Evolutionary models of synthetic Z-linked suppression gene drives

Luke Holman*
**luke.holman@unimelb.edu.au*

Abstract

My abstract text.

Keywords: *W*- and *X*-shredders, Lepidopteran pests, meiotic drive, population control, schistosomiasis, selfish genes.

*School of BioSciences, The University of Melbourne, Victoria, Australia.

Introduction

It is becoming increasingly technically feasible to genetically engineer natural populations in order to control pathogens, pathogen vectors, agricultural pests, and invasive species using 'gene drives' [1–6]. Gene drives assist the propagation of engineered genes through populations using a range of mechanisms including gene conversion, poison-antidote systems, segregation distortion, and genetic incompatibility [7]. For example, the CRISPR-Cas9 gene editing system allows one to produce a transgenic insertion that is transmitted to almost 100% of the offspring of heterozygous individuals instead of the usual 50% (due to targeted cutting of the homologous chromosome followed by repair using the gene drive as a template). CRISPR-Cas9 can be used to create 'replacement drives', which propagate a human-beneficial allele that would not otherwise spread, e.g. an allele that interferes with the tranmission of malaria by mosquitoes [1]. CRISPR-Cas9 can also be used to create suppression drives, which aim to cause extinction (or at least reduce the size) of the target population by introducing genes with lethal or sterilising effects (MORE HERE [2,5]), or by skewing the population sex ratio – typically towards males MORE HERE [8].

Given that gene drive technologies are largely new and untested, several recent theoretical papers have investigated their feasibility and possible consequencies. For example, Noble et al. [6] used models to argue that CRISPR-based gene drives are likely to be highly invasive and would readily spread to other populations, which may be undesirable. Additionally, models [9,10] illustrate that CRISPR gene drives may fail to spread in the presence of 'resistant' alleles that are immune to being cut and replaced by the drive allele. The issue of resistance is compounded because the standard implementation of CRISPR gene drives (but see [4,5,10]) tends to create its own resistance alleles (adding to those that might be naturally present) through a process called non-homologous end joining (NHEJ; [1,10,11]). Given the potential safety, ethical, and socio-political concerns surrounding gene drives, there is also much interest in creating gene drives that would go extinct once their job is done [12], would stay confined to particular populations [13], and/or could be reversed once they have spread [14].

Here, I focus on the evolutionary dynamics of a Z-linked suppression gene drive. The simulation is inspired by proposals for various types of CRISPR-Cas9 Z-linked gene drives that have been proposed by Kevin Esvelt and colleagues (see www.sculptingevolution.org/genedrives/current/schistosomiasis; these ideas are not yet published in a journal or pre-print archive). Various Z-linked suppression drives proposed by Esvelt and colleagues are shown schematically in Figure 1. The gene drive would enjoy a transmission advantage in ZW females, and optionally also in ZZ males. Esvelt proposes that Z-linked drives could be used to control the trematode parasites (*Schistosoma* spp.) responsible for the deadly disease schistosomiasis, though Z-linked drives could theoretically be used to control any organism with female-heterogametic sex determination (such as Lepidopteran pests or invasive birds).

A Z-linked gene drive could suppress populations by biasing transmission in females, for example by targetting and cutting unique sequences on the W chromosome; this gene drive would be a 'W-shredder', similar to the X-shredder proposed for XY species [15,16]. Females carrying the gene drive would thus produce relatively few W-bearing eggs, and therefore

produce mainly drive-carrying sons. Esvelt et al. point out that the total number of offspring produced by these drive-carrying females will depend on multiple factors – such as the timing of the drive mechanism, the strength of individual-level fitness costs to drive carriers, and the ecology of the target species – and will be critically important to the evolutionary dynamics of the drive. For example, for some W-shredders, drive females might produce roughly the same number of offspring as a wild-type female (only with a preponderance of sons) because the W chromosome is destroyed at an early enough stage in oogenesis that the number of mature ggs is not affected (Figure 1, 1A). Alternatively, drive-carrying females might produce half (or less than half) the number of offspring, e.g. if the drive works by destroying all ova or offspring that inherit a W chromosome, and females cannot compensate by producing more. As an alternative to W-shredders, Esvelt et al. also proposed that one could suppress populations using a Z-linked locus that caused sterility or lethality in females. If this female-harming gene was capable of gene drive in males (see below), it could perhaps reach high enough frequencies to suppress the population.

Esvelt and colleagues also note that if the W-shredder or female-harming Z-linked locus caused gene drive in males, it would probably spread through the population faster and be more likely to result in extinction. Male gene drive could be accomplished using 'standard' CRISPR-Cas9 gene conversion (REF), whereby the driving Z locus would convert the wild type locus using homing endonuclease activity followed by DNA repair inside male heterozygotes, leading these males to produce mostly drive-carrying sperm and offspring.

Impressed by the promise of these ideas, I wrote an evolutionary simulation that can accommodate all of these hypothesised types of Z-linked drives. We aimed to test which properties of the gene drive and the ecology of the target species are critical to determining the likelihood and speed with which the gene drive causes extinction. For example, the gene drive will presumably spread faster if it can bias transmission in both sexes, but perhaps a female-only gene drive (which might sometimes be easier to engineer) would be perfectly adequate. Also, since the population will become more male-biased as the gene drive invades, there will be eco-evo feedback (REF) that might affect the evolutionary outcome in non-intuitive ways. For example, the altered sex ratio might intensify the fitness advantage accruing to any resistant W chromosomes or autosomal modifiers that prevent W-shredding (due to Fisherian selection for an even sex ratio; [17]), relative to that observed in models dealing with resistance to gene drives carried on autosomes [9,10]. Moreover, the change in sex ratio could affect the demographics of the population, particularly if males and females contribute differentially to density-dependent population growth [18], or have different dispersal rates [19]. The model incorporates the possibility that Z-linked resistance alleles are sometimes formed through NHEJ in males that are heterozygous for the drive allele [1,10,11]. It is not clear *a priori* whether the creation of resistant Z-linked alleles by NHEJ is a equally problematic for a Z-linked gene drive as it is for an autosomal drive, because it would only hinder gene conversion in males, assuming that NHEJ does not occur in response to W-shredding (which is likely, because the W-shredder could be designed to target many repetitive regions of the W chromosome).

Table 1: List of variables, and their corresponding parameter(s) in the model, which were varied in order to study their effects on the likelihood of population extinction.

Variable	Parameter(s)	Outcome
Strength of gene drive in females (e.g. W-shredding)	p_{shred}	1.00
Strength of gene drive in males (e.g. gene conversion)	p_{conv}	1.00
Cost of gene drive allele to female fecundity	c_f	1.00
Cost of gene drive allele to male mating success	c_m	1.00
Frequency of W-linked resistance mutations	μ_W	1.00
Frequency of Z-linked resistance mutations and NHEJ	μ_Z and p_{nhej}	1.00
Frequency of autosomal resistance alleles	μ_A and μ_B	1.00
Patchiness of the population	k	1.00
Dispersal rate of males and females	x_m and x_f	1.00
Global versus local density-dependence of female fecundity	ψ	1.00
Contribution of males relative to females in density-dependence	δ	1.00
Number of gene drive carrier males released	$n_{release}$	1.00
Release strategy: all in one patch, or global	-	1.00
Fecundity of females at low population densities	r	1.00
Shape of density dependence	α	1.00

Methods

Overview

We model a finite population of dioecious diploids with ZW sex determination, living in j discrete habitat patches which are arranged linearly in a ring (preventing 'edge effects'). The model considers the demography and evolution of a population into which $n_{release}$ males carrying the Z-linked gene drive are released. The drive allele is capable of gene drive in females (e.g. through W-shredding) and optionally also gene drive in males (e.g. via gene consersion). The life cycle consists of discrete generations, each composed of the following phases: birth, dispersal between patches, breeding with patches, and death of the parental generation. The species has 3 loci with 2 or 3 alleles each, some of which potentially show non-Mendelian inheritance. The equilibrium population size was roughly 10,000 in all simulations prior to the release of the gene drive. The model is a stochastic individual-based simulation written in R 3.4.0 (REFERENCE) and was run on `Spartan`, a computer cluster at the University of Melbourne. An accompanying website presents the R scripts (with annotations) used to run the model and generate the figures (link).

Loci and alleles

Each male individual in the simulation carries one Z-linked locus and two autosomal loci, each with two alleles. Each female carries a single allele at the Z-linked locus plus a W chromosome, as well as two alleles at both of the autosomal loci.

There are three possible Z-linked alleles: a wild-type allele (denoted $Z+$) which is vulnerable to gene drive; a gene drive allele (Z^*), and a resistant allele (Zr) which is immune to gene

drive. Similarly, there are two possible types of W chromosomes: a wild-type W chromosome ($W+$) that is vulnerable to gene drive by the Z^* allele, and a resistant W chromosome (Wr) that is immune to gene drive.

The two autosomal loci are denoted A/a and B/b, and control immunity to W-shredding and gene conversion respectively. A/a and B/b could be called 'trans-acting' resistance loci, since they are at a different locus (indeed, a different chromosome) to the gene drive allele, in contrast to the 'cis-acting' resistance conferred by the Zr and Wr alleles. The A allele is dominant to a and confers immunity to Z-linked gene drive (e.g. W-shredding) in females. The B allele is dominant to b and confers immunity to Z-linked gene drive (e.g. gene conversion) in males.

Calculating female and male fitness

We assume that wild-type individuals (i.e. those lacking any drive or resistance) have a fitness (w) of 1, while other genotypes have $0 \leq w \leq 1$. The fecundity of females carrying the gene drive is reduced by a factor $1 - C_f$, where a small C_f implies minimal costs (e.g. because the Z-shredding occurs early enough that lost ova/offspring can be replaced), $C_f = 0.5$ could represent the case where all daughters die and are not replaced, and $C_f = 1$ means that gene drive females are completely sterile (which is useful for modelling a female-sterilising Z-linked suppression drive as opposed to a W-shredder). Similarly, the fitness of males carrying the gene drive is reduced by a factor $1 - C_m$; male fitness is used in the calculation of mating success (see below). Furthermore, the resistant chromosomes Wr and Zr are assumed to reduce fitness by a factor $1 - C_w$ and $1 - C_z$ respectively. All costs are multiplicative; for example, a Z^*Zr male would have fitness $(1-C_m)(1-C_z)$. Additionally, all costs are assumed to be dominant, meaning that having one drive or resistance allele is equally costly as having two.

AUTOSOMES?

Gamete production and gene drive

We assume that the A/a and B/b loci segregate independently during meiosis and display standard Mendelian inheritance. Inheritance of the sex chromosomes is also Mendelian, except for certain genotypes carrying one copy of Z^*.

Firstly, $Z^*W+aaBB$, $Z^*W+aaBb$, and $Z^*W+aabb$ females produce a fraction $\frac{1}{2}(1 + p_{shred})$ of Z-bearing gametes and $\frac{1}{2}(1 - p_{shred})$ W-bearing gametes. Therefore, these three female genotypes produce >50% sons when $p_{shred} > 0$, due to the shortage of W chromosomes in their gametes. Note that the gamete frequencies of Z^*Wr *females*, or of females carrying at least one A allele, conform to the standard Mendelian expectations due to resistance.

Secondly, $Z^*Z+AAbb$, $Z^*Z+Aabb$, and $Z^*Z+aabb$ males produce a fraction $\frac{1}{2}(1 + p_{conv} - p_{conv}p_{nhej})$ of gametes carrying the Z^* allele, $\frac{1}{2}(1 - p_{conv})$ gametes carrying the $Z+$ allele, and $\frac{1}{2}(p_{conv}p_{nhej}))$ gametes carrying the Zr allele. Thus, gene conversion occurs in males

if $p_{conv} > 0$, meaning that the Z^* allele is over-represented in the gametes of these three male genotypes. The parameter p_{nhej} represents non-homologous end joining, in which an endonuclease-based gene drive fails to copy itself to the homologous chromosome, and instead deletes its target site and thereby creates a resistant allele. The gamete frequencies of Z^*Zr males, or of males carrying at least one B allele, conform to the standard Mendelian expectations due to resistance.

Calculating female fecundity

In the breeding phase of the lifecycle, we first determine the number of offspring produced by each female in the population. We first calculate the expected fecundity of each female, which is affected by three factors: the female's genotype, the density of males and females in the local patch and/or in the full population, and some global parameters in the model.

Specifically, the expected fecundity of female i (F_i) is calculated as

$$F_i = (1 + w_i r(1 - (D_i/K)^{\alpha}))$$

where w_i is the relative fitness of female i (possible range: 0 to 1, where 1 is the fitness of the wild type $Z+W+aabb$ females), D_i is the 'density' experienced by female i, K is the carrying capacity, and r and c are constants that scale the maximum possible fecundity and the shape of density-dependence, respectively. Thus, we assume that offspring production is density-dependent according to a Richards model [20].

To ensure that the simulation captures various possible types of life history and ecology (see Introduction), we calculate the density D_i in various ways across different simulation runs. First, we define the 'global density' d_g, which is experienced equally by every female in every patch, as

$$d_g = \sum_{i=1}^{N_f} w_i + \delta N_m$$

where N_f and N_m is the number of females and males across all patches, the first term is the sum of the fitnesses of all these females, and δ is a constant (range: $0 - \infty$) that scales the effect of each male on d_g, relative to a female with fitness $w_i = 1$. This formulation means that females with high relative fitness (i.e. fecundity) have a stronger effect on the global density than do low-fitness females. We also assume that each male contributes a fixed amount to the global density, irrespective of his genotype/fitness (male fitness is only used to determine male mating success; see below). The parameter δ represents sex differences in ecological niche use and behaviours that affect female fecundity. For example, we might expect $\delta < 1$ in species where males and females utilise very different environmental niches, or $\delta > 1$ in species with strong inter-locus sexual conflict.

Second, we define the 'local density' d_j, which is experienced by every female in patch j, as

$$d_j = \sum_{i=1}^{n_{f,j}} w_i + \delta n_{m,j}$$

where $n_{f,j}$ and $n_{m,j}$ are the numbers of females and males in patch j. As before, this formulation means that d_j depends on the fitnesses of the females in the patch, as well as the number of males (scaled by the constant δ).

Finally, the overall density experienced by female i in patch j (D_i) is a composite of the global and local densities given by $D_i = \psi d_g + (1 - \psi) d_j$. The parameter ψ scales the importance of global and local density to female fecundity. When $\psi = 0$, only local density matters and selection on females is entirely "soft", while when $\psi = 1$ only global density matters and selection on females is completely "hard" (REFERENCE). Intermediate values of ψ produce a mixture of hard and soft selection on females, and the growth rate of population depends on density at both scales.

Once we have calculated the expected fecundity of each female (F_i), we generate the realised fecundity of the female by randomly sampling from a Poisson distribution with $\lambda = F_i$ (allowing for stochastic variation in fecundity between females with equal F_i). If the resulting number of offspring exceeds the global carrying capacity K, we randomly cull the offspring until K are left.

Competition between males

After determining how many offspring each female produces, we determine the fathers of each of these offspring. We assume that all breeding occurs within patches, such that males only compete for mating/fertilisation with males from the same patch (i.e. selection on males is always "soft"; REFERENCE). If the patch contains k different male genotypes and there are $n_1, n_2, ...n_k$ males of each genotype, the probability that a male of genotype k is the father of any given offspring is

$$p_j = \frac{n_k w_k}{\sum_{i=1}^{k} n_i w_i}$$

such that relatively common and/or high-fitness male genotypes are more likely to sire offspring. This formulation means that we assume that both sexes potentially reproduce with multiple different partners.

Reproduction, mutation and dispersal

After picking the parents, we randomly generate each offspring's genotype based on the gamete (and thus zygote) frequencies that are expected from the parental genotypes. Offspring are born in the same patch as their parents, and the parental generation is replaced by the offspring generation (i.e. we assume discrete, non-overlapping generations).

When an offspring is created, each *Z+* allele it carries has a chance μ_Z to mutate to a *Zr* allele, and *vice versa* (i.e. mutation in both directions is equally probable). Similarly, each *W+* allele has a chance μ_W to mutate to a *Wr* allele, and *vice versa*.

Female and male offspring disperse to another patch with probabilities x_f and x_m respectively. We model two types of dispersal, in separate simulations: local dispersal, in which offspring move to one of the two neighbouring patches with equal probability (recalling that the patches are arranged linearly in a ring), or global dispersal, in which dispersing offspring can land in any of the other patches.

One compete run of the simulation

We first initialised a population of 10,000 individuals (i.e. the carrying capacity, K) with low or zero frequencies of *Zr*, *Wr*, *A* and *B* alleles, higher frequencies of the wild type *Z+*, *W+*, *a*, and *b* alleles, and zero *Z** gene drive alleles. We then iterated the population for 50 generations of burn-in, to allow the population to approach demographic and genotypic equilibrium. We then introduced $n_{release}$ males with the genotype *Z*Z*aabb*, which represents the release into the wild of a laboratory-reared strain homozygous for the driving *Z* and for autosomal factors conferring susceptibility to drive. Males are released after density-dependent regulation of female fecundity, but before picking fathers for the offspring. In some simulations, all the *Z*Z*aabb* males were released in a single patch, while in others the $n_{release}$ males were randomly and evenly divided across all k patches. We continued to cycle through the lifecycle (birth, migration, breeding, death) until either A) the driving *Z** allele went extinct, B) the population went extinct, C) the *Wr* chromosome went to fixation (making population suppression impossible), D) the *Z** allele fixed, but failed to cause population extinction, or E) 900 generations had elapsed. We recorded which of these five outcomes occurred, as well as the allele frequencies, population size, and sex ratio at each generation.

Investigating the parameter space

For each of the parameters in Table 1, we selected two or more possible parameter values (e.g. high versus low rates of *W*-shredding p_{shred}; many versus few patches k). We then ran the model once for all possible combinations of these parameter values (n = XXX model runs). The aim was to measure the 'main effect' of each parameter across a background of assumptions for the other parameters, as well as to investigate all the possible 2-way interactions between the parameters (e.g. to test if the effect of p_{shred} depends on k and *vice versa*).

Results

Three illustrative simulation runs

Blah blah

Main effects of each parameter

Figure 2 shows the main effects of each of the parameters in the model, arranged in order of importance to the evolutionary outcome.

- Note that when females hardly migrate, the Wr is slow to spread across patches. It only has a good invasion probability if Z* is present, otherwise it's neutral or costly

Discussion

- schistos have large males, small females
- birds often have sex-biased dispersal
- females may be more demographically limiting leps, since it is them that lays the eggs

Acknowledgements

So long, and thanks for all the fish!

References

1. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, James AA. 2015 Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences* **112**, E6736–E6743.

2. Hammond A *et al.* 2016 A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology* **34**, 78.

3. Wang Z, Pan Q, Gendron P, Zhu W, Guo F, Cen S, Wainberg MA, Liang C. 2016 CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. *Cell Reports* **15**, 481–489.

4. Prowse TA, Cassey P, Ross JV, Pfitzner C, Wittmann TA, Thomas P. 2017 Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proc. Roy. Soc. B* **284**, 20170799.

5. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, Nolan T, Crisanti A. 2018 A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology*

6. Noble C, Adlam B, Church GM, Esvelt KM, Nowak MA. 2018 Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* **7**, e33423.

7. Lindholm AK *et al.* 2016 The ecology and evolutionary dynamics of meiotic drive. *Trends in Ecology & Evolution* **31**, 315–326.

8. Beaghton A, Beaghton PJ, Burt A. 2017 Vector control with driving Y chromosomes: modelling the evolution of resistance. *Malaria Journal* **16**, 286.

9. Drury DW, Dapper AL, Siniard DJ, Zentner GE, Wade MJ. 2017 CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science Advances* **3**, e1601910.

10. Unckless RL, Clark AG, Messer PW. 2017 Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**, 827–841.

11. Gantz VM, Bier E. 2015 The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, aaa5945.

12. Min J, Noble C, Najjar D, Esvelt KM. 2017 Daisyfield gene drive systems harness repeated genomic elements as a generational clock to limit spread. *BioRxiv*, 104877.

13. Maselko M, Heinsch SC, Das S, Smanski MJ. 2018 Genetic incompatibility combined with female-lethality is effective and robust in simulations of *Aedes aegypti* population control. *bioRxiv*, 316406.

14. Vella MR, Gunning CE, Lloyd AL, Gould F. 2017 Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific Reports* **7**, 11038.

15. North A, Burt A, Godfray HCJ. 2013 Modelling the spatial spread of a homing endonuclease gene in a mosquito population. *Journal of Applied Ecology* **50**, 1216–1225.

16. Galizi R, Doyle LA, Menichelli M, Bernardini F, Deredec A, Burt A, Stoddard BL, Windbichler N, Crisanti A. 2014 A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nature Communications* **5**, 3977.

17. Holman L, Price TA, Wedell N, Kokko H. 2015 Coevolutionary dynamics of polyandry and sex-linked meiotic drive. *Evolution* **69**, 709–720.

18. Rankin DJ, Kokko H. 2007 Do males matter? The role of males in population dynamics. *Oikos* **116**, 335–348.

19. Li X-Y, Kokko H. 2019 Sex-biased dispersal: A review of the theory. *Biological Reviews* **in press**.

20. Fowler CW. 1981 Density dependence as related to life history strategy. *Ecology* **62**, 602–610.

Supporting information