Supplementary Material

From: Holman 2019, *Evolutionary simulations of* Z*-linked suppression gene drives*

The R scripts used to run the model and generate all figures and tables can be viewed at **https://lukeholman.github.io/W_shredder/**, along with annotations explaining the code.

Supplementary Methods

Gamete production and gene drive

I 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 Z^* allele.

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. The gamete frequencies of resistant female genotypes – i.e. genotypes that include a W^r chromosome and/or at least one A allele – conform to the standard Mendelian expectation.

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 Z^r 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 the creation of resistance alleles via non-homologous end joining, in which the gene drive fails to copy itself to the homologous chromosome, and instead induces an indel mutation that creates a resistant allele. The gamete frequencies of Z^*Z^r males, or of males carrying at least one B allele, conform to the standard Mendelian expectations due to resistance.

Calculating female fecundity

Females with no Z^* alleles have an intrinsic quality or 'fitness' of $w = 1$, while the fitness of females carrying Z^* is $1 - c_f$. Small c_f implies minimal costs of Z^* to offspring production (e.g. because mothers replace lost gametes/offspring and/or sib-sib competition is intense), $c_f = 0.5$ could represent the case where all daughters die and are not replaced, and $c_f = 1$ means that females carrying Z^* are completely sterile. Setting $c_f = 1$ allows simulation of a female-sterilising Z-linked drive (Figure 1D).

In the breeding phase of the lifecycle, the simulation first determines the number of offspring produced by each female. The expected fecundity of female i (F_i) 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, as follows:

$$F_i = w_i\big(2 + r(1 - (D_i/K)^\alpha)\big) \tag{1}$$

where D_i is the 'density' experienced by female i, w_i is her fitness, K is the carrying capacity, r controls the maximum possible fecundity, and α determines the shape of density-dependence [1]. The function means that female fecundity declines to a maximum of two offspring as D_i approaches K, and that females with $w_i = 0$ produce no offspring. When $D_i > K$, we set $D_i = K$, preventing fecundity from going below $F_i = 2w_i$.

To ensure that the simulation captures various possible types of life history and ecology, I calculated density D_i in various ways in different simulation runs. First, I define the global density d_g, which acts equally on every female in every patch, as

$$d_g = \sum_{i=1}^{N_f} w_i + \delta N_m \tag{2}$$

where N_f and N_m are the numbers of females and males across all patches, the first term is the summed 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 (and thus higher fecundity) have a stronger effect on the global density than do low-fitness females. I also assume that each male contributes a fixed amount to the global density, irrespective of his genotype/fitness (since I assume that male fitness only affects male mating success; see below). The parameter δ represents sex differences in ecological niche use and behaviour. For example, one might expect $\delta < 1$ in species where males and females utilise very different environmental niches, or $\delta > 1$ in species where males are harmful to females.

Second, I define the local density d_j experienced by every female in patch j, as

$$d_j = \sum_{i=1}^{n_{f,j}} w_i + \delta n_{m,j} \tag{3}$$

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 summed 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 weighted sum of the global and local densities given by $D_i = \psi d_g + (1 - \psi) d_j$, where the parameter ψ weights 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' [2]. Intermediate values of ψ produce a mixture of hard and soft selection on females.

After calculating the *expected* fecundity of each female (F_i), the simulation generates each feamle's *realised* fecundity by randomly drawing an integer from a Poisson distribution with $\lambda = F_i$; this allows fecundity to vary stochastically between females with equal F_i. If the resulting number of offspring exceeded the global carrying capacity K, the model randomly selects K surviving offspring.

Competition between males

Similarly to females, I assume that males have intrinsic fitness determined by their genotype, except that male fitness determines mating success (see below) rather than fecundity. The fitness of males carrying Z^* is reduced by a factor $1 - c_m$. For simplicity, the costs of Z^* to males were assumed to be dominant, such that Z^*Z^+ males and Z^*Z^* males had equal fitness.

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 matings/fertilisations with males in the same patch. If the patch contains l different male genotypes and there are $n_1, n_2, ...n_l$ males of each genotype, the probability that a male of the l^{th} genotype is the father of any given offspring is

$$p_l = \frac{n_l w_l}{\sum_{i=1}^{l} n_i w_i} \tag{4}$$

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

Reproduction, mutation and dispersal

After picking the parents, the model randomly generates each offspring's genotype according to its parents' expected gamete (and thus zygote) frequencies. Offspring are born in the same patch as their parents, and the parental generation is replaced by the offspring generation.

When an offspring is created, each Z^+ allele it carries has a chance μ_Z to mutate to a Z^r 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 W^r 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 in a ring), or global dispersal, in which dispersing offspring can land in any of the other patches.

One compete run of the simulation

The simulation first initialises a population of 10,000 individuals (the carrying capacity, K) with low or zero frequencies of Z^r, W^r, A and B alleles, higher frequencies of the wild type Z^+, $W+$, a, and b alleles, and zero Z^* gene drive alleles. The simulation then runs 50 generations as 'burn-in', to allow the population to reach demographic and genotypic equilibrium. Next, $n_{release}$ males with the genotype Z^*Z^*aabb are added to the population just before fathers are selected, representing the release into the wild of a laboratory-reared strain homozygous for the driving Z. 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. The model continued until either A) the driving Z^* allele went extinct, B) the population went extinct, C) the W^r chromosome went to fixation (making population suppression impossible), D) the Z^* allele fixed without causing extinction, or E) 1000 generations had elapsed. The model 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

To investigate the effect of each parameter on the evolutionary outcome, I used Latin hypercube sampling (LHS; implemented via the R package `lhs`), which is an efficient way to select random parameter combinations that are uniformly distributed across a given multidimensional parameter space. I defined this space by choosing a range of biologically relevant values *a priori*, and/or by focusing on convenient values for plotting. For example, I set the range of possible values for c_f at 0-0.6, causing the model to consider W-shredders that have no cost to female fecundity (e.g. because all gametes/offspring lost to W-shredding are replaced; $c_f \approx 0$), a moderate cost (reflecting incomplete replacement, or costs of the W-shredder to survival or fecundity; $c_f \approx 0.1 - 0.3$), or a *c.* 50% cost (e.g. because all daughters are lost and not replaced). I also selected the maximum value, 0.6, because higher values always resulted in failure to cause extinction (Figure 3).

After defining the range of values to investigate for each parameter (shown along the x-axis of Figure S1), I ran a total of 1,810,247 simulation runs, yielding the data shown in Figure S1. I then fixed $p_{shred} = 1$ in light of the results shown in Figure S1, and ran a further 721,587 simulations (shown in Figures 3-4). Finally, to investigate the evolution of a female-sterilising gene drive, I $c_f = 1$ and ran a further 1,559,817 simulations – most of these simulations allowed all other parameters to vary freely (again, using LHS), while a subset of them had all parameters related to drive resistance set to zero, to give the Z^* allele the best possible chance to cause extinction (which it never managed to do).

Supplementary tables

Table S1: List of the parameter values used to generate the simulation runs shown in Figure 2.

	Panel A	Panel B	Panel C
Release strategy	Scattered over all patches	All in one patch	All in one patch
Gene drive in females (p_{shred})	0.99	0.84	0.84
Gene drive in males (p_{conv})	0.74	0.38	0.54
Rate of NHEJ (p_{nhej})	0.07	0.04	0.03
Z-linked resistance (μ_Z)	0.00	0.00	0.00
W-linked resistance (μ_W)	0.00	0.00	0.00
Cost of Z^* to females (c_f)	0.25	0.07	0.03
Cost of Z^* to males (c_m)	0.10	0.04	0.40
Male dispersal frequency (x_m)	0.23	0.46	0.38
Male dispersal frequency (x_f)	0.50	0.34	0.26
Dispersal type	Local	Global	Global
Number of patches (k)	15	3	29
Importance of local density (ψ)	0.65	0.37	0.08
Male effect on density (δ)	1.78	0.99	0.20
Shape of density dependence (α)	1.27	1.25	1.60
Maximum fecundity (r)	697	788	188
Intial freq W-shredding resistance allele A	0.00	0.00	0.05
Intial freq gene conversion resistance allele B	0.00	0.00	0.05

Supplementary figures

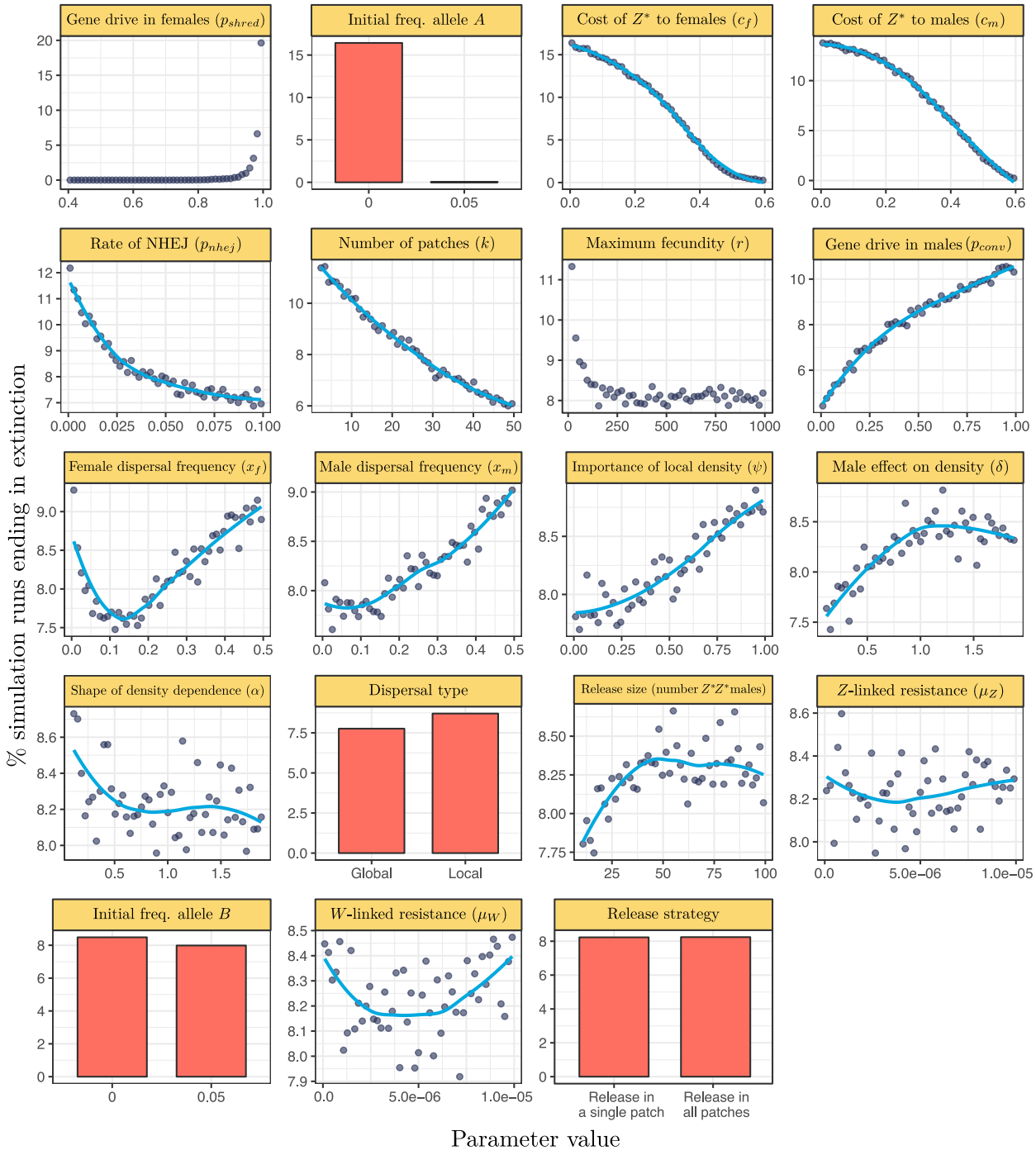

Figure S1: The plot shows the same information as in Figure 3. These data come from an independent set of 1,810,247 simulation runs in which p_{shred} was allowed to vary, in addition to all of the parameters shown in Figure 3. Given the decisive effect of p_{shred} on extinction (top-left panel), I elected to re-run the simulation while fixing $p_{shred} = 1$ (results in Figures 3-4).

References

1. Fowler, C. W. 1981 Density dependence as related to life history strategy. *Ecology* **62**, 602–610.

2. Li, X.-Y. & Holman, L. 2018 Evolution of female choice under intralocus sexual conflict and genotype-by-environment interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170425.