for all scenarios ($q = p = 0.95$). Where break repair does not occur via HDR, the formation of resistance alleles that preserve (δ) or disrupt ($1-\delta$) the function of the target gene with corresponding impacts on fitness of the organism are formed, denoted as *u* and *r*, respectively.

We modified existing models for homing-based gene drives by considering three possibilities that might occur prior to any homing events in the germline of the target organism. Successful self-elimination is given as (α), and results in the formation of an allele (*v*) that is resistant to the homing drive. We also considered that any SEM will break down and become permanently non-functional with probability (γ). If neither of these events occur, the transgene remains as is with probability (β), where $\alpha + \beta + \gamma = 1$. The six different alleles considered in the deterministic model are shown in Figure 1 and the structure and probabilities of our model and their relation to the various alleles are shown in Figure 2.

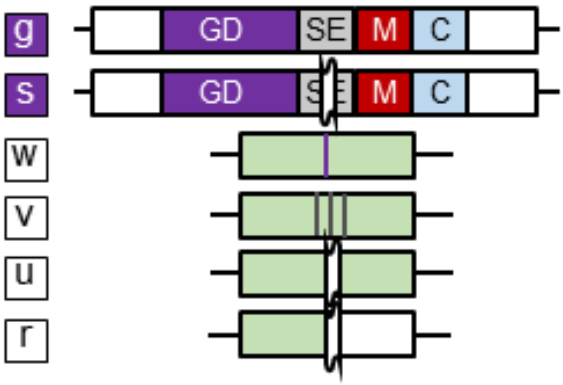

Figure 1. Six different allele types considered by our deterministic model. Two alleles contain the CRISPR/Cas9 gene drive (GD), Marker (M), Cargo (C) and either functional (g) or defective (s) self-elimination mechanism. Four alleles are free of transgenic sequences and include wild-type, CRISPR-susceptible (w), wild-type, pre-determined CRISPR-resistant (v), CRISPR-resistant in frame (u), CRISPR-resistant out-of-frame (r).

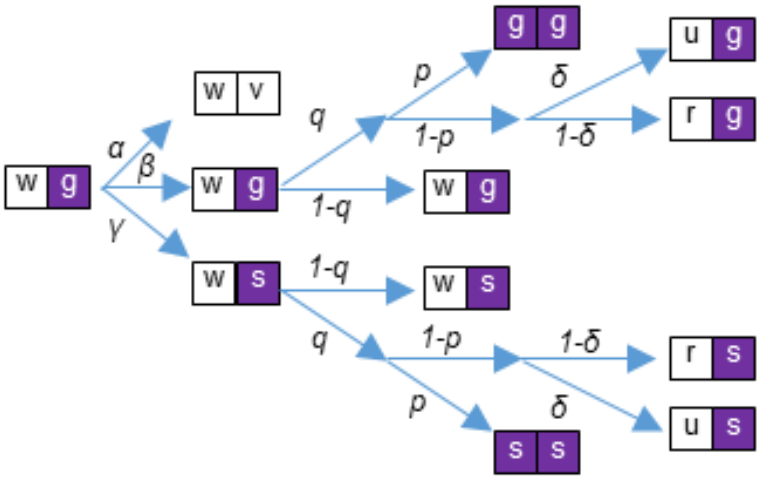

Figure 2. Structure and probabilities of the deterministic model and relation to the different alleles

CRISPR Single Self-Elimination gRNA Target Site with Single Homing Drive gRNA Target Site

By pairing a self-elimination mechanism with the CRISPR gene drive, we explore the possibility of stopping and reversing the effects of the homing drive propagating through the population. With a single gRNA target sight for successful excision, the SEM can only undergo one imperfect excision or mutation (with probability γ) before the transgene is permanently fixed. Likewise, the Cas9-induced homing drive can only undergo a single DSB repair by NHEJ before the homing drive target gRNA target site is lost, resulting in the formation of a resistance allele (*u* or *r*).

Methodology

Model Structure

For each of the gene drive mechanisms, we developed a system of delayed differential equations to predict the number of offspring generated during each time step. The results were generated by running system dynamics models utilizing the differential equations generated. Malthusian population growth was assumed with a daily time step through the models. The model structure was adapted for each SEM-GD scenario based on the work conducted by Najmitabrizi[84]. Differential equations were concatenated and analyzed using MATLAB 2017b. Plots were generated using Python 3.7.

The system dynamics models returned the number of adult and juvenile individuals of each genotype for every time step throughout the simulation. Initial model parameters are provided in Table 1.

Table 1. Variable definitions

Variable	Description	Value
λ	Female reproduction rate (per day)	7
σ	Proportion of female offspring	0.5
c_i	Fitness cost of genotype *i*	Varies
μ_A	Adult mortality rate (per day)	0.3
μ_J	Juvenile mortality rate (per day)	0.03
η	Development time (in days)	12

Using the fitness costs associated with each genotype and sex, adult and juvenile mortality rates were adjusted such that the mortality rate could not be more than 1, giving us:

$$\mu_{A_i} = \frac{\mu_A}{(1-c_i)} \text{ for } (1-c_i) \geq \mu_A, \text{ otherwise } \mu_A = 1$$

$$\mu_{J_i} = \frac{\mu_J}{(1-c_i)} \text{ for } (1-c_i) \geq \mu_J, \text{ otherwise } \mu_J = 1$$

Mortality rates were applied at each time step, where the surviving number adult individuals of each genotype $A_i(T)$ was calculated by reducing the number of adult individuals of each genotype at the previous time step $A_i(T-1)$ by the mortality rate, such that:

$$A_i(T) = (1-\mu_{A_i})A_i(T-1)$$

Juvenile mortality was applied at the time the juveniles became adults, where the number of juvenile individuals surviving the development period was defined as:

$$J_i(T-\eta)(1-\mu_{J_i})^{\eta}$$

Combining the surviving adults with the fully developed juveniles (also now adults), the number of adults with a particular genotype at time *T* can be defined as the number of adults surviving a single time increment (from time *T-1*) and the number of surviving juveniles (from time $T-\eta$), such that:

$$A_i(T) = (1-\mu_{A_i})A_i(T-1) + J_i(T-\eta)(1-\mu_{J_i})^{\eta}$$

The number of females with a particular genotype F_i was directly used in calculating the number of offspring produced. Since males do not directly produce offspring, the proportion of adult males with a particular genotype M_i was calculated such that:

$$M_i = A_{M_i}\sum_{i=1}^{n}\frac{1}{A_{M_i}}$$

Utilizing the equations generated for the calculation of the number of offspring of each genotype, the fitness costs, initial input, self-elimination (α, ß, γ), and the gene drive mechanism parameters (provided in Table 2), the number of offspring created for each time step were calculated.

Table 2. Gene drive mechanism parameters

Gene Drive Mechanism	Parameter	Description	Value
MEDEA	Ω	Proportion of non-transgenic offspring surviving from transgenic females	0.05
Under-dominance	Ω_A	Proportion of non-transgenic offspring surviving from toxin A	0.05
	Ω_B	Proportion of non-transgenic offspring surviving from toxin B	0.05
CRISPR	q	Probability of Cas9 cut to cause DSB	0.95
	p	Probability of successful HDR	0.95
	δ	Probability of functional allele formation through NHEJ	Varies

Equation Generation

A two-dimensional matrix was generated of all the possible genotypes of females (F_i) and males (M_i). A third dimension was added to capture every possible outcome of offspring (g_i). The value of each index within this three-dimensional matrix corresponded to the probability that the combination of the two parental genotypes would produce the respective offspring of the genotype. For example, a female with the *ww* genotype mating with a male with *ws* genotype would produce an offspring of genotype

ws with a probability of 0.5. Hence, the resulting value of the index that corresponded to (*ww* female, *ws* male, *ws* offspring) = 0.5, or $\Psi(g_i \mid F_i, M_i) = 0.5$. Iterating through all possible combinations of F_i, M_i, and g_i, a matrix of probabilities was generated. It was not uncommon that a particular index had a value of 0, since a particular mating combination of male and female genotypes could only result in certain offspring genotypes.

Once the matrix was fully populated, a string was concatenated with the parental genotypes and probability of producing an offspring, resulting in the form:

$$F_i * \Psi(g_i \mid F_i, M_i) * M_i$$

This would be utilized in the calculation of the number of offspring in the system dynamics model. All combinations of parental genotypes to create a particular offspring genotype *k* were concatenated in the form:

$$g_i = \sum_{j=1}^{l} \sum_{k=1}^{n} F_j * \Psi(g_i \mid F_j, M_k) * M_k$$

Equations were simplified using MATLAB's *str2sym* function to reduce the additional computations necessary when referencing and calculating equations from the system dynamics model. To calculate the daily number of offspring of genotype *i* that were being produced, daily reproduction rates, sex ratio, and fitness costs were additionally concatenated into the equation following the simplification of the equations, for females giving:

$$\frac{\partial g_i}{\partial t} = \lambda * \sigma * (1 - c_i) \sum_{j=1}^{l} \sum_{k=1}^{n} \left[F_j * \Psi\left(g_i \mid F_j, M_k\right) * M_k\right]$$

and for males:

$$\frac{\partial g_i}{\partial t} = \lambda * (1-\sigma) * (1-c_i) \sum_{j=1}^{l} \sum_{k=1}^{n} \left[F_j * \Psi\left(g_i \mid F_j, M_k\right) * M_k \right]$$

In the case of a single gRNA target site with a single chance for self-elimination, six alleles (*w, v, g, s, u*, and *r*) were utilized creating 21 possible genotypes. The presence of a *g* allele allowed for self-elimination, producing *v, g,* or *s* alleles with probabilities α, ß, and γ, respectively. If a transgenic allele (*g* or *s*) was present with a *w* allele, a DSB could occur with probability q. The *v, u,* and *r* alleles have lost the gRNA target site and therefore cannot be cut by Cas9. The probabilities of inheritance are initially presented by Noble et al.[63] and modified here to include the self-eliminating mechanism.

The probability that gametes are passed from parents to progeny is defined as follows:

- Individuals with resistant alleles (*v, u, r*) and the wild-type allele *w* will produce resistant and wild-type gametes equiprobably such that we have:

$$P_{wv,w} = P_{wu,w} = P_{wr,w} = P_{wv,v} = P_{wu,u} = P_{wr,r} = \frac{1}{2}$$

- Individuals with resistant alleles (*v, u, r*) and the permanently fixed transgenic allele *s* will produce resistant and permanently fixed transgenic gametes equiprobably such that we have:

$$P_{vs,v} = P_{us,u} = P_{rs,r} = P_{vs,s} = P_{us,s} = P_{rs,s} = \frac{1}{2}$$

- Individuals with the resistant allele *v* with transgenic allele *g* will produce resistant gamete *v* through inheritance and self-elimination such that:

$$P_{vg,v} = \frac{1+\alpha}{2}$$

- Individuals with wild-type allele *w* or resistant alleles *u* or *r* with transgenic allele *g* will produce resistant gamete *v* such that:

$$P_{wg,v} = P_{ug,v} = P_{rg,v} = \frac{\alpha}{2}$$

- Individuals with resistant alleles *u* or *r* with transgenic allele *g* will produce resistant gametes *u* or *r* such that:

$$P_{ug,u} = P_{rg,r} = \frac{1}{2}$$

- Individuals with resistant alleles (*v*, *u*, *r*) with transgenic allele *g* will produce transgenic gamete *g* when self-elimination and permanent fixation does not occur such that:

$$P_{vg,g} = P_{ug,g} = P_{rg,g} = \frac{ß}{2}$$

- Individuals with resistant alleles (*v*, *u*, *r*) with transgenic allele *g* will produce permanently fixed transgenic gamete *s* when permanent fixation occurs such that:

$$P_{vg,s} = P_{ug,s} = P_{rg,s} = \frac{\gamma}{2}$$

- Individuals with wild-type allele *w* and transgenic allele *g* will produce wild-type gamete *w* when successful self-elimination excision occurs or when self-elimination does not occur and no cutting of the target site occurs such that:

$$P_{wg,w} = \frac{\alpha + (1-q)(ß+\gamma)}{2}$$

- Individuals with wild-type allele *w* and transgenic allele *g* will produce transgenic gamete *g* when successful self-elimination excision does not occur or permanent transgene fixation does not occur. Through cutting and HDR, the transgenic material is copied onto the *w* allele such that:

$$P_{wg,g} = \frac{(1+qp)(ß)}{2}$$

- Individuals with wild-type allele *w* and transgenic gamete *g* will produce resistant allele *u* when successful self-elimination excision does not occur or permanent transgene fixation occurs. Through cutting and NHEJ a functional resistant allele is produced such that:

$$P_{wg,u} = \frac{(ß+\gamma)q(1-P)\delta}{2}$$

- Individuals with wild-type allele *w* and transgenic allele *g* will produce resistant gamete *r* when successful self-elimination excision does not occur or permanent transgene fixation occurs. Through cutting and NHEJ a nonfunctional resistant allele is produced such that:

$$P_{wg,r} = \frac{(ß+\gamma)q(1-p)(1-\delta)}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce wild-type gamete *w* if cutting does not occur such that we have:

$$P_{ws,w} = \frac{1-q}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce permanently fixed transgenic gamete *s* through cutting and HDR, when the transgenic material is copied onto the *w* allele such that we have:

$$P_{ws,s} = \frac{1+qp}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce resistant gamete *u* through cutting at the target site with probability *q* and NHEJ with probability *1-p*, repairing the cut to form a functional allele with probability δ, such that:

$$P_{ws,u} = \frac{q(1-p)\delta}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce resistant gamete *u* through cutting and NHEJ, repairing the cut to form a function allele with probability δ, such that:

$$P_{ws,r} = \frac{q(1-p)(1-\delta)}{2}$$

Each of the alleles and probabilities for males and females were calculated and stored. By combining the two alleles (one from each parent) and multiplying the probabilities of the two alleles together, the probability that an offspring with a particular genotype could be created from two parents was obtained.

Results

The single self-elimination target site with a single gRNA target site allows for a single chance for the self-elimination mechanism to remove the transgenic material and a single chance for Cas9 to cause a DSB. We assumed a fitness cost of 5% per transgenic allele *g* and *s*. A fitness cost due to the disruption of gene function associated with the homologous *r* alleles was set as 5%, according to the parameters established in Noble et al. [63]. A 1% release of homozygous transgenic *gg* males into a population consisting of 99% wild-type (*ww*) males and females was used.

When the SEM is inactive ($\alpha = \gamma = 0$) and there is a high probability for the formation of functional resistance alleles ($\delta = 0.33$), the transgene quickly drives into the population, reducing the proportion of wild-type individuals as shown in Figure 3. Since the resistance alleles *u* and *r* are not susceptible to reinvasion of the transgene and because the functional resistance alleles *u* do not have an associated fitness cost, the *u* alleles begin to restore the wild-type population due to their competitive advantage over the transgenic alleles. As the formation of functional resistance alleles *u* becomes more difficult (δ decreases), the transgenes last progressively longer in the population as shown in Figure 4.

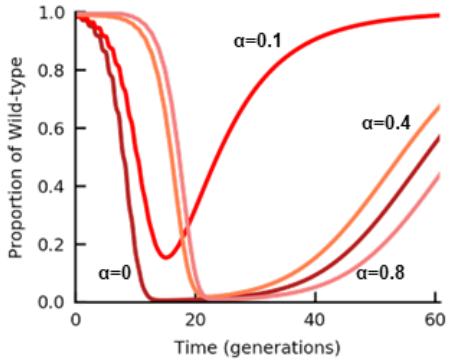

Figure 3. Proportion of transgene-free individuals with a simulated release of gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) where γ = 0.01

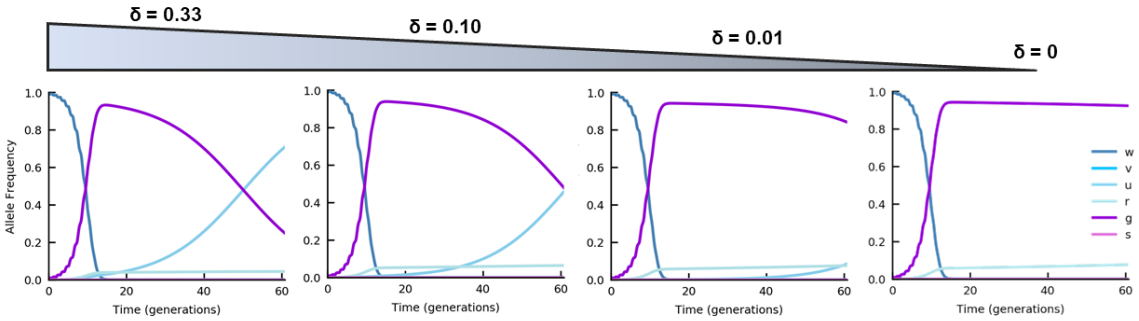

Figure 4. Allele frequency following a simulated release of gene drive containing individuals at 1% of the wild-type with no transgene self-elimination (α = γ = 0) across four values of functional resistance allele formation (δ = 0, 0.01, 0.10, 0.33)

With an active SEM (α > 0, γ > 0) and successful transgene excision, the transgenic material from the *g* alleles is removed, resulting in the formation of *v* alleles. Since α corresponds to the probability that the transgenic material will be successfully removed from the *g* allele, we expect to observe faster removal of the transgene and restoration of the wild-type population. However, we observe a counterintuitive phenomenon, where higher probabilities of transgenic removal (α) increase the time to restoration of the wild-

type population, as shown in Figure 5. In all cases, the restoration of the wild-type population depends on the creation of resistance alleles (*v*, *u*, *r*) that cannot be reinvaded by the homing drive. Because we cannot limit the formation of permanently-fixed transgenic alleles *s* that form with probability γ, we must assume γ > 0 which will result in the formation of *s* alleles.

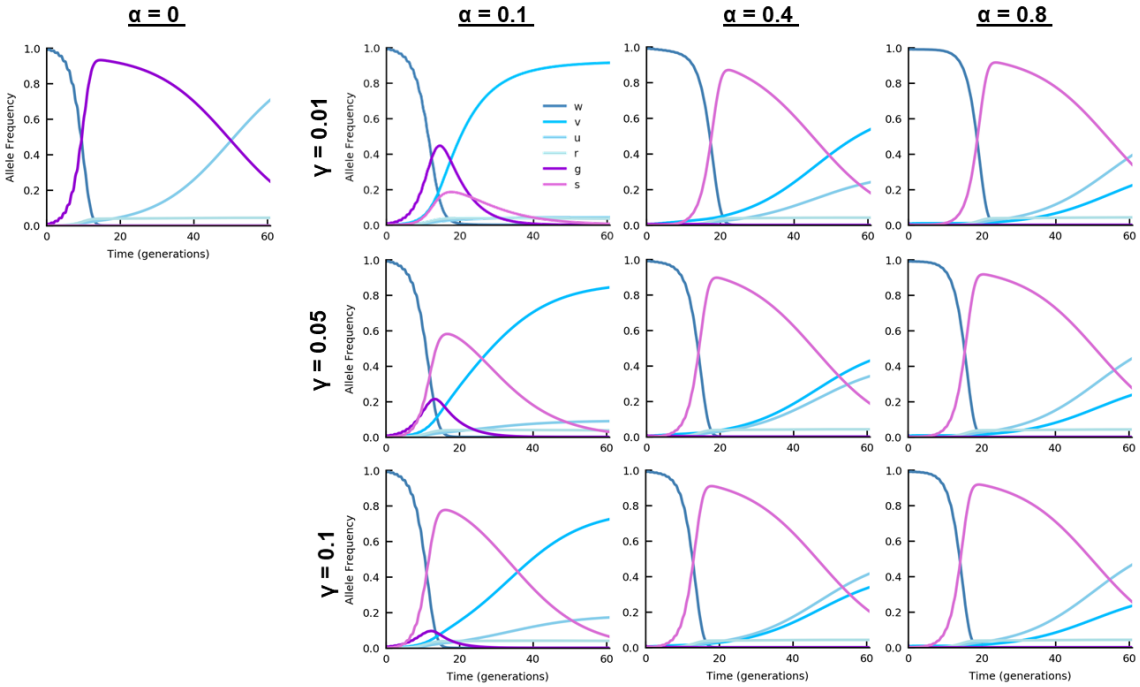

Figure 5. Allele frequency within a simulated release of gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10), where δ = 0.33

Under higher values of α, because the transgene *g* is removed very quickly, very few resistance alleles *v* are produced. Additionally, although the number of permanently-fixed transgenic *s* alleles produce is miniscule, due to the lack of resistance alleles *v*

present in the population, a large proportion of the population is not resistant to re-invasion of the GD. This creates a GD that can no longer be controlled by the SEM, resulting in the restoration of wild-type population depending increasingly on the formation of resistance alleles *u* and *r*. At lower levels of α, the transgene *g* persists for a longer period of time, creating more resistance alleles *v* at a slower rate. As the *g* allele mutates into the permanently-fixed transgene *s* over time, a significantly larger number of created resistance alleles *v* are present to stop the permanently-fixed GD and enable for a faster restoration the wild-type population. As δ is decreased, we observe a decrease in the number of functional resistance alleles *u* created and the restoration of the wild-type population can be increasingly attributed the created resistance allele *v*, as shown in Figure 6.

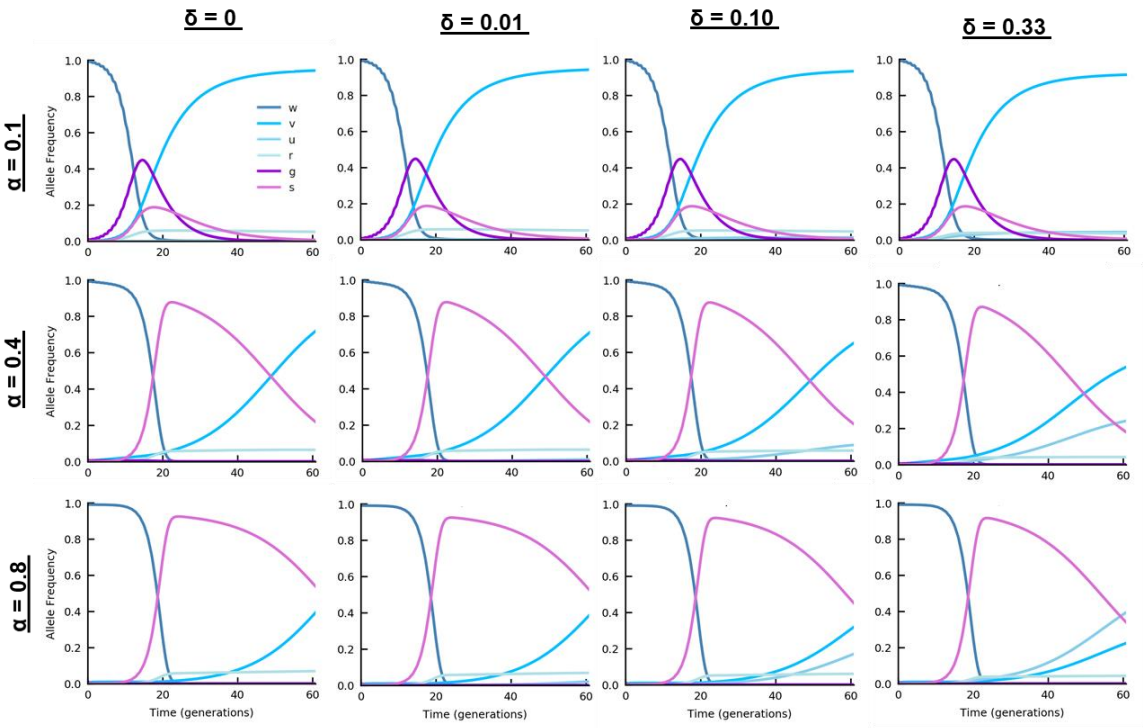

Figure 6. Allele frequency within a simulated release of gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and four rates of functional resistance allele formation (δ = 0, 0.01, 0.10, 0.33). For all plots, the rate of permanent failure of the self-elimination mechanism is set to γ = 0.01.

Evaluating the wild-type allele proportion of the population for 60 generations across a range of α and γ values, we confirmed that low α and low γ probabilities were the most effective for the restoration of the wild-type population. We further separated this parameter space with the original wild-type alleles *w* and created resistance alleles *v*, and compared this to the resistance alleles *u* and *r* that form through NHEJ following a DSB. As shown in Figure 7, at low α and γ values, a significant proportion of the wild-type population can be attributed to the creation of *v* alleles as a result of the transgenic excision induced by the SEM. As δ is decreased, we observe the proportion of the wild-type allele

w and the created resistance allele *v* resemble the total wild-type allele proportion while the proportion of resistance alleles *u* and *v* decreases.

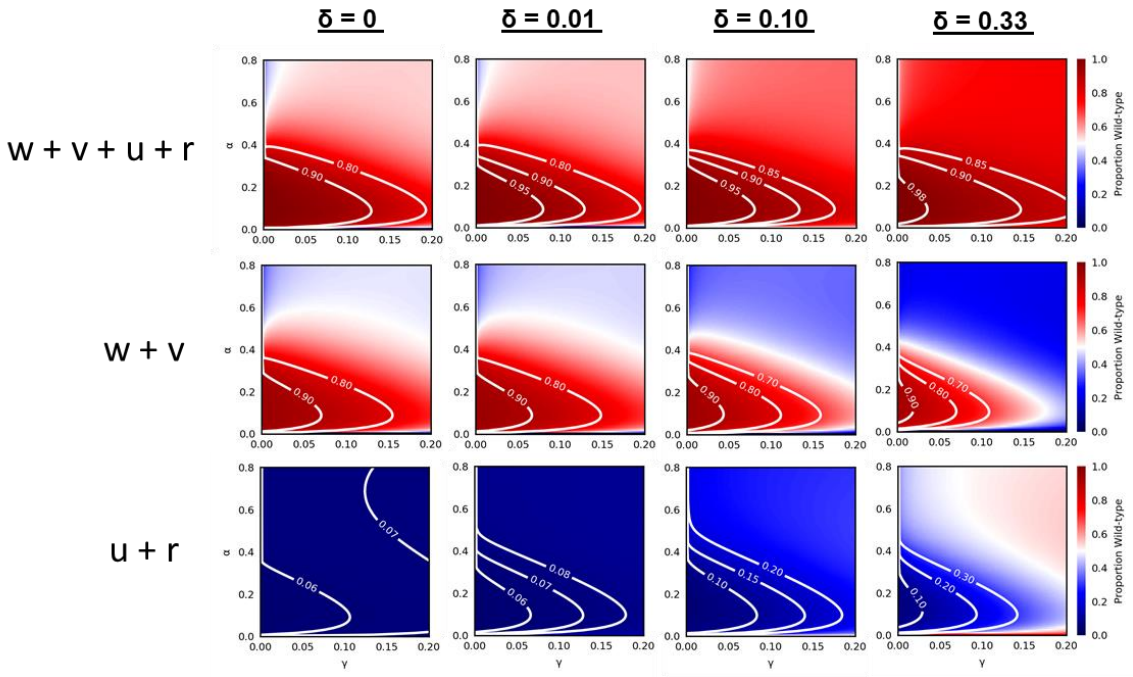

Figure 7. Proportion of wild-type alleles after 60 generations at four rates of functional resistance allele formation (δ = 0, 0.01, 0.10, 0.33)

A threshold for the maximum proportion of transgenic alleles in the population was generated to evaluate the potential bio-containment effectiveness of the SEM. Although the threshold increased as α increased, the level of containment was only increased to two orders of magnitude beyond the initial release proportion, as shown in Figure 8.

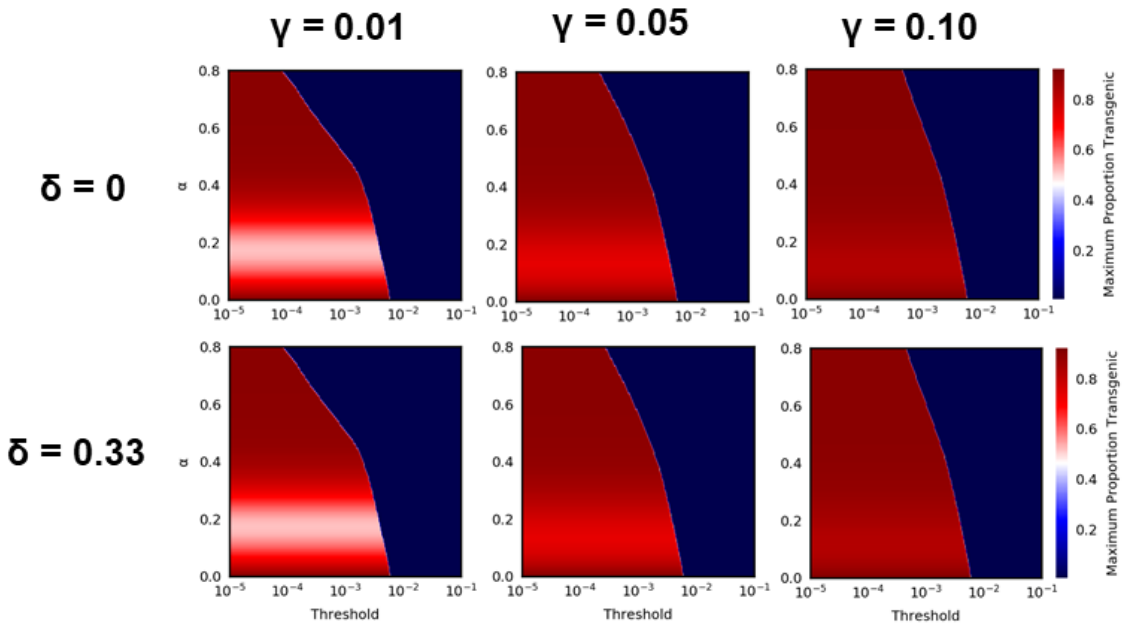

Figure 8. Threshold of the maximum proportion of transgenic alleles at three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10) and two rates of functional resistance allele formation (δ = 0, 0.33)

A second scenario was investigated based on the findings shown by Kyrou et al.[64], where the *doublesex* (dsx) gene was targeted. In this approach, females carrying two transgenic alleles (*gg, gs, ss*), two nonfunctional resistance alleles (*rr*), or a combination of the transgene and nonfunctional allele (*gr, sr*) will not survive. We applied a fitness cost of 100% to the females with these genotypes and retained the fitness costs from the prior approach for males. Additionally, δ = 0, since no functional resistance alleles *u* were observed.

An inactive SEM (α = γ = 0) resulted in a successful drive of the transgene into the population and ultimately caused the population to crash (shown in Figure 9). Since no functional resistance alleles were created because δ = 0, the transgene was capable of

driving into the population. As drive progresses, the population becomes increasing male-biased due to the mortality of females carrying two copies of the transgene or nonfunctional resistance allele, causing the number of individuals to decrease over time.

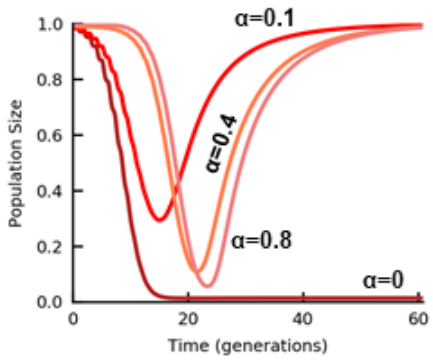

Figure 9. Proportion of transgene-free individuals with a simulated release of dsx gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) where γ = 0.01

When the SEM is active (α > 0, γ > 0), we observe similar trends of lower α restoring the wild-type population at a faster rate than higher α due to longer persistence of the transgenic allele *g* and the subsequent creation of more resistance alleles *v*. In this case, no functional resistance alleles *u* are created and due to the mortality of females with *rr*, *gr* and *sr* genotypes, the nonfunctional resistance alleles *r* are primarily carried by males. As a result, the restoration of the wild-type population becomes increasingly dependent on the created resistance allele *v*. Even at higher values of α, the created *v* allele is the primary driver in the restoration of the wild-type due to the competitive advantage (0% fitness cost), as shown in Figure 10. Allele frequency within a simulated release of dsx gene drive containing individuals at 1% of a wild-type population at four different

rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) where γ = 0.01. However, since the persistence of the *g* allele is increased under low values of α, more resistance alleles *v* are present to restore the wild-type population than under high values of α.

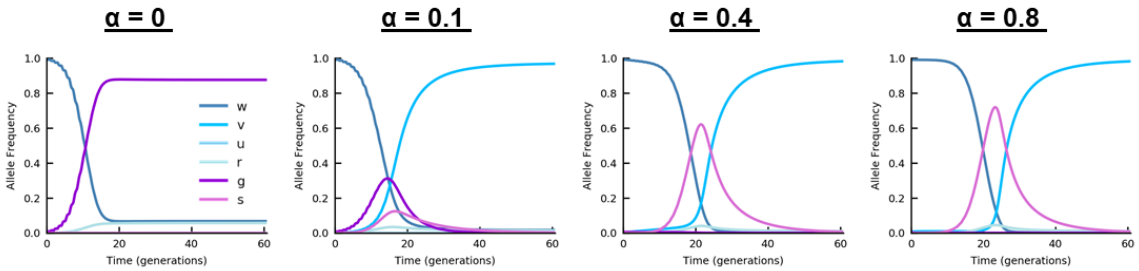

Figure 10. Allele frequency within a simulated release of dsx gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) where γ = 0.01.

Evaluating the parameter space across values of α and γ, we observe a complete restoration of the wild-type population under all combinations (with the exception of α = 0, or when the SEM is inactive). Furthermore, we observe that this restoration of the wild-type population is fully attributable to the natural wild-type and created resistance alleles *w* and *v*, as shown in Figure 11.

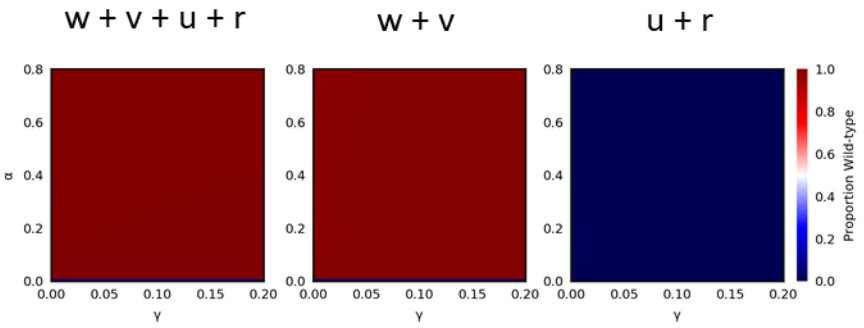

Figure 11. Proportion of wild-type alleles after 60 generations for the dsx gene drive (δ = 0)

The threshold plots show that lower thresholds are necessary for a successful drive of the transgene into the population at higher rates of transgene removal (shown in Figure 12). The results indicate that a gene drive may be prevented when the threshold is up to two orders of magnitude above the initial release proportion of gene drive carrying organisms.

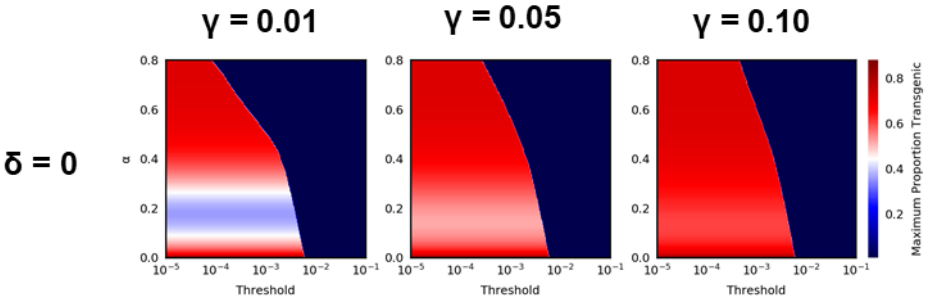

Figure 12. Threshold of the maximum proportion of transgenic alleles with the dsx gene drive at three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10) and one rate of functional resistance allele formation (δ = 0)

CRISPR Five Self-Elimination gRNA Target Sites with Single Homing Drive gRNA Target Site

To limit a subsequent homing drive of the permanently-fixed transgene *s* following the removal of all excision-competent transgenes *g*, a five self-elimination target site construct was proposed. With five chances to successfully excise the transgenic segment of DNA, the rate of permanently-fixed transgene formation γ is significantly reduced. We apply a 5% fitness cost per transgenic allele, with individuals carrying a nonfunctional allele and resistance allele (g_ir, *sr*) having a 10% fitness cost. A 5% fitness cost was applied to individuals with two nonfunctional alleles (*rr*). Fitness costs were identical in both males and females.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single homing drive target site was extended for multiple chances for successful self-elimination. The model structure and function were otherwise unchanged.

Equation Generation

Using the structure of the single self-elimination cut with a single gRNA target site model, we introduced five self-elimination target sites modeled as g_1-g_5 (where n = 5), allowing for five chances for a self-elimination cut to occur. Along with the wild-type and resistance alleles, this produces 55 possible genotypes. Transgene excision, no excision,

and permanent transgene fixation follow a multinomial distribution, while Cas9 cutting, HDR, and NHEJ remain unchanged.

The probability that gametes are passed to offspring from parents are as follows:

- Individuals with resistant alleles (*v*, *u*, *r*) and the wild-type allele *w* will produce resistant and wild-type gametes equiprobably such that we have:

$$P_{wv,w} = P_{wu,w} = P_{wr,w} = P_{wv,v} = P_{wu,u} = P_{wr,r} = \frac{1}{2}$$

- Individuals with resistant alleles (*v*, *u*, *r*) and the permanently fixed transgenic allele *s* will produce resistant and permanently fixed transgenic gametes equiprobably such that we have:

$$P_{vs,v} = P_{us,u} = P_{rs,r} = P_{vs,s} = P_{us,s} = P_{rs,s} = \frac{1}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce wild-type gamete *w* if cutting does not occur such that we have:

$$P_{ws,w} = \frac{1-q}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce permanently fixed transgenic gamete *s* through cutting and HDR, when the transgenic material is copied onto the *w* allele such that we have:

$$P_{ws,s} = \frac{1+qp}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce resistant gamete *u* through cutting at the target site with probability *q* and NHEJ with probability *1-p*, repairing the cut to form a functional allele with probability *δ*, such that:

$$P_{ws,u} = \frac{q(1-p)\delta}{2}$$

- Individuals with wild-type allele *w* and the permanently fixed transgenic allele *s* will produce resistant gamete *u* through cutting and NHEJ, repairing the cut to form a function allele with probability δ, such that:

$$P_{ws,r} = \frac{q(1-p)(1-\delta)}{2}$$

- Individuals with the resistant allele *v* with transgenic allele g_i will produce resistant gamete *v* through inheritance and self-elimination such that:

$$P_{vg_i,v} = \frac{1+(1-(1-\alpha)^{n-i})}{2}$$

- Individuals with the wild-type allele *w* or the resistant alleles *u* or *r* with transgenic allele g_i will produce resistant gamete *v* through inheritance and self-elimination such that:

$$P_{wg_i,v} = P_{ug_i,v} = P_{rg_i,v} = \frac{(1-(1-\alpha)^{n-i})}{2}$$

- Individuals with the resistant alleles *u* or *r* with transgenic allele g_i will produce resistant gametes *u* or *r* such that:

$$P_{ug_i,u} = P_{rg_i,r} = \frac{1}{2}$$

- Individuals with the resistant alleles (*v*, *u*, *r*) with transgenic allele g_i will produce transgenic gametes g_i such that:

$$P_{vg_i,g_i} = P_{ug_i,g_i} = P_{rg_i,g_i} = \frac{ß^{n-i}}{2}$$

- Individuals with the resistant alleles (*v*, *u*, *r*) with transgenic allele g_i will produce transgenic gametes g_k such that:

$$P_{vg_i,g_k} = P_{ug_i,g_i} = P_{rg_i,g_i} = \frac{\binom{n-i+1}{k-i} ß^{n-k+1} \gamma^{k-i}}{2}$$

- Individuals with the resistant alleles (*v*, *u*, *r*) with transgenic allele g_i will produce transgenic gamete *s* such that:

$$P_{vg_i,s} = P_{ug_i,s} = P_{rg_i,s} = \frac{\gamma^{n-i+i}}{2}$$

- Individuals with the wild-type alleles *w* with transgenic allele g_i will produce wild-type gamete *w* such that:

$$P_{wg_i,w} = \frac{\left(1-(1-\alpha)^{n-i}\right) + (1-q)(1-\alpha^{n-i})}{2}$$

- Individuals with the wild-type alleles *w* with transgenic allele g_i will produce functional resistant gamete *u* such that:

$$P_{wg_i,u} = \frac{(1-\alpha)^{n-i} q(1-p)\delta}{2}$$

- Individuals with the wild-type alleles *w* with transgenic allele g_i will produce nonfunctional resistant gamete *r* such that:

$$P_{wg_i,r} = \frac{(1-\alpha)^{n-i} q(1-p)(1-\delta)}{2}$$

Results

Since the number of self-elimination target sites does not have an impact when the SEM is inactive ($\alpha = \gamma = 0$), the construct of five self-elimination target sites with a single target site for the homing drive produces identical results to the single self-elimination target site with a single gRNA target site (shown in Figure 3). The proportion of transgene-free individuals is shown below in Figure 13 for the five self-elimination target site construct.

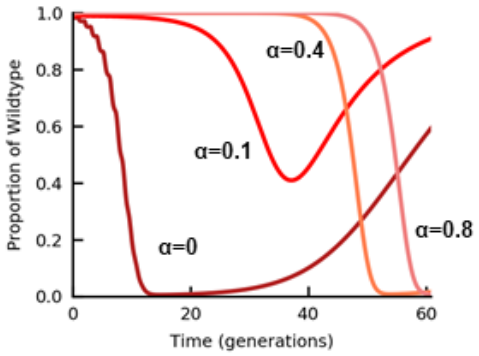

Figure 13. Proportion of transgene-free individuals with a simulated release of five self-elimination target site gene drive containing individuals at 1% of a wild-type population at four rates of transgene self-elimination ($\alpha = 0, 0.1, 0.4, 0.8$) for $\gamma = 0.01$ and $\delta = 0.33$

When the SEM is active ($\alpha > 0$, $\gamma > 0$), we observe a limited drive under $\alpha = 0.1$, where only up to 60% of the population carries a transgene. Under $\alpha = 0.4$ and 0.8, the drive is delayed beyond 40 generations from the time of the initial release (Figure 13). In Figure 14, the results presented from the single self-elimination target sites are intensified and we observe the transgene eliminated at much higher rates, creating an even smaller

number of created resistance alleles *v*. Through the use of five self-elimination target sites, we were able to drastically decrease the initial rate of formation of permanently-fixed transgenic alleles *s* from γ to γ^n. However, γ was not decreased to γ = 0, allowing for the creation of permanently-fixed transgenic alleles *s*. Due to the deterministic nature of the model, the presence of these transgenic alleles would be enough to drive into the population, as shown in Figure 14.

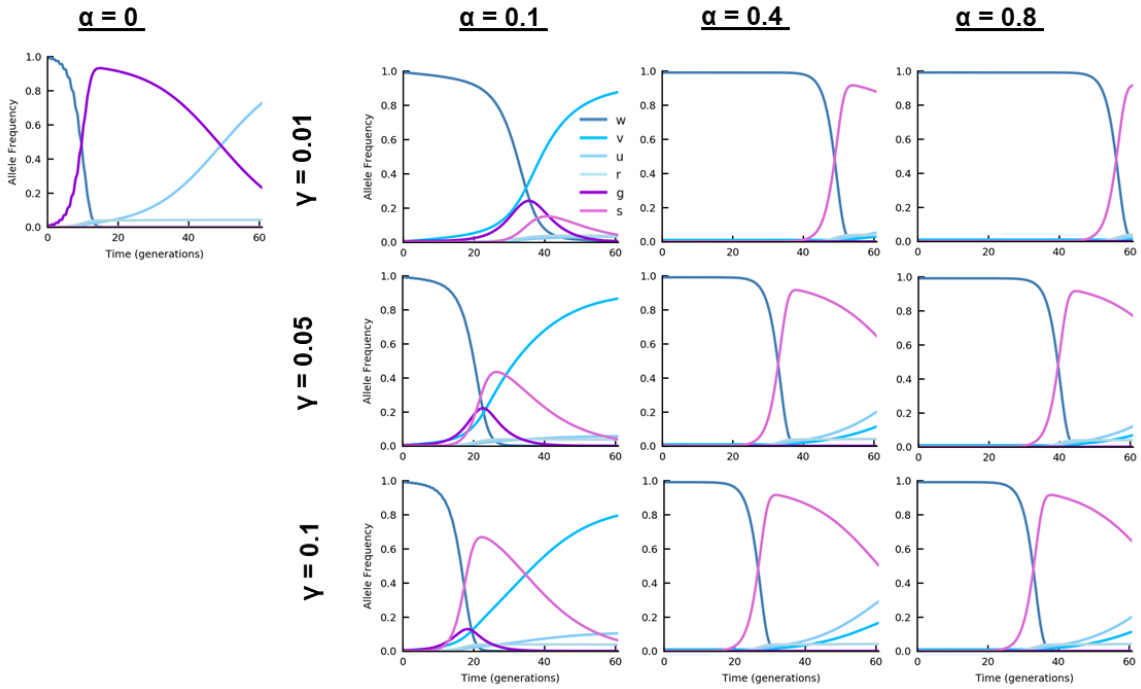

Figure 14. Allele frequency within a simulated release of five self-elimination target site gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10), where δ = 0.33

Parameter space evaluation of the proportion of wild-type alleles after 60 generations indicated that low α and γ rates were most effective for the elimination of

transgenic alleles from the population (shown in Figure 15). Reduction of δ decreased the number of functional resistance alleles *u* created and the wild-type population was restored primarily through the created resistance allele *v*.

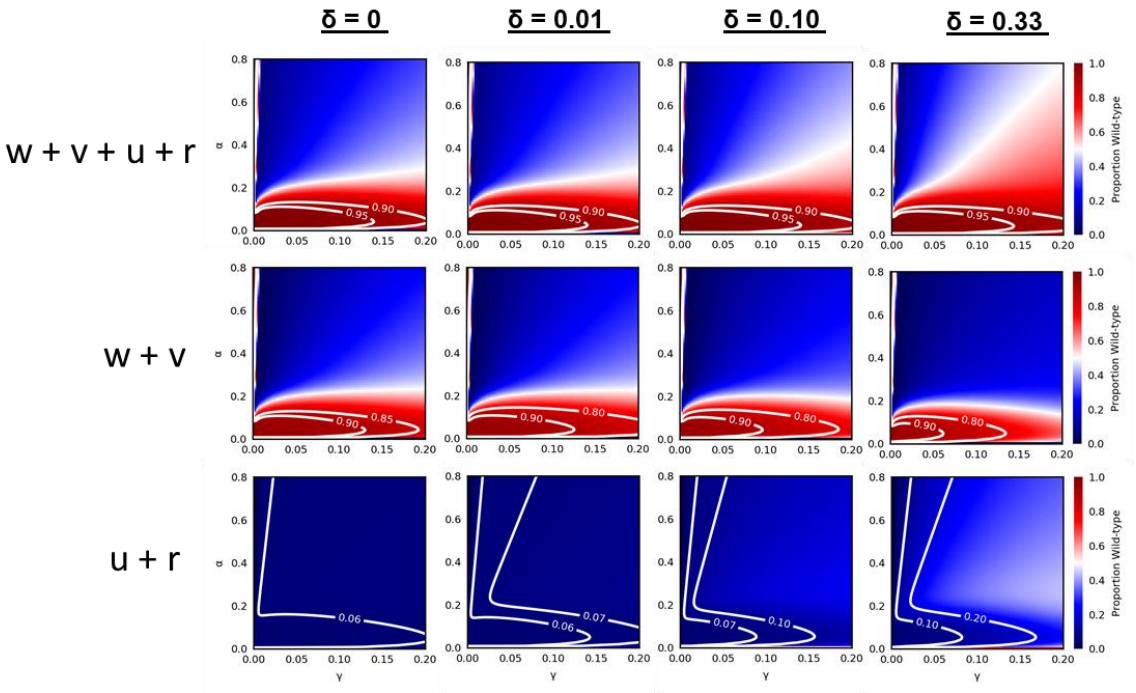

Figure 15. Proportion of wild-type alleles after 60 generations for five self-elimination target sites at four rates of functional resistance allele formation (δ = 0, 0.01, 0.10, 0.33)

While the addition of self-elimination target sites did not prevent the formation of permanently-fixed transgenes, the threshold of transgene elimination was significantly increased at higher rates of transgene excision (α), as shown in Figure 16. The thresholds established indicate the five self-elimination target site construct could be an effective means of biocontainment when the rate of transgene excision is high.

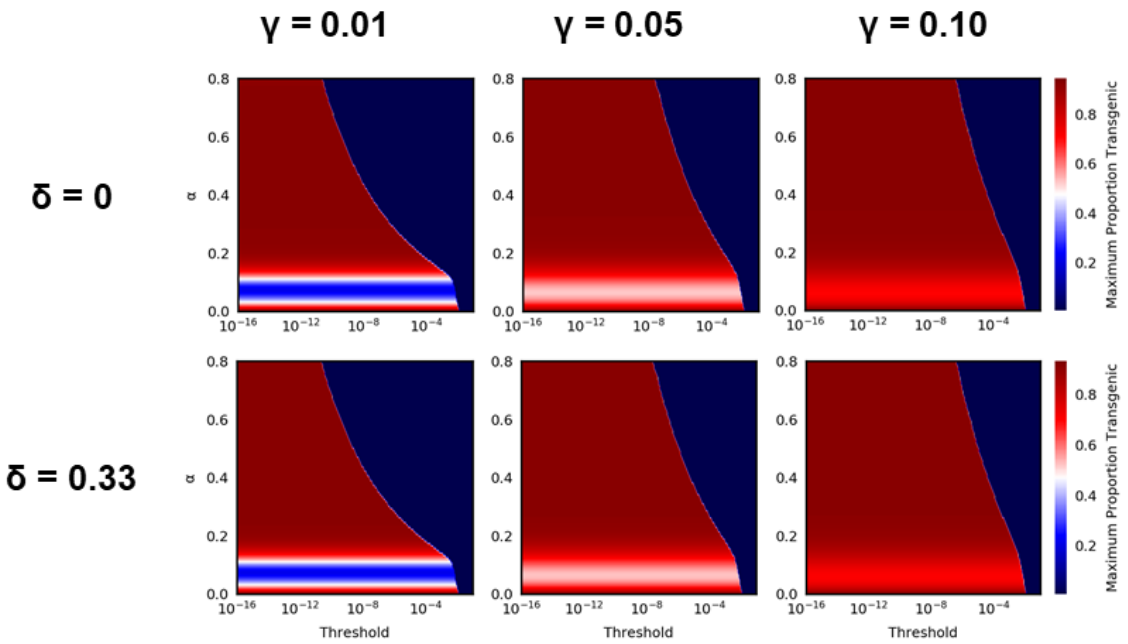

Figure 16. Threshold of the maximum proportion of transgenic alleles for five self-elimination target site gene drive at three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10) and two rates of functional resistance allele formation (δ = 0, 0.33)

Applying the five self-elimination target site construct to the dsx gene drive, we set the fitness cost for females carrying two transgenic, two nonfunctional resistance alleles, or a combination of transgenic and nonfunctional resistance alleles to 100%. Individuals carrying a single transgene were applied a 5% fitness cost and males carrying two transgenic alleles were assigned a fitness cost of 10%. A fitness cost of 5% was applied to males with two nonfunctional resistance alleles.

Like the previous scenario, when the SEM is inactive (α = γ = 0), the results for the five self-elimination target site dsx drive were identical to the single self-elimination target site dsx drive (Figure 17). In both cases, the transgene drove into the population and suppressed the total population by causing mortality in transgenic females. As a result, the

surviving population was biased towards transgenic males, stopping the growth of the population.

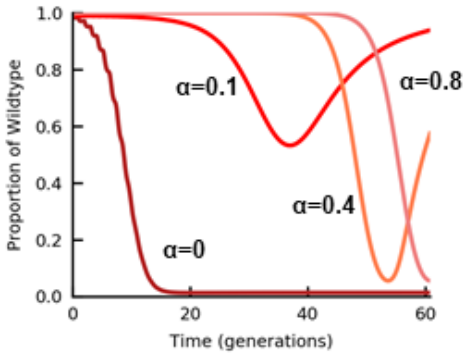

Figure 17. Proportion of transgene-free individuals with a simulated release of five self-elimination target site dsx gene drive containing individuals at 1% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) for γ = 0.01 and δ = 0

When the SEM was active ($\alpha > 0$, $\gamma > 0$), we observed a limited gene drive for low α values, where $\alpha = 0.1$ resulted in faster elimination of the transgenic alleles than $\alpha = 0.4$ and $\alpha = 0.8$. Compared to the previous (non-dsx) gene drive, the dsx gene drive demonstrated an increased proportion of transgene-free individuals when compared across respective α values. We observe for $\alpha = 0.4$ and 0.8, the proportion of created resistance alleles *v* is very small before the natural wild-type alleles *w* are suppressed. However, following the drive of permanently-fixed transgene *s*, the created resistance allele *v* quickly restores the proportion of transgene-free individuals (Figure 18). This can be attributed to the high mortality rate of females with two transgenic alleles, which limits the homing drive in males.

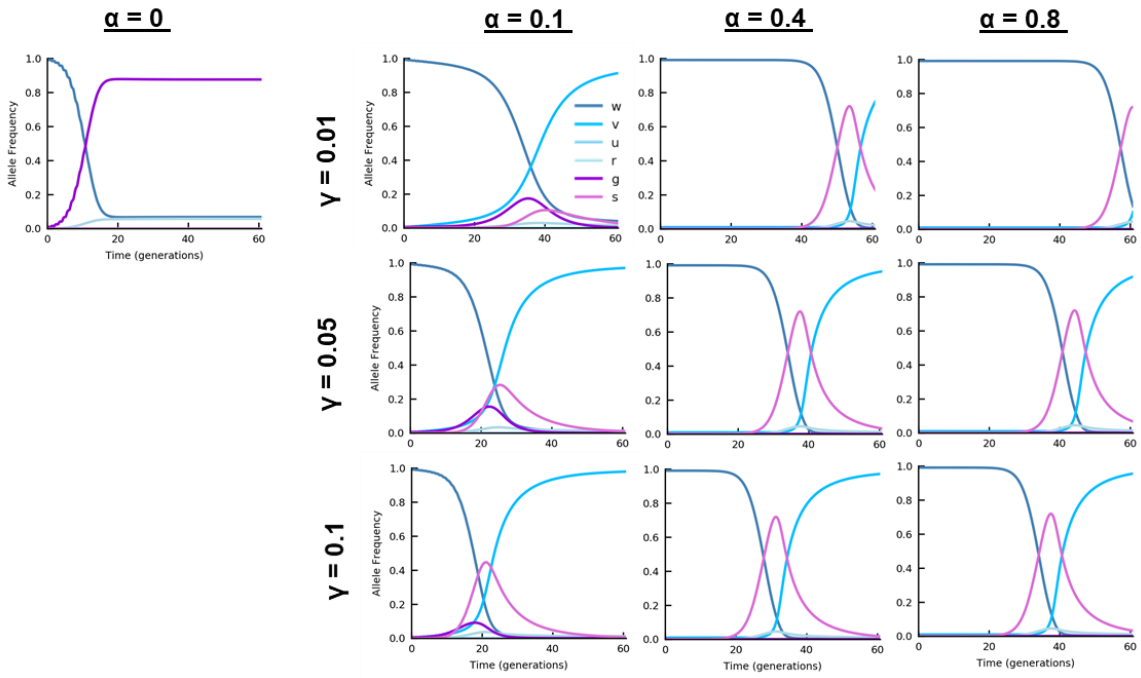

Figure 18. Allele frequency within a simulated release of five self-elimination target site dsx gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10), where δ = 0

The parameter space evaluation for the proportion of the wild-type alleles at 60 generations indicated the wild-type population would be restored for the majority of α and γ values (Figure 19). Only the very low γ values did not indicate a full restoration of the wild-type alleles, since the drive of the permanently-fixed transgene *s* had not fully occurred at the time of 60 generations.

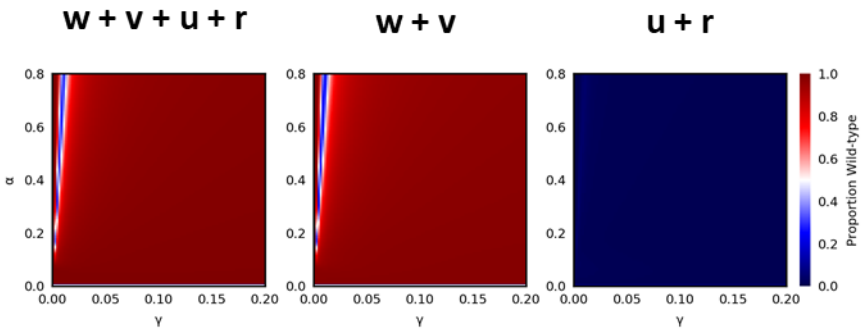

Figure 19. Proportion of wild-type alleles after 60 generations for five self-elimination target sites dsx gene drive for δ = 0

As in the previous gene drive for the five self-elimination target sites construct, the threshold for the elimination of the transgene was significantly higher than the initial release proportion, as shown in Figure 20, at high rates of transgene excision. This indicates the five self-elimination target site dsx gene drive could be utilized for biocontainment of the gene drive if the rate of transgene excision (α) was substantially high.

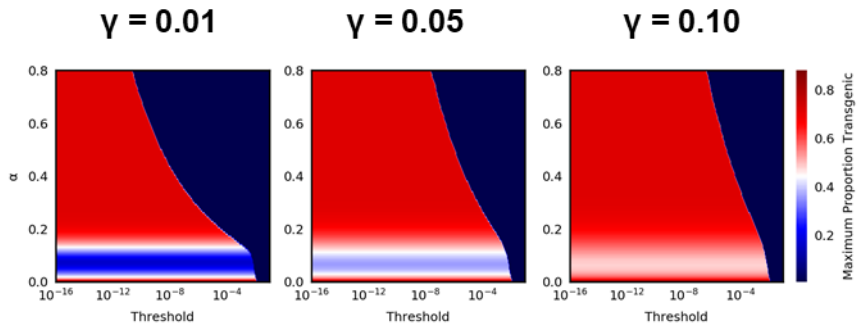

Figure 20. Threshold of the maximum proportion of transgenic alleles for five self-elimination target site dsx gene drive at three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10) where δ = 0

CRISPR Single Self-Elimination gRNA Target Site with Five Homing Drive gRNA Target Sites

A proposed method by Noble et al.[63] explored the use of multiple gRNA target sites to reduce the formation of resistance alleles. By utilizing multiple targets sites for the homing drive, the DNA of heterozygous individuals would have multiple chances for successful copying of the transgenic material onto the non-transgenic allele through HDR. We explored the multiple gRNA target site approach proposed by Noble et al. for five gRNA target sites, paired with a single self-elimination target site. Utilizing the fitness costs established by Noble et al., we set the fitness costs of all individuals carrying transgenes (heterozygous and homozygous) to 5% and individuals with two nonfunctional resistance alleles (*rr*) to 99%.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single homing drive target site was extended for multiple chances for successful HDR. The model structure and function were otherwise unchanged.

Equation Generation

Adding a single chance for self-elimination to the proposed drive presented in Noble et al.[63], we introduced multiple resistance alleles u_1-u_5 and r_1-r_5, where u_1-u_4 and r_1-r_4 provide additional target sites for the drive mechanism to cut. Combined with the

self-elimination mechanism, this produces 105 possible genotypes. As shown by Noble et al., a binomial distribution $P_K(k \mid n, i, q)$, is defined as the probability given there are n total target sites, i are resistant to cutting and each of the non-resistant sites can be cut with probability q, such that:

$$P_K(k \mid n, i, q) = \binom{n-i}{k} q^k (1-q)^{n-i-k} \qquad \text{for } 0 \le k \le n\text{-}i$$

When two or more cuts occurred, Noble et al. assume the loss of all target sites between the outermost cut locations, defining $P_L(l \mid k, n, i)$ as the probability that l target sites are lost given k cuts, n target sites, and i resistant sites, where:

$$P_L(l \mid k, n, i) = (n-i-l+1)\binom{l-2}{k-2} / \binom{n-i}{k} \qquad \text{for } 2 \le k \le l \le n\text{-}i$$

The probability that alleles are passed from parents to offspring are defined by the following:

- Individuals with resistance alleles r_i, u_i or the wild-type allele w and the permanently fixed transgenic allele s will produce permanently fixed transgenic gamete s through cutting at one or more target sites and HDR such that:

$$P_{r_i s,s} = P_{u_i s,s} = P_{ws,s} = \frac{1 + p(1-(1-q)^{n-i})}{2}$$

- Individuals with resistance alleles r_i, u_i or the wild-type allele w and the permanently fixed transgenic allele s will produce gametes r_i, u_i, and w, respectively, when no cutting occurs:

$$P_{r_i s,r_i} = P_{u_i s,u_i} = P_{ws,w} = \frac{(1-q)^{n-i}}{2}$$

- Individuals with resistance alleles r_i and the permanently fixed transgenic allele *s* will produce resistant gamete r_{i+1} when only one cut occurs and repair by NHEJ:

$$P_{r_i s, r_i} = \frac{(1-p)(n-i)q(1-q)^{n-i-1}}{2}$$

- Individuals with resistance alleles r_i and the permanently fixed transgenic allele *s* will produce resistant gamete r_k when *k-i* target sites are lost and repair by NHEJ occurs (for $i+2 \leq k \leq n$) such that:

$$P_{r_i s, r_k} = \frac{1-p}{2} \sum_{j=2}^{k-i} P_L(k-i \mid j, n, i)\, P_k(j \mid n, i, q)$$

- Individuals with resistance alleles u_i or wild-type allele *w* (equivalent to u_0) and the permanently fixed transgenic allele *s* will produce functional resistant gametes u_{i+1} through cutting at one target site and NHEJ with probability δ, such that:

$$P_{u_i s, u_{i+1}} = \frac{1-p}{2} \delta(n-i)q(1-q)^{n-i-1}$$

- Individuals with resistance alleles u_i or wild-type allele *w* (equivalent to u_0) and the permanently fixed transgenic allele *s* do not produce u_k gametes when $k \geq i+2$, as cutting at two or more target sites would result in the loss of large segment of DNA, resulting in a nonfunctional resistance allele. Therefore:

$$P_{u_i s, u_{i+2}} = \cdots = P_{u_i s, u_n} = 0$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the permanently fixed transgenic allele s produce r_{i+1} alleles when a single cut occurs and is repaired by NHEJ and creates a nonfunctional resistance gamete, such that:

$$P_{u_i s, r_{i+1}} = \frac{1-p}{2}(1-\delta)(n-i)q(1-q)^{n-i-1}$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the permanently fixed transgenic allele s produce r_k gametes when a k-i (for $i+2 \leq k \leq n$) target sites are lost and repair by NHEJ creates a nonfunctional resistance allele, such that:

$$P_{u_i s, r_k} = \frac{1-p}{2}\sum_{j=2}^{k-i} P_L(k-i \mid j, n, i)\, P_k(j \mid n, i, q)$$

- Individuals with resistance alleles v and the permanently fixed transgenic allele s produce resistant gamete v and transgenic gamete s equiprobably, such that:

$$P_{vs,v} = P_{vs,v} = \frac{1}{2}$$

- Individuals with resistance alleles r_i, u_i or the wild-type allele w and the transgenic allele g will produce the transgenic gamete g if no collapse of permanent fixation of the transgene, cutting at one or more target sites and HDR occurs such that:

$$P_{r_i g, g} = P_{u_i g, g} = P_{wg,g} = \frac{ß\left(1 + p\left(1-(1-q)^{n-i}\right)\right)}{2}$$

- Individuals with resistance alleles r_i, u_i or the wild-type allele w and the transgenic allele g will produce the permanently fixed transgenic gamete s if permanent fixation of the transgene, cutting at one or more target sites and HDR occurs such that:

$$P_{r_i g,s} = P_{u_i g,s} = P_{wg,s} = \frac{\gamma\left(1 + p\left(1 - (1-q)^{n-i}\right)\right)}{2}$$

- Individuals with resistance alleles r_i, u_i or the wild-type allele w and the transgenic allele g will produce gametes r_i, u_i, and w, respectively, when no excision of the transgene and no cutting occurs such that:

$$P_{r_i g,r_i} = P_{u_i g,u_i} = P_{wg,w} = \frac{(ß + \gamma)(1-q)^{n-i}}{2}$$

- Individuals with resistance alleles r_i and the transgenic allele g will produce resistant gamete r_{i+1} when there is no excision of the transgene, only one cut occurs and repair by NHEJ occurs such that:

$$P_{r_i g,r_i} = \frac{(ß + \gamma)(1-p)(n-i)q(1-q)^{n-i-1}}{2}$$

- Individuals with resistance alleles r_i and the transgenic allele g will produce resistant allele r_k when there is no excision of the transgene, k-i target sites are lost and repair by NHEJ occurs (for $i+2 \leq k \leq n$) such that:

$$P_{r_i g,r_k} = \frac{(ß + \gamma)(1-p)}{2} \sum_{j=2}^{k-i} P_L(k - i \mid j, n, i)\, P_k(j \mid n, i, q)$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the fixed allele g will produce functional resistant gametes u_{i+1} when there is no excision of the transgene, cutting at one target site and NHEJ with probability δ, such that:

$$P_{u_i g, u_{i+1}} = \frac{(ß + \gamma)(1-p)}{2} \delta (n-i) q (1-q)^{n-i-1}$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the transgenic allele g do not produce u_k gametes when $k \geq i+2$, as cutting at two or more target sites would result in the loss of large segment of DNA, resulting in a nonfunctional resistance allele. Therefore:

$$P_{u_i g, u_{i+2}} = \cdots = P_{u_i g, u_n} = 0$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the fixed allele g produce r_{i+1} gametes when there is no excision of the transgene, a single cut followed by repair through NHEJ which creates a nonfunctional resistance allele, such that:

$$P_{u_i g, r_{i+1}} = \frac{(ß + \gamma)(1-p)}{2} (1-\delta)(n-i) q (1-q)^{n-i-1}$$

- Individuals with resistance alleles u_i or wild-type allele w (equivalent to u_0) and the transgenic allele g produce r_k gametes when there is no transgene excision, k-i (for $i+2 \leq k \leq n$) target sites are lost and repair by NHEJ creates a nonfunctional resistance allele, such that:

$$P_{u_i s, r_k} = \frac{(ß + \gamma)(1-p)}{2} \sum_{j=2}^{k-i} P_L(k-i \mid j, n, i)\, P_k(j \mid n, i, q)$$

- Individuals with resistant allele *v* and the transgenic allele *g* produce resistant gamete *v*, when the excision of the transgene occurs such that:

$$P_{vg,v} = \frac{1+\alpha}{2}$$

- Individuals with resistant alleles r_i, u_i or the wild-type allele *w* and the transgenic allele *g* produce resistant gamete *v*, when the excision of the transgene occurs such that:

$$P_{r_i g,v} = P_{u_i g,v} = P_{wg,v} = \frac{\alpha}{2}$$

Results

With an inactive SEM (α = γ = 0), the five target site GD is much stronger than the single target site case, driving into the population and dominates the wild-type population within 20 generations, as shown in Figure 21.

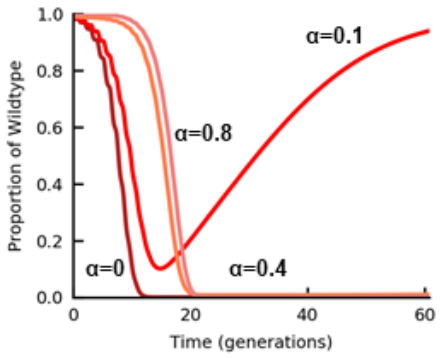

Figure 21. Proportion of transgene-free individuals with a simulated release of five target site gene drive containing individuals at 1% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) where γ = 0.01 and δ = 0.33

When the SEM mechanism is active ($\alpha > 0$, $\gamma > 0$), we observe a similar trend as in the single target site GD. As in the single target site scenario, as the rate of transgene excision (α) increases, fewer of the resistant alleles *v* are created, leaving a majority of the population susceptible to reinvasion by a subsequent drive of the permanently fixed transgene *s*. Furthermore, because the formation of functional resistance alleles *u* is significantly decreased through the use of five gRNA targets sites for the homing drive, the restoration of the wild-type population is dependent on the created resistance allele *v*. If the transgenic allele *g* is removed from the population too quickly, there are not sufficient numbers of created resistance alleles *v* to remove the permanently-fixed transgene *s* and restore the wild-type population within 60 generations (Figure 21). The allele frequencies in Figure 22 illustrate the limited number of resistance alleles *u* and *r* that are created and show the restoration of the wild-type individuals can be attributed to the increase in the created resistance allele *v* over time.

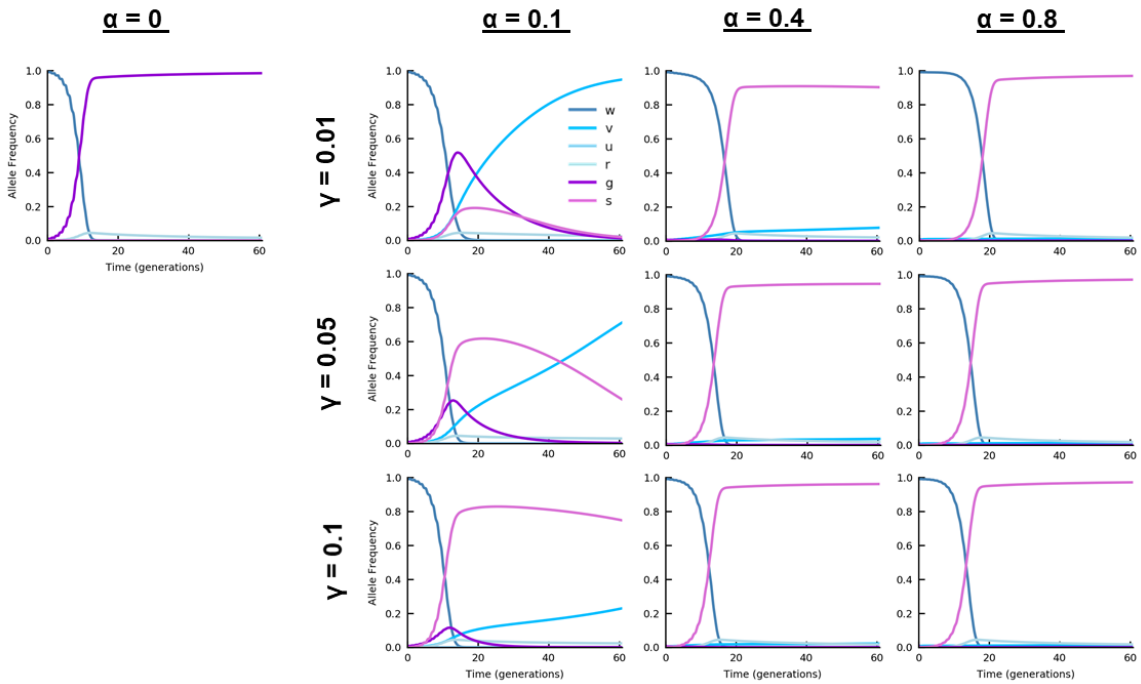

Figure 22. Allele frequency within a simulated release of five gRNA target site gene drive containing individuals at 1% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10), where δ = 0.33

Parameter space evaluation highlights low values of α and γ to be the most effective for the restoration of the wild-type population and removal of the transgenic organisms within 60 generations. Further evaluation of the parameter spaces for varying rates of functional resistance allele (δ) show no distinguishable difference between δ = 0 and δ = 0.33, as shown in Figure 23.

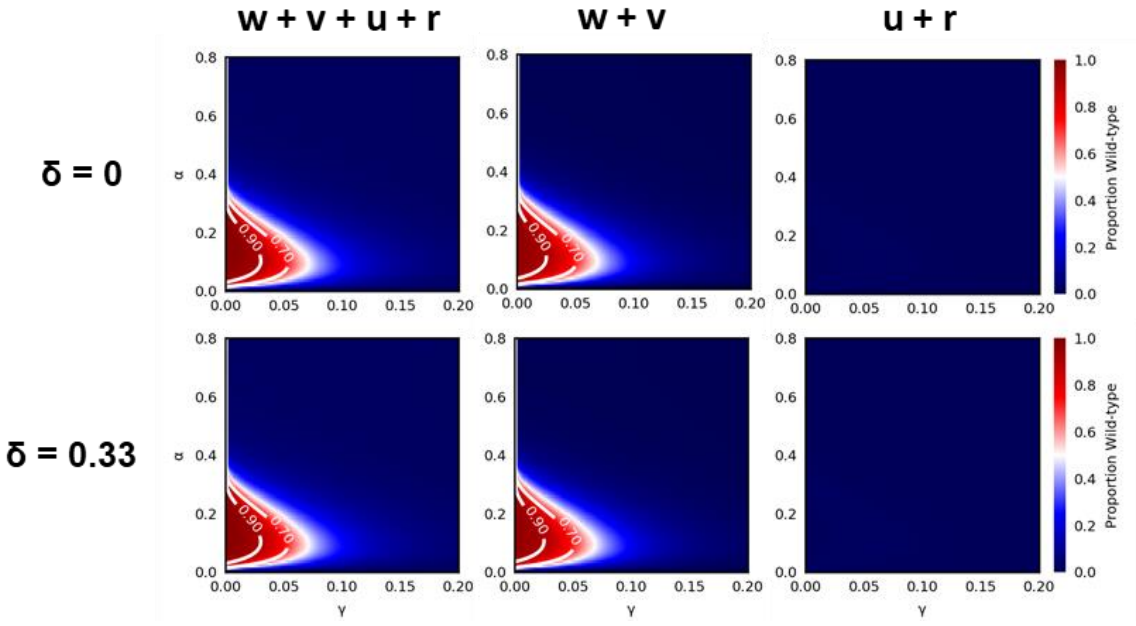

Figure 23. Proportion of wild-type alleles after 60 generations for a five target site GD-SEM at two rates of functional resistance allele formation (δ = 0, 0.33)

The thresholds for removing transgenic alleles are not significantly decreased from the initial levels of released transgenic individuals and would allow for the transgene to drive through the population across all rates of transgene excision α (as shown in Figure 24). While higher α values do eliminate the transgene at smaller thresholds, these threshold levels indicate this GD mechanism would not be suitable for containing an accidental release of gene drive carrying individuals.

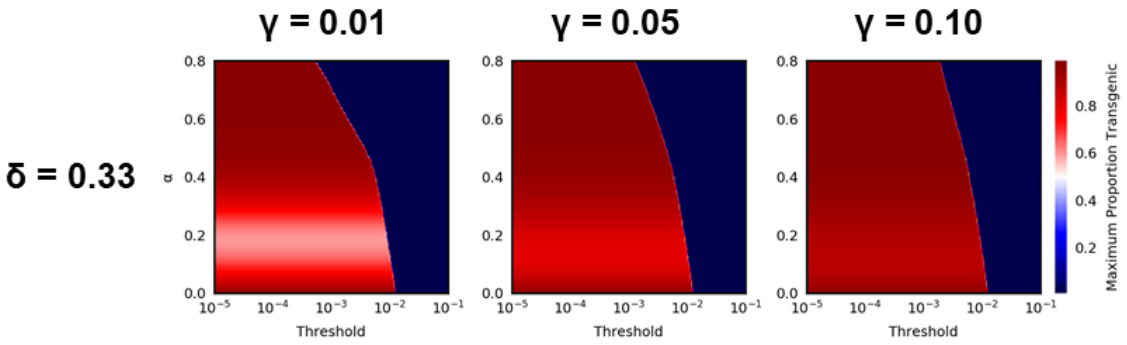

Figure 24. Threshold of the maximum proportion of transgenic alleles with the five gRNA target site gene drive at three rates of permanent transgene fixation (γ = 0.01, 0.05, 0.10) and one rate of functional resistance allele formation (δ = 0.33)

Discussion

We have modeled the functionality of an SEM paired with three types of CRISPR gene drive constructs that evaluate combinations for multiple chances for self-elimination and drive of the transgene. We also explore the dynamics of a low-cost and a high-cost, sex-biasing gene drive paired with the SEM. The results highlight feasible parameter values for the rates of successful excision and rates of permanent fixation of the transgene for restoring the wild-type population within 60 generations for each of the gene drive constructs evaluated.

By inducing a DSB and repair through HDR, the CRISPR gene drive increases the rates at which transgenic gametes are inherited by progeny. The creation of resistance alleles in cases where HDR fails would eventually limit the spread of the transgene due to the fitness cost associated with the transgene[63,81]. When the SEM was inactive, our results reflect the findings shown by modeling by Noble et al.[63], where the drive would not become permanently fixed in the population due the creation of the functional resistance

alleles at a high rate δ. However, when the rate that functional resistance alleles are created decreases, the transgene persistence increases as fewer functional resistance alleles are created. By activating the SEM, our results show the transgene can be removed due to the creation of the resistance allele *v* that results following the excision of the transgene. As a result, regardless of how few resistance alleles are formed from NHEJ, the removal of the gene drive can be facilitated through the formation of the resistance allele *v.*

While we expected high rates of transgene excision (α) to result in the most rapid restoration of the wild-type population, lower rates of transgene excision were most effective. Due to the permanent transgene fixation that could occur as a result of mutations in the target site of the enzyme or the enzyme binding site[85], the rate of permanent fixation was assumed to be $\gamma > 0$. As a result, only alleles that have lost the target site of the drive-inducing protein Cas9 will be resistant to cutting[63,81,86]. Since the more of the resistance allele *v* is created when the excision-competent transgene is present for a longer period of time, lower rates of transgene excision ensure more of the resistance allele *v* is created. Once a permanently-fixed transgene is formed, the drive can no longer be contained through the SEM, but relies on the formation of resistance alleles to restore the wild-type population.

The proposed drive by Noble et al.[63] suggested multiplexing of five gRNA target sites for the homing drive to reduce the rate at which functional resistance alleles were being created. As a result, the strength of the gene drive is significantly increased, allowing for the establishment of the transgenic alleles in the population[83]. Applying this approach, our model produced equivalent results when the SEM was not active, where the transgene

drove quickly into the population and the number of functional resistance alleles was significantly reduced. However, once active, the SEM was capable of restoring the transgene-free population through the creation of the resistance allele *v*. The results indicate the use of a single gRNA target site SEM is effective in reversing powerful gene drive mechanisms, without relying on the misrepair of DNA to form resistance alleles.

Results from Kyrou et al.[64] indicate the formation functional resistance alleles can be significantly decreased, allowing for the dsx allele to spread throughout the population. Additionally, since the two copies of the transgene or nonfunctional resistance allele results in the death of females, the population size decreases as the number of reproducing females diminishes. With an inactive SEM, our high-cost sex-biasing gene drive models demonstrated the spread of the transgene and the subsequent decline in the number of individuals in the population. An active SEM was capable in reversing the dsx gene drive through the creation of resistance alleles *v* that would enable the restoration of transgene-free individuals. We also observed a rapid restoration of the wild-type population following a drive of the transgene for all rates of transgene excision. This was primarily because the only surviving females following the drive of the transgene carried the resistance allele *v*, allowing for a rapid spread of this resistance allele.

The CRISPR-SEM demonstrated the ability to reverse the drive of a transgene. While higher rates of transgene excision removed the excision-competent transgene quickly, this was not able to prevent a subsequent drive of the permanently-fixed transgene. In the single SEM gRNA target site models, since the threshold of transgene removal was not significantly higher than the initial release of transgenic individuals, the

method to prevent transgenic fixation was driven by the creation of a homing drive resistance allele that would allow transgene-free individuals to outcompete the transgenic individuals. Because the threshold for prevention of the homing drive was approximately two orders of magnitude below the initial release threshold at the highest rate of transgene excision, this may not be a sufficient construct for biocontainment of a gene drive carrying organism.

When five SEM gRNA target sites were modeled, the drive of the permanently-fixed transgene could not be prevented, however, the significant delay of the transgenic drive indicated a significant amount of the transgene had been eliminated through the SEM mechanism. Depending on the threshold for elimination of the gene drive[87], this may be a suitable construct for biocontainment and preventing a gene drive from escaping into the environment when the rate of transgene excision is high.

The CRISPR-SEM showed a trend of partial gene drive under low rates of excision. This enabled the transgene to partially drive into the population, before quickly decreasing and restoring the transgene-free population. While the impact of a fully transgenic population on the environment may be unknown, through this approach, the impacts could be limited to a fraction of the individuals for a short period. The subsequent removal of the transgene further diminishes the hazards of potential permanent effects the release of the gene drive individuals may have. This approach could serve as a useful tool for the pilot-testing of large transgenic populations in the environment and observing the short-term impacts transgenic populations have on the environment.

Since gene-drive technology is designed to spread, the transformation of entire wild-type populations could be permanent and uncontainable, policy regulations must consider the possible public and environment health risks[58–60,88]. Our results demonstrate the potential for limiting the presence of gene drive organisms in the environment through the use of an SEM. While many of the hazards associated with the transgenic transformation of an entire natural population cannot be quantified, our proposed approach could serve as a measure of containing the accidental or deliberate release of gene drive organisms into the environment. Furthermore, as demonstrated by the pairing of the SEM and CRISPR gene drive mechanism, the transgenic transformation of large proportions of the population before the complete restoration of the wild-type could be utilized for the pilot-testing of transgenic organism behavior and short-term effects on the environment that transgenic organisms may have.

The use of deterministic models provided a foundation for evaluating the dynamics of releasing a small proportion of CRISPR-SEM carrying individuals into a wild-type population. However, the assumptions of randomly mating populations, unlimited population growth, and no migration should be addressed in future work. Small-scale stochastic modeling will allow for the simulation of individual scenarios, but may not be representative of the behavior of populations in the environment. Through multiple stochastic simulations of cage trials, the likelihood of a permanently-fixed transgene that is not controllable by the SEM spreading throughout the population can be evaluated across multiple rates of transgene excision and permanent fixation.

CHAPTER III

MEDEA-SEM

Utilizing a maternally-linked toxin, the MEDEA gene drive mechanism is biased against the wild-type offspring of transgenic females carrying the MEDEA construct by causing mortality to all offspring that do not inherit a transgenic allele. Pairing the GD with a SEM, we explore two possible outcomes following the excision of the transgenic DNA. We consider a non-resistant transgene collapse, where the transgene excision results in the formation of a wild-type allele *w* that does not provide the antidote to the maternally-linked toxin. We also consider the formation of a resistance allele *v* that is created following the excision of the transgene that provides the antidote to the maternal toxin. For both scenarios, a fitness cost of 5% per transgene was assigned and only 5% of transgene-free offspring would be viable from transgenic females.

Non-Resistant Transgene Collapse

We consider the allele that is formed as a result of a successful excision of the transgene that is identical to the wild-type allele. This allele is considered non-resistant as it does not provide the antidote to the maternal toxin and will result in a nonviable offspring.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single CRISPR homing drive gRNA target site was modified for the non-resistant collapse MEDEA-SEM. The model structure and function were otherwise unchanged.

Equation Generation

For the formation of a non-resistant allele following the collapse of the transgene, three alleles are present: wild-type (non-resistant) *w*, excision-competent transgene *g*, and the permanently-fixed transgene *s*. In subsequent generations, the excision-competent transgene *g* is capable of excision, where the resulting allele formed is *w* (with probability α), and permanent fixation, resulting in the *s* allele (with probability γ). If neither of these occur, the transgene will remain excision-competent (with probability ß).

The probability that gametes are passed from parents to progeny is defined as follows:

- Individuals with a wild-type allele (*w*) and a permanently-fixed transgenic allele *s* will produce wild-type and permanently-fixed gametes equiprobably such that we have:

$$P_{ws,w} = P_{ws,s} = \frac{1}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele *g* will produce wild-type gametes *w* such that we have:

$$P_{wg,w} = \frac{1+\alpha}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele *g* will produce excision-competent transgenic gametes *g* such that we have:

$$P_{wg,g} = \frac{ß}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele *g* will produce permanently-fixed transgenic gametes *s* such that we have:

$$P_{wg,s} = \frac{\gamma}{2}$$

- Individuals with two excision-competent transgenic alleles *g* will produce wild-type gametes *w* such that we have:

$$P_{gg,w} = \alpha$$

- Individuals with two excision-competent transgenic alleles *g* will produce excision-competent gametes *g* such that we have:

$$P_{gg,g} = ß$$

- Individuals with two excision-competent transgenic alleles *g* will produce permanently-fixed gametes *s* such that we have:

$$P_{gg,s} = \gamma$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce wild-type gametes *w* such that we have:

$$P_{gs,w} = \frac{\alpha}{2}$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce excision-competent gametes *g* such that we have:

$$P_{gs,g} = \frac{ß}{2}$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce permanently-fixed transgenic gametes *s* such that we have:

$$P_{gs,s} = \frac{1+\gamma}{2}$$

To capture the MEDEA gene drive mechanism, an additional parameter was introduced to the model to induce mortality in the offspring from transgenic females. By capturing the female genotype, the decision to induce mortality could be applied accordingly. If the female was transgenic, the offspring genotype was evaluated for the presence of a transgenic allele (*g* or *s*). In the event that a transgenic allele was present in the offspring, no subsequent mortality was applied. However, if the offspring did not inherit a transgenic allele, a survival rate parameter was concatenated into the probability for the specific genotype, such that only a fraction (if any) of those offspring from the particular combination of male and female genotypes would survive.

Results

We modeled a variety of starting populations to understand the impact an SEM would have in controlling or reversing the MEDEA gene drive when the resulting allele of a successful transgene excision did not provide the antidote to the maternal toxin. We define *w* as the wild-type allele, *g* as the excision-competent transgenic alleles, and *s* as the permanently-fixed transgenic allele. With a release of 15% transgenic homozygous (*gg*) males into a population of wild-type males and females, the wild-type population was restored quickly when the SEM was active across all values of permanent transgene fixation (γ). Although an active SEM with higher rates of transgene excision (α) correlated with the rate at which the wild-type population was restored, an inactive SEM was also sufficient to begin the restoration of the wild-type population within 60 generations, as shown in Figure 25. Allele frequencies in Figure 26 show a decrease in the transgenic alleles for an inactive SEM (where $\alpha = 0$), due to the fitness cost associated with each transgenic allele.

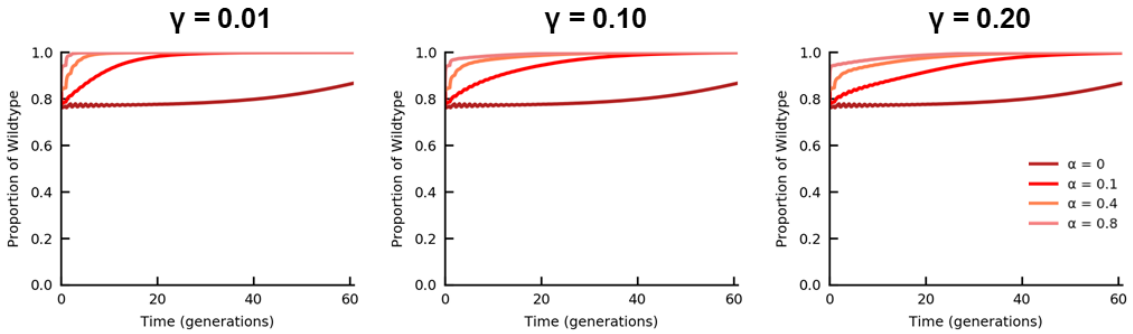

Figure 25. Proportion of transgene-free individuals with a simulated release of MEDEA-SEM containing males at 15% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

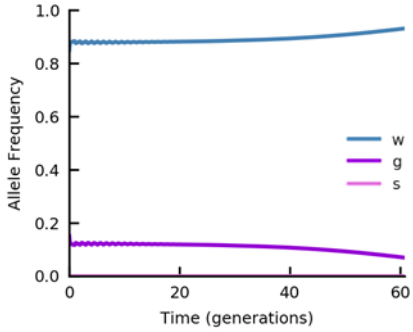

Figure 26. Allele frequencies of a simulated release of MEDEA-SEM containing males at 15% of a wild-type population with an inactive SEM (α = 0, γ = 0)

As the starting population of transgenic males was increased to 25% and 33%, an active SEM with a low rate of permanent transgene fixation was necessary in order to restore the wild-type population. Increasingly higher rates of transgene excision were necessary as the rate of permanent fixation increased (Figure 27). An increase in the allele frequency of the permanently-fixed transgene *s* in Figure 28 and Figure 29 shows that the

rate of transgene excision α must be sufficiently higher to counteract the rate of permanent transgene fixation γ to restore the wild-type population.

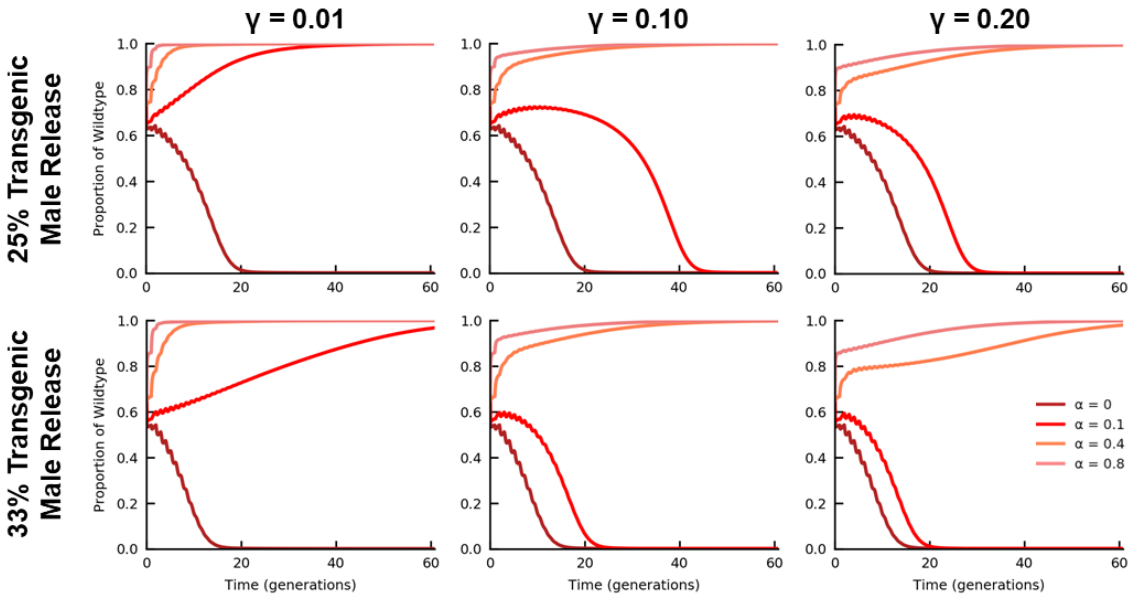

Figure 27. Proportion of transgene-free individuals with a simulated release of MEDEA-SEM containing males at 25 and 33% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

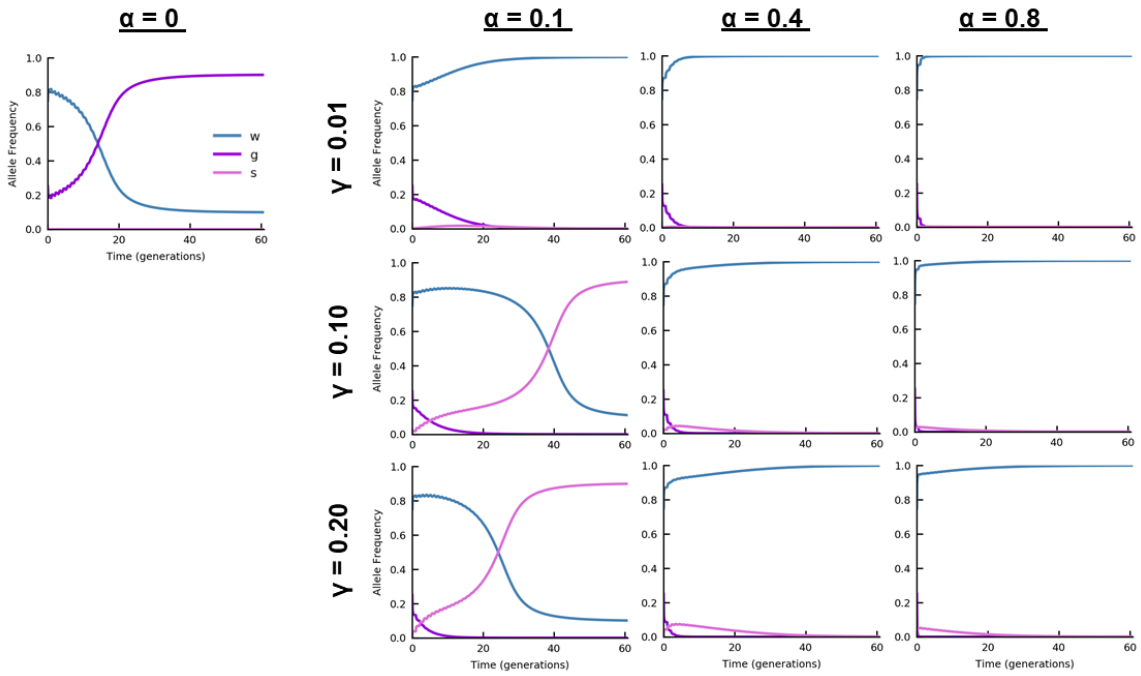

Figure 28. Allele frequencies of a simulated release of MEDEA-SEM containing males at 25% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

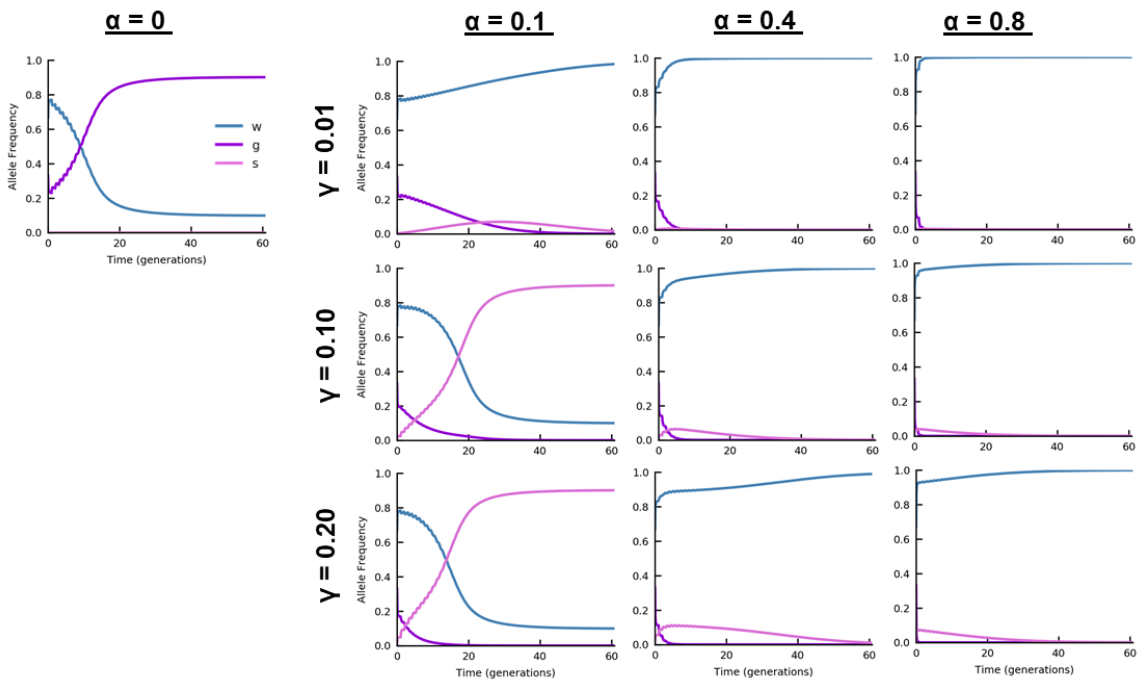

Figure 29. Allele frequencies of a simulated release of MEDEA-SEM containing males at 33% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

Exploring the parameter space between the rate of transgene excision and the rate of permanent transgene fixation, we observe a distinct threshold between the restoration of the wild-type and removal of all wild-type individuals within 60 generations. As the initial starting population of males increased from 25% to 33%, the parameter space for restoring the wild-type population decreased, as shown in Figure 30.

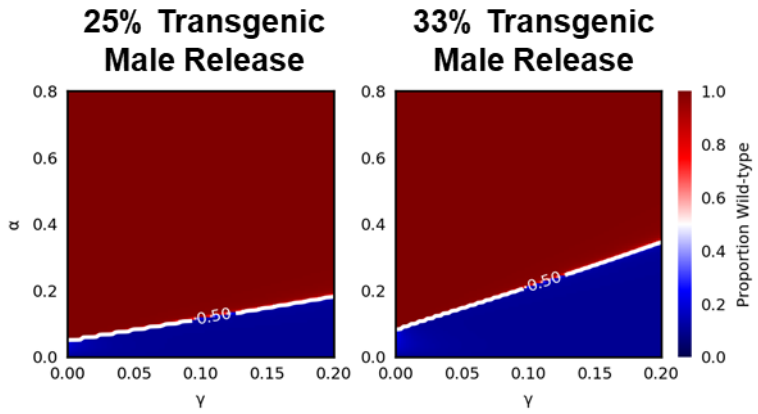

Figure 30. Wild-type allele frequencies of a simulated release of MEDEA-SEM containing males at 25 and 33% of a wild-type population after 60 generations

Populations of larger releases of transgenic males and females were also explored, including starting populations consisting of 50, 75, and 100% transgenic males and females. As the starting proportion of transgenic individuals increased, the acceptable ranges of α and γ for the restoration of wild-type population decreased significantly, as shown in Figure 31. Figure 32-Figure 34 provide the allele frequencies for starting population of 50, 75, and 100% transgenic individuals, respectively. The parameter spaces of proportion of wild-type alleles after 60 generations for starting populations of 50, 75, and 100% transgenic individuals are provided in Figure 35.

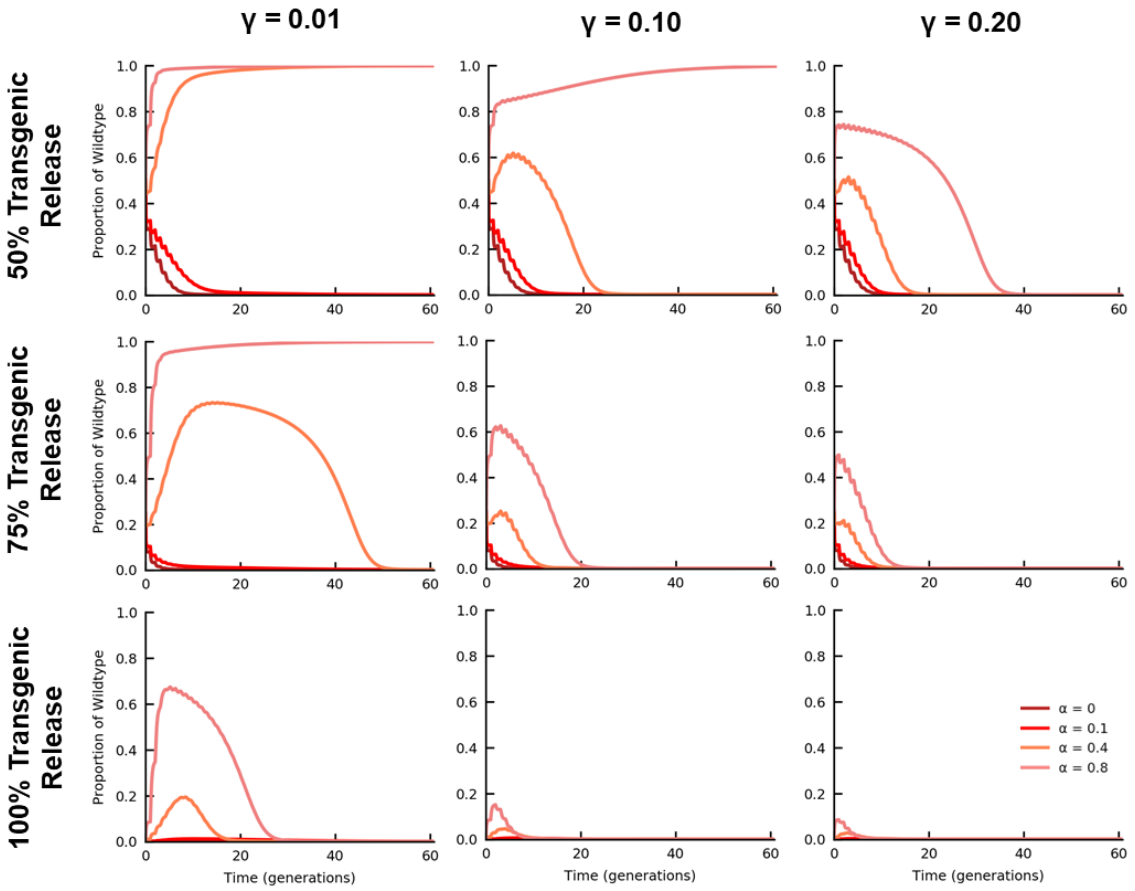

Figure 31. Proportion of transgene-free individuals with a simulated release of MEDEA-SEM containing population at 50, 75, and 100% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

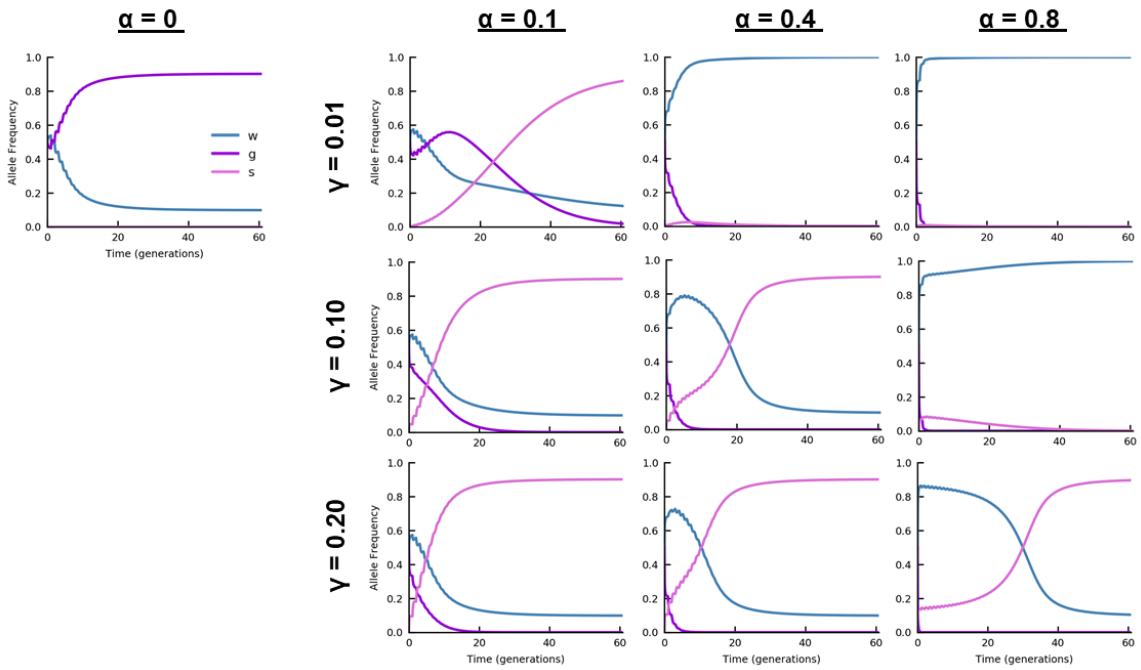

Figure 32. Allele frequencies of a simulated release of MEDEA-SEM containing 50% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

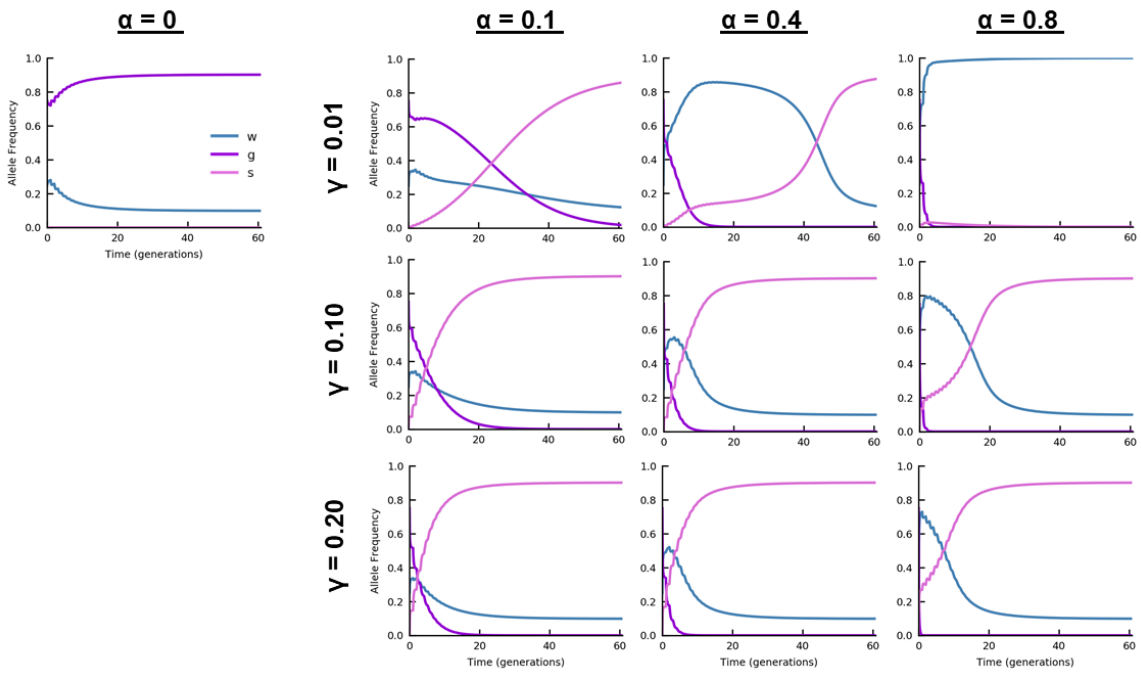

Figure 33. Allele frequencies of a simulated release of MEDEA-SEM containing population at 75% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

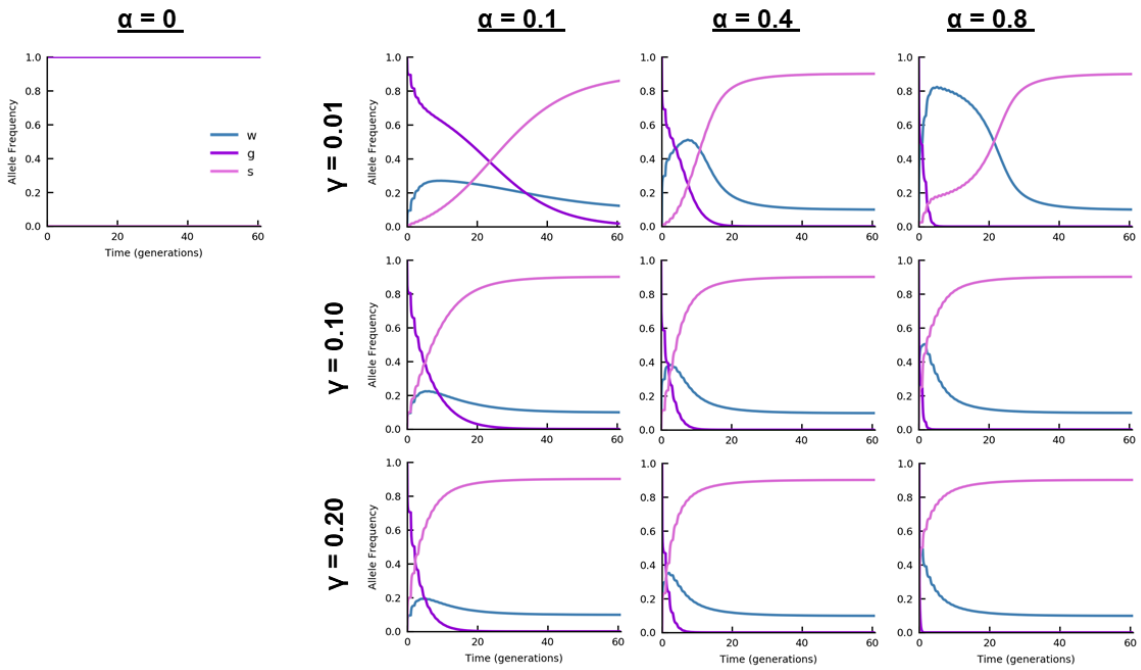

Figure 34. Allele frequencies of a simulated MEDEA-SEM containing population at 100% transgenic individuals at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

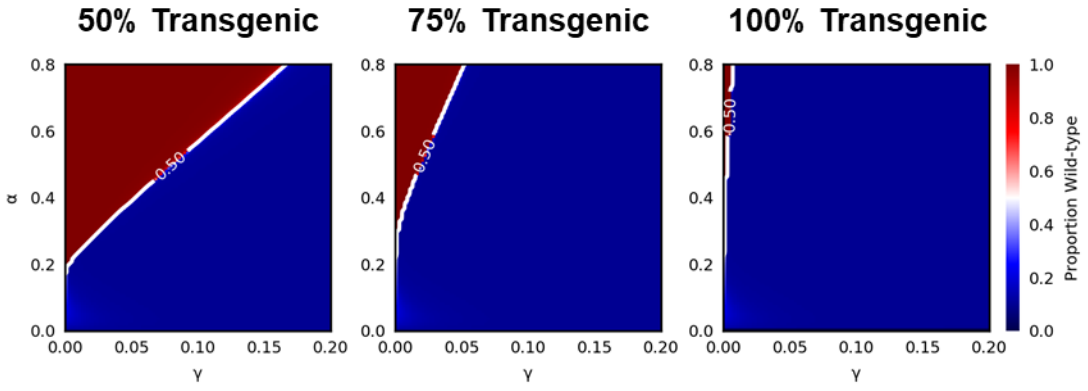

Figure 35. Wild-type allele frequencies of a simulated release of MEDEA-SEM containing males at 50, 75, and 100% transgenic individuals of a wild-type population after 60 generations

Resistant Transgenic Collapse

We consider the allele that is formed as a result of a successful excision of the transgene that is equivalent to the wild-type allele, but provides the antidote to the maternally-linked toxin. This allele is considered as resistant as it provides the antidote to the maternal toxin and will result in a nonviable offspring.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single CRISPR homing drive gRNA target site was modified for the resistant collapse MEDEA-SEM. The model structure and function were otherwise unchanged.

Equation Generation

For the formation of a resistant allele following the collapse of the transgene, four alleles are present: wild-type allele (non-resistant) *w*, resistance allele *v*, excision-competent transgene *g*, and the permanently-fixed transgene *s*. In subsequent generations, the excision-competent transgene *g* is capable of excision, where the resulting allele formed is *v* (with probability α), and permanent fixation, resulting in the *s* allele (with probability γ). If neither of these occur, the transgene will remain excision-competent (with probability ß).

The probability that gametes are passed from parents to progeny is defined as follows:

- Individuals with a wild-type allele (*w* or *v*) and the a permanently-fixed transgenic allele *s* will produce wild-type and permanently-fixed gametes equiprobably such that we have:

$$P_{ws,w} = P_{ws,s} = P_{wv,w} = P_{wv,v} = P_{vs,v} = P_{vs,s} = \frac{1}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele *g* will produce wild-type gametes *w* such that we have:

$$P_{wg,w} = \frac{1}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele *g* will produce resistant gametes *v* such that we have:

$$P_{wg,v} = \frac{\alpha}{2}$$

- Individuals with a resistant allele *v* and an excision-competent transgenic allele *g* will produce resistant gametes *v* such that we have:

$$P_{vg,v} = \frac{1+\alpha}{2}$$

- Individuals with a wild-type allele *w* or resistant allele *v* and an excision-competent transgenic allele *g* will produce excision-competent transgenic gametes *g* such that we have:

$$P_{wg,g} = P_{vg,g} = \frac{ß}{2}$$

- Individuals with a wild-type allele *w* or resistant allele *v* and an excision-competent transgenic allele *g* will produce permanently-fixed transgenic gametes *s* such that we have:

$$P_{wg,s} = P_{vg,s} = \frac{\gamma}{2}$$

- Individuals with two excision-competent transgenic alleles *g* will produce resistant gametes *v* such that we have:

$$P_{gg,v} = \alpha$$

- Individuals with two excision-competent transgenic alleles *g* will produce excision-competent gametes *g* such that we have:

$$P_{gg,g} = ß$$

- Individuals with two excision-competent transgenic alleles *g* will produce permanently-fixed gametes *s* such that we have:

$$P_{gg,s} = \gamma$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce resistance gametes *v* such that we have:

$$P_{gs,v} = \frac{\alpha}{2}$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce excision-competent gametes *g* such that we have:

$$P_{gs,g} = \frac{ß}{2}$$

- Individuals with one excision-competent transgenic allele *g* and a permanently-fixed transgene *s* will produce permanently-fixed transgenic gametes *s* such that we have:

$$P_{gs,s} = \frac{1+\gamma}{2}$$

As in the MEDEA non-resistant scenario, if the female was transgenic, a mortality rate would be applied to all of the offspring that were not transgenic or did not inherit the *v* allele. Offspring that were not transgenic, but inherited the *v* allele would not have the MEDEA-induce mortality rate applied.

Results

We explore the creation of a resistance allele *v*, which provides immunity to the maternally-linked toxin, resulting from the successful transgene excision from the transgenic *g* allele in the MEDEA gene drive. We model the release of 15, 25, and 33% of the starting population as transgenic (*gg*) males, with the remaining population consisting of wild-type males and females. Similar to the non-resistant case, an inactive SEM ($\alpha = 0$) results in a gradual restoration of the wild-type in the starting population of 15% transgenic males and a successful drive of the transgene in the starting populations of 25 and 33% transgenic males. We observe the removal of the transgenic population in all cases when the SEM is active across all rates of permanent transgene fixation, as shown in Figure 36. Figure 37 provides the allele frequency plots across rates of transgene excision.

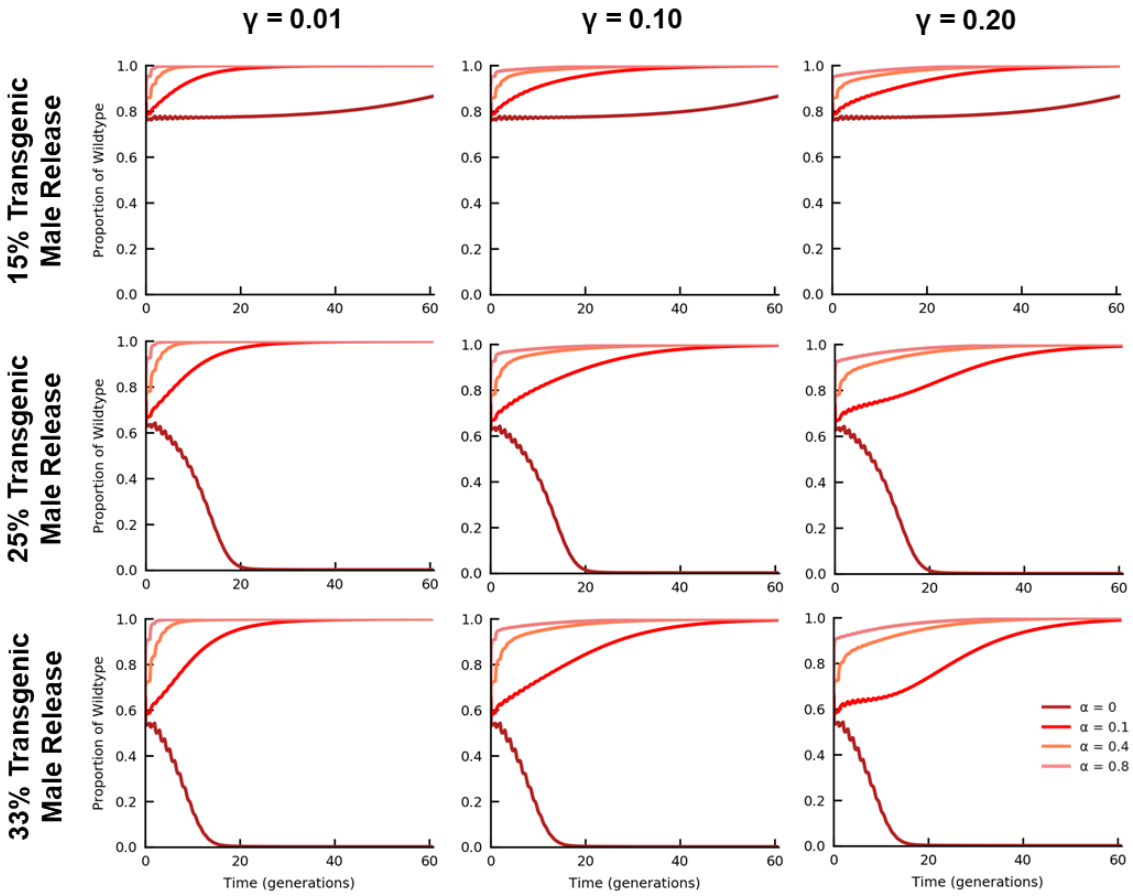

Figure 36. Proportion of transgene-free individuals with a simulated release of resistant MEDEA-SEM containing males at 15, 25, and 33% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

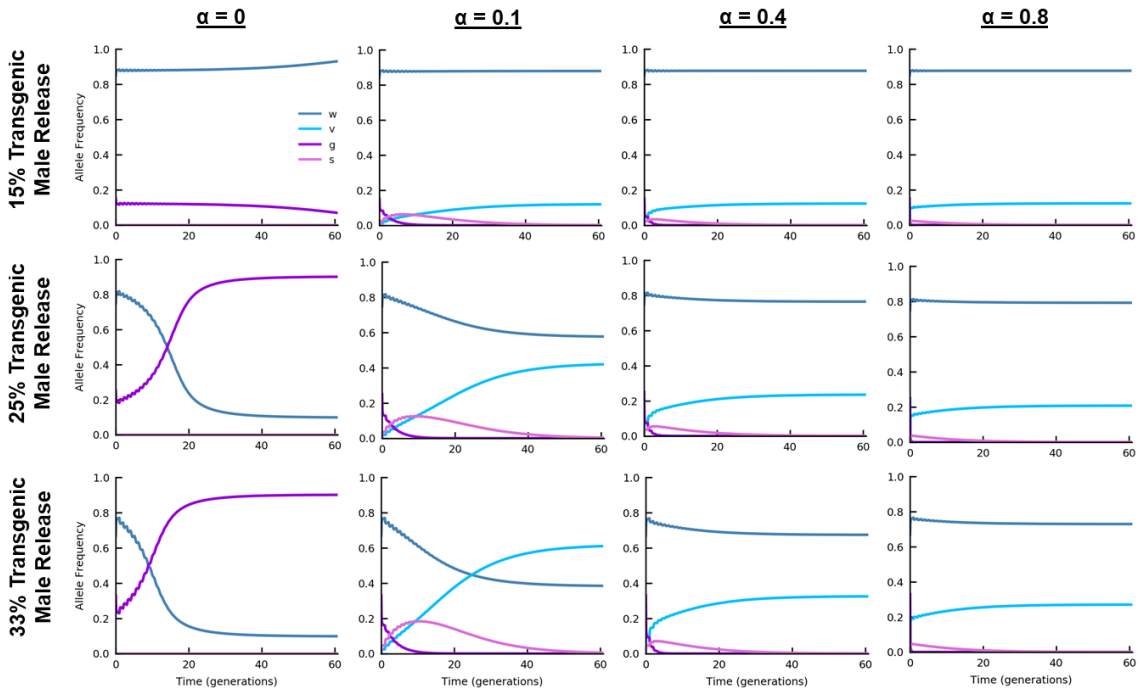

Figure 37. Allele frequencies of a simulated release of resistant MEDEA-SEM containing males at 15, 25, and 33% of a wild-type population at four different rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and where γ = 0.20

Modeling the release of transgenic males and females, the starting populations of 50, 75, and 100% transgenic individuals demonstrated the ability of the SEM to restore the wild-type population when active across all starting populations (as shown in Figure 38). Unlike the non-resistance transgene collapse scenario, the creation of the resistance allele *v* in the resistance transgene collapse case enables the restoration of the wild-type population across all starting populations. Figure 39 indicates the prevalence of the resistance allele *v* as the transgene is collapsed and the wild-type population is restored.

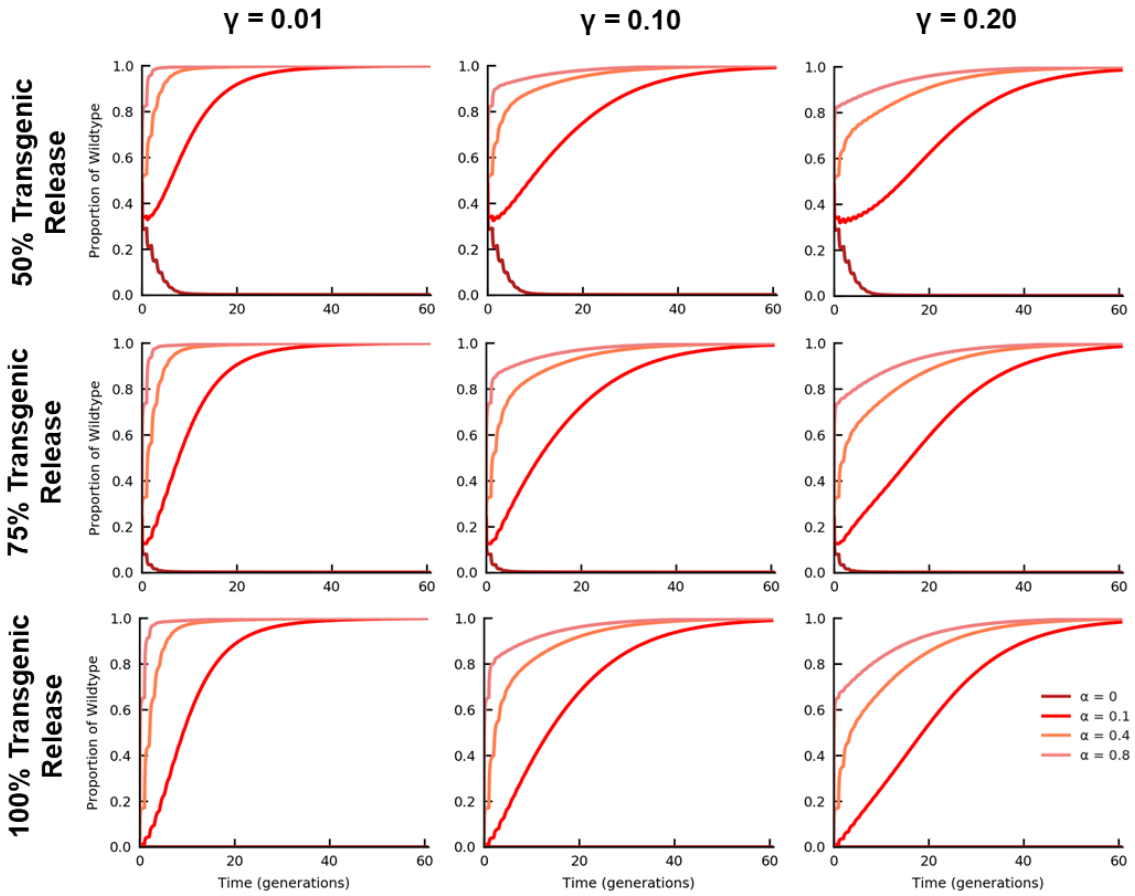

Figure 38. Proportion of transgene-free individuals with a simulated release of resistant collapse MEDEA-SEM containing population at 50, 75, and 100% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

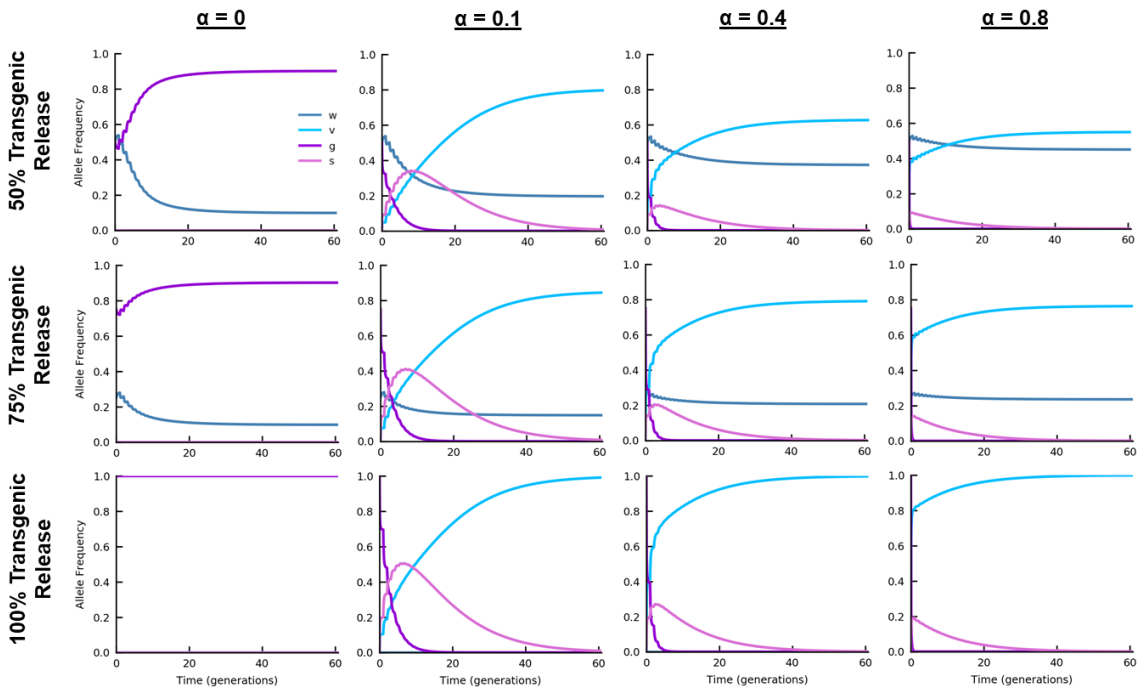

Figure 39. Allele frequency with a simulated release of resistant collapse MEDEA-SEM containing population at 50, 75, and 100% at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and where γ = 0.20

Discussion

The approach proposed by Marshall et al.[41] explored the ability to confine the gene drive to a single population, using a conservative estimate of 1% migration per generation and showed confinement was achievable in the presence of a high fitness cost of c=0.5. As noted in the study, this is not a reliable mechanism to achieve containment of the gene drive, since studies have shown fitness cost to be lower than c=0.2 in *Anopheles* mosquitoes[89]. Chen et al.[42] demonstrated the importance of fitness costs in models with starting proportions of transgenic males of 25%, where fitness costs of c=0.2 would result in the decrease of the MEDEA allele within 20 generations. Experimental results also indicated this fitness cost is likely to approach c=0 in *Drosophila* and result in the spread

of the MEDEA transgene throughout the population[42]. When utilizing the fitness costs and maternal toxin survival rates proposed by Chen et al.[42], the release of 25% MEDEA bearing males into a population consisting of 25% wild-type males and 50% wild-type females showed identical trends when the SEM was inactive across all fitness costs.

The non-resistant transgene collapse using a MEDEA gene drive mechanism demonstrated that a threshold starting population of more than 15% transgenic males was necessary to achieve transgene fixation in the population when the SEM mechanism was not active. Although our models did not explore the impacts of migration between multiple populations, the results indicate a release of 15% MEDEA bearing males will not be successful in establishing the transgene in the wild-type population, however larger starting populations of transgenic males would cause a drive. Similar results have been observed in previous modeling studies for equivalent fitness costs and release ratios[90]. At higher introduction rates of MEDEA-bearing males, fixation into the population was achieved when the SEM was not active.

In the non-resistant scenario, the removal of the transgene would result in the formation of a wild-type allele that would not provide the antidote to the maternally-linked toxin, allowing for the drive to reoccur in future generations. While the rate at which MEDEA transgene was removed from the population depended on a combination of the rate of transgene excision and the rate of permanent transgene fixation. Since the permanently-fixed transgene behaved similarly to the cases of an inactive SEM, once the frequency of the permanently-fixed transgene exceeded 15%, a subsequent drive was achieved. Because the permanently-fixed transgene was produced from the excision-

competent transgene, more permanently-fixed transgenes were produced when the excision-competent transgenes remained longer in the population (due to a lower rate of transgene excision). As a result, scenarios considering high rates of transgene fixation required high rates of transgene excision to remove the transgene from the population. Likewise, increasing the starting proportion of transgenic individuals also required high rates of transgene excision to eliminate the transgene and restore the wild-type population. Due to the high starting proportions of transgenic individuals, the abundance of excision-competent transgenes would cause an increased rate of permanently-fixed transgene formation that could not be controlled by the SEM. As a result, the excision-competent transgene had to be removed from the population before the permanently-fixed transgene frequency increased beyond 15% to ensure the prevention of the drive into the population.

In the MEDEA-SEM scenario where the collapse of the transgene resulted in the formation of an allele resistant to the maternal toxin, the results were identical to those of the non-resistant MEDEA-SEM when the SEM was inactive. When the SEM was active, the population was restored under all starting proportions of transgenic individuals. Since the collapsed transgene provides immunity to the maternally-linked toxin, the progeny of transgenic females inheriting the resistance allele are resistant against drive and can proliferate the resistance alleles in future generations. Hence, even when the permanently fixed-transgene exceeded 15% (the level at which the drive was not preventable for an inactive SEM), the presence of the resistance allele created from the excision of the excision-competent transgene would allow for the restoration of the wild-type population. This was consistent across all starting proportions of transgenic individuals.

Containment of a gene drive mechanism is a major concern for public policy[58,88]. Along with previous studies[41,90], our study results indicate a large number of individuals carrying the MEDEA construct would be required to cause a transformation of the population when a fitness cost is applied. Furthermore, with the introduction of a SEM with a probability for transgene excision in subsequent generations, the MEDEA gene drive construct could serve a powerful method of biocontainment to limit the hazards of an accidental gene drive carrying individual release into the environment.

Utilizing a SEM would further reduce the likelihood of transgene fixation in the population when the probability of transgene excision is high and the rate of permanent transgene fixation is low when a non-resistant MEDEA construct is utilized and in across all rates of transgene excision and permanent fixation when a resistant MEDEA construct is used. In cases where the transgene was observed to drive into the population, the permanently-fixed transgenes (not controllable by the SEM) were the primary drivers. By reducing the likelihood that the transgene is permanently fixed, the drive of the transgene into the population is further reduced.

Unlike previous studies[41,90], we did not consider migration between multiple populations and limited our models to a single population. Future work will focus on modeling the spread of the MEDEA-SEM construct across multiple populations and the release, transgene excision, and permanent fixation parameters required for successful containment of the drive.

CHAPTER IV

UD-SEM

The maternal effect lethal under-dominance gene drive utilizes two maternally-linked toxin-antidote pairs (referred to as *toxin A* and *toxin B*). Similar to MEDEA, transgenic females provide the toxins they express to their offspring, causing mortality within the offspring unless they inherit the complementary toxin-antidote allele. Offspring from transgenic females with both toxin-antidotes will need to inherit both toxin-antidote pairs to survive. Since individuals that inherit only one transgenic allele will suffer high mortality rates, the release of small proportions of transgenic individuals is not adequate to cause a drive of the transgene into the population. We assume the survival rate of each toxin to be 5%, where the survival rate is multiplicative when two toxin-antidotes are present in the female parent (hence the survival rate is 0.05 x 0.05 = 0.0025). Additionally, since there are two transgenes corresponding to *toxin A* and *toxin B*, we differentiate between the two excision-competent transgenes g_A and g_B, where the permanent fixation of the transgenes will result in permanently-fixed transgenes s_A and s_B, respectively.

Single-Locus SEM

We evaluate the dynamics of an UD-SEM construct where both toxin-antidote pairs are placed at the same locus. A total of 15 allele combinations (genotypes) were possible from the five alleles utilized. We assume a single chance for self-elimination of the transgenic DNA.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single CRISPR homing drive gRNA target site was modified for the single-locus UD-SEM. The model structure and function were otherwise unchanged.

Equation Generation

For the dynamics of a single-locus UD-SEM model, five alleles are present: wild-type *w*, excision-competent transgenes g_A and g_B, and the permanently-fixed transgenes s_A and s_B. In subsequent generations, the excision-competent transgenes g_A and g_B *are* capable of excision, where the resulting allele formed is *w* (with probability α), and permanent fixation, resulting in the s_A and s_B alleles (with probability γ), respectfully. If neither of these occur, the transgene will remain excision-competent (with probability ß).

The probability that gametes are passed from parents to progeny is defined as follows:

- Individuals with a wild-type allele (*w*) and a permanently-fixed transgenic allele s_A or s_B will produce wild-type and permanently-fixed gametes equiprobably such that we have:

$$P_{ws_A,w} = P_{ws_A,s_A} = P_{ws_B,w} = P_{ws_B,s_B} = \frac{1}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele g_A or g_B will produce wild-type gametes *w* such that we have:

$$P_{wg_A,w} = P_{wg_B,w} = \frac{1+\alpha}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele g_A or g_B will produce excision-competent transgenic gametes g_A and g_B such that we have:

$$P_{wg_A,g_A} = P_{wg_B,g_B} = \frac{ß}{2}$$

- Individuals with a wild-type allele *w* and an excision-competent transgenic allele g_A or g_B will produce permanently-fixed transgenic gametes s_A or s_B such that we have:

$$P_{wg_A,s_A} = P_{wg_B,s_B} = \frac{\gamma}{2}$$

- Individuals with two excision-competent transgenic alleles g_A or g_B will produce wild-type gametes *w* such that we have:

$$P_{g_Ag_A,w} = P_{g_Ag_B,w} = P_{g_Bg_B,w} = \alpha$$

- Individuals with two identical excision-competent transgenic alleles g_A or g_B will produce excision-competent gametes g_A or g_B such that we have:

$$P_{g_Ag_A,g_A} = P_{g_Bg_B,g_B} = ß$$

- Individuals with two excision-competent transgenic alleles g_A and g_B will produce excision-competent gametes g_A and g_B such that we have:

$$P_{g_Ag_B,g_A} = P_{g_Ag_B,g_B} = \frac{ß}{2}$$

- Individuals with two identical excision-competent transgenic alleles g_A or g_B will produce permanently-fixed gametes s_A or s_B such that we have:

$$P_{g_A g_A, s_A} = P_{g_B g_B, s_B} = \gamma$$

- Individuals with two excision-competent transgenic alleles g_A and g_B will produce excision-competent gametes s_A and s_B such that we have:

$$P_{g_A g_B, s_A} = P_{g_A g_B, s_B} = \frac{\gamma}{2}$$

- Individuals with one excision-competent transgenic allele g_A or g_B and a permanently-fixed transgene s_A or s_B will produce wild-type gametes w such that we have:

$$P_{g_A s_A, w} = P_{g_A s_B, w} = P_{g_B s_A, w} = P_{g_B s_B, w} = \frac{\alpha}{2}$$

- Individuals with one excision-competent transgenic allele g_A or g_B and a permanently-fixed transgene s_A or s_B will produce excision-competent transgenic gametes g_A and g_B such that we have:

$$P_{g_A s_A, g_A} = P_{g_A s_B, g_A} = P_{g_B s_A, g_B} = P_{g_B s_B, g_B} = \frac{ß}{2}$$

- Individuals with one excision-competent transgenic allele g_A and a permanently-fixed transgene s_A will produce permanently-fixed transgenic gametes s such that we have:

$$P_{g_A s_A, s_A} = \frac{1 + \gamma}{2}$$

- Individuals with one excision-competent transgenic allele g_B and a permanently-fixed transgene s_B will produce permanently-fixed transgenic gametes s such that we have

$$P_{g_B s_B, s_B} = \frac{1 + \gamma}{2}$$

- Individuals with one excision-competent transgenic allele g_A or g_B and the opposite permanently-fixed transgene s_A or s_B will produce permanently-fixed transgenic gametes s_A or s_B such that we have:

$$P_{g_A s_B, s_A} = P_{g_B s_A, s_B} = \frac{\gamma}{2}$$

Once the final genotypes were derived, the female parent genotype was evaluated. If the female genotype was transgenic, the presence of toxins *A* and *B* were identified. Because the mortality rate could be impacted differently by each toxin, Ω_A and Ω_B were assigned as the survival rates of offspring exposed to toxins *A* and *B*, respectively. If the female was transgenic with the toxin *A* (g_A, s_A), the offspring must inherit the g_B or s_B alleles to survive. Otherwise, Ω_A was concatenated to the probability of the offspring. Likewise, if the female was transgenic with the toxin B, the offspring must inherit the g_A or s_A alleles to survive; Ω_B was concatenated to the probability otherwise. If the female was transgenic with both toxins, the offspring must inherit both transgenes to provide both antidotes. Inheritance of a single transgenic antidote would still result in the application of the missing toxin survival rate Ω_A or Ω_B.

Results

The single-locus scenario establishes the toxin-antidote pairs on the same chromosome. We assign a fitness cost of 5% per transgenic allele. The starting population consisting of 50% transgenic males and females ($g_A g_B$) with a wild-type population of males and females showed the removal of the transgenic individuals was possible across all rates of transgene excision and permanent fixation (Figure 40). Increasing the starting

population of transgenic individuals to 75%, the restoration of the wild-type population was only possible with an active SEM. Lastly, a starting population consisting solely of transgenic individuals was only possible under combinations of high rates of transgene excision and low rates of permanent transgenic fixation. In a completely transgenic population, the rate permanent fixation of the transgene is the primary inhibitor of the wild-type population restoration. However, if the transgene is rapidly removed successfully through the SEM, the wild-type population is fully restored within 20 generations. Figure 40 shows the proportion of wild-type individuals for the three starting populations across multiple levels of transgene fixation and excision, while Figure 41 indicates the wild-type allele frequency after 60 generations for α and γ. Progressions of individual alleles for 50 and 75% transgenic starting populations are provided in Figure 42 and allele frequencies for 100% transgenic starting populations are shown in Figure 43.

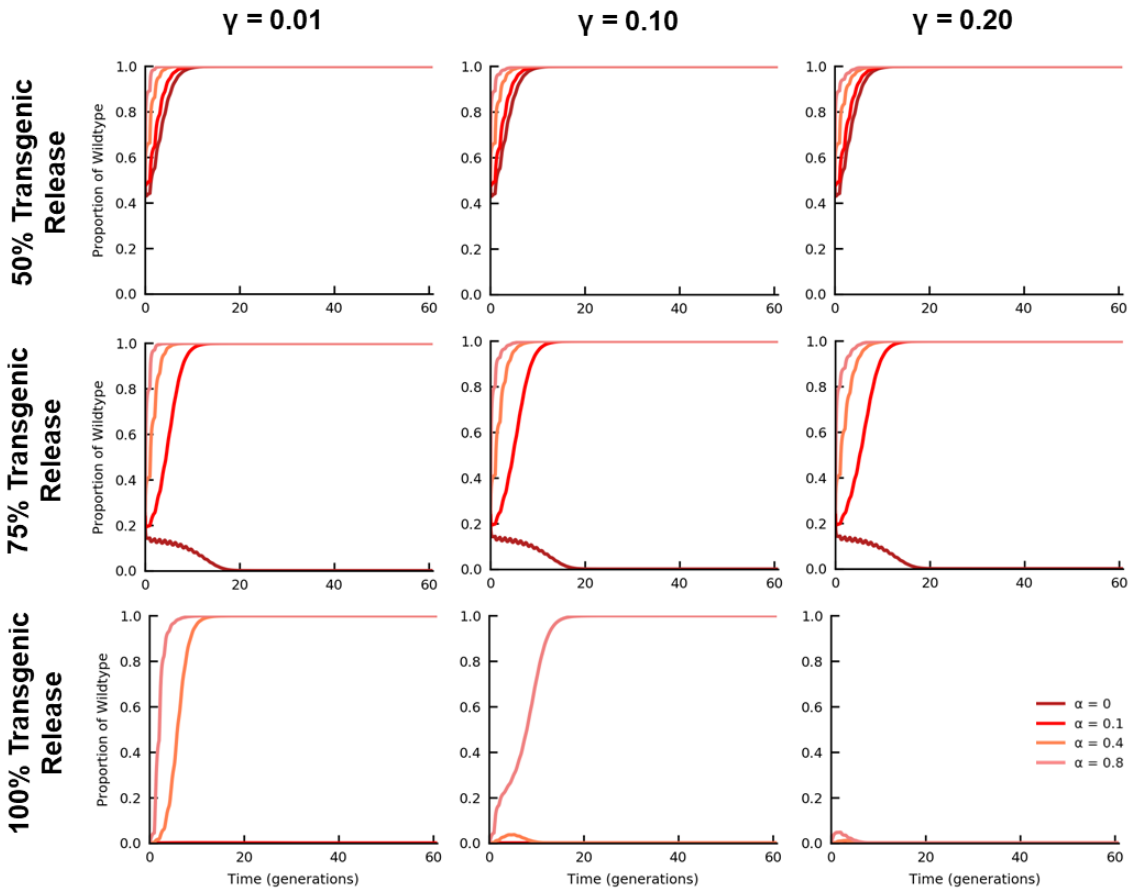

Figure 40. Proportion of transgene-free individuals with a simulated release of resistant collapse single locus UD-SEM containing population at 50, 75, and 100% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

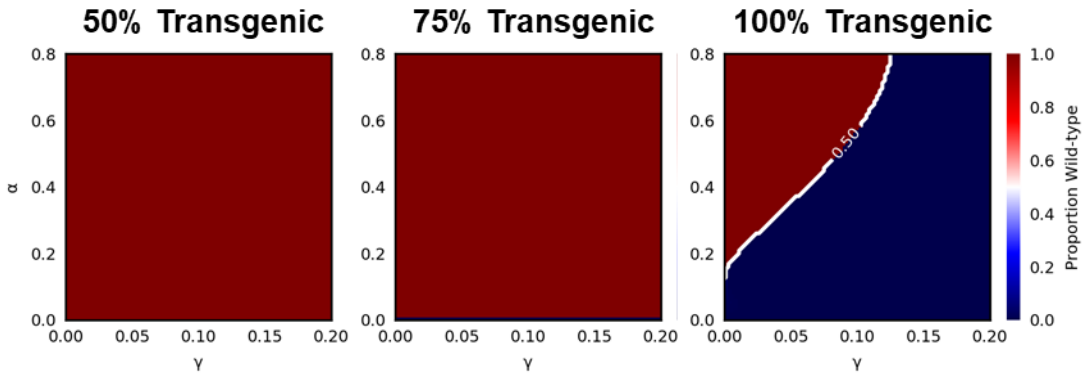

Figure 41. Wild-type allele frequencies of a simulated release of single-locus UD-SEM containing individuals at 50, 75, and 100% transgenic individuals of a wild-type population after 60 generations

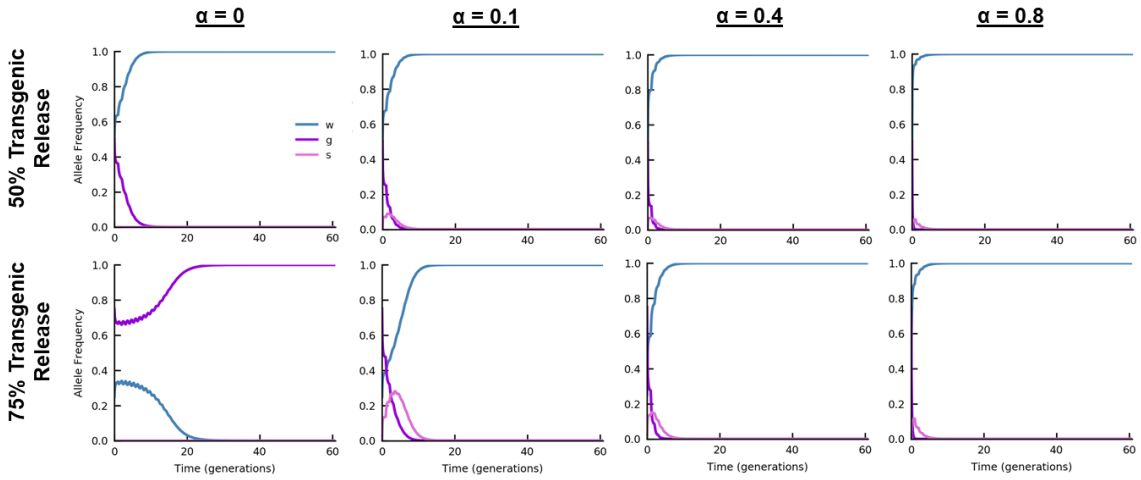

Figure 42. Allele frequency with a simulated release of single-locus UD-SEM containing population at 50 and 75% transgenic individuals at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and where γ = 0.20

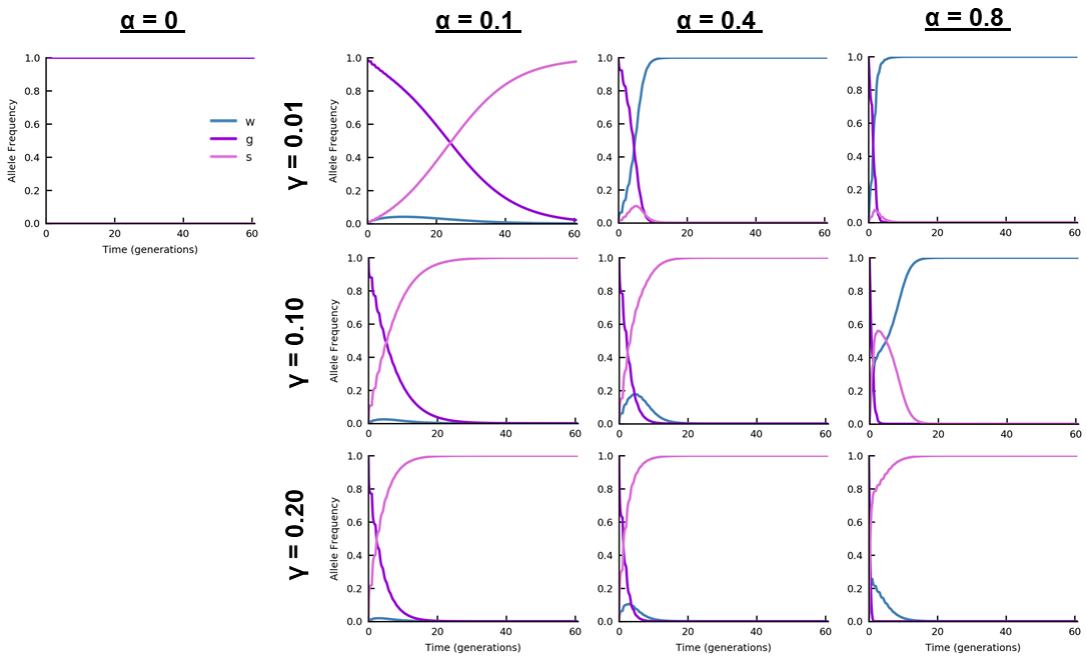

Figure 43. Allele frequency with a simulated release of single-locus UD-SEM containing population at 100% transgenic individuals at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

Two-Locus SEM

We explore the dynamics of a UD-SEM system where the toxin-antidotes *A* and *B* are located on separate loci. Six genotypes were possible per locus, yielding 36 allele combinations (genotypes) in both loci. As a result, several more viable transgenic genotypes were included. Similar to the single-locus case, we assume a single chance for self-elimination of the transgenic DNA.

Methodology

Model Structure

The system dynamics model detailed in the methodology of the single SEM gRNA target site and single CRISPR homing drive gRNA target site was modified for the two-locus UD-SEM. The model structure and function were otherwise unchanged.

Equation Generation

Compared to the single-locus UD model, the two-locus UD model has the same types of alleles present. However, we limit the *A* and *B* toxin-antidotes to separate loci, which limits the number of outcomes per locus. The equation generation for the two-locus model follows the single-locus equation generation for each locus independently, where applicable. For example a parent with a g_Ag_A genotype on the first locus (in the two-locus case) will still produce the same gametes at the same rates as the g_Ag_A parent in the single-locus case, but a g_Ag_B genotype would not be possible in the two-locus case since the toxin-antidotes are on separate loci. Instead we have *ww*, wg_A, ws_A, g_Ag_A, g_As_A, and s_As_A present on the first locus and *ww*, wg_B, ws_B, g_Bg_B, g_Bs_B, and s_Bs_B present on the second locus.

The probability that a combination of gametes passed to progeny is then calculated as the product of each gamete from each locus. For example, the rate at which an s_A gamete is given from a g_As_A locus is defined as:

$$P_{g_As_A,s_A} = \frac{1+\gamma}{2}$$

and the rate at which a *w* gamete is given from a wg_B locus is defined as:

$$P_{wg_B,w} = \frac{1 + ß}{2}$$

Combining these two rates together, we can define the rate at which s_A-*w* gametes will be inherited from g_As_A-wg_B parents as:

$$P_{g_As_Awg_B,s_Aw} = (P_{g_As_A,s_A})(P_{wg_B,w}) = (\frac{1+\gamma}{2})(\frac{1+ß}{2})$$

Like the single-locus UD-SEM model, once the probabilities were calculated, the parental female genotype was evaluated. If the female was transgenic, the created genotype in the offspring was evaluated for the respective antidote(s) that were present in the female. If the antidote for the female toxin was not present in the offspring, only a fraction (if any) of the offspring would survive (applied through the survival rates Ω_A and Ω_B).

Results

In the two-locus case, the toxin-antidote pairs are placed on different chromosomes. As a result, the individuals can have multiple combinations of the possible toxin-antidotes that allow for survival. We assign a fitness cost of 2.5% per transgenic allele, such that individuals with four transgenic alleles (two on each chromosome) have a maximum fitness cost of 10%. The survival rate of transgene-free progeny of transgenic females remained unchanged.

A starting population of 50% transgenic males and females was able to completely restore the wild-type population when the rate of transgene excision was high. Lower rates

of transgene excision resulted in the spread of the transgene throughout the population, decreasing the number of wild-type individuals as shown in Figure 44. When the starting population of transgenic individuals was increased to 75%, high rates of excision were also capable of restoring the wild-type population. A starting population of 100% transgenic individuals was only able to restore the wild-type population when the rate of permanent transgene fixation was very small ($\gamma < 0.01$) and where the rate of transgene excision was high. Allele frequencies are provided in Figure 44-Figure 47 for 50, 75 and 100% transgenic starting populations and with parameter spaces are provided in Figure 48.

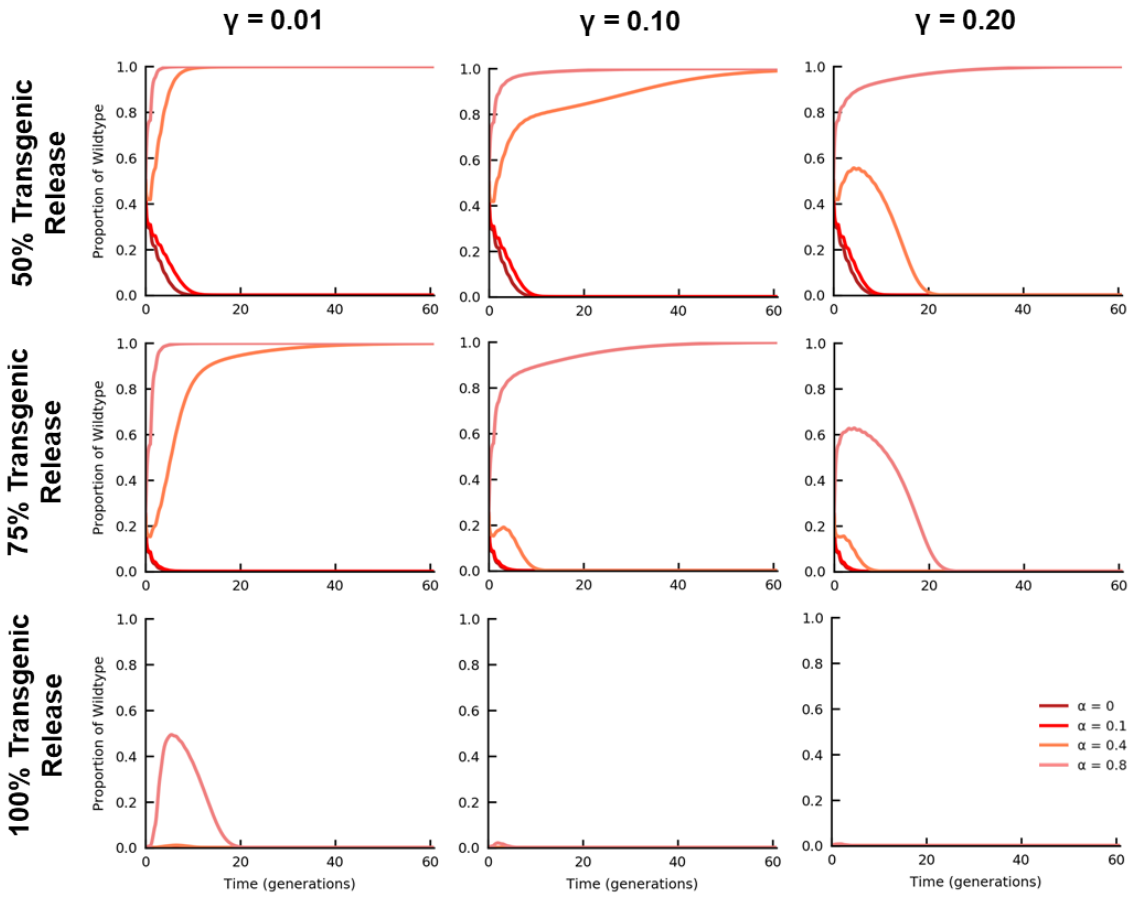

Figure 44. Proportion of transgene-free individuals with a simulated release of resistant collapse two-locus UD-SEM containing population at 50, 75, and 100% of a wild-type population at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

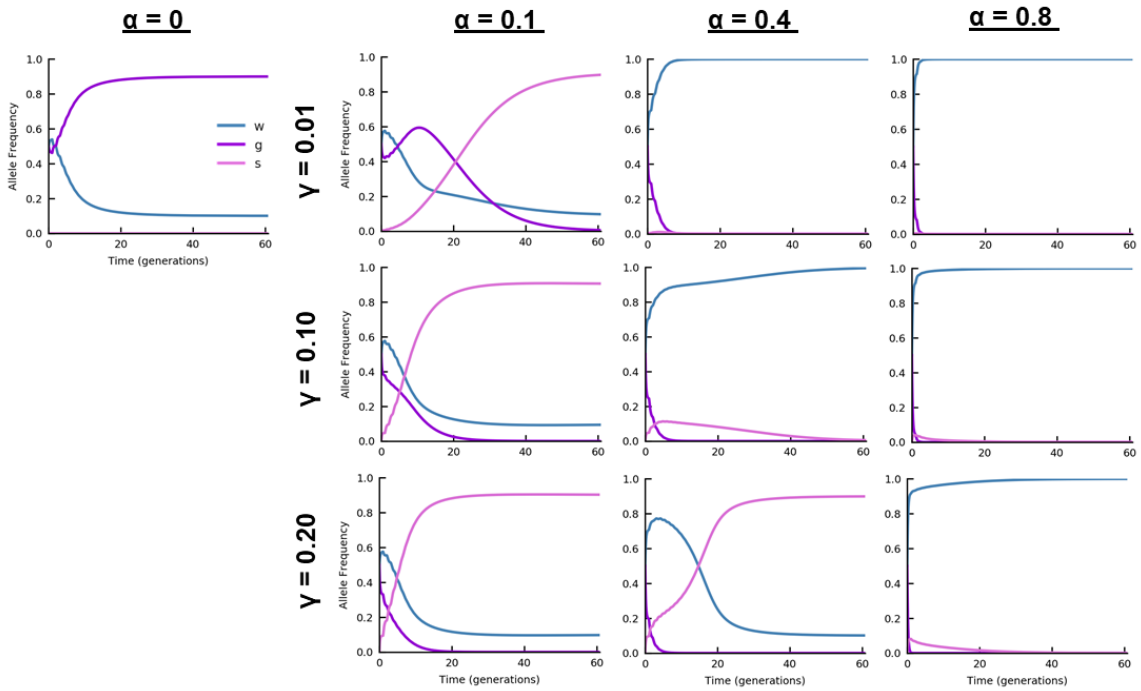

Figure 45. Allele frequency with a simulated release of two-locus UD-SEM containing population of 50% transgenic individuals at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

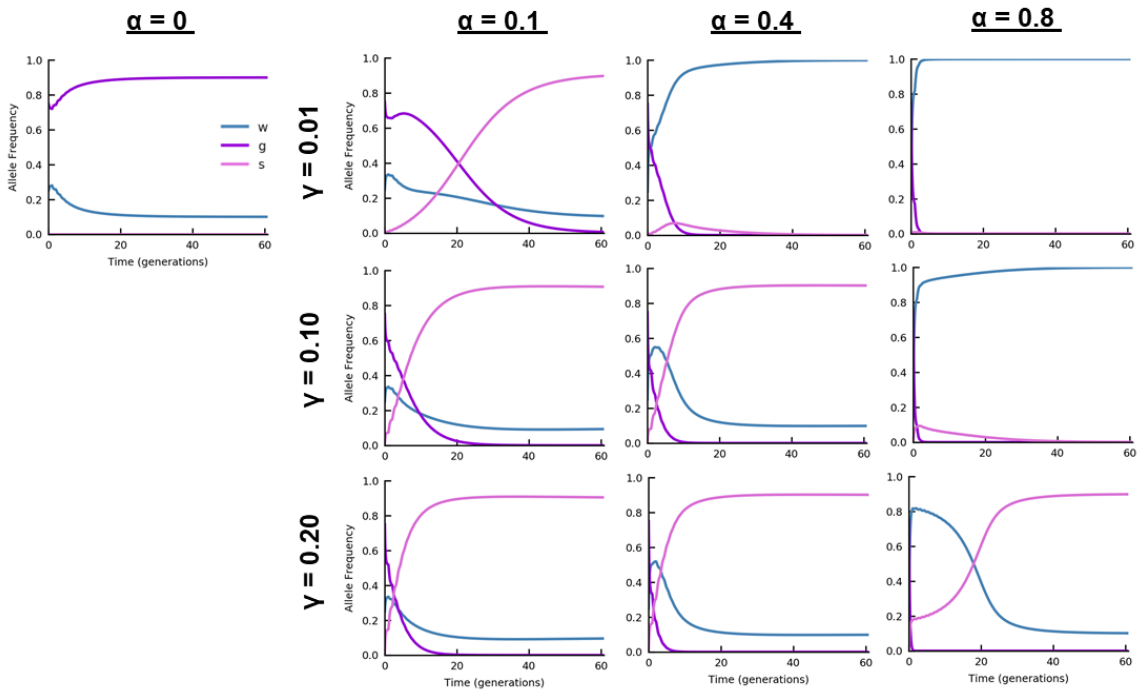

Figure 46. Allele frequency with a simulated release of two-locus UD-SEM containing population of 75% transgenic individuals at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

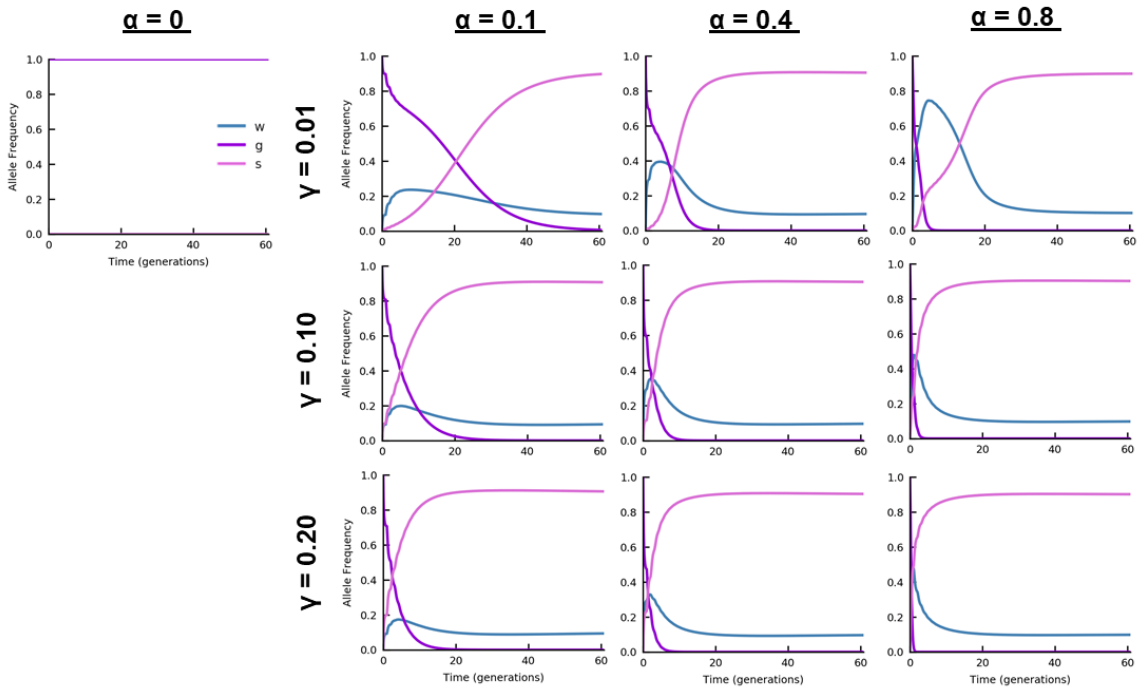

Figure 47. Allele frequency with a simulated release of two-locus UD-SEM containing population of 100% transgenic individuals at four rates of transgene self-elimination (α = 0, 0.1, 0.4, 0.8) and three rates of permanent transgene fixation (γ = 0.01, 0.10, 0.20)

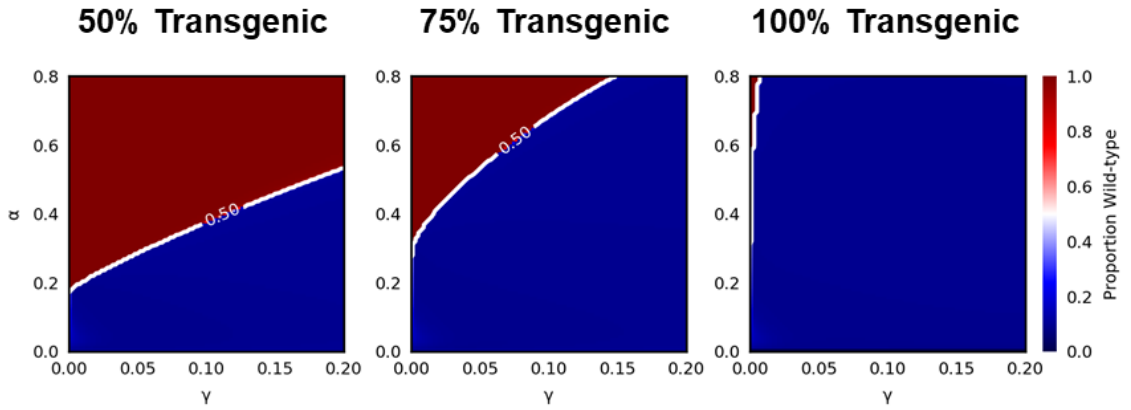

Figure 48. Wild-type allele frequencies of a simulated release of two-locus UD-SEM containing individuals at 50, 75, and 100% transgenic individuals of a wild-type population after 60 generations

Discussion

The underdominance system relies on a pair of maternally-linked toxin-antidote constructs that only ensure that offspring inheriting the antidote to the maternal toxin will survive. When both toxin-antidote pairs are located on the same locus, there are few offspring genotypes that are viable from the transgenic female. If the female is heterozygous transgenic (or homozygous transgenic with two copies of one toxin), the survival of the offspring becomes dependent on corresponding the paternal gamete that is inherited to provide the toxin to the antidote. Similarly, offspring of females carrying a copy of each transgene survive only if the gametes inherited from both parents will provide the antidote to both maternal toxins.

Hence, assuming a high mortality rate induced by the toxin in the offspring of more than 95%[43], the release of transgenic females would not be capable of inducing a drive of the transgene into the population. In order to spread throughout the population, a substantial number of transgenic males and females must be released, such that the number of surviving progeny from the mating between transgenic parents exceeds the number of progeny from wild-type parents. In the single-locus scenario, starting populations consisting of 50% transgenic males and females showed no drive of the transgene into the population across all rates of transgene excision and permanent transgene fixation. As we increased the starting population to 75% transgenic individuals, an inactive SEM was capable of inducing a drive. Once the SEM was active, the wild-type population was restored at all rates of transgene excision, with higher rates corresponding to faster removal of the transgene.

Evaluating an entirely transgenic population, we observed that high rates of transgene excision and low rates of transgene fixation were necessary to completely remove the transgene from the population. In this case, we observed the formation of the permanently-fixed transgene as the primary inhibitor of the restoration of the wild-type population. As we observed in the 75% transgenic starting population, where only the inactive SEM resulted in a successful drive of the transgene, once the permanently-fixed transgenic allele frequency exceeded approximately 75%, the wild-type population was not restored. Since the formation of the permanently-fixed transgene could only result from the excision-competent transgene, a high rate of transgene excision was required to limit the formation of the permanently-fixed transgene to restore the wild-type population. Our results indicate that the SEM could be a potential mechanism for increasing the threshold at which transgenic individuals must be released to ensure a drive into the environment and may be sufficient for containing a transgenic population within an environmental release test site. Furthermore, if the SEM could be chemically suppressed[91,92] until the transgene is permanently established in a population, subsequent chemical removal could trigger high rates of transgene excision and low rates of permanent transgene fixation to restore the wild-type population and remove the transgene.

When each toxin-antidote was located on separate loci, multiple viable genotypes were created. In the two-locus case, starting populations of 50% transgenic individuals were capable of causing a drive of the transgene into the population when the SEM was inactive or the rate of transgene excision was low. Similar to the single-locus case, the

drive was primarily due to the creation of the permanently-fixed transgenic allele. Since the permanently-fixed allele could only result from the presence of the excision-competent allele, the longer the excision-competent allele was present (due to a slower rate of transgene excision), the more permanently-fixed alleles were created. Further increases in the rate of permanent transgene fixation increased the number of permanently-fixed transgenes and wild-type population restoration was only possible at higher rates of transgene excision. This became more apparent when the starting transgenic population was increased to 75% and 100%. When the starting population was entirely transgenic, wild-type population restoration was only possible when the rate of permanent transgene fixation was very low ($\gamma < 0.01$) and the transgene excision rate was very high ($\alpha > 0.70$). As demonstrated by previous work[43], we expected the two-locus case to be a more invasive gene drive mechanism due to the increased number of viable genotypes. While the transgene is more difficult to remove, the addition of an SEM could be a viable approach to limiting the spread of the gene drive transgene.

The results of both the single- and two-locus UD-SEM models reflected the findings demonstrated by Akbari et al.[43] when similar fitness costs and starting populations were used and the SEM was inactive. Because each toxin provides a high mortality rate to the offspring that do not inherit the corresponding antidote, the UD gene drive mechanism requires a large release of both transgenic males and females to drive into the population and may be an ideal gene drive for testing in the environment[41,43]. While our model did not observe the impacts of migration, the high release thresholds required for a successful drive of the transgene and previous studies show that low

migration rates are containable. A major concern of the policy surrounding gene drives is the ability to contain or prevent the spread of the transgene into wide spread areas[58,60,88]. Furthermore, the UD mechanism may provide a higher level of biocontainment (compared to MEDEA) in the case of a small release of transgenic organisms and could serve as a method of rapid wild-type population restoration (through chemical triggers) once a transgene was fixed in a population.

CHAPTER V

CONCLUSIONS

Gene drive mechanisms are an efficient approach to spreading transgenes that may limit the spread and transmission of vector-borne diseases in endemic areas[34]. However, due to the unknown and irreversible long-term consequences that gene drive mechanisms may pose to environment, the release of organisms possessing gene drive constructs has been highly debated and strictly limited[58,60]. Self-eliminating mechanisms have been used in agricultural applications to remove transgenic material from the pollen of transgenic plants to limit the hazards of cross-pollination with closely related species[44] and in human gene therapy to limit deleterious reintegration of transgenic DNA[45,46]. We explored the limiting of a gene drive mechanism through the use of an SEM.

By pairing a pre-programmed self-elimination mechanism with a gene drive construct, we identified effective probabilities of transgene removal and permanent transgenic fixation that would allow for the mitigation of the hazards associated with the permanent establishment of the gene drive transgene in the environment. We observed that the CRISPR GD mechanism could be delayed through the use of an SEM, however we could not prevent the drive from occurring. Counterintuitively, lower rates of transgene excision resulted in the restoration of the wild-type population faster than higher rates of transgene excision because the creation of resistance alleles *v* slowed when the transgenic population was eliminated too quickly. Multiplexing the SEM significantly increased the rate of transgene excision and proved to be a potential biocontainment mechanism at high rates of transgene excision. A five gRNA target sight SEM was capable of raising the

threshold of a potential drive occurring by 6-7 orders of magnitude from the initial release of 1%.

The MEDEA and UD GD mechanisms also showed promising results when paired with the SEM mechanisms. The initial releases of less than 50% transgenic individuals showed the wild-type population could be restored across a range of permanent fixation and excision rates. While increasing the starting populations did limit the acceptable excision and permanent fixation rates, high rates of transgene excision demonstrated that a GD-SEM construct could be utilized as a biocontainment measure.

The use of a GD-SEM construct shows promise for being able to restore the wild-type population and remove the transgene presence following the release of transgenic individuals into the environment. Gene drive technology can be a powerful and efficient tool for introducing a wanted trait into a population. However, this technology must be appropriately tested prior to its use in the environment. Public acceptance of gene drives will also play a key role in whether gene drives are utilized and the policy surrounding their use. Lastly, the regulation of this biotechnology will need establish guidelines for how testing and application in the environment will be conducted. While there are numerous technological, social and political aspects that must be addressed before the application of GD technology is utilized in the environment, the GD-SEM may play a major role in addressing the some of the concerns associated with the release of GD organisms in the environment.

Future Work

While results of this study provided a groundwork for the proof-of-concept that a GD-SEM construct would be effective in removing the presence of a transgene from the environment, there are several limitations that should be addressed with continuing research in this area. The models developed in this study were strictly deterministic as a theoretical foundation for the function of the GD-SEM system. Additionally, because the single population models utilized were unbounded and allowed to grow exponentially, the spread of the gene drive transgene was not limited by an availability of resources in the environment and may have propagated faster than a more constrained population with a finite carrying capacity.

Future studies should also focus on the evaluation of stochastic models to explore the likelihood that the SEM construct will behave as expected when uncertainty is present in the model as shown by previous studies[41,43,64]. Additionally, utilizing a carrying capacity or density-dependent mortality rate similar to those used by Marshall et al.[83] to limit the number of individuals in the population and to more accurately reflect the behavior of organisms in the environment may slow the rate at which the gene drive transgene is spread into the population. While a single population was used, migration between multiple populations should be utilized to explore how the transgene could spread geographically or be contained as shown by Akbari[43] and others[41]. Environmental factors such as temperature have been shown to impact the development rates of mosquitoes and other vectors[93,94]. Incorporation of these factors across a multi-staged population

structure[95–97] would more closely resemble the population dynamics of mosquitoes in the environment.

Applications to public health should also be addressed. If this system was applied to limit the ability of mosquitoes or ticks to transmit pathogens, the impact on the prevalence of the disease in human and animal populations can be explored[98–100]. Multi-population models driven by environmental factors would be significant to understanding the impacts this technology may have public health. The future direction of this research can explore a variety of applications for public health and policy to potentially eliminate the prevalence of vector-borne diseases throughout endemic areas.

REFERENCES

1. Medlock, J. M. & Leach, S. A. Effect of climate change on vector-borne disease risk in the UK. *Lancet Infect. Dis.* **15**, 721–730 (2015).
2. Rosenberg, R. *et al.* Vital signs: trends in reported vectorborne disease cases—United States and territories, 2004--2016. *Morb. Mortal. Wkly. Rep.* **67**, 496 (2018).
3. Beard, C. B. *et al. Ch. 5: Vectorborne Diseases*. (2016).
4. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095–2128 (2012).
5. Murray, C. J. L. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990--2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
6. Organization, W. H. & others. *Eliminating malaria*. (2016).
7. Benelli, G. & Mehlhorn, H. Declining malaria, rising of dengue and Zika virus: insights for mosquito vector control. *Parasitol. Res.* **115**, 1747–1754 (2016).
8. Jensen, M. & Mehlhorn, H. Seventy-five years of Resochin®in the fight against malaria. *Parasitol. Res.* **105**, 609 (2009).
9. Mehlhorn, H., Al-Rasheid, K. A. S., Al-Quraishy, S. & Abdel-Ghaffar, F. Research and increase of expertise in arachno-entomology are urgently needed. *Parasitol. Res.* **110**, 259–265 (2012).
10. World Health Organization. *World malaria report 2018*. (2018).

11. O'meara, G. F., Evans Jr, L. F., Gettman, A. D. & Cuda, J. P. Spread of Aedes albopictus and decline of Ae. aegypti (Diptera: Culicidae) in Florida. *J. Med. Entomol.* **32**, 554–562 (1995).

12. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of Aedes aegypti-A Review. *Mem. Inst. Oswaldo Cruz* **108**, 11–17 (2013).

13. Gubler, D. J. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol.* **10**, 100–103 (2002).

14. Powell, J. R. Mosquitoes on the move. *Science (80-.).* **354**, 971–972 (2016).

15. Hartelt, K., Pluta, S., Oehme, R. & Kimmig, P. Spread of ticks and tick-borne diseases in Germany due to global warming. *Parasitol. Res.* **103**, 109–116 (2008).

16. Jinek, M. *et al.* A programmable dual-RNA--guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

17. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).

18. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).* **339**, 819–823 (2013).

19. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957 (2013).

20. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346**, 1258096 (2014).

21. Beaty, B. J. Genetic manipulation of vectors: a potential novel approach for control of vector-borne diseases. *Proc. Natl. Acad. Sci.* **97**, 10295–10297 (2000).

22. Alphey, L. *et al.* Malaria control with genetically manipulated insect vectors. *Science (80-.).* **298**, 119–121 (2002).

23. Wilke, A. B. B. & Marrelli, M. T. Paratransgenesis: a promising new strategy for mosquito vector control. *Parasit. Vectors* **8**, 342 (2015).

24. Ito, J., Ghosh, A., Moreira, L. A., Wimmer, E. A. & Jacobs-Lorena, M. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417**, 452 (2002).

25. Kokoza, V. *et al.* Engineering blood meal-activated systemic immunity in the yellow fever mosquito, Aedes aegypti. *Proc. Natl. Acad. Sci.* **97**, 9144–9149 (2000).

26. Franz, A. W. E. *et al.* Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. *Proc. Natl. Acad. Sci.* **103**, 4198–4203 (2006).

27. Knipling, E. F. Possibilities of insect control or eradication through the use of sexually sterile males. *J. Econ. Entomol.* **48**, 459–462 (1955).

28. Harris, A. F. *et al.* Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nat. Biotechnol.* **30**, 828 (2012).

29. Phuc, H. K. *et al.* Late-acting dominant lethal genetic systems and mosquito control. *BMC Biol.* **5**, 11 (2007).

30. Dame, D. A., Lowe, R. E. & Williamson, D. L. Assessment of released sterile Anopheles albimanus and Glossina morsitans morsitans. in *Cytogenetics and genetics of vectors: proceedings of a symposium of the XVIth International Congress of Entomology/editors, R. Pal, JB Kitzmiller, T. Kanda* (1981).

31. Laven, H. Eradication of Culex pipiens fatigans through cytoplasmic incompatibility. *Nature* **216**, 383 (1967).

32. Patterson, R. S., Weidhaas, D. E., Ford, H. R. & Lofgren, C. S. Suppression and elimination of an island population of Culex pipiens quinquefasciatus with sterile males. *Science (80-.).* **168**, 1368–1369 (1970).

33. Lofgren, C. S. *et al.* Release of chemosterilized males for the control of Anopheles albimanus in El Salvador. *Am. J. Trop. Med. Hyg.* **23**, 288–297 (1974).

34. Curtis, C. F. Possible Use of Translocations to fix Desirable Genes in Insect Pest Populations. *Nature* **218**, 368–369 (1968).

35. Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34**, 78 (2016).

36. Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci.* **112**, E6736--E6743 (2015).

37. Champer, J. *et al.* Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **13**, e1006796 (2017).

38. Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3**, e03401 (2014).

39. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. London B Biol. Sci.* **270**, 921–928 (2003).

40. Deredec, A., Burt, A. & Godfray, H. C. J. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**, 2013–2026 (2008).

41. Marshall, J. M. & Hay, B. A. Confinement of gene drive systems to local populations: a comparative analysis. *J. Theor. Biol.* **294**, 153–171 (2012).

42. Chen, C.-H. *et al.* A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science (80-.).* **316**, 597–600 (2007).

43. Akbari, O. S. *et al.* A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Curr. Biol.* **23**, 671–677 (2013).

44. Keenan, R. J. & Stemmer, W. P. C. Nontransgenic crops from transgenic plants. *Nat. Biotechnol.* **20**, 215 (2002).

45. Li, X. *et al.* piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci.* **110**, E2279--E2287 (2013).

46. Wang, G. *et al.* Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. *Nat. Protoc.* **12**, 88 (2017).

47. Nishizawa-Yokoi, A., Endo, M., Ohtsuki, N., Saika, H. & Toki, S. Precision genome editing in plants via gene targeting and piggy B ac-mediated marker excision. *Plant J.* **81**, 160–168 (2015).

48. Elick, T. A., Bauser, C. A. & Fraser, M. J. Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **98**, 33–41 (1996).

49. Nishizawa-Yokoi, A., Endo, M., Osakabe, K., Saika, H. & Toki, S. Precise marker excision system using an animal-derived piggyBac transposon in plants. *Plant J.* **77**, 454–463 (2014).

50. O'Hare, K. & Rubin, G. M. Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. *Cell* **34**, 25–35 (1983).

51. Häcker, I. *et al.* Cre/lox-recombinase-mediated cassette exchange for reversible site-specific genomic targeting of the disease vector, Aedes aegypti. *Sci. Rep.* **7**, 43883 (2017).

52. Éva, C. *et al.* Cold inducible promoter driven Cre-lox system proved to be highly efficient for marker gene excision in transgenic barley. *J. Biotechnol.* **265**, 15–24 (2018).

53. Verweire, D., Verleyen, K., De Buck, S., Claeys, M. & Angenon, G. Marker-free transgenic plants through genetically programmed auto-excision. *Plant Physiol.* **145**, 1220–1231 (2007).

54. Hare, P. D. & Chua, N.-H. Excision of selectable marker genes from transgenic plants. *Nat. Biotechnol.* **20**, 575 (2002).

55. Luo, K. *et al.* 'GM-gene-deletor': fused loxP-FRT recognition sequences dramatically improve the efficiency of FLP or CRE recombinase on transgene excision from pollen and seed of tobacco plants. *Plant Biotechnol. J.* **5**, 263–374 (2007).

56. Traver, B. E., Anderson, M. A. E. & Adelman, Z. N. Homing endonucleases catalyze double-stranded DNA breaks and somatic transgene excision in Aedes aegypti. *Insect Mol. Biol.* **18**, 623–633 (2009).

57. Aryan, A., Anderson, M. A. E., Myles, K. M. & Adelman, Z. N. Germline excision of transgenes in Aedes aegypti by homing endonucleases. *Sci. Rep.* **3**, 1603 (2013).

58. Rudenko, L., Palmer, M. J. & Oye, K. Considerations for the governance of gene drive organisms. *Pathog. Glob. Health* **112**, 162–181 (2018).

59. Meghani, Z. & Kuzma, J. Regulating animals with gene drive systems: lessons from the regulatory assessment of a genetically engineered mosquito. *J. Responsible Innov.* **5**, S203--S222 (2018).

60. Oye, K. A. *et al.* Regulating gene drives. *Science (80-.).* **345**, 626–628 (2014).

61. Windbichler, N. *et al.* A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473**, 212 (2011).

62. Gantz, V. M. & Bier, E. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science (80-.).* **348**, 442–444 (2015).

63. Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. & Nowak, M. A. Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3**, e1601964 (2017).

64. Kyrou, K. *et al.* A CRISPR--Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* (2018).

65. Beeman, R. W., Friesen, K. S. & Denell, R. E. Maternal-effect selfish genes in flour beetles. *Science (80-.).* **256**, 89–92 (1992).

66. Beeman, R. W. & Friesen, K. S. Properties and natural occurrence of maternal-effect selfish genes (‘Medea’factors) in the Red Flour Beetle, Tribolium castaneum. *Heredity (Edinb).* **82**, 529–534 (1999).

67. Spradling, A. C. & Rubin, G. M. Transposition of cloned P elements into Drosophila germ line chromosomes. *Science (80-.).* **218**, 341–347 (1982).

68. Sun, N. & Zhao, H. Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol. Bioeng.* **111**, 1048–1053 (2014).

69. Hannan, N. R. F. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 392–394 (2011).

70. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci.* **108**, 1531–1536 (2011).

71. Haghighat-Khah, R. E. *et al.* Site-specific cassette exchange systems in the Aedes aegypti mosquito and the Plutella xylostella moth. *PLoS One* **10**, e0121097 (2015).

72. Pradhan, S. *et al.* Marker-free transgenic rice expressing the vegetative insecticidal protein (Vip) of Bacillus thuringiensis shows broad insecticidal properties. *Planta* **244**, 789–804 (2016).

73. Chen, H. *et al.* Application of Cre-lox gene switch to limit the Cry expression in rice green tissues. *Sci. Rep.* **7**, 14505 (2017).

74. Zuo, J., Niu, Q.-W., Møller, S. G. & Chua, N.-H. Chemical-regulated, site-specific DNA excision in transgenic plants. *Nat. Biotechnol.* **19**, 157 (2001).

75. Ge, J. *et al.* Intein-mediated Cre protein assembly for transgene excision in hybrid progeny of transgenic Arabidopsis. *Plant Cell Rep.* **35**, 2045–2053 (2016).

76. Wang, B., Zhang, Y., Zhao, J., Dong, M. & Zhang, J. Heat-Shock-Induced Removal of Transgenes Using the Gene-Deletor System in Hybrid Aspen (Populus tremula$\times$ P. tremuloides). *Genes (Basel).* **9**, 484 (2018).

77. Somleva, M. N. *et al.* Transgene autoexcision in switchgrass pollen mediated by the Bxb1 recombinase. *BMC Biotechnol.* **14**, 79 (2014).

78. Huang, M. Y. & Mitchell, A. P. Marker recycling in Candida albicans through CRISPR-Cas9-induced marker excision. *mSphere* **2**, e00050--17 (2017).

79. Solis-Escalante, D. *et al.* Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in Saccharomyces cerevisiae. *FEMS Yeast Res.* **14**, 741–754 (2014).

80. Unckless, R. L., Messer, P. W., Connallon, T. & Clark, A. G. Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics* **201**, 425–431 (2015).

81. Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**, 827–841 (2017).

82. Vella, M. R., Gunning, C. E., Lloyd, A. L. & Gould, F. Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Sci. Rep.* **7**, 11038 (2017).

83. Marshall, J. M., Buchman, A., Akbari, O. S. & others. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* **7**, 3776 (2017).

84. Najmitabrizi, N. Making Gene Bio-Degradable: Parameter Estimation and Mathematical Modeling of Spread of Anti-Pathogen Gene in Mosquitoes. (Texas A&M University, 2017).

85. Graf, U. *et al.* Somatic mutation and recombination test in Drosophila melanogaster. *Environ. Mutagen.* **6**, 153–188 (1984).

86. Sinkins, S. P. & Gould, F. Gene drive systems for insect disease vectors. *Nat. Rev. Genet.* **7**, 427 (2006).

87. Marshall, J. M. & Akbari, O. S. Can CRISPR-based gene drive be confined in the wild? A question for molecular and population biology. *ACS Chem. Biol.* **13**, 424–430 (2018).

88. National Academies of Sciences, Engineering & and Medicine. *Gene drives on the horizon: advancing science, navigating uncertainty, and aligning research with public values*. (National Academies Press, 2016).

89. Catteruccia, F., Godfray, H. C. J. & Crisanti, A. Impact of genetic manipulation on the fitness of Anopheles stephensi mosquitoes. *Science (80-.).* **299**, 1225–1227 (2003).

90. Hay, B. A. *et al.* Engineering the genomes of wild insect populations: challenges, and opportunities provided by synthetic Medea selfish genetic elements. *J. Insect Physiol.* **56**, 1402–1413 (2010).

91. Endoh, T., Kawasaki, Y. & Sugimoto, N. Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. *Angew. Chemie Int. Ed.* **52**, 5522–5526 (2013).

92. Pendurthi, U. R. & Rao, L. V. M. Suppression of transcription factor Egr-1 by curcumin. *Thromb. Res.* **97**, 179–189 (2000).

93. Carrington, L. B., Armijos, M. V., Lambrechts, L., Barker, C. M. & Scott, T. W. Effects of Fluctuating Daily Temperatures at Critical Thermal Extremes on Aedes aegypti Life-History Traits. *PLoS One* **8**, 1–9 (2013).

94. Zapletal, J., Erraguntla, M., Adelman, Z. N., Myles, K. M. & Lawley, M. A. Impacts of diurnal temperature and larval density on aquatic development of Aedes aegypti. *PLoS One* **13**, e0194025 (2018).

95. Erickson, R. A., Presley, S. M., Allen, L. J. S., Long, K. R. & Cox, S. B. A stage-structured, Aedes albopictus population model. *Ecol. Modell.* **221**, 1273–1282 (2010).

96. Jia, P. *et al.* A climate-driven mechanistic population model of Aedes albopictus with diapause. *Parasit. Vectors* **9**, 175 (2016).

97. Tran, A. *et al.* A rainfall-and temperature-driven abundance model for Aedes albopictus populations. *Int. J. Environ. Res. Public Health* **10**, 1698–1719 (2013).

98. Eckhoff, P. A. & Wenger, E. A. A mathematical model of the impact of present and future malaria vaccines. *Malar. J.* **12**, 126 (2013).

99. Erickson, R. A., Presley, S. M., Allen, L. J. S., Long, K. R. & Cox, S. B. A dengue model with a dynamic Aedes albopictus vector population. *Ecol. Modell.* **221**, 2899–2908 (2010).

100. Mpeshe, S. C., Haario, H. & Tchuenche, J. M. A mathematical model of Rift Valley fever with human host. *Acta Biotheor.* **59**, 231–250 (2011).