Slow and steady wins the race: spatial and stochastic processes and the failure of suppression gene drives

Jeff F. Paril* and Ben L. Phillips

School of BioSciences, University of Melbourne, Parkville VIC 3010

*Author and address for correspondence:

jeff.paril@unimelb.edu.au

February 10, 2022

Abstract

Gene drives that skew sex ratios offer a new management tool to suppress or eradicate pest populations. Early models and empirical work suggest that these suppression drives can completely eradicate well-mixed populations,

but models that incorporate stochasticity and space (i.e., drift, and founder events) often result in loss or failure of the drive. We developed a stochastic model to examine these processes in a simple 1-dimensional space. This simple space allows us to map the events and outcomes that emerged and examine how properties of the drive's wave of invasion affect outcomes. Our simulations, across a biologically-realistic section of parameter space, suggest that drive failure might be a common outcome in spatially explicit, stochastic systems, and that properties of the drive wave appear to mediate outcomes. Surprisingly, the drives that would be considered fittest in an aspatial model were strongly associated with failure in the spatial setting. The fittest drives cause fast moving, narrow drive waves that have a high chance of being penetrated by founder events, leading to failure. Our results also show that high rates of dispersal reduce the chance of failure because drive waves get disproportionately wider as dispersal rates increase. Overall, wide, slow-moving drive waves were much less prone to failure. Our results point to the complexity inherent in using a genetic system to effect demographic outcomes and speak to a clear need for ecological and evolutionary modelling to inform the drive design process.

Keywords

gene drive chasing, invasion waves, spatial simulations, suppression gene drive, W-shredder, X-shredder.

1 Introduction

The discovery of CRISPR-Cas9 will soon revolutionise our capacity to manipulate populations. By enabling precise gene editing of individuals, CRISPR-Cas9 allows us to introduce novel genes into populations (Chang et al. 2013; DiCarlo et al. 2013; Friedland et al. 2013; Horvath and Barrangou 2013; Miao et al. 2013; Shen et al. 2013; Yu et al. 2013; Kyrou et al. 2018). Such novel genes are, like all functional genes, subject to natural selection. But by pairing the novel genes we wish to introduce with selfish genetic elements, we can introduce genetic constructs that will tend to increase in frequency despite fitness costs. Such "gene drives" allow us to drive traits into populations, against the force of natural selection (Noble et al. 2018). In principle, we can use such constructs to change the very demographics of a population: to reduce population fitness, and even to drive populations extinct (Hammond et al. 2021).

The possibility that we might use gene drives for population suppression is of clear interest in the management of pests and diseases. Globally, we spend billions of dollars annually to control of pests and diseases (*e.g.*, annual ryegrass and rodents, Bradshaw et al. 2021), and many of the tools we use are blunt ones: they involve environmental impacts and raise ethical questions we prefer to look away from (Hough 2021). Gene drives offer a new alternative. Various suppression gene drive systems have now been proposed, most of which make use of CRISPR-Cas9 (Sinkins and Gould 2006; Macias

et al. 2017; Champer et al. 2020a; Champer et al. 2020b). These gene drives aim to eradicate populations by forcing the inheritance of fitness-reducing alleles leading to sex ratio distortion, sterility, and/or lethality. Such systems, in principle, offer highly targeted control and completely avoid the environmental and welfare impacts of many existing pest control strategies.

The key qualifier in both previous paragraphs is "in principle". There remains substantial work to be done to develop effective, safe, targeted, and socially acceptable control options using gene drives. On the technical front, an effective gene drive needs to invade the target population and attain a frequency high enough to effect management aims. To do so, it will need to push hard against the force of natural selection. This is not a trivial challenge, but several constructs now exist that prevent drive resistant alleles emerging, and which can be used to skew sex ratios until the population collapses. Such constructs work in theory and have also been shown to drive laboratory populations extinct. For example, CRISPR-Cas9 gene drives have driven caged populations of *Anopheles gambiae* mosquitoes to extinction (Kyrou et al. 2018; Simoni et al. 2020) and have also been demonstrated to cause significant sex distortion in *Ceratis capitata* fruit flies under controlled laboratory conditions (Meccariello et al. 2021). These laboratory trials are an encouraging proof of concept, but whether such constructs will work in a real setting remains to be seen.

One of the major challenges to the effectiveness of drive constructs for population suppression arises when we move these systems from the lab and

into more realistic spatially explicit settings. Here, the very success of a construct might work against it. In a spatially explicit setting, population suppression at one location immediately sets up a density gradient between suppressed and non-suppressed areas. This gradient causes an asymmetry in gene flow, such that there is a net flow of genes from non-suppressed areas into the suppressed areas. Thus, in space, a drive construct needs to push against both natural selection, and gene flow (Girardin et al. 2019).

To guide our understanding of how gene drives might behave in real settings, analytical and simulation models are valuable tools. The earliest gene drive models take the sensible simplifying step of ignoring space: they assume panmixia. These mathematical models and population-based stochastic simulations routinely show that suppression gene drives will rapidly crash populations to extinction (Burt and Deredec 2018; Deredec et al. 2008; Prowse et al. 2017). But when we introduce space, and stochastic processes, things become complicated. The clearest stochastic evolutionary process is drift, in which alleles can be lost from small populations despite deterministic expectations that they should increase in frequency. When space is included in models, we also have to contend with founder events, which are essentially a spatial manifestation of drift (Slatkin and Excoffier 2012). These two processes can lead to surprising emergent dynamics that are absent from deterministic and aspatial models. Founder events, for example, may allow wild-type individuals to recolonise previously cleared space. When such events are sufficiently common, the result can be a complex dynamic in which

the gene drive endlessly chases the wild-type population and so fails to effect eradication (Champer et al. 2021).

With spatial and stochastic processes in play, we take a drive system that is capable of crashing a large, panmictic population, and we observe it fail to crash a finite, spatially-structured population. If we are ever to apply gene drive technology to the control of pest species, it is clear that we need to understand these stochastic and spatial effects. What are the conditions under which they manifest? Are some drive systems less susceptible to spatial effects than others? Answering these questions will bring us a step closer to the effective control of spatially-structured populations such as cane toads in Australia (Urban et al. 2008), and lepidopteran pest populations in large-acre farming (Anderson et al. 2016; Jones et al. 2019).

While analytical models often offer more general conclusions, they rapidly become intractable when bent to describe gene drives with demographic effects (Girardin and Débarre 2021), and become more complex still with the inclusion of stochasticity. In this paper, we gain some insight from analytical models, but largely approach from the opposite direction, with an individual-based simulation model in which complex dynamics can emerge. We sacrifice general conclusions in favour of phenomenological description of emergent patterns. In our model, we consider the waves of invasion of both wild-type and gene-drive populations. We commence with the expectation that there are situations (i.e., parts of parameter and model space) in which stochastic effects are negligible, and situations in which they are not. Failure

typically does not occur where stochastic effects are negligible. Importantly, the impact of stochastic effects are in large part controlled by two characteristics of the emergent invasion waves. First, the width of these waves matter: relatively narrow drive waves are more likely to drift to extinction or to be penetrated by founder events than wide drive waves. Second, once a drive wave has been penetrated, the relative velocity of the drive wave matters: drive waves that move faster than wild type waves are more prone to drive loss following founder events. Both the emergent property of wave shape and velocity is strongly affected by the choice of the suppression drive system. Thus, careful choice of drive systems can avoid the worst stochastic outcomes by generating a relatively wide invasion wave whose velocity is calibrated to be similar to that of the wild type.

2 Simulation Methods

We performed spatially explicit individual-based stochastic simulations of suppression gene drive introduction to eradicate a wild-type (WT) population. We complement these simulations with the derivation of the recursive gene drive allele frequency equations for the most popular suppression gene drive systems, the sex distorter drives: X-shredder and W-shredder (Galizi et al. 2016; Burt and Deredec 2018; Kyrou et al. 2018; Holman 2019; Prowse et al. 2017; Simoni et al. 2020). From these equations we built naive partial differential equations (which ignore advection) describing gene drive and

WT dynamics across 1-dimensional space, and derived Fisher's asymptotic spreading velocities for the two suppression gene drive systems. Because advection is ignored, these Fisher velocities can be thought of as the maximum velocity that a drive wave is likely to achieve.

2.1 Local dynamics

We simulated a population with discrete non-overlapping generations, inhabiting a continuous 1-dimensional space with a total length of 2,500 spatial units. This bounded 1D space has absorbing boundaries simulating the phenomenon where individuals dispersing beyond their range fail to survive. Bounded space is a pragmatic choice in that it reflects the reality of species ranges. It is important to note that some of the pathways to success of the drive depend upon this feature of space (see "Emergent dynamics" below).

Individuals are assumed to spread their density across space according to a univariate normal distribution with standard deviation of one. Local density experienced by an individual, N_i is calculated as the sum of densities across all individuals at the reference individual's location:

$$N_i = \sum_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-d_{i,j}^2/2}, \tag{1}$$

where d_{ij} is the Euclidean distance between individuals i and j, and n is the total number of individuals in the population at any given time.

Females choose a mate from their local neighbourhood, within 3 spatial

units. This constraint provides a spatial Allee effect: individuals that cannot find a mate die without reproducing. A male within this range is chosen at random, but the probability of choosing a particular male scales with the density that male provides to the female's location (given by the individual density described in equation (1)). Mated females produce offspring with an expected reproductive output defined by the Beverton-Holt growth function:

$$E(W_i) = \frac{R}{1 + aN_i}, \tag{2}$$

where R is the maximum expected number of offspring a female can produce which ranges from 2 to 10, and a determines the response to density. In our case, we are modelling discrete sexes with an expected equal sex ratio, so we set $a = \frac{R-2}{2N^*}$, where N^* denotes the equilibrium population density and was set to 5 in most simulations. A female's realised number of offspring is given by a draw from a Poisson distribution, $W_i \sim \text{Poiss}(E(W_i))$.

2.2 Dispersal dynamics

Offspring are born at the mother's location and immediately disperse. Dispersal is described by a univariate normal distribution with standard deviation, $\sigma \in [2, 20]$. This dispersion variable defines the root mean squared dispersal distance as σ spatial units in any given simulation. Thus, σ, sets the scale of dispersal (relative to the scale of local dynamics), in each simulation. This range translates to an average movement of 0.1% to 1% of the

landscape's width per individual per generation, which is realistic for many established invasive species, *e.g.*, cane toads (Phillips et al. 2006; Kearney et al. 2008; Urban et al. 2008).

2.3 Gene drives and population suppression

We studied two sex-distorter suppression gene drives, X-shredder and W-shredder. X-shredder works on the XY sex-determination system, where individuals with XX chromosomes are females while those with XY are males. On the other hand, W-shredder works on the ZW sex-determination system, where ZW individuals are females and ZZ individuals are males. Both these systems have the ability to eradicate populations by skewing the sex ratio in favour of males, hence reducing the overall fecundity of the population until the population becomes 100% male and becomes extinct.

In the X-shredder system, the Y chromosome carries the homing endonuclease gene drive (Y^d) which creates breaks along the X chromosome rendering it dysfunctional. As a consequence, XY^d males will only produce Y^d gametes. We assume that the volume of male gametes is unaffected by the drive. Similarly in W-shredder, the Z chromosome carries the homing endonuclease gene drive (Z^d) which likewise "shreds" the W chromosome. As a consequence, Z^dW females will only produce Z^d gametes. We also assume that the Z^dW females have unaltered fecundity.

For both W- and X-shredders, we assume 100% shredding efficiency. We also assume that the gene drive alleles do not reduce the fecundity or the total

gametic output of the individuals carrying them. Additionally, we ignore the possibility of evolution of resistance to shredding through mutated shredding target sites. Thus, we modelled an ideal gene drive system (which is yet to be achieved in the lab or in the field).

2.4 Simulated scenarios

For each of our drive types, we simulated a total of 100 different scenarios, each replicated 100 times. Each scenario is a combination of the two variables: maximum female fecundity, $R_i \in [2, 10]$ and dispersion parameter, $\sigma_i \in [2, 20]$ for $i \in 1, 10$. In all scenarios, we set $N^* = 5$. This sets our maximum population size across the entire spatial domain (2,500 units) to 12,500 individuals. Within the more limited area defining local population dynamics (within 3 units of a focal individual) this carrying capacity implies a maximum local population size of approximately 30 individuals.

We initialised the wild-type population to uniformly occupy space between positions $x = 250$ to $x = 1,250$ (10% to 50% of the spatial domain). The suppression gene drive-carriers (i.e., heterozygote for the drive allele) were introduced at 1% of the wild-type population size. These drive-carriers were all placed at $x = 250$, i.e., the left-most border of the initialised population. This initialisation established two invasion waves: one for the wild-type population invading $x > 1,250$, and the other for the drive invading $x > 250$. Population dynamics proceeded for at most 1,000 generations, or until the population was eradicated, or after an additional 10 generations following

the loss of the drive allele, whichever occurred first. At each time step, we recorded the following:

- population size,
- proportion of females,
- population density (mean of the local densities of each individual),
- standard deviation of the population density (over space),
- drive allele frequency, and
- mean number of offspring (total, and separately for both wild-type and drive-carrier parents).

Additionally, the following wave properties were measured for both wild-type and drive waves when drive wave penetration has not yet occurred, and the waves have not yet reached the edge of the landscape:

- wave velocities: defined as the difference in location between furthest leading individuals between t and $t-1$ (averaged across generations),
- wave heights: defined as the maximum density of the respective wave, i.e., the wave peak (measured after 1 generation of mating and dispersion following gene drive introduction), and
- wave widths (measured after 1 generation of mating and dispersion following gene drive introduction): *WT wave width* defined as twice

the distance between the farthest individuals and the peak of the wave at the leading edge; *drive wave leading half width*: distance between the leading individual and the peak of the wave; and *drive wave trailing half width*: distance between the trailing individual and the peak of the wave; *drive wave width*: *drive wave leading half width* plus *drive wave trailing half width*.

2.5 Emergent dynamics and outcomes

We define suppression gene drive "success" as the complete eradication of the wild-type population on or before $t = 1000$ generations. "Failure", is defined as the population being extant at $t = 1000$, or the complete loss of the drive allele any time before then. With regard to loss of the drive, there are two pathways by which this might happen (Figure 1). First, in an event we call "drift loss", the drive might be lost very early in the simulation through drift. This is a global loss of the drive and leads to immediate failure. The other process through which we might see loss of the drive is a founder event, in which a drive-free population is foundered into space recently cleared by the suppression drive. This only causes local loss of the drive but has several possible downstream outcomes. Local loss might lead to global loss in an event we call "wild-type escape". Here, the founder event leaves a wild-type population isolated in space from the drive population. The drive then goes on to eradicate the remainder of the wild-type population and then goes extinct itself, leaving the isolated population to recolonise

empty space. This outcome should be particularly likely if the drive wave is faster than the wild-type wave. Where the drive wave is slower than the wild-type wave, we often see "drive chasing" in which a foundered wild-type population is reinvaded by the drive, causing the drive wave to spread in both directions. In some parts of parameter space, this occurred only occasionally and so would lead to eradication as the drive corners the wild-type against the spatial boundaries of the population. In many parts of parameter space, however, founder events were frequent enough to allow the population to persist to $t = 1000$. For purposes of scoring our various simulations, we define chasing as the reinvasion of the drive following a founder-induced local loss. Foundering, or drive wave penetration, by wild-types has occurred if at least one wild-type individual moved at least 1σ units beyond the trailing edge of the gene drive wave (defined as the drive-carrying individual with the lowest value of x). These events and outcomes are depicted in Figure 1.

3 How often do we expect these suppression gene drives to fail? And why do they fail?

Across the parameter space we explored, we found that the suppression gene drives, X-shredder and W-shredder, failed to eradicate the WT population in more than half (50.7%) of our simulations. The X-shredder drive was notably more prone to failure (58%) than W-shredder (44%).

In Figure 2, we present the estimated probabilities of success and failure

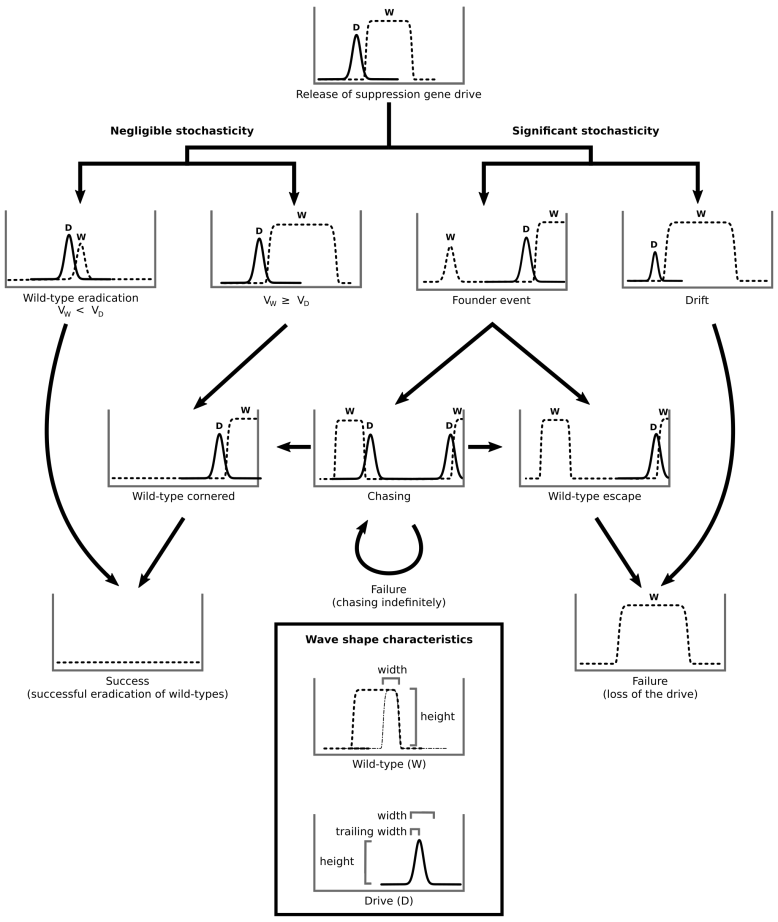


Figure 1: Events and outcomes of a spatially explicit individual-based stochastic simulation of suppression gene drives. Figure panels show population density through space for both wild-type (W) and gene drive (D) alleles at a snapshot in time. In a bounded 1-dimensional space, the simulation events which occur prior to the success or failure of the drive can be classified by characteristics of relative wave velocity, and the stochastic events of drift and foundering. Our simulations involve an absorbing boundary such that the wild-type population can be cornered by the drive wave. The box at the bottom illustrates how the wave shape characteristics, i.e., wave heights and widths were measured for the wild-type and drive waves.

of the gene drive across our $\{R_{\max}, \sigma\}$ parameter space, examining the effect of maximum female fecundity and dispersion parameter on failure probability. Failure becomes less likely as dispersion increases. This is expected because populations approach panmixia as dispersion increases (we are moving towards an aspatial system), but failures appear to become rare well before panmixis is achieved. The range of maximum female fecundity we tested appeared to have a weaker effect on the outcomes; failure tended to be more likely at both high and very low values of maximum growth rate. At no growth ($R_{\max} = 2$) failure is markedly more likely even at the upper range of the dispersion parameter. Otherwise, there is a slight tendency towards increased failure as growth rate increases. Drive failure is notably more common in X-shredder than W-shredder. These relationships are expressed concisely in Figure S1 as violin plots, where we assumed completely additive effects of the drive type, maximum female fecundity, and dispersion parameter.

These broad results raise a number of questions. These are drives that, in an aspatial simulation will send the population extinct. Why are they failing at such a high rate in this spatial setting? Why did growth rate or female fecundity only weakly affect the success of the drive? What properties differentiate W-shredder from X-shredder so much so that success rates between them are significantly different? In the following, we attempt to answer these questions and identify the characteristics of an ideal suppression gene drive. Finally, we put forth recommendations on key areas of study required before

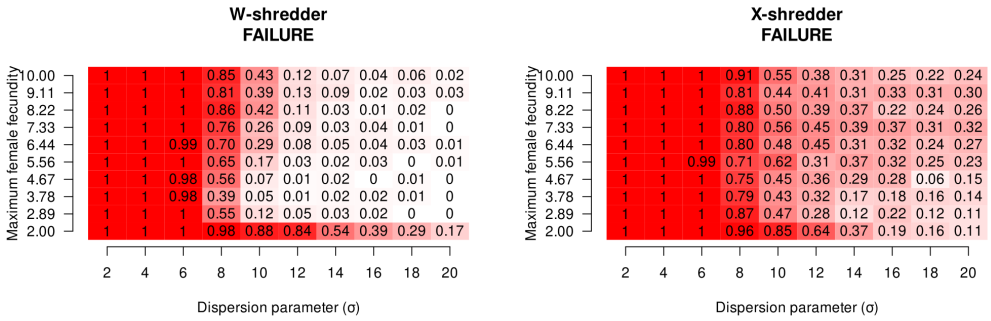


Figure 2: Estimated probability of failure of W-shredder and X-shredder suppression gene drives to eradicate the target population. Failure probability is shown as a function of the maximum female fecundity and dispersion parameters.

we can successfully use suppression gene drives to control invasive species, pests and diseases.

The fundamental reasons for the failure of suppression gene drives to eradicate the population can be broadly divided into two: indefinite chasing, and drive loss. Drive loss (i.e., loss of drive alleles despite the presence of wild-types) can occur through either drift early in the simulation or through wild-type escape following a later founder event (Figure 3). Drive loss must always result in the failure of the drive, and it occurred in $\sim 37\%$ of our simulations. The remaining failures we observed were from populations reaching 1000 generations with both wild-type and drive alleles still in coexistence. These failures are caused by the complex chasing dynamic in which founder events are sufficiently common that cornering never happens; chasing continues indefinitely. The necessary event for nearly all these failure types is chasing; an outcome that occurred in 68% of our simulations.

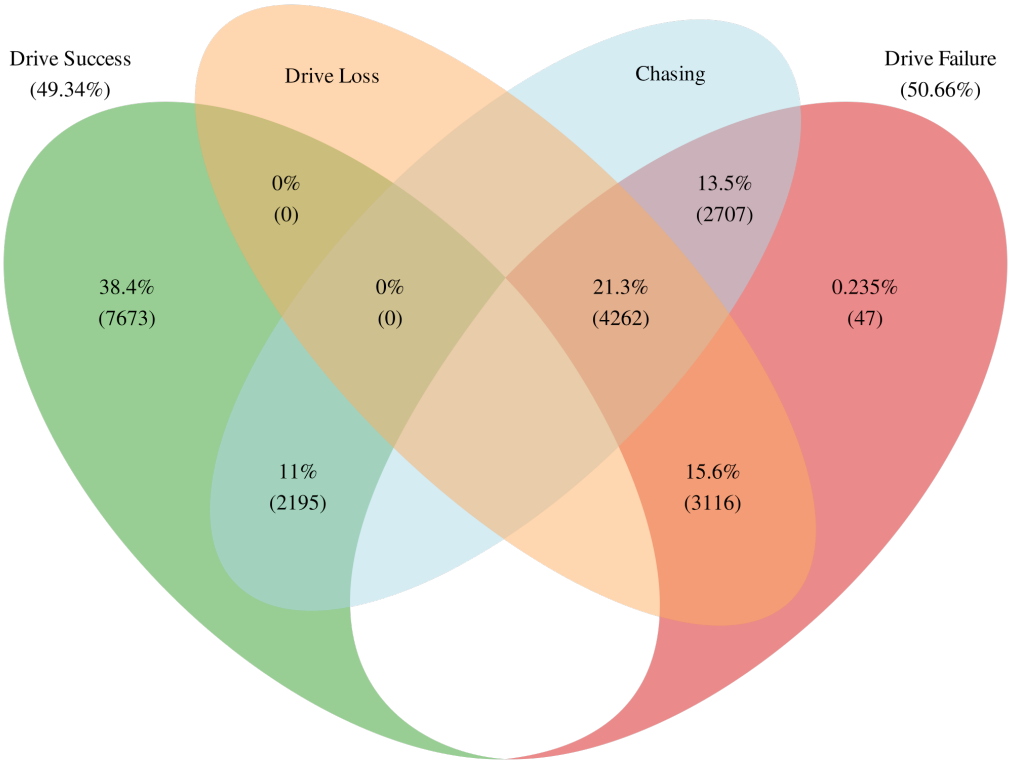


Figure 3: Frequencies of the simulation outcomes (suppression gene drive success and failure), and their intersections with the loss of suppression gene drive allele and chasing.

Under our definition, chasing occurs if the drive reinvades a foundered wild-type population. Where founder events are common, the process of extinction is continuously undermined by local wild-type escapes. This can lead to a dynamic in which global coexistence occurs despite complete local turnover of allele frequencies; a situation that occurred in 14% of our simulations. Deterministic models of suppression gene drives do not deliver this dynamic (Beaghton et al. 2016; Tanaka et al. 2017). It has, however, been observed in individual-based stochastic simulations in 2-dimensional space (Champer et al. 2021), and here we see it occur in a bounded 1-dimensional space also. It is clear that this global coexistence is a fundamentally stochastic phenomenon rooted in drift and foundering on the retreating population margin.

4 Relative wave velocity and chasing

Although we imagine a bounded system, allowing wildtypes to be cornered by the drive, it is clear that the relative speeds of the wild-type and drive waves are important. In particular, they are likely to be important following a founder-induced local loss of the drive. Under some assumptions, it is possible to derive the expected speeds from a deterministic model, and we examine this now to see what insight can be drawn. The mathematical models used in these derivations are deterministic. Drive loss due to drift and foundering cannot occur. Further, the drive wave also assumes homogeneous population

density (i.e., it does not account for the asymmetric gene flow caused by suppression and so will be an upper bound on the drive speed). We derive these analytic speeds and compare our results to stochastic simulations in which we measured the wave velocities of the wild-types and drive-carriers.

4.1 Velocities of WT and drive waves

We described population density and allele frequency dynamics across time and 1-dimensional space. We derived the velocities of the wild-types invading empty space, and the suppression gene drive alleles invading the wild-type population using a system of partial differential equations (PDE). To define these PDEs, we derived the recursive allele frequency functions, $q_{t+1} = f(q_t)$ for W-shredder and X-shredder, as well as the recursive population density function, $n_{t+1} = g(n_t)$. These functions were used to define the PDEs assuming overlapping generations, i.e., $\frac{dq}{dt} = \Delta q = q_{t+1} - q_t$ and $\frac{dN}{dt} = \Delta N = N_{t+1} - N_t$. The advection term (Girardin et al. 2019; Girardin and Débarre 2021) in the PDE for drive allele frequency was ignored because it massively complicates the calculation of velocity. Ignoring this term is effectively ignoring asymmetrical gene flow caused by the density gradient arising from suppression and so in this case yields a velocity that is higher than the velocity that would be realised if we accounted for asymmetrical gene flow. Without the advection term we can calculate velocities using Fisher's definition of asymptotic spreading speed (Fisher 1937; Lewis et al. 2016).

We derived the recursive suppression gene drive allele frequencies, q_{t+1} for W-shredder to be:

$$q_{t+1} = \frac{8q_t - 3cq_t - cq_t^2}{8 - 4c}, \tag{3}$$

and for X-shredder it is:

$$q_{t+1} = \frac{2q_t}{2 - c + cq_t}, \tag{4}$$

where c is the chromosomal shredding efficiency of the gene drive (refer to Supplementary Information for details of the derivations, i.e., subsections "Derivation of W-shredder equations" and "Derivation of X-shredder equations").

We derived the recursive equation for population density, N_{t+1} to be:

$$N_{t+1} = N_t \phi_t b_t, \tag{5}$$

where the frequency of females at generation t is $\phi_t = \frac{2-c-cq_t}{4-2c}$, and the female fecundity, b_t is the Beverton-Holt growth function (equation (2)). These deterministic recursive equations approximate the output of the stochastic simulations under approximate panmixia (Figure 4). However, the deterministic equations have a tendency to overestimate the rates of decrease in population size and the rates of increase in drive allele frequencies. This is largely what we would expect given the lack of accounting for advection, but Figure 4 gives us a sense that the bias is mild relative to the variance

introduced by stochasticity.

To derive the asymptotic gene drive and wild-type velocities, we assumed overlapping generations yielding the gene drive allele frequency differential for W-shredder as:

$$\frac{dq}{dt} = \frac{cq(1-q)}{8-4c}, \tag{6}$$

and for X-shredder as:

$$\frac{dq}{dt} = \frac{cq(1-q)}{2-c+cq}. \tag{7}$$

When we assume overlapping generations, the Beverton-Holt function is analogous to the logistic growth function (Bohner and Warth 2007):

$$\frac{dN}{dt} = rN\left(1-\frac{N}{K}\right), \tag{8}$$

where r is the intrinsic growth rate (also called the low-density growth rate), and K is the carrying capacity (the equilibrium density of individuals along the spatial x-axis, in the absence of suppression and migration).

So far, these equations (equations (3) to (8)) are not spatially explicit. To account for the spatial dynamics, we formulated the PDE for population density $\frac{\partial N}{\partial t}$ in 1-dimensional space to be:

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2} + \text{growth rate} + \text{suppression rate}, \tag{9}$$

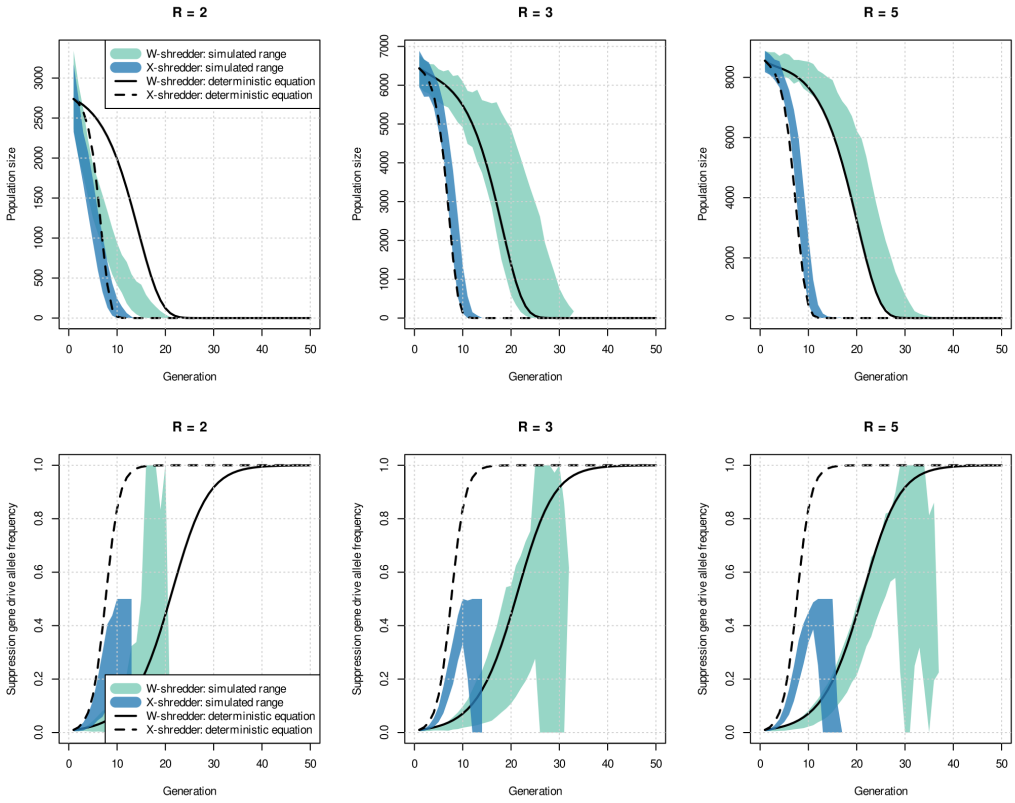


Figure 4: Time-series plots of total population size and suppression gene drive allele frequency across generations using the stochastic simulation data and the deterministic recursive equations for W-shredder and X-shredder suppression gene drives. R is the maximum female fecundity, and the dispersion parameter was set to $\sigma = 10$.

where:

$$\text{growth rate} = \frac{dN}{dt}, \tag{10}$$

D is the diffusion coefficient. We can calculate the velocity by examining parts of space a long way from the invasion front, where $\frac{N}{K} \approx 0$ (low density) and suppression rate ≈ 0, then the PDE simplifies to:

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2} + rN. \tag{11}$$

The asymptotic velocity (as $t \to \infty$) of the wild-type individuals in occupying the open space is therefore:

$$v_{\text{wild type}} = 2\sqrt{Dr}. \tag{12}$$

Similar logic can be used to derive the asymptotic velocity of the invading suppression gene drive, but in this case we need to make an additional assumption that the density gradient resulting from the suppression drive is effectively zero in the vicinity of the leading edge of the gene drive invasion. Ignoring the advection term effectively gives us an upper bound on the gene drive velocity of:

$$\widehat{v}_{\text{gene drive allele}} = 2\sqrt{Dm}, \tag{13}$$

where the fitness of W-shredder is:

$$
\begin{aligned}
m_{\text{W-shredder}} &= \frac{dq}{dt}(q(1-q))^{-1} \\
&= \frac{c}{8-4c},
\end{aligned}
\tag{14}
$$

and the fitness of X-shredder is:

$$
\begin{aligned}
m_{\text{X-shredder}} &= \frac{dq}{dt}(q(1-q))^{-1} \\
&= \frac{c}{2-c+cq}.
\end{aligned}
\tag{15}
$$

This yields the upper bound velocity of the W-shredder drive as:

$$
\widehat{v}_{\text{W-shredder}} = 2\sqrt{\frac{D}{4}}, \tag{16}
$$

and the upper bound velocity of the X-shredder assuming $q \approx 0$ (i.e., very low frequency at the wave front) as:

$$
\widehat{v}_{\text{X-shredder}} = 2\sqrt{D}. \tag{17}
$$

Thus, we find that the wild-types are faster than the suppression gene drives at $r > \frac{1}{4}$ for W-shredder, and at $r > 1$ for X-shredder. Mapping the low-density growth rate, r into the non-overlapping case, assuming non-skewed sex ratio, $\phi = 1/2$ and low-density such that $b = R_{\max}$, then:

$$\begin{aligned} \frac{dN}{dt} &= \frac{\Delta N}{\Delta t} \\ &\rightarrow rN = N\phi b - N \\ &\rightarrow rN = \frac{NR_{\max}}{2} - N \\ &\rightarrow R_{\max} = 2(r+1). \end{aligned} \tag{18}$$

Hence, if the maximum female fecundity, $R_{\max}$ is greater than 2.5 for W-shredder and greater than 4 for X-shredder then the wild-type wave is expected to be faster than gene drive. We note that these critical values of $R_{\max}$ do not appear to have a major impact on the probability of failure when viewed across all simulation (Fig. 2).

4.2 Penetrating the wave of invading drives

While the relative velocity of the drive wave is reasonably easy to identify, the conditions promoting penetration through foundering are less clear. The likelihood of a founder event causing local loss of drive alleles might depend upon both the velocity and shape of the gene drive wave front. All else being equal, a drive front that is slower than the wild-type, and very narrow should be more easily penetrated than one that is fast and wide. Indeed, we observed this in our simulations (Figure 5 top-left and top-centre panels): the probability of wild-types penetrating the drive wave (simply referred to as "penetration") was negatively correlated with the relative wave velocity

($v_{\text{drive}} - v_{\text{WT}}$) and the width of the drive wave relative to the dispersion parameter (width/σ). By contrast, the wave height is not as important.

The violin plots in Figure 5 bottom panels show that the X-shredder wave is faster than the W-shredder wave (as expected from their respective drive fitnesses: eqns 14, 15); however, X-shredder has thinner and shorter waves than W-shredder. Despite being slower, W-shredder is less likely to result in drive failure after penetration (Supplementary Figure S2 top left violin plot). This suggests that in bounded space, the ability of the drive wave to resist penetration is a more significant factor to drive success than the velocity of the drive wave. The thicker wave of W-shredder compared with that of X-shredder is because $m_{\text{X-shredder}} > m_{\text{W-shredder}}$, and assuming a standard Fisher wave, the wave width is predicted to be $\sqrt{D/m}$ (Fisher 1937). Counter-intuitively, then, the higher fitness of the X-shredder drive explains its overall higher failure rate, through higher rates of penetration.

Naively, we might expect penetration to be unaffected by higher dispersion: an increase in dispersion might be expected to affect the wild type and drive wave widths equally. But this is not what we observed. Indeed the probability of penetration drops rapidly as dispersion increases, and this occurs at values of σ well below those approaching panmixia (Supplementary Figure S3, left plot). Assuming a Fisher wave, the wild type wave should have a width equal to $\sqrt{D/r}$, where $D \propto \sigma^2$. Ignoring advection, the drive wave should have a width equal to $\sqrt{D/m}$. That is, we would naively expect both waves to scale linearly with σ. To examine this, we used a log-log

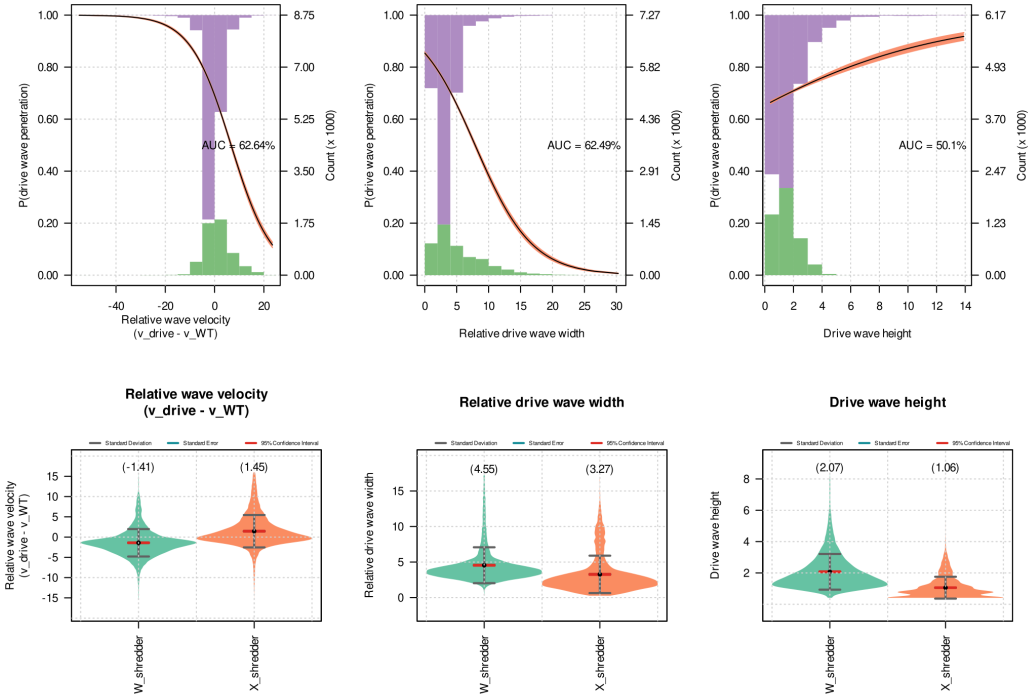


Figure 5: Suppression gene drive wave characteristics. Top panels: logistic regression plots showing the relationship between the probability of wild-types penetrating the wave of suppression gene drive-carriers and the three wave characteristics, i.e., relative wave velocity (drive velocity - WT velocity), width of the drive wave relative to the dispersion parameter, σ (i.e., width/σ), and height of the drive wave. Bottom panels: violin plots of the wave characteristics of W-shredder and X-shredder suppression gene drives.

regression to estimate the scaling exponent for each wave width against σ. We estimated the scaling exponent for the wild-type wave width to be 0.95 ($\pm$0.0021); sub-linear, but approximately linear, as expected. By contrast, our simulations yielded scaling exponents of 1.12 ($\pm$0.0084) for the full width of the drive wave: 1.09 ($\pm$0.0132) for the width of the leading half, and 1.21 ($\pm$0.0129) for the width of the trailing half (refer to Figure 6). This means that, relative to the wild type, the drive wave gets disproportionately wider as dispersion increases (Supplementary Figure S3, right plot). This result is almost certainly related to advection, i.e., the asymmetric gene flow caused by the density gradient on the trailing edge of the population, and likely explains the rapid drop in penetration probability with increasing dispersion. In our model, σ sets the scale of dispersal relative to local dynamics; our results suggest that when the scale of dispersal is more than an order of magnitude greater than the scale of local dynamics, penetration rapidly becomes an unlikely event. The likely explanation for this transition is the super-linear scaling of the drive width with σ. Whether our observation here is general, or a manifestation of the particular parameter space we examined, remains to be seen, but our results point to a fruitful avenue for further analytical exploration.

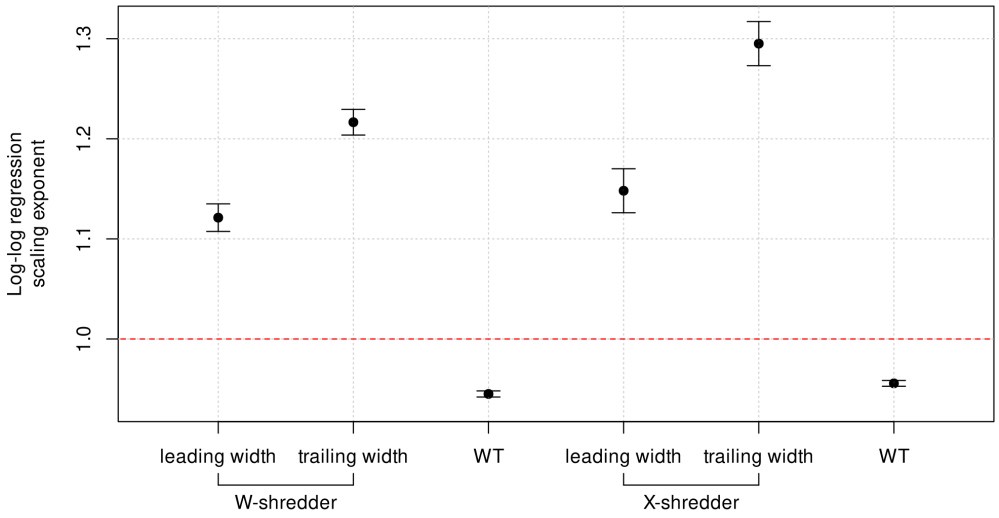


Figure 6: Scaling exponents for the width of the wild-type wave, as well as the trailing and leading widths of the suppression gene drive waves. Estimated from log-log regression against the dispersion parameter, σ (i.e., $ln(\text{width}) = 1 + ln(\sigma)$). Scaling exponents are shown as points, while the corresponding standard errors depicted as error bars.

5 Drive loss

Loss of the suppression gene drive allele is a major reason why suppression gene drives fail to eradicate the target population, occurring in 37% of our simulations. Drive loss results from drift and founder events, both of which are stochastic processes whose strength scales with population size. The probability of drive allele extinction through drift, P(drive loss | no penetration) is likely a function of the population size and the introduction frequency. In our simulations, we largely avoided this outcome by using a relatively large number of introduced drive-carriers (1% of the wild-type population size). As a consequence, early loss from drift occurred in only 16 simulations. Instead, most of the drive loss events we recorded occurred after penetration events. We would expect the probability of successful wild-type escape following founder events, P(drive loss | penetration) to be a function of population size as well as the velocity and width of the drive wave.

5.1 Drift: P(drive loss | no penetration)

Drift is the loss of an allele purely by chance. Its probability scales negatively with the increase in population size. Small populations are more likely to lose an allele than large ones. In the case of suppression gene drives, the introduction frequency of the drive allele should be proportional to the likelihood of its fixation or loss. Low introduction frequency is more likely to result in drive loss than high initial frequency. We confirmed this using our

simulations by varying the equilibrium population density, N^* from 2 to 5 (refer to Supplementary Figure S5).

The effect of drift suggests that species or populations with high inbreeding to outbreeding ratio (*e.g.*, self-pollinating plants) are less likely to incorporate a newly introduced suppression gene drive allele unless the introduction frequency is a significant fraction of target population. Regardless of the inbreeding ratio, however, the density of a population plummets along the edge of the wave. Hence, populations are subject to stronger drift along its borders. It follows that deterministic reaction-diffusion-advection models have limited applicability in understanding these boundaries where we expect high stochasticity.

5.2 Wild-type escape: P(drive loss | penetration)

Founder events are the spatial equivalent of drift, and can lead to a drive wave being penetrated by a founder population composed entirely of wildtypes. In many cases, this penetration is followed by wild type escape in which local loss of the drive through penetration becomes global loss following the extinction of the population in other parts of space. Similar to its precursor event (penetration), escape likely depends on the velocity and the shape of the drive wave. In contrast to penetration, however, escape is more likely if the drive wave is faster than the wild-type wave (Figure 7 left plot). This stands to reason: if the drive wave is faster than the wild-type, then a newly formed wild-type population is unlikely to overtake the drive wave and so

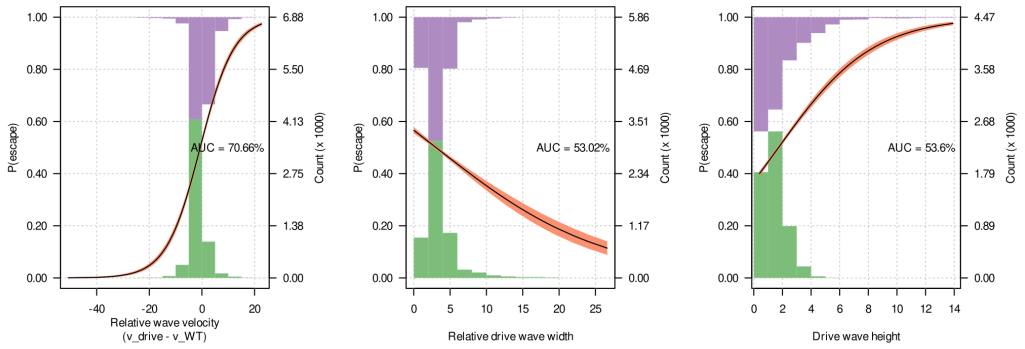


Figure 7: Logistic regression plot of the relationship between the probability of wild-type escape (i.e., the probability of drive loss after drive wave penetration by wild-types) and the three wave characteristics, i.e., relative wave velocity (drive velocity - WT velocity), width of the drive wave relative to the dispersion parameter, σ (i.e., width/σ), and height of the drive wave.

be reinvaded by it. Since faster drives generate thinner waves which are easier to penetrate than slower thicker ones, and given that 99.8% of the simulations which concluded with the loss of the gene drive allele were due to wild-type escape; it follows that slower drives are generally more effective; they decrease both the probability of penetration, and the probability of subsequent wildtype escape.

Of course, after a penetration event, a complex chasing dynamics might ensue, which over multiple generations generates a complex landscape with multiple wild-type and drive waves growing, merging, dying, and interacting with each other. During these scenarios, measuring wave characteristics became intractable after multiple wild-type and drive waves emerged. Exploring and discovering avenues to accurately measure these wave characteristics and interactions is an area we believe also worthy of future work.

6 Conclusion and recommendation

Our simulations are in agreement with those of others (Champer et al. 2021): there is significant risk that suppression gene drives will fail in real-world settings through spatial and stochastic processes. Previous stochastic models have revealed these processes in 2-dimensional spaces; we show that they are also present in the simplest 1-dimensional space. Our simpler 1-dimensional space allows increased clarity as to the events and possible outcomes that emerge. It also allows us to focus on the fundamental aspects of the invasion waves: their width, height, and speed, that are more difficult to measure in higher-dimensional spaces. We show that these aspects of the drive and wildtype waves have a strong bearing on outcomes, but in a complex way. Overall, and somewhat counter-intuitively, it is clear that slower and wider drive waves are substantially more effective than fast, steep waves. This contrasts very clearly with an aspatial view, in which faster (i.e. fitter) drives will effect eradication more rapidly, and with lower chance of resistance evolving.

Our results also point to an important role for dispersal in mediating outcomes. In our simulations (and those of others; Champer et al. 2021), higher dispersal rates led to lower failure probability, and this occurred well below dispersal rates that would be consistent with panmixis. Closer examination revealed that this result is driven by a different scaling between dispersal rate and the widths of wildtype and drive waves. Drive wave width appears

to scale super-linearly with dispersal, whereas wildtype waves are linear or sub-linear. Thus, for a given increase in dispersal, drive waves get wider disproportionately faster than wildtype waves. The precise reason for this remains unknown, but likely rests with the advection and asymmetric gene flow that emerges along the trailing edge of the drive wave. In the meantime, it seems that many of the spatial causes of drive failure (Figure 1) become negligible for species in which dispersal distances are very large relative to the scale of local dynamics.

Given both the promise and risk of suppression gene drives, their dynamics are worthy of careful examination. Since most populations exhibit spatial structure, finite population sizes, and bounded ranges these considerations are important in the overall design of a drive system. Our results very clearly demonstrate that a system optimised assuming an aspatial scenario is quite likely a long way from optimal in a spatially explicit scenario. In parallel with solving the molecular genetic challenges in developing suppression gene drive constructs in model and non-model species, we also need to study novel gene drive systems under stochastic and spatially explicit scenarios. Doing so can very quickly make a promising candidate (for example X-shredder) appear far from promising. Ongoing work in this space will clearly benefit from close collaboration between geneticists, evolutionary biologists, and ecologists.

7 Code and data availability

The code and the simulation output (.Rds R data format) are available in a public GitHub repository: drivechasingecol.

8 Bibliography

References

Anderson, C. J. et al. (2016). "Population structure and gene flow in the global pest, Helicoverpa armigera". In: *Molecular Ecology* 25.21. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.13841, pp. 5296–5311. ISSN: 1365-294X. DOI: 10.1111/mec.13841. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13841 (visited on 12/09/2021).

Beaghton, Andrea, Pantelis John Beaghton, and Austin Burt (Apr. 2016). "Gene drive through a landscape: Reaction-diffusion models of population suppression and elimination by a sex ratio distorter". In: *Theor Popul Biol* 108, pp. 51–69. ISSN: 1096-0325. DOI: 10.1016/j.tpb.2015.11.005.

Bohner, Martin and Howard Warth (Aug. 2007). "The Beverton–Holt dynamic equation". In: *Applicable Analysis* 86.8, pp. 1007–1015. ISSN: 0003-6811, 1563-504X. DOI: 10.1080/00036810701474140. URL: http://www.tandfonline.com/doi/abs/10.1080/00036810701474140 (visited on 07/07/2021).

Bradshaw, Corey J. A. et al. (July 29, 2021). "Detailed assessment of the reported economic costs of invasive species in Australia". In: *NeoBiota* 67. Publisher: Pensoft Publishers, pp. 511–550. ISSN: 1314-2488. DOI: 10.3897/neobiota.67.58834. URL: https://neobiota.pensoft.net/article/58834/ (visited on 10/25/2021).

Burt, Austin and Anne Deredec (July 25, 2018). "Self-limiting population genetic control with sex-linked genome editors". In: *Proc Biol Sci* 285.1883, p. 20180776. ISSN: 0962-8452. DOI: 10.1098/rspb.2018.0776. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6083257/ (visited on 07/18/2021).

Champer, Jackson et al. (Feb. 27, 2020a). "A toxin-antidote CRISPR gene drive system for regional population modification". In: *Nat Commun* 11.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: CRISPR-Cas9 genome editing;Genetic engineering;Population genetics;Protein design Subject_term_id: cas9-endonuclease;genetic-engineering;population-genetics;protein-design, p. 1082. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14960-3. URL: https://www.nature.com/articles/s41467-020-14960-3 (visited on 07/01/2021).

Champer, Jackson et al. (Mar. 12, 2020b). "Performance analysis of novel toxin-antidote CRISPR gene drive systems". In: *BMC Biology* 18.1, p. 27. ISSN: 1741-7007. DOI: 10.1186/s12915-020-0761-2. URL: https://doi.org/10.1186/s12915-020-0761-2 (visited on 07/01/2021).

— (Feb. 2021). "Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles". In: *Mol Ecol* 30.4, pp. 1086–1101. ISSN: 0962-1083, 1365-294X. DOI: 10.1111/mec.15788. URL: https://onlinelibrary.wiley.com/doi/10.1111/mec.15788 (visited on 07/02/2021).

Chang, Nannan et al. (Apr. 2013). "Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos". In: *Cell Res* 23.4, pp. 465–472. ISSN: 1001-0602. DOI: 10.1038/cr.2013.45. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3616424/ (visited on 12/08/2021).

Deredec, Anne, Austin Burt, and H. C. J. Godfray (Aug. 1, 2008). "The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management". In: *Genetics* 179.4. Publisher: Genetics Section: Investigations, pp. 2013–2026. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.108.089037. URL: https://www.genetics.org/content/179/4/2013 (visited on 04/05/2021).

DiCarlo, James E. et al. (Apr. 2013). "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems". In: *Nucleic Acids Res* 41.7, pp. 4336–4343. ISSN: 0305-1048. DOI: 10.1093/nar/gkt135. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627607/ (visited on 12/08/2021).

Fisher, R. A. (1937). "The Wave of Advance of Advantageous Genes". In: *Annals of Eugenics* 7.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1937.tb02153.x, pp. 355–369. ISSN: 2050-1439. DOI: https://doi.

org/10.1111/j.1469-1809.1937.tb02153.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1937.tb02153.x (visited on 04/14/2021).

Friedland, Ari E. et al. (Aug. 2013). "Heritable genome editing in C. elegans via a CRISPR-Cas9 system". In: *Nat Methods* 10.8, pp. 741–743. ISSN: 1548-7091. DOI: 10.1038/nmeth.2532. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3822328/ (visited on 12/08/2021).

Galizi, Roberto et al. (Aug. 3, 2016). "A CRISPR-Cas9 sex-ratio distortion system for genetic control". In: *Sci Rep* 6.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Biotechnology;Genetics Subject_term_id: biotechnology;genetics, p. 31139. ISSN: 2045-2322. DOI: 10.1038/srep31139. URL: https://www.nature.com/articles/srep31139 (visited on 07/18/2021).

Girardin, Léo, Vincent Calvez, and Florence Débarre (Dec. 1, 2019). "Catch Me If You Can: A Spatial Model for a Brake-Driven Gene Drive Reversal". In: *Bull Math Biol* 81.12, pp. 5054–5088. ISSN: 1522-9602. DOI: 10.1007/s11538-019-00668-z. URL: https://doi.org/10.1007/s11538-019-00668-z (visited on 06/24/2021).

Girardin, Léo and Florence Débarre (Jan. 27, 2021). "Demographic feedbacks can hamper the spatial spread of a gene drive". In: *arXiv:2101.11255 [math, q-bio]*. arXiv: 2101.11255. URL: http://arxiv.org/abs/2101.11255 (visited on 07/01/2021).

Hammond, Andrew et al. (July 28, 2021). "Gene-drive suppression of mosquito populations in large cages as a bridge between lab and field". In: *Nat Commun* 12.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Behavioural ecology;Environmental biotechnology;Population dynamics;Synthetic biology Subject_term_id: behavioural-ecology;environmental-biotechnology;population-dynamics;synthetic-biology, p. 4589. ISSN: 2041-1723. DOI: 10.1038/s41467-021-24790-6. URL: https://www.nature.com/articles/s41467-021-24790-6 (visited on 12/08/2021).

Holman, Luke (Oct. 9, 2019). "Evolutionary simulations of Z-linked suppression gene drives". In: *Proceedings of the Royal Society B: Biological Sciences* 286.1912. Publisher: Royal Society, p. 20191070. DOI: 10.1098/rspb.2019.1070. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2019.1070 (visited on 04/06/2021).

Horvath, Philippe and Rodolphe Barrangou (June 2013). "RNA-guided genome editing à la carte". In: *Cell Res* 23.6, pp. 733–734. ISSN: 1001-0602. DOI: 10.1038/cr.2013.39. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3674385/ (visited on 12/08/2021).

Hough, Rupert Lloyd (Apr. 2021). "A world view of pesticides". In: *Nat. Geosci.* 14.4. Number: 4 Publisher: Nature Publishing Group, pp. 183–184. ISSN: 1752-0908. DOI: 10.1038/s41561-021-00723-2. URL: https:

//www.nature.com/articles/s41561-021-00723-2 (visited on 02/03/2022).

Jones, Christopher M. et al. (2019). "Movement Ecology of Pest Helicoverpa: Implications for Ongoing Spread". In: *Annual Review of Entomology* 64.1. _eprint: https://doi.org/10.1146/annurev-ento-011118-111959, pp. 277–295. DOI: 10.1146/annurev-ento-011118-111959. URL: https://doi.org/10.1146/annurev-ento-011118-111959 (visited on 12/09/2021).

Kearney, Michael et al. (2008). "Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates". In: *Ecography* 31.4. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0906-7590.2008.05457.x, pp. 423–434. ISSN: 1600-0587. DOI: 10.1111/j.0906-7590.2008.05457.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0906-7590.2008.05457.x (visited on 09/03/2021).

Kyrou, Kyros et al. (Nov. 2018). "A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes". In: *Nat Biotechnol* 36.11. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 11 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Genetic engineering;Genetic techniques Subject_term_id: genetic-engineering;genetic-techniques, pp. 1062–1066. ISSN: 1546-1696. DOI: 10.1038/nbt.4245. URL: https://www.nature.com/articles/nbt.4245 (visited on 07/15/2021).

Lewis, Mark, Sergei V. Petrovskii, and Jonathan Potts (2016). *The Mathematics Behind Biological Invasions*. Interdisciplinary Applied Mathematics, Mathematics of Planet Earth Collection. Springer International Publishing. ISBN: 978-3-319-32042-7. DOI: 10.1007/978-3-319-32043-4. URL: https://www.springer.com/gp/book/9783319320427 (visited on 07/01/2021).

Macias, Vanessa M., Johanna R. Ohm, and Jason L. Rasgon (Sept. 2, 2017). "Gene Drive for Mosquito Control: Where Did It Come from and Where Are We Headed?" In: *Int J Environ Res Public Health* 14.9, E1006. ISSN: 1660-4601. DOI: 10.3390/ijerph14091006.

Meccariello, Angela et al. (Apr. 16, 2021). "Engineered sex ratio distortion by X-shredding in the global agricultural pest Ceratitis capitata". In: *BMC Biology* 19.1, p. 78. ISSN: 1741-7007. DOI: 10.1186/s12915-021-01010-7. URL: https://doi.org/10.1186/s12915-021-01010-7 (visited on 07/18/2021).

Miao, Jin et al. (Oct. 2013). "Targeted mutagenesis in rice using CRISPR-Cas system". In: *Cell Res* 23.10, pp. 1233–1236. ISSN: 1001-0602. DOI: 10.1038/cr.2013.123. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3790239/ (visited on 12/08/2021).

Noble, Charleston et al. (June 19, 2018). "Current CRISPR gene drive systems are likely to be highly invasive in wild populations". In: *eLife* 7. Ed. by Michael Doebeli. Publisher: eLife Sciences Publications, Ltd, e33423.

731 ISSN: 2050-084X. DOI: 10.7554/eLife.33423. URL: https://doi.org/
732 10.7554/eLife.33423 (visited on 12/08/2021).
733 Phillips, Benjamin L. et al. (Feb. 2006). "Invasion and the evolution of speed
734 in toads". In: *Nature* 439.7078. Bandiera_abtest: a Cg_type: Nature Re-
735 search Journals Number: 7078 Primary_atype: Research Publisher: Na-
736 ture Publishing Group, pp. 803–803. ISSN: 1476-4687. DOI: 10.1038/
737 439803a. URL: https://www.nature.com/articles/439803a (visited
738 on 09/03/2021).
739 Prowse, Thomas A. A. et al. (Aug. 16, 2017). "Dodging silver bullets: good
740 CRISPR gene-drive design is critical for eradicating exotic vertebrates".
741 In: *Proceedings of the Royal Society B: Biological Sciences* 284.1860. Pub-
742 lisher: Royal Society, p. 20170799. DOI: 10.1098/rspb.2017.0799. URL:
743 https://royalsocietypublishing.org/doi/10.1098/rspb.2017.
744 0799 (visited on 07/01/2021).
745 Shen, Bin et al. (May 2013). "Generation of gene-modified mice via Cas9/RNA-
746 mediated gene targeting". In: *Cell Res* 23.5, pp. 720–723. ISSN: 1001-0602.
747 DOI: 10.1038/cr.2013.46. URL: https://www.ncbi.nlm.nih.gov/
748 pmc/articles/PMC3641603/ (visited on 12/08/2021).
749 Simoni, Alekos et al. (Sept. 2020). "A male-biased sex-distorter gene drive
750 for the human malaria vector Anopheles gambiae". In: *Nat Biotechnol*
751 38.9. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research
752 Journals Number: 9 Primary_atype: Research Publisher: Nature Publish-
753 ing Group Subject_term: Gene regulation;Genetics;Molecular engineering

Subject_term_id: gene-regulation;genetics;molecular-engineering, pp. 1054–1060. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0508-1. URL: https://www.nature.com/articles/s41587-020-0508-1 (visited on 07/18/2021).

Sinkins, Steven P. and Fred Gould (June 2006). "Gene drive systems for insect disease vectors". In: *Nat Rev Genet* 7.6, pp. 427–435. ISSN: 1471-0056. DOI: 10.1038/nrg1870.

Slatkin, Montgomery and Laurent Excoffier (May 1, 2012). "Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift". In: *Genetics* 191.1. Publisher: Genetics Section: Investigations, pp. 171–181. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.112.139022. URL: https://www.genetics.org/content/191/1/171 (visited on 11/18/2021).

Tanaka, Hidenori, Howard A. Stone, and David R. Nelson (Aug. 8, 2017). "Spatial gene drives and pushed genetic waves". In: *PNAS* 114.32. Publisher: National Academy of Sciences Section: Physical Sciences, pp. 8452–8457. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1705868114. URL: https://www.pnas.org/content/114/32/8452 (visited on 04/12/2021).

Urban, Mark C. et al. (Mar. 1, 2008). "A Toad More Traveled: The Heterogeneous Invasion Dynamics of Cane Toads in Australia." In: *The American Naturalist* 171.3. Publisher: The University of Chicago Press, E134–E148. ISSN: 0003-0147. DOI: 10.1086/527494. URL: https://www.journals.uchicago.edu/doi/full/10.1086/527494 (visited on 12/09/2021).

Yu, Zhongsheng et al. (Sept. 2013). "Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in Drosophila". In: *Genetics* 195.1, pp. 289–291. ISSN: 0016-6731. DOI: 10.1534/genetics.113.153825. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761309/ (visited on 12/08/2021).

9 Supplementary Information

9.1 Derivation of W-shredder equations

In ZW/ZZ sex determination systems, *e.g.*, birds, reptiles, and lepidopteran pests, where ZW are females and ZZ are males.

$$Z^dW - c \to Z^d \tag{19}$$

Let:

- $P(\text{shredded W} \mid \text{W shredding}) = c$, i.e., the probability of shredding the W chromosome given W-shredding be c;
- $P(\text{intact W} \mid \text{W shredding}) = 1 - c$, i.e., the probability of intact W chromosome given W-shredding be $1 - c$;
- $P(\text{shredded Z} \mid \text{W shredding}) = 0$, i.e., the Z chromosome be unaffected by the W-shredding;
- $P(\text{intact Z} \mid \text{W shredding}) = 1$, i.e., the Z chromosome be intact;

- f be female;
- m be male;
- $P(d) = P(d \mid m) = P(d \mid f) = q$, i.e., the probability of the drive allele in the Z chromosome be q;
- $P(D) = P(D \mid m) = P(D \mid f) = 1 - q$, i.e., the probability of the wild-type or non-drive allele in the Z chromosome be $1 - q$.

Probability of Z and W chromosomes per genotype and per sex

Female sex chromosomes:

$$P(W \mid Z^dW \mid f) = \frac{P(\text{intact W} \mid \text{W shredding})}{P(\text{intact W} \mid \text{W shredding}) + P(\text{intact Z} \mid \text{W shredding})} = \frac{1-c}{2-c} \tag{20}$$

$$P(Z \mid Z^dW \mid f) = \frac{P(\text{intact Z} \mid \text{W shredding})}{P(\text{intact W} \mid \text{W shredding}) + P(\text{intact Z} \mid \text{W shredding})} = \frac{1}{2-c} \tag{21}$$

$$P(W \mid ZW \mid f) = 1/2 \tag{22}$$

$$P(Z \mid ZW \mid f) = 1/2 \tag{23}$$

$$
\begin{aligned}
P(W \mid f) &= P(d \mid f)\ P(W \mid Z^dW \mid f) + P(D \mid f)\ P(W \mid ZW \mid f) \\
&= q\frac{1-c}{2-c} + (1-q)\frac{1}{2} = \frac{2-c-cq}{4-2c}
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
P(Z \mid f) &= P(d \mid f)\ P(Z \mid Z^dW \mid f) + P(D \mid f)\ P(Z \mid ZW \mid f) \\
&= q\frac{1}{2-c} + (1-q)\frac{1}{2} = \frac{2-c+cq}{4-2c}
\end{aligned}
\tag{25}
$$

Male sex chromosomes:

$$
\begin{aligned}
P(\text{W chromosome} \mid ZZ \mid m) &= P(\text{W chromosome} \mid ZZ^d \mid m) \\
&= P(\text{W chromosome} \mid Z^dZ \mid m) \\
&= P(\text{W chromosome} \mid Z^dZ^d \mid m) \\
&= 0
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
P(\text{Z chromosome} \mid ZZ \mid m) &= P(\text{Z chromosome} \mid ZZ^d \mid m) \\
&= P(\text{Z chromosome} \mid Z^dZ \mid m) \\
&= 1
\end{aligned}
\tag{27}
$$

$$
P(\text{W chromosome} \mid m) = 0 \tag{28}
$$

$$
P(\text{Z chromosome} \mid m) = 1 \tag{29}
$$

Probability of offspring genotypes at the drive locus in the Z chromosome

Female genotypes:

$$
\begin{aligned}
P(D0) = P(Z^D W) &= P(D \cap Z \mid m) \times P(\text{W chromosome} \mid f) \\
&= P(D \mid m) \times P(\text{Z chromosome} \mid m) \times P(\text{W chromosome} \mid f) \\
&= (1-q) \times 1 \times \frac{2-c-cq}{4-2c} = \frac{2-c-2q+cq^2}{4-2c}
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
P(d0) = P(Z^d W) &= P(d \cap Z \mid m) \times P(W \mid f) \\
&= P(d \mid m) \times P(Z \mid m) \times P(\text{W chromosome} \mid f) \\
&= q \times 1 \times \frac{2-c-cq}{4-2c} \\
&= \frac{2q-cq-cq^2}{4-2c}
\end{aligned} \tag{31}
$$

805 Male genotypes:

$$
\begin{aligned}
P(DD) = P(Z^D Z^D) &= P(D \cap Z \mid m) \times P(D \cap Z \mid f) \\
&= P(D \mid m) \times P(Z \mid m) \times P(D \mid f) \times P(Z \mid ZW \mid f) \\
&= (1-q) \times 1 \times (1-q) \times 1/2 \\
&= \frac{1-2q+q^2}{2}
\end{aligned} \tag{32}
$$

$$
\begin{aligned}
P(Dd) = P(Z^D Z^d) &= P(D \cap Z \mid m) \times P(d \cap Z \mid f) \\
&= P(D \mid m) \times P(Z \mid m) \times P(d \mid f) \times P(Z \mid Z^d W \mid f) \\
&= (1-q) \times 1 \times q \times \frac{1}{2-c} \\
&= \frac{q-q^2}{2-c}
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
P(dD) = P(Z^d Z^D) &= P(d \cap Z \mid m) \times P(D \cap Z \mid f) \\
&= P(d \mid m) \times P(Z \mid m) \times P(D \mid f) \times P(Z \mid ZW \mid f) \\
&= q \times 1 \times (1-q) \times 1/2 \\
&= \frac{q - q^2}{2}
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
P(dd) = P(Z^d Z^d) &= P(d \cap Z \mid m) \times P(d \cap Z \mid f) \\
&= P(d \mid m) \times P(Z \mid m) \times P(d \mid f) \times P(Z \mid Z^d W \mid f) \\
&= q \times 1 \times q \times \frac{1}{2-c} \\
&= \frac{q^2}{2-c}
\end{aligned}
\tag{35}
$$

Frequency of the drive allele in the Z chromosome of the offspring

$$
\begin{aligned}
P(d \mid offspring) &= \frac{P(Dd)/2 + P(dD)/2 + P(dd) + P(d0)}{P(DD) + P(Dd) + P(dD) + P(dd) + P(D0) + P(d0)} \\
&= \frac{\frac{q-q^2}{2(2-c)} + \frac{q-q^2}{2\times 2} + \frac{q^2}{2} + \frac{2q-cq-cq^2}{4-2c}}{\frac{1-2q+q^2}{2} + \frac{q-q^2}{2-c} + \frac{q-q^2}{2} + \frac{q^2}{2} + \frac{2-c-2q+cq^2}{4-2c} + \frac{2q-cq-cq^2}{4-2c}} \\
&= \frac{\frac{8q-3cq-cq^2}{4(2-c)}}{1} \\
&= \frac{8q - 3cq - cq^2}{8 - 4c}
\end{aligned}
\tag{36}
$$

In other words:

$$
q_{t+1} = \frac{8q_t - 3cq_t - cq_t^2}{8 - 4c} \tag{37}
$$

Derivative of drive allele frequency with respect to time

If we were to assume overlapping generations, then the derivative with respect to time can be expressed as:

$$\begin{aligned}\frac{dq}{dt} &= \frac{q_{t+1} - q_t}{t + 1 - t} \\ &= \frac{8q_t - 3cq_t - cq_t^2}{8 - 4c} - q_t \\ &= \frac{cq(1-q)}{8-4c},\end{aligned} \tag{38}$$

and the solution can be expressed as:

$$q(t) = \frac{q_0 e^{ct/(8-4c)}}{1 + q_0(e^{ct/(8-4c)} - 1)} \tag{39}$$

This can be thought of as an approximation recursive equation q_{t+1} above.

Population suppression

Population suppression is proportional to the frequency of females. The frequency of females in the population is equivalent to the frequency of the W chromosome:

$$\begin{aligned}P(f) &= P(\text{W chromosome} \mid f) \\ &= \frac{2 - c - cq}{4 - 2c}\end{aligned} \tag{40}$$

The population size in the next generation, N_{t+1} can be computed as the frequency of the females, $P(f)$ multiplied by the population size in the current generation, N_t and the fecundity of each female which we define as the Beverton-Holt growth rate, $r_t = R_0/(1 + \alpha n_t)$, where $n_t = \frac{N_t}{\text{area}}$ is the population density, $\alpha = (R_{\max} - 2)/2N^*$ where $R_{\max}$ is the maximum fecundity and N^* is the carrying capacity given $R_{\max}$:

$$\begin{aligned} N_{t+1} &= N_t \times P(f) \times r_t \\ &= N_t \left(\frac{2 - c - cq_t}{4 - 2c} \right) \left(\frac{R_{\max}}{1 + n_t \frac{R_{\max} - 2}{2N^*}} \right). \end{aligned} \tag{41}$$

9.2 Derivation of X-shredder equations

Re-derivation or a verbose derivation with inputs from Deredec et al. 2008:

In XX/XY sex determination systems, *e.g.*, mammals, where XX are females and XY are males.

$$XY^d - c \rightarrow 0Y^d \tag{42}$$

Let:

- $P(\text{shredded X} \mid \text{X shredding}) = c$, i.e., the probability of shredding the X chromosome given X-shredding be c;
- $P(\text{intact X} \mid \text{X shredding}) = 1 - c$, i.e., the probability of intact X

chromosome given X-shredding be $1 - c$;

- $P(\text{shredded Y} \mid \text{X shredding}) = 0$, i.e., the Y chromosome be unaffected by the X-shredding;
- $P(\text{intact Y} \mid \text{X shredding}) = 1$, i.e., the Y chromosome be intact;
- $P(d) = P(d \mid m) = q$, i.e., the probability of the drive allele in the Y chromosome (in males) be q; and
- $P(D) = P(D \mid m) = 1 - q$, i.e., the probability of the wild-type or non-drive allele in the Y chromosome (in males) be $1 - q$.

Probability of X and Y chromosomes per genotype and per sex

Female sex chromosomes:

$$\begin{aligned} P(X \mid f) &= P(X \mid XX) \\ &= 1 \end{aligned} \tag{43}$$

$$\begin{aligned} P(Y \mid f) &= P(Y \mid XX) \\ &= 0 \end{aligned} \tag{44}$$

Male sex chromosomes:

$$
\begin{aligned}
P(X \mid XY^d \mid m) &= \frac{P(\text{intact X} \mid \text{X shredding})}{P(\text{intact X} \mid \text{ X shredding}) + P(\text{intact Y} \mid \text{X shredding})} \\
&= \frac{1-c}{2-c}
\end{aligned}
\tag{45}
$$

$$
\begin{aligned}
P(Y \mid XY^d \mid m) &= \frac{P(\text{intact Y} \mid \text{X shredding})}{P(\text{intact X} \mid \text{X shredding}) + P(\text{intact Z} \mid \text{W shredding})} \\
&= \frac{1}{2-c}
\end{aligned}
\tag{46}
$$

$$
\begin{aligned}
P(X \mid XY^D \mid m) &= 1/2 \; P(Y \mid XY^D \mid m) \\
&= 1/2
\end{aligned}
\tag{47}
$$

$$
\begin{aligned}
P(X \mid m) &= P(D \mid m) \times P(X \mid XY^D \mid m) + P(d \mid m) \times P(X \mid XY^d \mid m) \\
&= q\frac{1-c}{2-c} + (1-q)\frac{1}{2} \\
&= \frac{2-c-cq}{4-2c}.
\end{aligned}
\tag{48}
$$

$$
\begin{aligned}
P(Y \mid m) &= P(D \mid m) \times P(Y \mid XY^D \mid m) + P(d \mid m) \times P(Y \mid XY^d \mid m) \\
&= P(Y^D) + P(Y^d) \\
&= (1-q)\ \frac{1}{2} + q\ \frac{1}{2-c} \\
&= \frac{2-c+cq}{4-2c}.
\end{aligned}
\tag{49}
$$

Probability of offspring genotypes in the drive allele locus (i.e., in males only)

$$\begin{aligned} P(XY^D) &= P(X \mid f) \times P(Y^D \mid m) \\ &= P(X \mid f) \times P(D \mid m) P(\text{Y chromosome} \mid XY^D \mid m) \\ &= (1) \times (1-q)\frac{1}{2} \\ &= \frac{1-q}{2}. \end{aligned} \tag{50}$$

$$\begin{aligned} P(XY^d) &= P(X \mid f) \times P(Y^d \mid m) \\ &= P(X \mid f) \times P(d \mid m) P(\text{Y chromosome} \mid XY^d \mid m) \\ &= (1) \times q\left(\frac{1}{2-c}\right) \\ &= \frac{q}{2-c}. \end{aligned} \tag{51}$$

Frequency of the drive allele in the Y chromosome of the offspring

$$\begin{aligned} P(d \mid \text{Y chromosome})_{t+1} &= \frac{P(XY^d)}{P(XY^D) + P(XY^d)} \\ &= \left(\frac{q}{2-c}\right)\left(\frac{1-q}{2} + \frac{q}{2-c}\right)^{-1} \\ &= \frac{2q}{2-c+cq}. \end{aligned} \tag{52}$$

In other words,

$$q_{t+1} = \frac{2q}{2-c+cq}. \tag{53}$$

If we were to assume overlapping generations, then its derivative with respect to time can be expressed as:

$$\begin{aligned}\frac{dq}{dt} &= \frac{q_{t+1} - q_t}{t + 1 - t} \\ &= \frac{2q_t}{2 - c + cq_t} - q_t \\ &= \frac{cq(1-q)}{2 - c + cq}.\end{aligned} \tag{54}$$

Population suppression

Population suppression is proportional to the frequency of females. The frequency of males in the population is equivalent to the frequency of the Y chromosome:

$$\begin{aligned}P(m) &= P(\text{Y chromosome} \mid m) \\ &= \frac{2 - c + cq}{4 - 2c},\end{aligned} \tag{55}$$

then the frequency of females is:

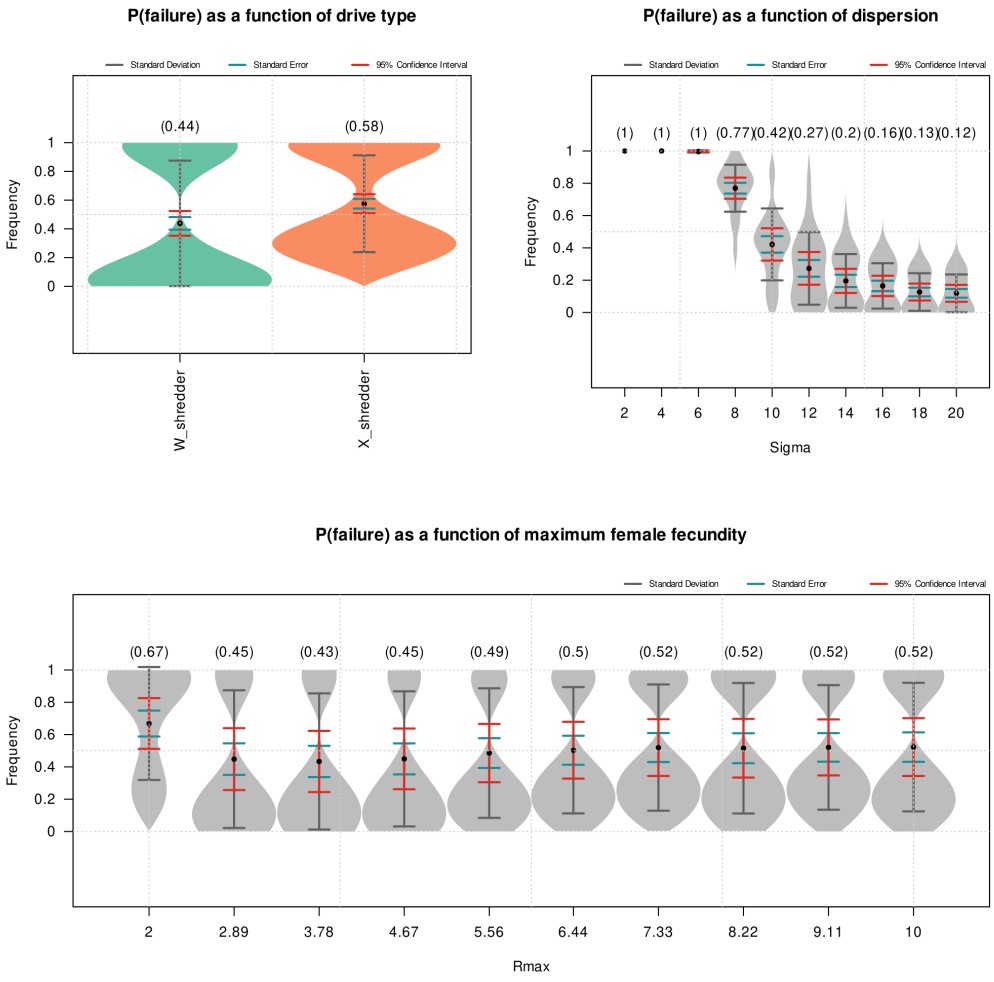
$$\begin{aligned}P(females) &= 1 - P(males) \\ &= 1 - \frac{2 - c + cq}{4 - 2c} \\ &= \frac{2 - c - cq}{4 - 2c}.\end{aligned} \tag{56}$$

This is exactly the same as in W-shredder. Hence, the population size in the next generation can be expressed as:

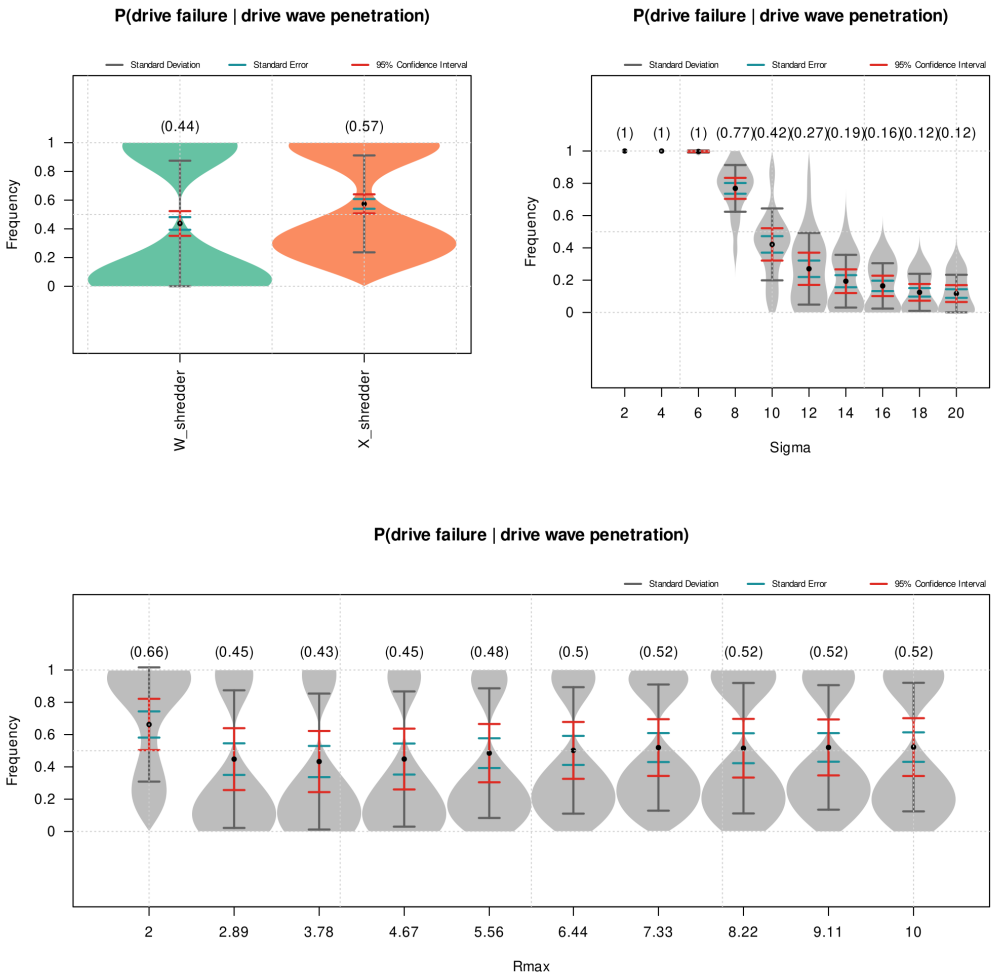
$$
\begin{aligned}
N_{t+1} &= N_t \times P(female) \times r_t \\
&= N_t \left(\frac{2 - c - cq_t}{4 - 2c} \right) \left(\frac{R_{\max}}{1 + n_t \frac{R_{\max} - 2}{2N^*}} \right).
\end{aligned} \tag{57}
$$

.

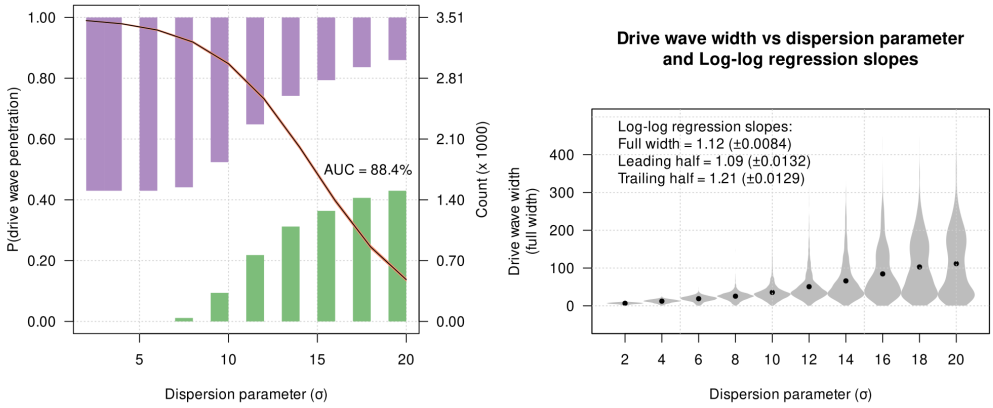
9.3 Supplementary Figures



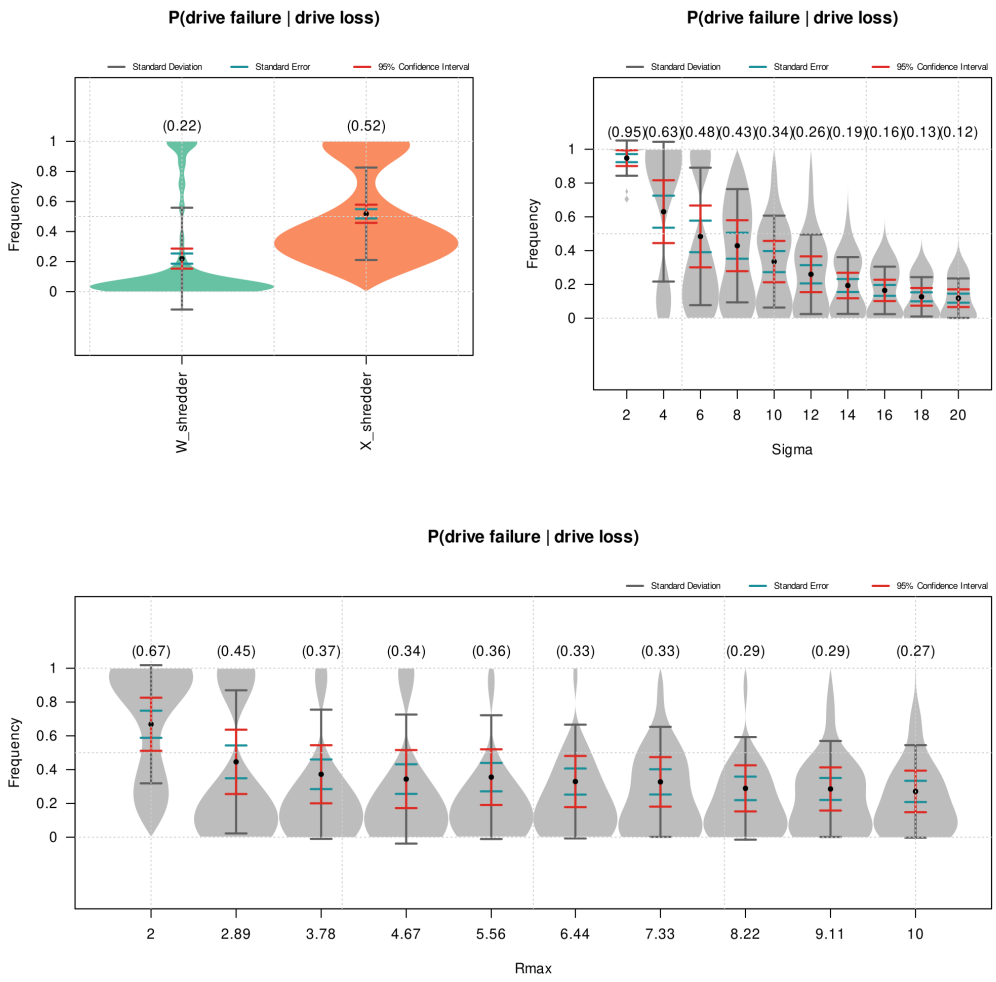
Supplementary Figure 1: Violin plots of the frequency of suppression gene drive failure as a function of drive type, dispersion parameter (Sigma), and maximum female fecundity (Rmax).



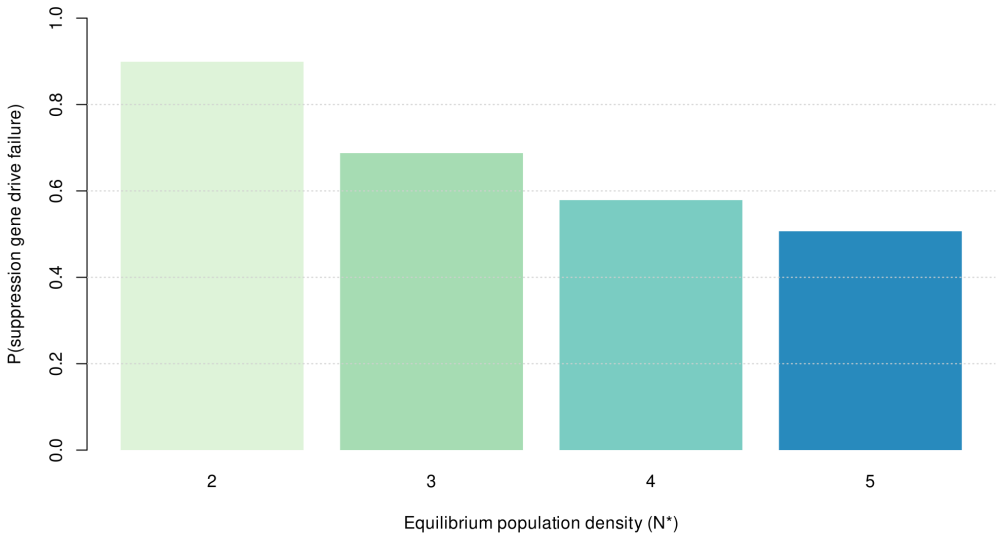
Supplementary Figure 2: Violin plots of the frequency of suppression gene drive failure due to drive wave penetration as a function of drive type, dispersion parameter (Sigma), and maximum female fecundity (Rmax).



Supplementary Figure 3: Relationships between the probability of suppression gene drive wave penetration by wild-types, dispersion parameter, and drive wave widths. Left panel: logistic regression plot relating the penetration probability with the dispersion parameter. Right panel: violin plot of the full width of the drive wave across the dispersion parameter space. Log-log regression: $\ln(\text{width}) = 1 + \ln(\sigma)$ was performed for the full width, leading half width, and trailing half width of the gene drive wave.



Supplementary Figure 4: Violin plots of the frequency of suppression gene drive failure due to the loss of the drive allele as a function of drive type, dispersion parameter (Sigma), and maximum female fecundity (Rmax).



Supplementary Figure 5: Barplot of the frequency of suppression gene drive failure as a function of the equilibrium population density (N^*).