

GENE DRIVES ACROSS ENGINEERED FITNESS VALLEYS: MODELING A DESIGN TO PREVENT DRIVE SPILLOVER

Frederik J.H. de Haas¹ and Sarah P. Otto¹

¹*Biodiversity Research Center, Department of Zoology,
University of British Columbia, Canada*

October 29, 2020

Keywords

Gene drive, CRISPR, Homing, Daisy-chain, Underdominance, Population Alteration

Contents

1	Abstract	3
2	Notes from reviewers	3
3	Introduction	3
4	Methods and results	6
4.1	Drive design	7
4.2	Dynamics in an isolated population	8
4.3	Invasion analysis	11
4.4	Spatial spread of the gene drive	12
5	Discussion	14
6	Acknowledgements	15
7	Supplementary Material	18
7.1	Daisy-chain length	18
7.2	Mating tables	18
7.3	Alternative payload fitness implementation	18

7.4	Individual-based simulations	21
7.5	Invasion analysis	25

1 Abstract

2 Engineered gene drive techniques for population replacement and/or suppression have potential for tackling
 3 complex challenges, including reducing the spread of diseases and invasive species. Unfortunately, the self-
 4 propelled behavior of drives can lead to the spread of transgenic elements beyond the target population,
 5 which is concerning. Gene drive systems with a low threshold frequency for invasion, such as homing-
 6 based gene drive systems, require initially few transgenic individuals to spread and are therefore easy to
 7 implement. However their ease of spread presents a double-edged sword; their low threshold makes these
 8 drives much more susceptible to spread outside of the target population (spillover). We model a proposed
 9 drive system that transitions in time from a low threshold drive system (homing-based gene drive) to
 10 a high threshold drive system (underdominance) using daisy chain technology. This combination leads
 11 to a spatially restricted drive strategy, while maintaining an attainable release threshold. We develop
 12 and analyze a discrete-time model as proof of concept and find that this technique effectively generates
 13 stable local population suppression, while preventing the spread of transgenic elements beyond the target
 14 population under biologically realistic parameters.

2 Notes from reviewers

16 3 Introduction

Engineered gene drives are self-replicating genetic constructs that can bias transmission of desired alleles
 18 (usually called payload genes) to progeny, allowing them to rapidly increase in frequency even when the
 19 alleles are selected against [Burt, 2003]. Introgression of these transgenic genes into sexually reproducing
 20 species could potentially solve many pressing environmental and humanitarian problems, ranging from
 21 public health and agriculture to conservation [National Academies of Sciences et al., 2016]. Examples
 22 include modifications of the mosquito genome to reduce its capacity to serve as a disease vector (e.g.
 23 reducing the number of female mosquitoes, their longevity or their ability to support development and
 24 transmission of the pathogen) [Isaacs et al., 2012]. Other examples include driving genes that protect species
 25 at risk by eradicating invasive species or reducing pest damage in agriculture by replacing resistant alleles
 26 with their ancestral equivalents to restore vulnerability to pesticide or herbicide [National Academies of
 27 Sciences et al., 2016].

28 There are two general approaches to gene drive: population suppression and population alteration.
 Population suppression makes use of gene drive systems to purposefully sterilise or kill the target population

while population alteration uses gene drive systems to alter the genomic composition without the aim to sterilise or kill the target population. A successful example of population suppression in the laboratory was shown by Kyrou et al. [2018] for *Anopheles gambiae*. They used a CRISPR-Cas9 gene drive targeting a doublesex gene (Agdsx) that lead to a bias in the sex ratio by rendering homozygous females completely sterile. Within 7–11 generations egg production was reduced to the point of total population collapse.

Extreme caution with population suppression is required as it could send a ripple through ecosystems, endangering many other plants and animals than just the target species. For example, mosquitoes that live in the arctic of Canada & Russia fly around in thick swarms and make up a huge part of the biomass there. They pollinate arctic plants and are a major food source for migrating birds so their loss could be disastrous to the ecosystem [Thien, 1969]. A much less invasive approach to gene drive is population alteration. This approach potentially has less ecological harm, but it might require multiple drives for each disease carried by the vector. Isaacs et al. [2012] discovered that *Anopheles stephensi* mosquitoes expressing m1C3, m4B7, or m2A10 single-chain antibodies (scFvs) have significantly lower levels of infection compared to controls when challenged with *Plasmodium falciparum*, a human malaria pathogen. Further research needs to show if expression of a single copy of a dual scFv transgene can completely inhibit parasite development, potentially without a fitness cost on the mosquito.

Although gene drives hold much promise, they have the strong possibility of unwanted ecological effects. Major risks include the unintended spread of the payload beyond the area of interest due to unwanted migration of individuals and the spread of the payload beyond the target species due to hybridization. Low threshold gene drive systems such as homing-based endonuclease gene drive (HEG) are especially susceptible to such spillovers. These systems exploit homology directed repair (HDR) to replace a targeted, naturally occurring genetic sequence with an engineered construct. Homing-based gene drives work by transcribing an endonuclease (often Cas9) and a guide RNA (gRNA) in the germline. The endonuclease then triggers a double stranded break at the complementary site of the gRNA which can be repaired by either one of two pathways: non-homologous end joining (NHEJ) or homologous end joining (HEJ). In NHEJ the break ends are directly ligated without the use of a homologous template. This often leads to mutations arising in the process, preventing the gRNA from recognizing the sequence for future double stranded breaks. However, when the double stranded break is repaired by HEJ, which requires the homologous chromosome carrying the construct, the result is a copy of the homologous chromosome at the double stranded break: creating effectively homozygous cells from heterozygous cells. The result is equivalent to meiotic drive, δ , leading to a greater fraction of gametes carrying the construct than expected under Mendelian inheritance (specifically, heterozygotes produce a fraction $\frac{1}{2}(1 + \delta)$ of gametes carrying

the driven allele).

High conversion rates have already been achieved in the lab by the CRISPR-Cas9 system for yeast ($\delta > 0.995$, [DiCarlo et al., 2015]), fruit flies ($\delta > 0.97$, [Gantz and Bier, 2015]), and the malaria vector mosquito, *Anopheles stephensi*, with engineered malaria resistance ($\delta > 0.98$, [Gantz et al., 2015]). These drives require few transgenic individuals for gene drive to occur (low threshold), making them very capable of spreading between subpopulations following introduction by migration or following hybridization between species [Marshall, 2009]. Numerous countermeasures have been proposed to reverse gene drives such as synthetic CRISPR-cas9 based reversal drives like CATCHA [Wu et al., 2016] and ERACR [Gantz and Bier, 2016]. However, reversal drives are techniques to fully eliminate rather than locally confine the drive, which is often the desired outcome.

To address the issue of local confinement, Noble et al. [2019] analyzed a new self-exhausting drive technique that relies on CRISPR-mediated homing-based gene drive called a “daisy-chain drive”. The daisy-chain drive separates multiple interdependent homing-based drive components across the organism’s genome, each bearing a cost (drive load). Each genetic element drives the next in the chain. The final element carries the payload and is driven to higher and higher frequencies due to the preceding drive elements. Because no element can drive itself and the first element is not driven at all, the eventual outcome is that all elements will be lost from the population over time due to natural selection against the drive load, and the population returns to its original wildtype state. Unfortunately, as the drive disappears, so does the payload gene, rendering this system ineffective for a persistent payload introgression.

Persistent and localized introgression would be possible if the daisy chain alleles drive a high threshold drive system (such as underdominance) past its high threshold frequency. The advantage of such a construct is that it would be eliminated from adjacent populations once the daisy chain is exhausted due to their low frequency (below the threshold). This design was first proposed conceptually by Min et al., under the name “daisy quorum drive” ([Min et al., 2017]) but has not yet been modeled, which is our goal here. This new drive design is predicted to spread through a population until all of its daisy elements have been lost, at which point its fitness becomes frequency-dependent. The result is an engineered population surrounded by wildtype populations with limited mixing at the boundaries due to the reduced fitness of crosses between wildtype individuals and individuals carrying the engineered alleles.

Drives that exhibit a high invasion threshold, such as engineered underdominance (EU) and Wolbachia drives, provide a natural means to achieve localization. These drives exhibit frequency-dependent dynamics where the drive can only spread if its frequency exceeds the invasion threshold. Below this level, the frequency of the drive will decrease, leading to its loss from the population. Spread of a threshold drive

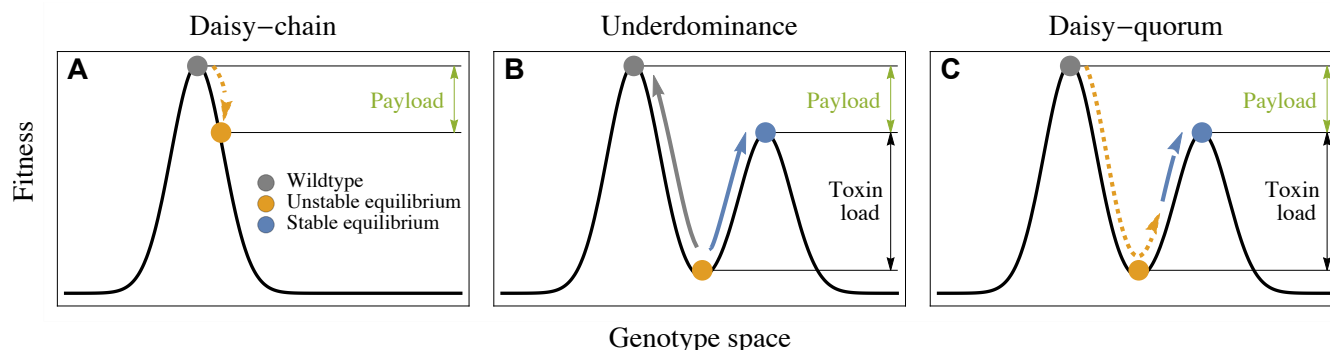


Figure 1: Conceptual overview of the different control systems. (A) A daisy-chain drive introgresses a payload temporarily, ultimately reverting back to the wildtype state (grey). (B) The dynamics of an underdominance gene drive is frequency dependent: left of the valley the population is driven back to a wildtype composition, while right of the valley it evolves toward the lower peak expressing the payload. (C) The proposed design (daisy-quorum drive) includes both systems and uses the temporary low-threshold daisy chain to drive an underdominant construct past the valley, after which it evolves toward the other peak with the desired payload.

across a patchy environment is more difficult and becomes highly unlikely or even impossible when the invasion threshold is 50 percent or higher [Barton and Turelli, 2011]. Threshold gene drive mechanisms were among the first forms of gene drive proposed and the only one to be deployed in the wild [Buchman et al., 2018]. Because organisms that migrate to other populations will be in the minority and consequently are selected against, underdominance with high thresholds poses a substantially lower risk of unintended spread.

An ideal gene drive system to alter wild populations would exclusively affect organisms within the boundaries of consenting communities. Here we present a simple population genetics model to elucidate the evolutionary dynamics of a new design that meets this criterion by combining daisy-chain drive with a two-locus underdominance system (Figure 1). The daisy-chain drive system provides a short window of time to drive the frequency-dependent underdominance system before it disappears, effectively transitioning from a low to a high threshold control system. By exploring spatial models in the face of recurrent migration, we find theoretical support that this new design could reduce the risk of gene drive spillovers while providing persistent payload introgression to the target population(s).

4 Methods and results

We formulate a deterministic model that considers the evolution of a large population of diploid organisms. We first introduce the model for a single population and then build up to an array of populations under recurrent migration to explore the spatial spread of the engineered construct. We refer to the loci in our model by bold letters (e.g. **A**) and to the alleles by the italic letters (e.g. *A* or *a*)

Daisy-chain Component				Underdominance Component			
Locus A		Locus B		Locus C		Locus D	
<i>a</i>	Wildtype	<i>b</i>	Wildtype	<i>c</i>	Wildtype	<i>d</i>	Wildtype
<i>A</i>	CRISPR-Cas9 gRNA targetting <i>b</i> Drive load (s_d)	<i>B</i>	CRISPR-Cas9 gRNA targetting <i>c</i> gRNA targetting <i>d</i> Drive load (s_d)	<i>C</i>	Payload (s_p) Suppressor of <i>D</i> Promoter Toxin load (s_t)	<i>D</i>	Payload (s_p) Suppressor of <i>C</i> Promoter Toxin load (s_t)

Table 1: Four-locus daisy-quorum drive. Loci **A** and **B** are part of the daisy-chain component and loci **C** and **D** make up the two-locus underdominance component. The daisy-chain transgenic alleles (**A** and **B**) have a fitness cost of s_d (drive load) and the engineered underdominance transgenic alleles have a toxin-load s_t and a payload s_p , which interact according to Table 2. The gRNA of locus **A** is complementary to the wildtype allele *b*, and locus **B** carries two gRNAs to cut the wildtype alleles *c* and *d* at the underdominance loci.

4.1 Drive design

We model a multilocus bi-allelic population with random mating using discrete-time dynamics. Our proposed drive system consists of two components. The first component encodes a daisy-chain gene drive, which can be of arbitrary length, and the second component is a two-locus engineered underdominance system with an added payload gene on each transgenic allele. We explore here a four-locus genetic architecture as an example case (Table 1), although the daisy chain could contain more loci if a longer acting drive is needed [Min et al., 2017] [Noble et al., 2019]. Unless otherwise stated, we assume that the loci involved are freely recombining ($r = \frac{1}{2}$) and that all dynamical systems are initialized at a proportion $1 - f_0$ of *abcd* gametes and f_0 of *ABCD* gametes.

The first component (daisy-chain) consists of a linear series of n loci arranged such that each element drives the next in the chain. The final element in the chain drives the “cargo”, which in our model are the two underdominance loci **C** and **D**. Each transgenic allele of the daisy chain consists of a CRISPR-cas9 complex and a drive load, which bears a fitness cost of s_d . The drive load helps ensure that all CRISPR-cas9 elements are counterselected and eventually disappear from the system.

In the case of the daisy-chain, the gRNA at locus i in the chain guides the Cas9 nuclease to the wildtype at locus $i + 1$, inducing a double stranded break, which will be repaired by homologous recombination with probability δ . We do not incorporate natural resistance alleles to the daisy-drive elements because their effects are minimal due to the transient nature of the daisy chain drive (see supplementary material).

The two-locus underdominance component is structured following the design proposed by Davis et al., [Davis et al., 2001]. The two transgenic alleles consist of four elements each: a payload, a suppressor, a promoter and a toxin. The two engineered alleles are carried on separate non-homologous pairs of chromosomes, each of which produces a toxin unless the suppressor on the other transgenic allele is also present. Specifically, in the absence of the matching suppressor, the promoter will cause the toxin gene to

be expressed, while individuals who carry one or more copies of both constructs are viable because both suppressors are present in the genome and neither toxin gene is expressed (Table 2). This scheme creates a frequency-dependent fitness regime, promoting the state of carrying both engineered constructs or none and selecting against individuals carrying one but not the other transgenic allele.

	cd	cD	Cd	CD
cd	1	$(1 - s_t)(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$
cD	$(1 - s_t)(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$
Cd	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$	$(1 - s_t)(1 - s_p)$	$(1 - s_p)$
CD	$(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$	$(1 - s_p)$

Table 2: Relative fitness for each diploid genotype in the two-locus underdominant component. The toxin s_t is expressed in individuals carrying either a C or D but not both. The payload s_p is expressed if the individual carries at least 1 copy (dominantly expressed). Both the toxin and the payload affect individual fitness multiplicatively.

The effects on fitness (toxin load s_t , drive load s_d and payload s_p) are assumed to determine individual fitness multiplicatively. Table 2 illustrates the effect of locus **C** and **D** on individual fitness where the payload (s_p) is dominantly expressed. In the supplementary material we explore the effect of different payload fitness regimes: multiplicative dominance and a recessive case. To derive the relative fitness for the full genotype, the fitness of the **CD** genotype is multiply by $(1 - s_d)^\rho$ where ρ equals the number of A and B alleles.

4.2 Dynamics in an isolated population

The recursion equations for the 16 (2^4) gamete types are too elaborate so we report them in the supplementary *Mathematica* file. Instead we focus on the dynamical equations of the first and second component in isolation (Table 1). We use recursion equations to take the population through diploid selection followed by meiosis with drive and recombination.

Assuming random mating in a single isolated population, the recursion equations for the frequency X_{ij} of the gamete ij at loci **A** and **B** are:

$$\begin{aligned}
X'_{ab} &= \frac{1}{\bar{W}}(X_{ab}^2 + X_{aB}X_{Ab}(1-s_d)^2(1-\delta)r + X_{ab}X_{aB}(1-s_d) + X_{ab}X_{Ab}(1-s_d) \\
&\quad + X_{ab}X_{AB}(1-s_d)^2(1-\delta)(1-r)) \\
X'_{aB} &= \frac{1}{\bar{W}}(X_{aB}^2(1-s_d)^2 + X_{aB}X_{Ab}(1-s_d)^2(1-r(1-\delta)) + X_{aB}X_{AB}(1-s_d)^3 + X_{ab}X_{aB}(1-s_d) \\
&\quad + X_{ab}X_{AB}(1-s_d)^2(1-r(1-\delta))) \\
X'_{Ab} &= \frac{1}{\bar{W}}(X_{Ab}^2(1-s_d)^2 + X_{aB}X_{Ab}(1-s_d)^2(1-\delta)(1-r) + X_{Ab}X_{AB}(1-s_d)^3(1-\delta) \\
&\quad + X_{ab}X_{Ab}(1-s_d) + X_{ab}X_{AB}(1-s_d)^2(1-\delta)r) \\
X'_{AB} &= \frac{1}{\bar{W}}(X_{AB}X_{ab}(1-s_d)^2(1-r(1-\delta)) + X_{AB}X_{Ab}(1-s_d)^3(1-\delta) + X_{AB}^2(1-s_d)^4 \\
&\quad + X_{aB}X_{Ab}(1-s_d)^2(r+\delta(1-r)) + X_{aB}X_{AB}(1-s_d)^3)
\end{aligned} \tag{1}$$

where $X_{ab} + X_{aB} + X_{Ab} + X_{AB} = 1$ and \bar{W} is the average fitness. In the supplementary material we add a mating table for these recursion equations (Table S1). An automated algorithm for deriving the dynamical equations for longer daisy chains is provided in the supplementary *Mathematica* file.

The recursion equations for the second underdominance component, describing the frequency X_{ij} of the gamete ij at loci **C** and **D**, are:

$$\begin{aligned}
X'_{cd} &= \frac{1}{\bar{W}}(X_{cd}^2 + X_{cd}X_{cD}(1-s_t)(1-s_p) + X_{cd}X_{Cd}(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)r \\
&\quad + X_{cd}X_{CD}(1-s_p)(1-r)) \\
X'_{cD} &= \frac{1}{\bar{W}}(X_{cd}X_{cD}(1-s_t)(1-s_p) + X_{cD}^2(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)(1-r) + X_{cd}X_{CD}(1-s_p)r \\
&\quad + X_{cD}X_{CD}(1-s_p)) \\
X'_{Cd} &= \frac{1}{\bar{W}}(X_{cd}X_{Cd}(1-s_t)(1-s_p) + X_{Cd}^2(1-s_t)(1-s_p) + X_{cD}X_{Cd}(1-s_p)(1-r) + X_{cd}X_{CD}(1-s_p)r \\
&\quad + X_{Cd}X_{CD}(1-s_p)) \\
X'_{CD} &= \frac{1}{\bar{W}}(X_{cd}X_{CD}(1-s_p)(1-r) + X_{cD}X_{CD}(1-s_p) + X_{Cd}X_{CD}(1-s_p) + X_{CD}^2(1-s_p) \\
&\quad + X_{cD}X_{Cd}(1-s_p)r)
\end{aligned} \tag{2}$$

where $X_{cd} + X_{cD} + X_{Cd} + X_{CD} = 1$ and \bar{W} is the average fitness (the sum of the numerators).

We assume that $X_{cD} = X_{Cd}$ at $t = 0$, which implies that their frequencies will remain the same at all time points in the future (Equation 2). By inspection of Equation 2, $X_{cd} = 1$ and $X_{CD} = 1$ are fixed point solutions to the underdominant component of the dynamical equations and a local stability analysis indicates that both are locally stable. A third fixed point represents the co-existence of alleles *C* and *c*, as well as *D* and *d*, and is locally unstable (see Supplementary Material for ternary plots of alternative payload fitness regimes). In Figure 2, we illustrate the location of the stable equilibria (blue and grey vertices) and the separatrix (the boundary separating two basins of attraction in a dynamical model) for different payloads s_p when the toxin is relatively ineffective ($s_t = 0.1$) or fully effective ($s_t = 1.0$). The

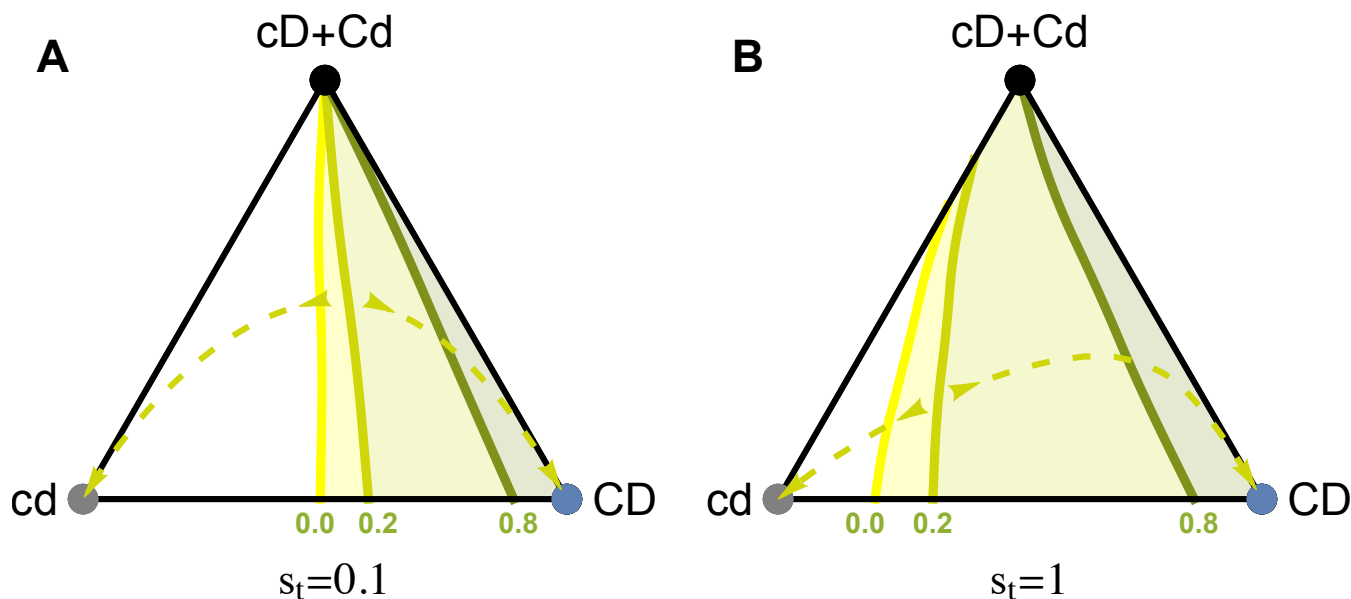


Figure 2: Ternary plot illustrating the separatrix (the boundary separating basins of attraction to different fixed points) for the two-locus underdominance construct with a toxin load of (A) $s_t = 0.1$ or (B) $s_t = 1.0$ and a payload of $s_p = 0.0, 0.2$ and 0.8 (green curves). Locus **C** and **D** are recombining freely ($r = \frac{1}{2}$). The dashed curves are example dynamics for $s_p = 0.2$.

separatrix is calculated numerically. Example dynamics confirm that the system converges to either fixed point depending on whether initialized to the left or right side of the separatrix.

As the payload increases (larger s_p), the separatrix moves to the right in Figure 2, reducing the basin of attraction of the engineered underdominance construct (haplotype CD). Higher payloads thus make it harder to drive such constructs into a population, but they also provide stronger fitness effects if the construct is able to fix. By contrast, as the toxin load increases (larger s_t), the basin of attraction to CD increases. While somewhat counterintuitive, this occurs because the toxin load can be suppressed by even one copy of the alternate allele, making haplotypes Cd and cD strongly selected against when cd is common (the wildtype) but only weakly selected against when CD is common, causing the fitness surface to fall faster near cd but not near CD when the load is stronger.

To prevent unintentional spread into neighbouring populations, it is desirable to design a drive where the separatrix is shifted far towards the transgenic haplotype so invasion will only occur once a significant number of transgenic haplotypes have been introduced into a new population. However, shifting the unstable equilibrium too far to the transgenic haplotype will cause the population to revert to the wildtype state more easily due to drift and incoming wildtype migrants from neighbouring populations. We therefore suggest that a robust system of localization is designed such that the unstable equilibrium is near the center of Figure 2, which can be achieved by adjusting the toxin load s_t relative to the payload s_p .

To derive the dynamics for all gamete types, the two components are linked by allele B driving allele C and D , just like allele A drives B .

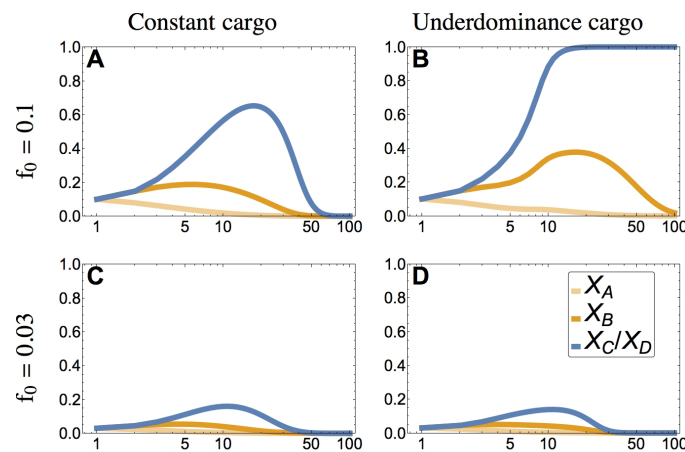


Figure 3: Top row illustrates dynamics of a daisy chain pushing a constant payload $s_p = 0.15$ while bottom row shows the daisy chain pushing an underdominance construct, also with a payload of $s_p = 0.15$. The panels are initialized with gametes $ABCD$ at frequency $f_0 = 0.1$ (left) and $f_0 = 0.03$ (right). Other parameters are: $s_t = 0.9$, $d = 0.95$, $s_d = 0.05$ and $r = \frac{1}{2}$.

4.3 Invasion analysis

We next perform an invasion analysis to determine the drive rate δ_c needed for the transgenic alleles C and D to spread. Here δ_c is treated as a constant value, equalling the frequency of individuals carrying the last daisy-chain locus $X_B^2 + 2X_B(1 - X_B)$ times the drive strength δ . The conditions for the spread of the cargo when rare are that the following condition holds (assuming $r = \frac{1}{2}$ and B , C and D are unlinked):

$$1 + \delta_c > \min\left[\frac{1}{(1 - s_p)(1 - s_t)}, \sqrt{\frac{2}{1 - s_p}}\right] \quad (3)$$

The minimum term describes the conditions under which either allele C or D could spread on their own (first criterion) or the combined CD haplotype could spread (second criterion). With a low toxin load (s_t), the first criterion is easier to satisfy, and we expect the alleles to spread individually. With a high toxin load, however, only the combined haplotype CD will spread, as long as drive is strong enough to counter the haplotype being broken down by recombination.

Given that $0 \leq \delta_c \leq 1$, a two-locus underdominant construct carrying a dominant payload with $s_p > 0.5$ can never invade, no matter the strength of the drive. With multiplicative dominance or recessive payload fitness regimes, a higher payload can be invaded (see Supplementary Material), however an interior basin of attraction prevents the payload from reaching fixation under these fitness regimes.

Figure 3 illustrates the contrasting effects of a constant “cargo” load compared to a two-locus underdominant “cargo” for different initial release frequencies f_0 . A constant cargo load is modeled as a third locus C' that only contains a payload, as is generally considered with daisy chain drives. Such a constraint cargo does not spread the payload to fixation as it is always being selected against while the underdominant system allows the fixation of the payload provided that it is released at a high enough initial release

frequency. Due to the daisy-chain drive, an underdominance construct carrying a payload of $s_p = 0.2$ can spread at a far lower initial release frequency than predicted for the underdominance component on its own (threshold of 0.24 when introducing CD , Figure 2). Panel D illustrates the scenario for a lower initial release frequency where the daisy drive proved to be not powerful enough to push the underdominance construct past the separatrix of Figure 2.

4.4 Spatial spread of the gene drive

We next model a stepping-stone model of M interconnected populations with migration rate m between adjacent patches. Migration is followed by dynamics within each path, as described in section 2.2. The frequency of gametes of type i after migration for a non-boundary population j is given by: $p_i^{j'} = (1 - m)p_i^j + \frac{1}{2}mp_i^{j-1} + \frac{1}{2}mp_i^{j+1}$. The boundary populations only give and receive migrants from the interior: $p_i^{0'} = (1 - \frac{1}{2}m)p_i^0 + \frac{1}{2}mp_i^1$ and $p_i^{M'} = (1 - \frac{1}{2}m)p_i^M + \frac{1}{2}mp_i^{M-1}$.

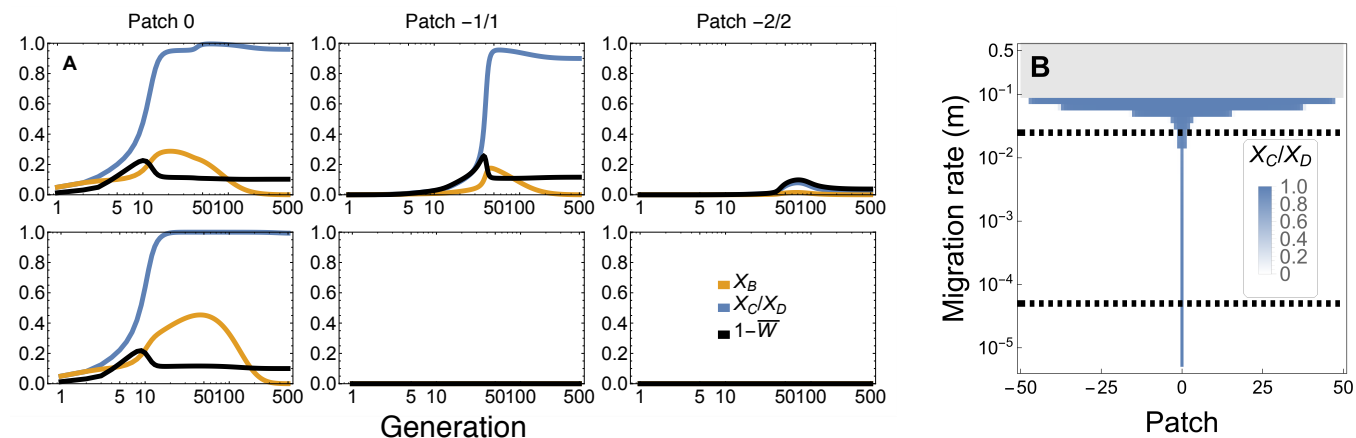


Figure 4: Dynamics of spread of a payload for different migration rates when $ABCD$ gametes are introduced into the center patch (patch 0) at an initial frequency of $f_0 = 0.05$. Top and bottom row of panel (A) show 3 interconnected patches and their dynamics for a migration rate of $m = 0.05$ and $m = 0.00005$, respectively (see Table 3 for other parameters). For a relatively high migration rate (top row), spread to neighbouring populations is more likely than for lower migration rates (bottom row). As predicted, the drive disappears from the system due to the drive load $s_d = 0.02$ (orange curves). Panel (B) shows the frequency of the payload alleles, X_C or X_D , at 1000 generations (blue shading) across all patches (x axis) for a variety of migration rates (y axis). In the gray area of panel (B), migration into the center patch (patch 0) is high enough to swamp the engineered construct causing no spread to occur anywhere.

Migration rate m is a crucial parameter to determine the spread of the payload through interconnected populations. In Figure 4 we show that our gene drive construct can be locally maintained in the face of recurrent migration and yet remain spatially restricted because migration into neighbouring demes does not introduce enough of the daisy chain to reach the threshold for invasion. Figure 4 (panel B) shows the range of migration rates where the construct will remain spatially restricted ($m < 10^{-1.3}$). For increasing migration rates, the payload spreads further (e.g., to patches +1 and -1 with $m = 0.05$) until migration rate reaches a critical value above which the drive construct is swamped within the initial population

and fails to spread. The payload frequency shows an almost stepwise function at the boundary, from a frequency of 1 to a frequency of 0 (gray zone).

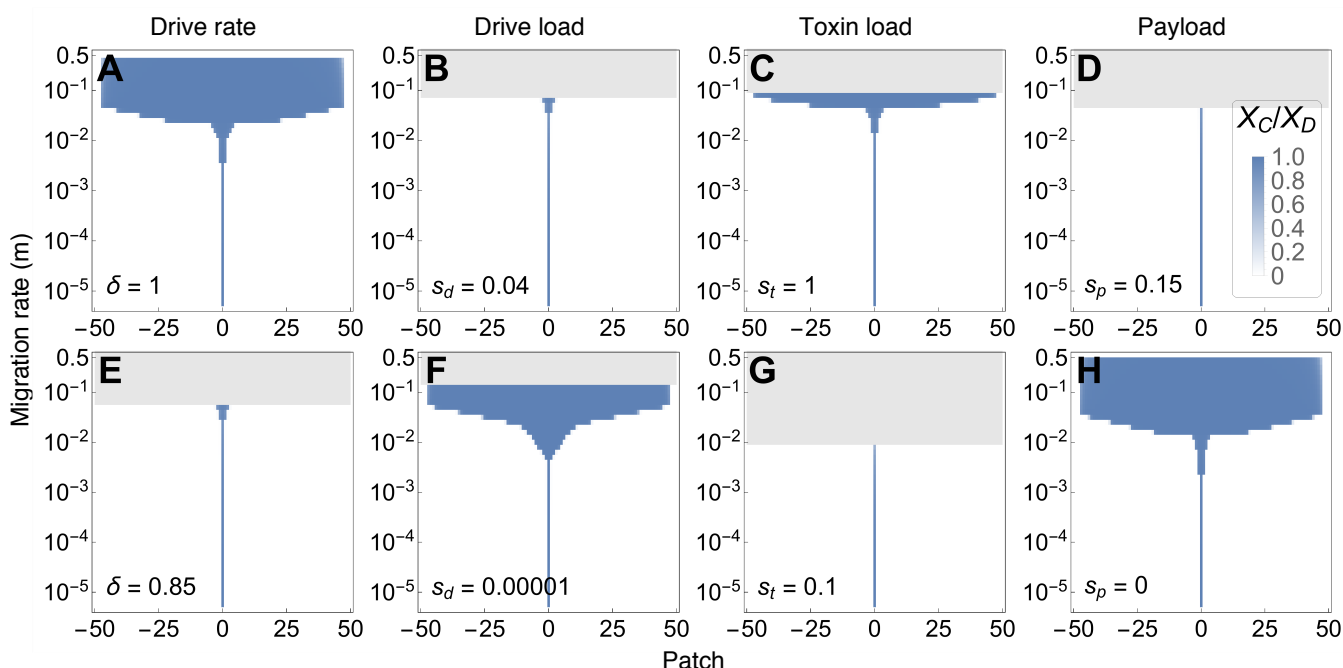


Figure 5: Illustration of the effect of changing a single parameter on the results shown in Figure 4 (panel B) for the default parameters (Table 3). The frequency of the payload is shown in blue after 1000 generations when $ABCD$ gametes are introduced into the center patch (patch 0) at frequency $f_0 = 0.05$. Patches are linearly connected and exchange migrants every generation at a rate m . Subfigures from left to right change the following single parameter up (top figures) or down (bottom figures): drive rate δ , drive load s_d , toxin load s_t and the payload s_p .

Given a particular migration rate m within a given natural system, the power of the daisychain combined with the location of the separatrix determine how far the drive can spread through neighbouring populations. Parameters that affect the daisychain are the drive load s_d , the drive rate δ , and the length of the daisy chain n . Parameters affecting the separatrix are the payload s_p and the toxin load s_t . These parameters could be adjusted during the engineering of the construct, affecting the risk of spread as shown in Figure 5.

Lowering the drive load s_d , increasing the drive rate δ or lengthening the daisy chain n (see supplementary material) results in a daisy chain that persists for longer and spreads the underdominant construct to more populations. Interestingly, a decrease in the toxin-load reduces the risk of spread of the payload to neighbouring populations, even under high migration rates. This is because the lower the toxin load, the closer the separatrix becomes to the state where the underdominant construct CD is fixed (Figure 3). Thus, somewhat counterintuitively, a lower toxin load provides a higher degree of safety in isolating the underdominance construct from spreading into neighbouring patches once the daisy-chain drive has exhausted itself.

Stochastic simulations confirm that these results are robust to finite population sizes in cases where

Parameter	default value
Initial frequency (f_0)	0.05
Drive rate (δ)	0.9
Recombination rate (r)	0.5
Drive load (s_d)	0.02
Payload (s_p)	0.1
Toxin load (s_t)	0.9

Table 3: Default parameter values for Figure 4 and 5

the population persists (Figure S4). In cases where the payload drives the local population extinct, recolonization eventually occurs from neighbouring populations (Figure S5).

5 Discussion

Low threshold gene drives such as homing-based gene drives risk spillover to neighbouring populations due to migrating individuals. The gene drive design that we study here transitions in time from a low threshold drive (homing-based gene drive), which is especially susceptible to these spillovers, to a high threshold drive (two-locus underdominance), which is more resilient to spillover. This transition is achieved by using a daisy-chain technique recently developed by Noble et al. [2019]. We successfully show that this design prevents spread of the payload under biologically realistic scenarios (Figures 4 + 5).

Previous models of homing-based gene drives have focused on the rise of resistance alleles that would prevent the drive from completely reaching fixation in the target population and ultimately revert the system back to the wildtype state. Here we do not explicitly take resistance alleles into account because the daisy-chain drive system, by its nature, decreases in effectiveness. Adding resistance alleles does not drastically change the dynamics of our results (see Supplementary Material). However, we note that the two-locus underdominance construct is susceptible to mutations that decrease or compensate for the payload, which would affect the long-term persistence of the engineered suppression. By contrast mutations that reduce the toxin-load have little effect because the toxin-load is transient and disappears once the underdominance constructs (loci C and D) are fixed. Unlike Dhole et al. [2018], we recommend that both transgenic alleles C and D carry the payload s_p which might better resist the evolution of a compensatory mutation within the payload genotype due to the existence of multiple gene copies.

We also tried an advantage to population alteration, compared to population suppression, that emerges in localized drive systems like the one explored here. If the payload causes local populations to decline to extinction (see Figure S5), wildtype individuals from neighbouring patches will eventually recolonize and return the system to its original state. By contrast, strategies that involve population alteration persist

for longer because migrants continue to face competition and toxic interactions with individuals in the target population carrying the underdominance construct (see Figure S4). Thus the engineered construct investigated here are most suited for population alteration, rather than population suppression, when a persistent and localized effect is sought.

The possibility that gene drive systems provide to alter the genome of species is concerning to many people. Implementing gene drives, accidentally or on purpose, can have long lasting effects on biological systems. It is clear that gene drives have the potential to solve many ecological challenges, however, all consequences of triggering a gene drive cannot be foreseen. Models like the one we present here are important first steps, but they can only be used as a rough guide, while much more experiments and refined models are required before release of gene drives in the wild [National Academies of Sciences et al., 2016].

In summary, we modeled a potential solution to the spillover and persistence problems posed by gene drive systems under realistic biological parameters using a combination of homing based gene drive and two-locus underdominance. Lab-based research is needed to determine the feasibility of developing an underdominant daisy-quorum drive. In particular, controlled experiments are needed to determine the likelihood of mutations reversing the payload (reducing persistence) and the risk of unintended side effects, such as self-driving elements arising from the daisy-chain genes (increasing spillover risk).

6 Acknowledgements

We would like to thank the Otto lab for helpful feedback and for their valuable input to this project. This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada to SPO (NSERC RGPIN-2016-03711) and stipend support to FJHdH from an NSERC CREATE grant (BIOS2).

References

- 286 N. H. Barton and M. Turelli. Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of allee effects. *The American Naturalist*, 178(3):E48–E75, 2011.
- 288 A. B. Buchman, T. Ivy, J. M. Marshall, O. S. Akbari, and B. A. Hay. Engineered reciprocal chromosome translocations drive high threshold, reversible population replacement in drosophila. *ACS synthetic biology*, 7(5):1359–1370, 2018.
- 290 A. Burt. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1518):921–928, 2003.
- 292 S. Davis, N. Bax, and P. Grewe. Engineered underdominance allows efficient and economical introgression of traits into pest populations. *Journal of Theoretical Biology*, 212(1):83–98, 2001.
- 294 S. Dhole, M. R. Vella, A. L. Lloyd, and F. Gould. Invasion and migration of spatially self-limiting gene drives: A comparative analysis. *Evolutionary applications*, 11(5):794–808, 2018.
- 296 J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, and G. M. Church. Safeguarding crispr-cas9 gene drives in yeast. *Nature biotechnology*, 33(12):1250, 2015.
- 298 V. M. Gantz and E. Bier. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, 348(6233):442–444, 2015.
- 300 V. M. Gantz and E. Bier. The dawn of active genetics. *Bioessays*, 38(1):50–63, 2016.
- 302 V. M. Gantz, N. Jasinskiene, O. Tatarenkova, A. Fazekas, V. M. Macias, E. Bier, and A. A. James. Highly efficient cas9-mediated gene drive for population modification of the malaria vector mosquito anopheles stephensi. *Proceedings of the National Academy of Sciences*, 112(49):E6736–E6743, 2015.
- 304 A. T. Isaacs, N. Jasinskiene, M. Tretiakov, I. Thiery, A. Zettor, C. Bourgouin, and A. A. James. Transgenic anopheles stephensi coexpressing single-chain antibodies resist plasmodium falciparum development. *Proceedings of the National Academy of Sciences*, 109(28):E1922–E1930, 2012.
- 306 K. Kyrou, A. M. Hammond, R. Galizi, N. Kranjc, A. Burt, A. K. Beaghton, T. Nolan, and A. Crisanti. A crispr-cas9 gene drive targeting doublesex causes complete population suppression in caged anopheles gambiae mosquitoes. *Nature biotechnology*, 36(11):1062, 2018.
- 310 J. M. Marshall. The effect of gene drive on containment of transgenic mosquitoes. *Journal of theoretical biology*, 258(2):250–265, 2009.
- 312

J. Min, C. Noble, D. Najjar, and K. Esvelt. Daisy quorum drives for the genetic restoration of wild
 314 populations. *BioRxiv*, page 115618, 2017.

E. National Academies of Sciences, Medicine, et al. *Gene drives on the horizon: advancing science,
 316 navigating uncertainty, and aligning research with public values*. National Academies Press, 2016.

C. Noble, J. Min, J. Olejarz, J. Buchthal, A. Chavez, A. L. Smidler, E. A. DeBenedictis, G. M. Church,
 318 M. A. Nowak, and K. M. Esvelt. Daisy-chain gene drives for the alteration of local populations. *Pro-
 ceedings of the National Academy of Sciences*, 116(17):8275–8282, 2019.

L. B. Thien. Mosquito pollination of *habenaria obtusata* (orchidaceae). *American Journal of Botany*, 56
 320 (2):232–237, 1969.

B. Wu, L. Luo, and X. J. Gao. Cas9-triggered chain ablation of cas9 as a gene drive brake. *Nature
 322 biotechnology*, 34(2):137, 2016.

7 Supplementary Material

7.1 Daisy-chain length

The length of the daisy chain is an important determinant of the maximum attainable frequency of the payload. Here we illustrate this effect for daisy chains of different lengths by tracking the maximal frequency attained by the last element in the chain. When the number of loci involved in the daisy chain increases, there is a stronger and longer force pushing the last element to a higher frequency (Figure S1). The inline figure shows an example of the dynamics of a daisy chain system for 3 loci.

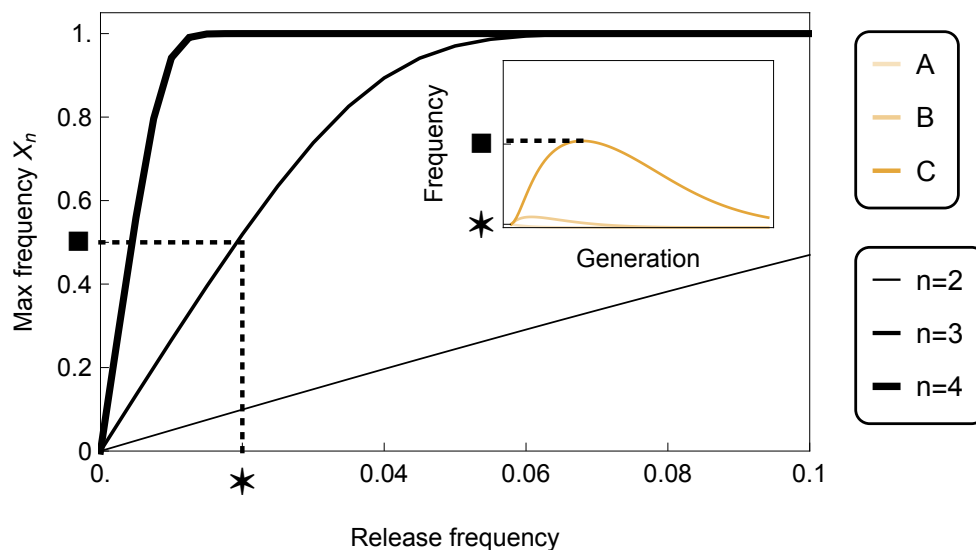


Figure S1: Maximum frequency of the last element in an n length daisy-chain construct like locus **A** and **B** in Table 1 in a single population. All elements of the daisy-chain carry a drive load $s_d = 0.05$. The frequency of the n^{th} transgenic allele is denoted by X_n . Other parameters are $\delta = 1$, $r = 0.5$.

7.2 Mating tables

Here we display the mating tables associated with the 4-locus daisy-quorum system that we are exploring in the main text (Table 1). Table S1 gives the mating table for the dynamics at loci **A** and **B** (the daisy chain component). Table S2 provides the mating table for the dynamics at loci **C** and **D** (the underdominance component).

7.3 Alternative payload fitness implementation

In the main text we focus on a dominant payload expression (Table 2), because under this fitness regime, fixation of the *CD* haplotype is locally stable. When we assume different payload regimes (multiplicative dominance or recessive effects on individual fitness), then fixation of *CD* haplotype is no longer a stable equilibrium, as described below.

Gamete 1	Gamete 2	Fitness	Freq	Gametes produced			
				<i>ab</i>	<i>aB</i>	<i>Ab</i>	<i>AB</i>
<i>ab</i>	<i>ab</i>	1	X_{ab}^2	1			
<i>ab</i>	<i>aB</i>	$(1 - s_d)$	$2X_{ab}X_{aB}$	$\frac{1}{2}$	$\frac{1}{2}$		
<i>ab</i>	<i>Ab</i>	$(1 - s_d)$	$2X_{ab}X_{Ab}$	$\frac{1}{2}$		$\frac{1}{2}$	
<i>ab</i>	<i>AB</i>	$(1 - s_d)^2$	$2X_{ab}X_{AB}$	$\frac{1}{2}(1 - \delta)(1 - r)$	$\frac{1}{2}(\delta + (1 - \delta)r)$	$\frac{1}{2}(1 - \delta)r$	$\frac{1}{2}(1 - (1 - \delta)r)$
<i>aB</i>	<i>aB</i>	$(1 - s_d)^2$	X_{aB}^2		1		
<i>aB</i>	<i>Ab</i>	$(1 - s_d)^2$	$2X_{aB}X_{Ab}$	$\frac{1}{2}(1 - \delta)r$	$\frac{1}{2}(1 - (1 - \delta)r)$	$\frac{1}{2}(1 - \delta)(1 - r)$	$\frac{1}{2}(\delta + (1 - \delta)r)$
<i>aB</i>	<i>AB</i>	$(1 - s_d)^3$	$2X_{aB}X_{AB}$		$\frac{1}{2}$	$\frac{1}{2}$	
<i>Ab</i>	<i>Ab</i>	$(1 - s_d)^2$	X_{Ab}^2			1	
<i>Ab</i>	<i>AB</i>	$(1 - s_d)^3$	$2X_{Ab}X_{AB}$			$\frac{1}{2}(1 - \delta)$	$\frac{1}{2}(1 + \delta)$
<i>AB</i>	<i>AB</i>	$(1 - s_d)^4$	X_{AB}^2				1

Table S1: Mating table for locus **A** and **B**, illustrating the gametes that come together to make a diploid individual (first two columns), their fitness (third column), at birth frequency (fourth column), and the gametes they produce (fifth-eighth columns)

Gamete 1	Gamete 2	Fitness	Freq	Gametes produced			
				<i>cd</i>	<i>cD</i>	<i>Cd</i>	<i>CD</i>
<i>cd</i>	<i>cd</i>	1	X_{cd}^2	1			
<i>cd</i>	<i>cD</i>	$(1 - s_t)(1 - s_p)$	$2X_{cd}X_{cD}$	$\frac{1}{2}$	$\frac{1}{2}$		
<i>cd</i>	<i>Cd</i>	$(1 - s_t)(1 - s_p)$	$2X_{cd}X_{Cd}$	$\frac{1}{2}$		$\frac{1}{2}$	
<i>cd</i>	<i>CD</i>	$(1 - s_p)$	$2X_{cd}X_{CD}$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1 - r)$
<i>cD</i>	<i>cD</i>	$(1 - s_t)(1 - s_p)$	X_{cD}^2		1		
<i>cD</i>	<i>Cd</i>	$(1 - s_p)$	$2X_{cD}X_{Cd}$	$\frac{1}{2}r$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}(1 - r)$	$\frac{1}{2}r$
<i>cD</i>	<i>CD</i>	$(1 - s_p)$	$2X_{cD}X_{CD}$		$\frac{1}{2}$	$\frac{1}{2}$	
<i>Cd</i>	<i>Cd</i>	$(1 - s_t)(1 - s_p)$	X_{Cd}^2			1	
<i>Cd</i>	<i>CD</i>	$(1 - s_p)$	$2X_{Cd}X_{CD}$			$\frac{1}{2}$	$\frac{1}{2}$
<i>CD</i>	<i>CD</i>	$(1 - s_p)$	X_{CD}^2				1

Table S2: Mating table for locus **C** and **D**, illustrating the gametes that come together to make a diploid individual (first two columns), their fitness (third column), at birth frequency (fourth column), and the gametes they produce (fifth-eighth columns)

Multiplicative dominance of payload

342 Table S3 shows the individual fitness under a multiplicative dominance regime for the payload.

	<i>cd</i>	<i>cD</i>	<i>Cd</i>	<i>CD</i>
<i>cd</i>	1	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_p)^{1/2}$
<i>cD</i>	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_t)(1 - s_p)^{1/2}$	$(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$
<i>Cd</i>	$(1 - s_t)(1 - s_p)^{1/4}$	$(1 - s_p)^{1/2}$	$(1 - s_t)(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$
<i>CD</i>	$(1 - s_p)^{1/2}$	$(1 - s_p)^{3/4}$	$(1 - s_p)^{3/4}$	$1 - s_p$

Table S3: Payload (s_p) is multiplicative within and between loci

The governing equations for the underdominance component with multiplicative payload selection in

344 Table S3 are:

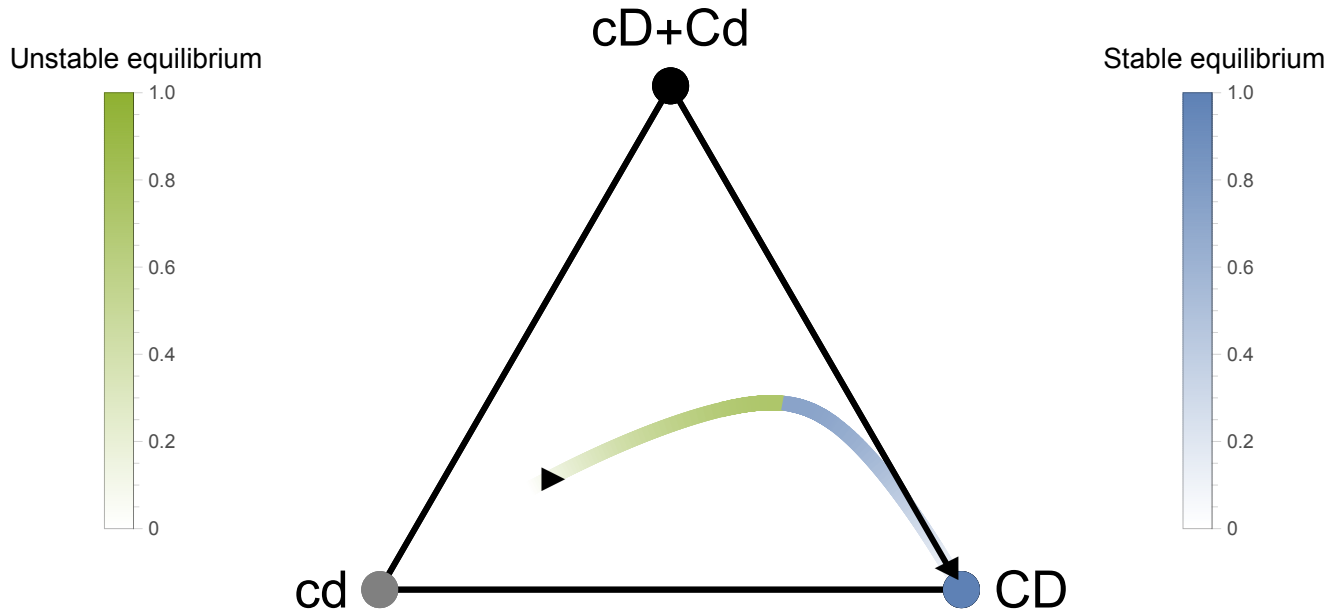


Figure S2: Equilibria of the multiplicative fitness regime of Table S3. The parametric lines indicate the internal unstable (green) and stable (blue) equilibria for different payload values s_p . The toxin load $s_t = 1$ and $r = \frac{1}{2}$. For a given payload, s_p , starting to the right of the green point will lead the system to approach the blue point with the same shade (s_p value).

$$\begin{aligned}
 X'_{cd} &= \frac{1}{W} (X_{cd}X_{cD} \sqrt[4]{1-s_p}(1-s_t) + X_{cd}X_{Cd} \sqrt[4]{1-s_p}(1-s_t) + X_{cd}X_{CD}(1-r) \sqrt[2]{1-s_p} + X_{cd}X_{cd} + X_{cD}X_{Cd}r \sqrt[2]{1-s_p}) \\
 X'_{cD} &= \frac{1}{W} (X_{cd}X_{CD}r \sqrt[2]{1-s_p} + X_{cd}X_{cD}(1-s_t) \sqrt[4]{1-s_p} + X_{cD}X_{Cd}(1-r) \sqrt[2]{1-s_p} + X_{cD}X_{CD} \sqrt[3]{1-s_p}) \\
 X'_{CD} &= \frac{1}{W} (X_{cd}X_{CD}r \sqrt[2]{1-s_p} + X_{cd}X_{Cd}(1-s_t) \sqrt[4]{1-s_p} + X_{cD}X_{Cd}(1-r) \sqrt[2]{1-s_p} + X_{cD}X_{CD} \sqrt[3]{1-s_p} +) \\
 X'_{CD} &= \frac{1}{W} (X_{CD}X_{cd}(1-r) \sqrt[2]{1-s_p} + X_{CD}X_{Cd} \sqrt[3]{1-s_p} - X_{CD}X_{CD}(1-s_p) + X_{cD}X_{Cd}r \sqrt[2]{1-s_p} + X_{cD}X_{CD} \sqrt[3]{1-s_p})
 \end{aligned}
 \tag{4}$$

Analysis of the equilibria of this dynamical system reveals two internal equilibria, plus $X_{cd} = 1$, which is stable, and $X_{CD} = 1$, which is unstable. In Figure S2 we illustrate the location of the various stable and unstable equilibria. As the payload increases, the stable and unstable equilibrium shift towards each other. This means that a drive is harder to invade and will spread to a lower frequency when the payload increases in strength.

The same invasion analysis as in section 3.3 reveals that the conditions for the drive rate δ_c needed for the transgenic alleles C and D to spread are:

$$\delta_c > \min \left[\frac{\frac{1}{\sqrt[4]{1-s_p}} + 1 - s_t}{1 - s_t}, \frac{\frac{\sqrt{-r(2s_p + \sqrt{1-s_p}-2) + s_p + \sqrt{1-s_p}-1}}{\sqrt[4]{1-s_p}} - r}{1 - r} \right]
 \tag{5}$$

This shows that the maximum payload that can be invaded by a two-locus underdominance system with multiplicative payload is $s_p = 15/16$ with maximal drive rate $\delta = 1$ and $s_t \approx 0$.

Recessive payload

Table S4 presents the relative fitness of individuals when the payload is recessive, by which we mean an individual needs two copies of C and two copies of D alleles before expressing the payload.

	cd	cD	Cd	CD
cd	1	$1 - s_t$	$1 - s_t$	1
cD	$1 - s_t$	$1 - s_t$	1	1
Cd	$1 - s_t$	1	$1 - s_t$	1
CD	1	1	1	$1 - s_p$

Table S4: Payload (s_p) is recessive

The governing equations for the underdominance component with a recessive payload are:

$$\begin{aligned}
 X'_{cd} &= \frac{1}{\bar{W}} X_{cd} X_{cD} (1 - s_t) + X_{cd} X_{CD} (1 - r) + X_{cd}^2 + X_{cD} X_{Cd} r \\
 X'_{cD} &= \frac{1}{\bar{W}} X_{cD} X_{cd} (1 - s_t) + X_{cD}^2 (1 - s_t) + X_{cD} X_{Cd} (1 - r) + X_{CD} X_{cd} r + X_{CD} X_{cD} \\
 X'_{Cd} &= \frac{1}{\bar{W}} X_{cd} X_{CD} r + X_{cd} X_{Cd} (1 - s_t) + X_{Cd} X_{cD} (1 - r) + X_{Cd}^2 (1 - s_t) + X_{Cd} X_{CD} \\
 X'_{CD} &= \frac{1}{\bar{W}} X_{CD} X_{cd} (1 - r) + X_{CD} X_{Cd} + X_{CD}^2 (1 - s_p) + X_{cD} X_{Cd} r + X_{cD} X_{CD}
 \end{aligned} \tag{6}$$

In Figure S3 we illustrate the location of the equilibria of this model for $s_t = 1.0$ and $s_t = 0.1$. We see that similar to the analysis in the main text for a dominant payload, a reduction in toxin efficiency brings the equilibrium further from the wildtype (Figure S3 (a,b)).

The results of the invasion analysis show that this construct can only invade when:

$$\delta_c > \min\left[\frac{s_t}{1 - s_t}, \frac{\sqrt{r}}{\sqrt{r} + 1}\right] \tag{7}$$

7.4 Individual-based simulations

Individual-based simulations (implemented in C^{++}) were developed to explore whether our results hold for finite population sizes. We use the same fitness regime as in the main text (dominant payload expression). We implement the following life cycle in an isolated population:

1. Calculate the expected frequency of genotypes predicted under the deterministic model.
2. Calculate mean fitness \bar{W} , and draw from a Poisson random variate with rate parameter $\lambda = \bar{W} \times F \times N_t$ the number of offspring in the next generation \hat{N}_{t+1} . Here, F represents the average

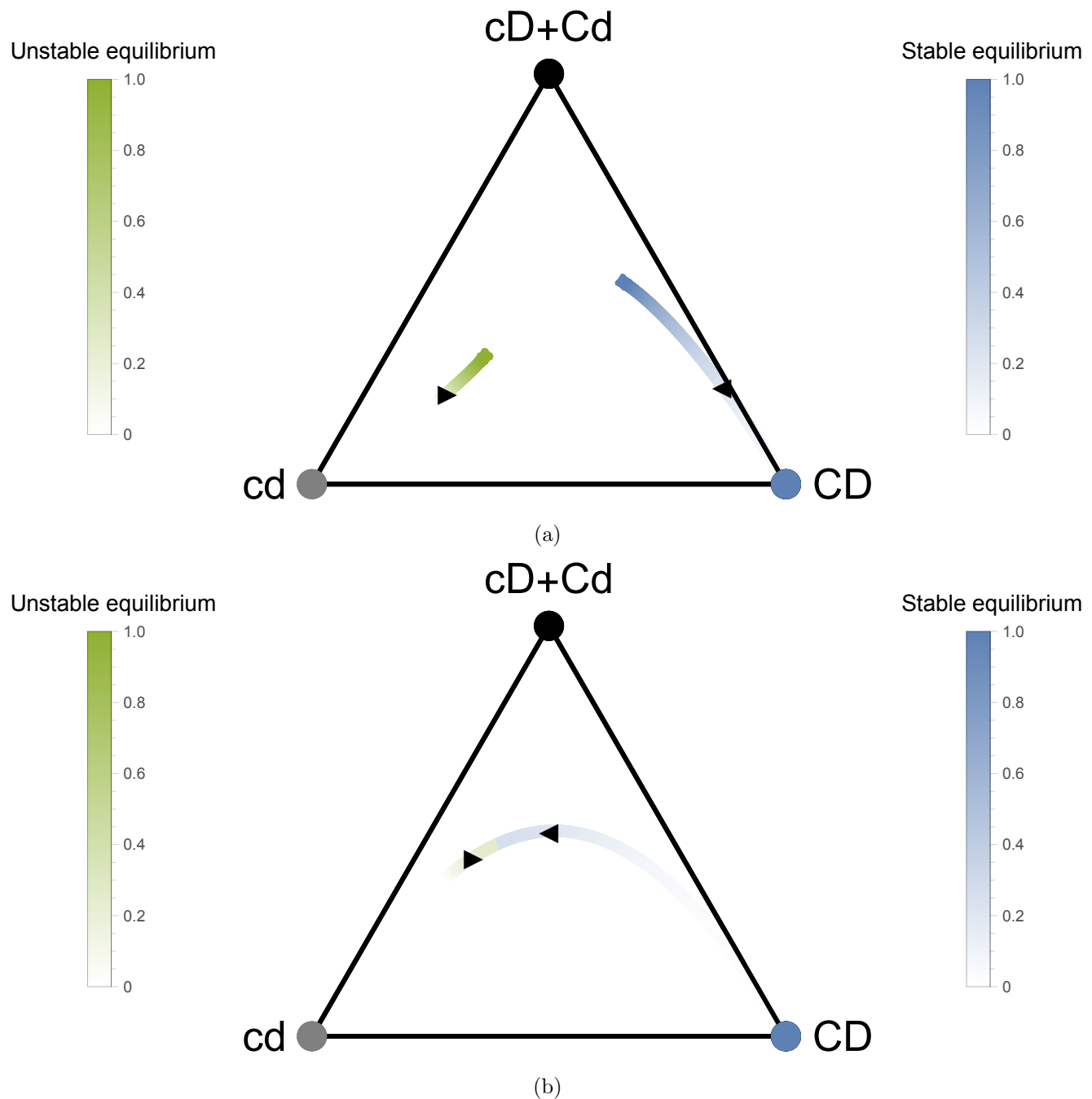


Figure S3: Equilibria of the recessive fitness regime of Table S4 for (a) $s_t = 1.0$ and (b) $s_t = 0.1$. The parametric lines indicate the internal unstable (green) and stable (blue) equilibria for different payload values s_p . The toxin load $s_t = 1$ and $r = \frac{1}{2}$. For a given payload, s_p , starting to the right of the green point will lead the system to approach the blue point with the same shade (s_p value).

number of offspring per individual. The total number of offspring in the next generation was then set to $N_{t+1} = \min[\hat{N}_{t+1}, K]$.

3. Draw N_{t+1} offspring genotypes from a multinomial distribution with the deterministic predicted genotype frequencies.

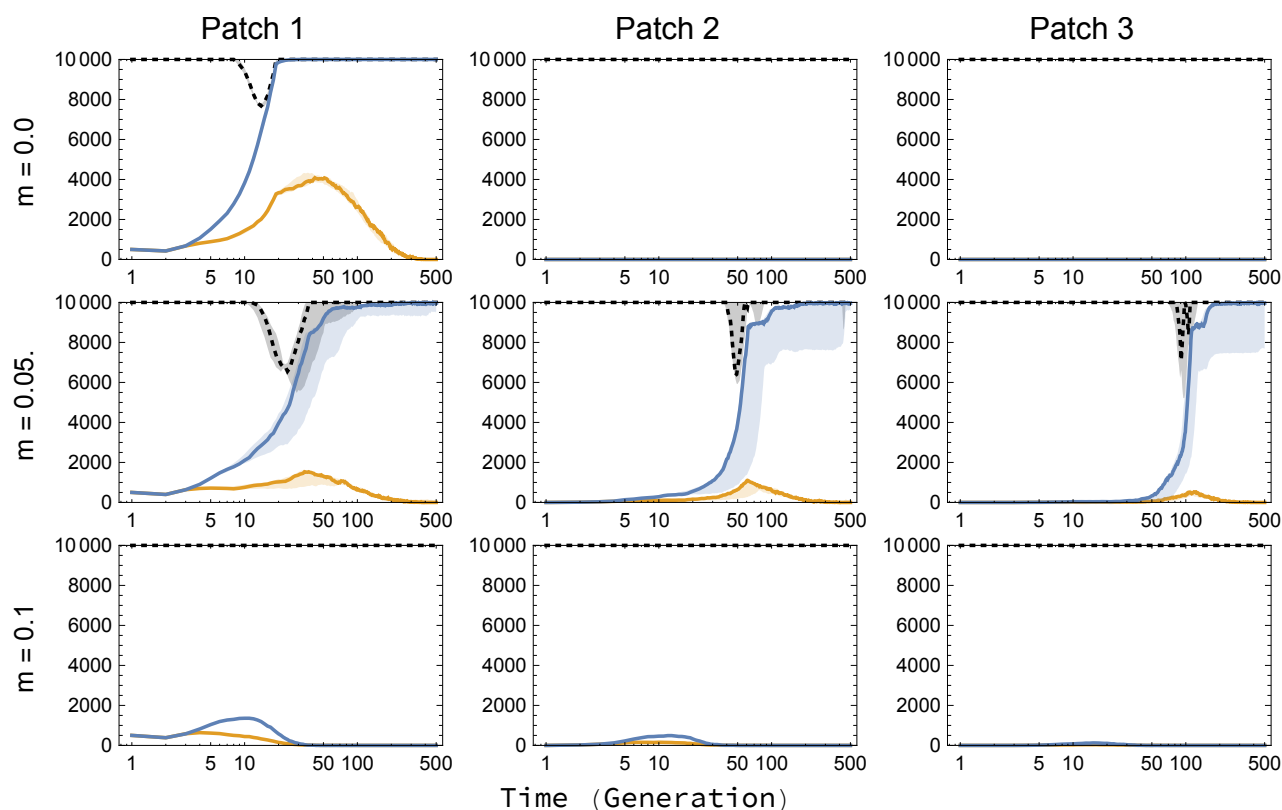


Figure S4: Finite population demographic dynamics. Black dotted line is the total population size at time t , blue solid line represents the number of individuals carrying the payload allele and orange solid line represents the number of individuals carrying the B allele. In this case, fertility is set to $F = 1.2$ and the payload to $s_p = 0.1$, so that patches can persist even if the underdominance construct fixes. Each patch has a carrying capacity of $K = 10000$. Rows have a migration probability of 0, 0.05, 0.1 from top to bottom. Drive is released on the left. Results are shown for 50 replicates. The solid line represents the median of the replicates, and the shaded region the first and third quantiles.

We repeat these steps for t_{max} generations and show results in Figure S4 and S5 for $F = 1.2$ and $F = 1.05$, respectively. In Figure S4, with $F = 1.2$ and a payload of $s_p = 0.1$, the average number of offspring per individual carrying the payload remains above unity: $\lambda = 0.9 \times 1.2 = 1.08$. However, the toxin load $s_t = 0.9$ causes a dip in population size (black dashed line) as the payload increases in frequency. Once the payload has fixed, the toxin load will not be expressed, and the population recovers to its carrying capacity. Figure S5 illustrates dynamics for $F = 1.05$. Here $\lambda = 0.9 \times 1.05 = 0.945$, which is below unity, meaning that individuals carrying the payload will not be able to replace themselves on average. We see that in this scenario the payload for $m = 0.05$ does not go to fixation, unlike Figure S4. This is because the population size is more severely affected by the payload. As the population declines, proportionally more

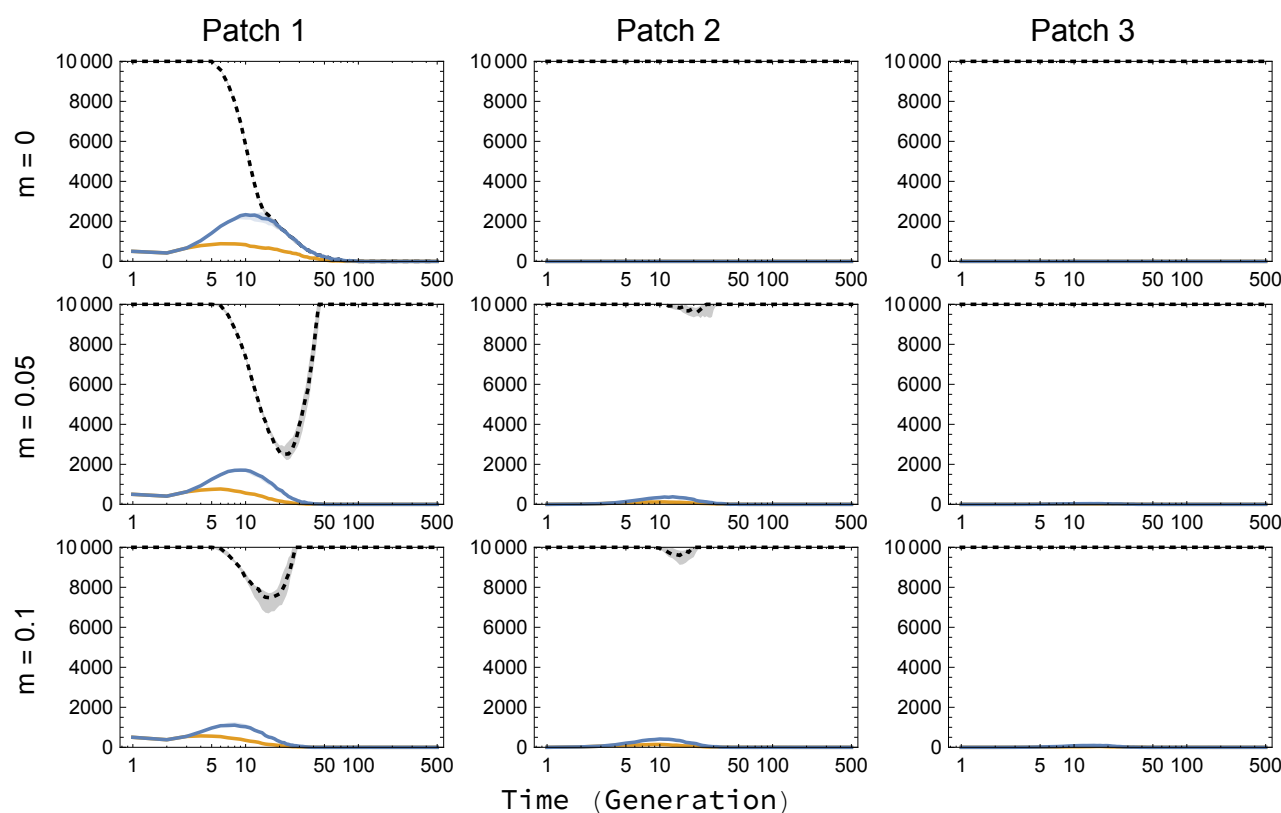


Figure S5: Finite population demographic dynamics. Black dotted line is the total population size at time t , blue solid line represents the number of individuals carrying the payload allele and orange solid line represents the number of individuals carrying the B allele. In this case, fertility is set to $F = 1.05$ and the payload to $s_p = 0.1$, so that patches are driven extinct if the payload fixes after which recolonization by the wildtype eventually occurs. Each patch has a carrying capacity of $K = 10000$. Rows have a migration probability of 0, 0.05, 0.1 from top to bottom. Drive is released on the left. Results are shown for 50 replicates. The solid line represents the median of the replicates, and the shaded region the first and third quantiles.

migrants enter the patch than when the population is large. This proportionally high inflow of wildtype individuals prevents fixation of the payload and eventually helps the population recover to its carrying capacity with only wildtype individuals.

In our deterministic analysis of spread in Figures 4 and 5, we noted that a decrease in the toxin load s_t can help prevent spillover of the construct by shifting the separatrix further from the wildtype state (Figure 2). However, decreasing the toxic load in order to shift the separatrix comes at a cost in terms of persistence of the construct in finite population sizes. In the limit as the toxic load goes to 0, the valley in Figure 1 (panel B) disappears, which results in directional selection towards the wildtype. Lowering the toxin load to center the separatrix (the location of the valley) will simultaneously reduce the depth of the fitness valley. Such a shallow fitness landscape means that the effects of drift in finite population sizes can influence the dynamics significantly.

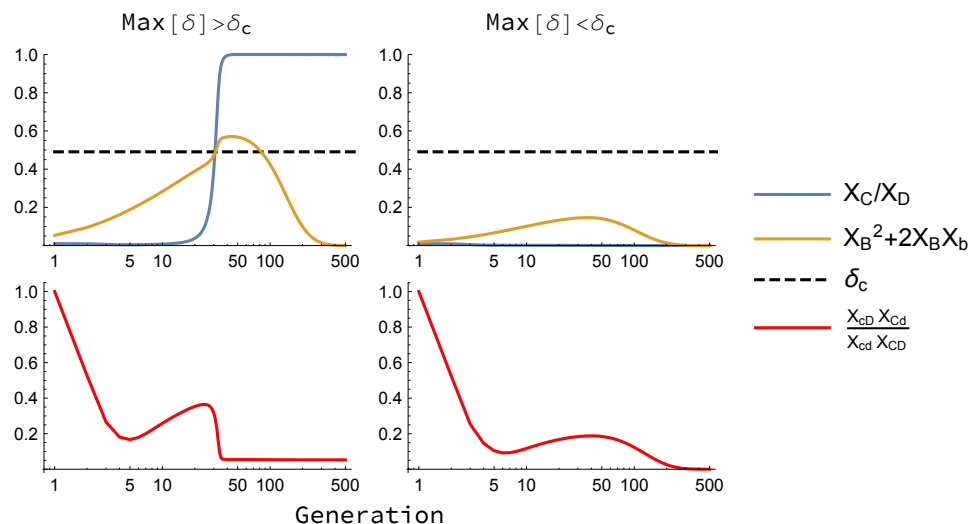


Figure S6: $\delta = 0.9$, $r = 0.5$, $s_p = 0.1$, $s_t = 0.9$, $f_{drive} = 0.01|0.03$ and $f_{EUD} = 0.01$

7.5 Invasion analysis

Our invasion analysis relies on there not being any linkage disequilibrium between the loci involved in the daisy chain and the underdominance component. In Figure S6 we illustrate the critical drive threshold above which the $X_{cd} = 1$ equilibrium becomes unstable assuming there is no linkage disequilibrium between the loci. The left column shows a starting frequency of the CD alleles at $f_{CD} = 0.01$ and the right column for $f_{CD} = 0.03$. We initiate the simulations without linkage disequilibrium between the **AB** and **CD** loci. The initial frequency for CD is $f_{CD} = 0.01$. The top graphs show the dynamics of the simulation, where the left column illustrates a scenario where the drive strength δ exceeds the critical threshold δ_c while the right does not. The bottom graphs show that significant linkage disequilibrium builds up soon after the simulations start. As a consequence of these genetic associations, the threshold drive does not precisely predict when spread will occur. For example, in the top middle panel, the drive frequency (orange) rises above the threshold (dashed line) but the underdominance construct fails to spread to fixation. Nevertheless, equation (3) provides an approximate guide to when the underdominance can (left) and will not (right) spread.