Supplementary Material

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

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

Supplementary Methods

Calculating fitness

Individuals with no Z^* alleles have an intrinsic fitness of $w = 1$, while other genotypes have $0 \leq w \leq 1$. The fecundity of females carrying Z^* is reduced by a factor $1 - c_f$. Small c_f implies minimal costs (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. Similarly, the fitness of males carrying Z^* is reduced by a factor $1 - c_m$; male fitness determines mating success (see below). For simplicity, I assume that the resistance alleles W^r, Z^r, A and B are cost-free. Also, the costs of Z^* to males were assumed to be dominant, such that Z^*Z^+ males and Z^*Z^* males had equal fitness.

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 Z^*W^r 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 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

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:

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

where D_i is the 'density' experienced by female i, w_i is her fitness, K is the carrying capacity, and r and α are constants that control the maximum possible fecundity and the shape of density-dependence respectively (function from [@fowler1981de]).

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 (i.e. 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, we 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' (as in [@li2018ev]). Intermediate values of ψ produce a mixture of hard and soft selection on females.

After calculating 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 exceeded the global carrying capacity K, the model randomly selects K surviving offspring.

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 matings/fertilisations with males in the same patch. 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

$$p_j = \frac{n_k w_k}{\sum_{i=1}^{k} n_i w_i} \tag{4}$$

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 model 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. It then runs 50 generations of 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

For each of the parameters in Table 1, I selected two or more possible parameter values (e.g. high versus low rates of W-shredding p_{shred}; many versus few patches k). I then ran the model once for all possible combinations of these parameter values (n = 6,287,654 model runs). The aim was to measure the effect of each parameter across various assumptions for the other parameters, as well as to investigate all 2-way interactions between the parameters. To gauge the relative importance of the various features of the Z^* allele and the species' ecology to the extinction probability, I fit a binomial generalised linear model (GLM) with extinction as the dependent variable, and all the model parameters and their 2-way interactions as predictors. The predictors were scaled and centred before running the GLM, allowing for a meaningful ranking of the predictors by their absolute effects on extinction.

Supplementary tables

Table S1: The number and percentage of simulation runs (out of 5,657,700 total) that ended with the five possible outcomes, for the subset of simulation runs focusing on a W-shredder gene drive.

Outcome	Number of simulations	%
Z* fixed without causing extinction	2,330,324	41.2
Z* went extinct	1,584,409	28.0
Population went extinct	917,328	16.2
Wr fixed	689,487	12.2
Timer expired	136,152	2.4

Table S2: The number and percentage of simulation runs (out of 629,954 total) that ended with the five possible outcomes, for the subset of simulation runs focusing on a female-sterilising Z-linked gene drive.

Outcome	Number of simulations	%
Z* went extinct	539,978	85.7
Timer expired	69,942	11.1
Population went extinct	12,992	2.1
Wr fixed	7,042	1.1

Supplementary figures

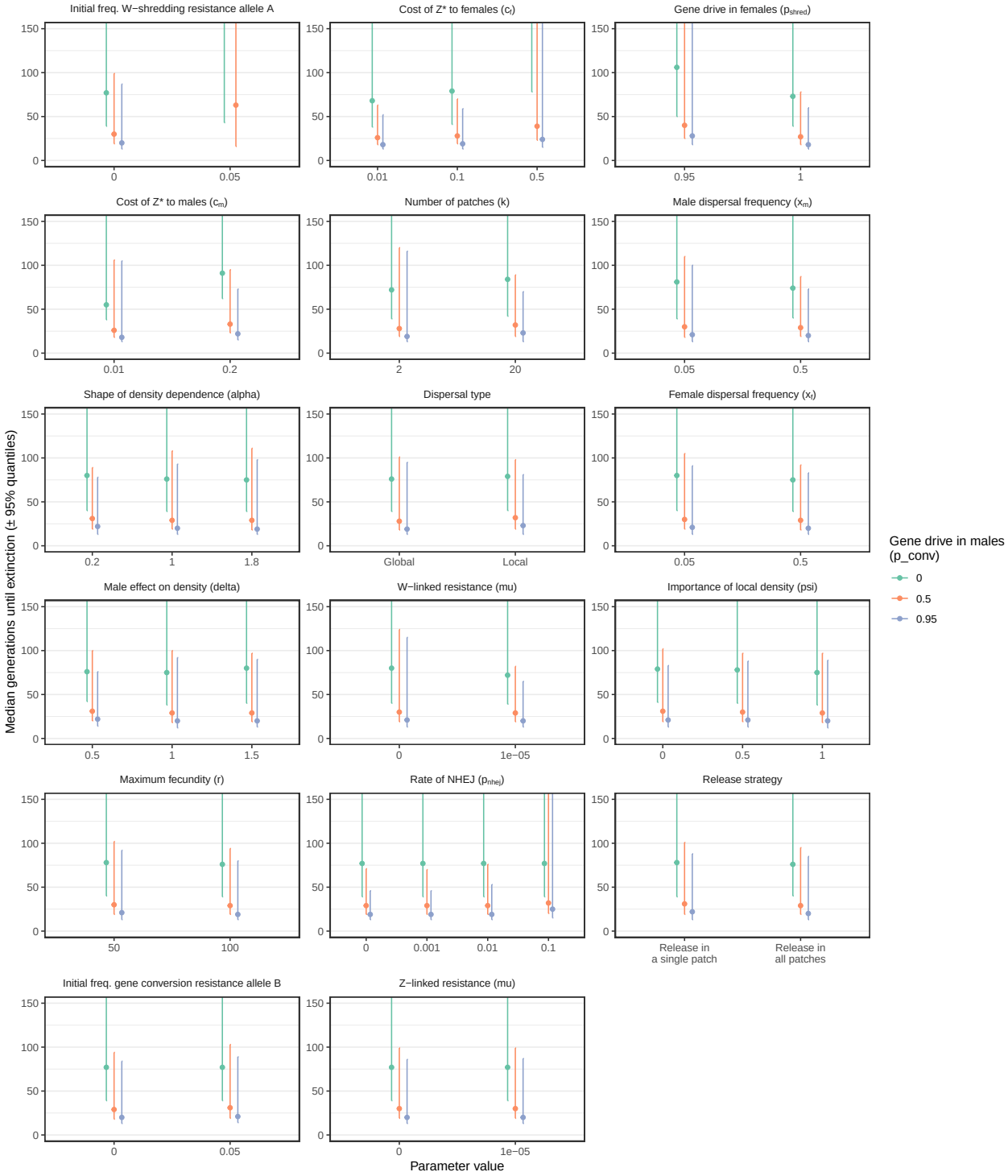

Figure S1: Similar plot to Figure 3, except that the y-axis shows the median number of generations until the W-shredder caused extinction, among just the subset of simulations in which extinction actually occurred (n = 917,328 simulation runs). The median was only calculated if at least 40 simulation runs reached extinction, and the y-axis is truncated at 150 generations.

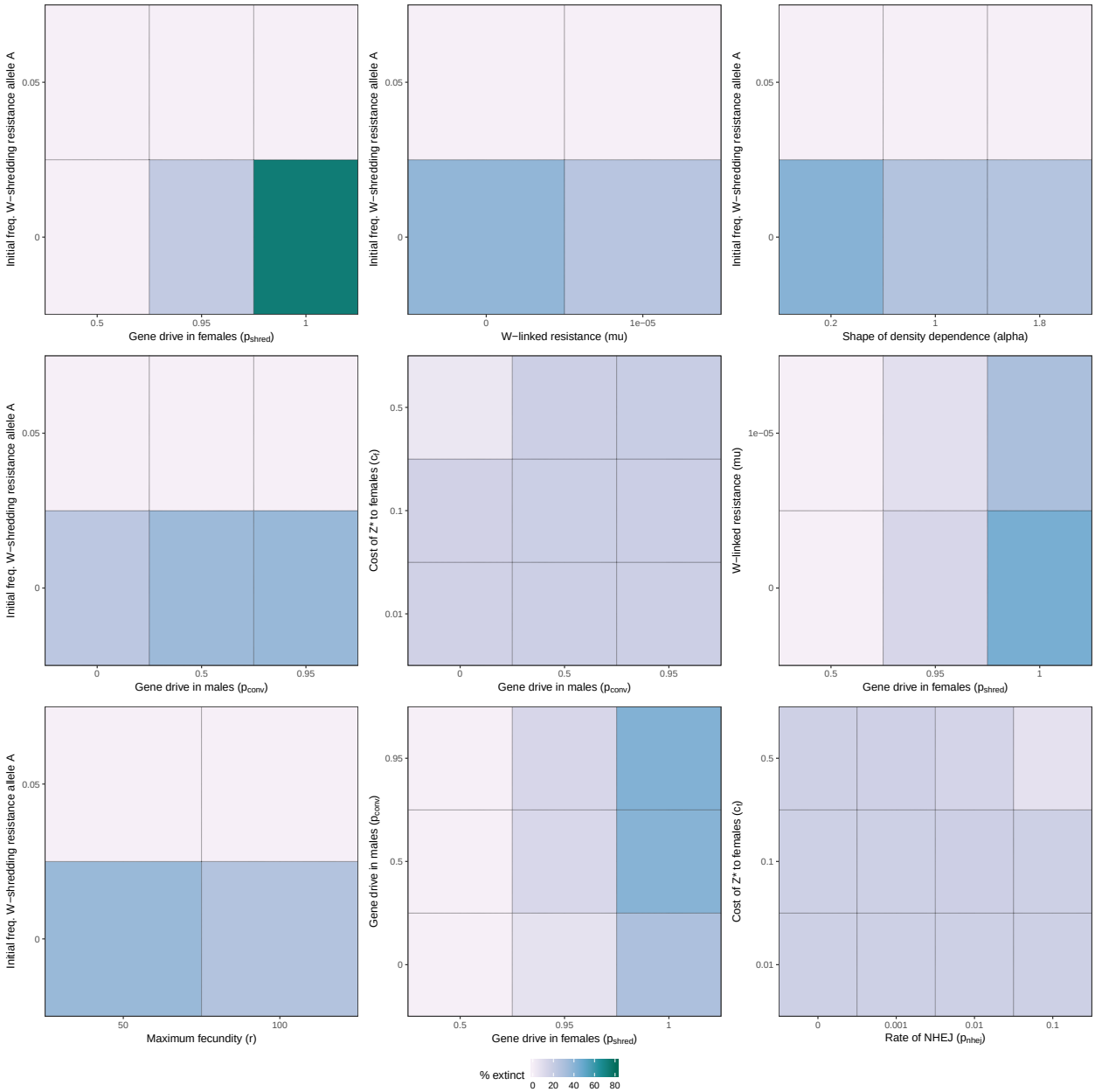

Figure S2: Heatmap illustrating the twelve strongest two-way interactions for simulations of a W-shredder, as determined by the effect sizes from the GLM plotted in Figure 4 (n = 5,657,700 simulation runs).

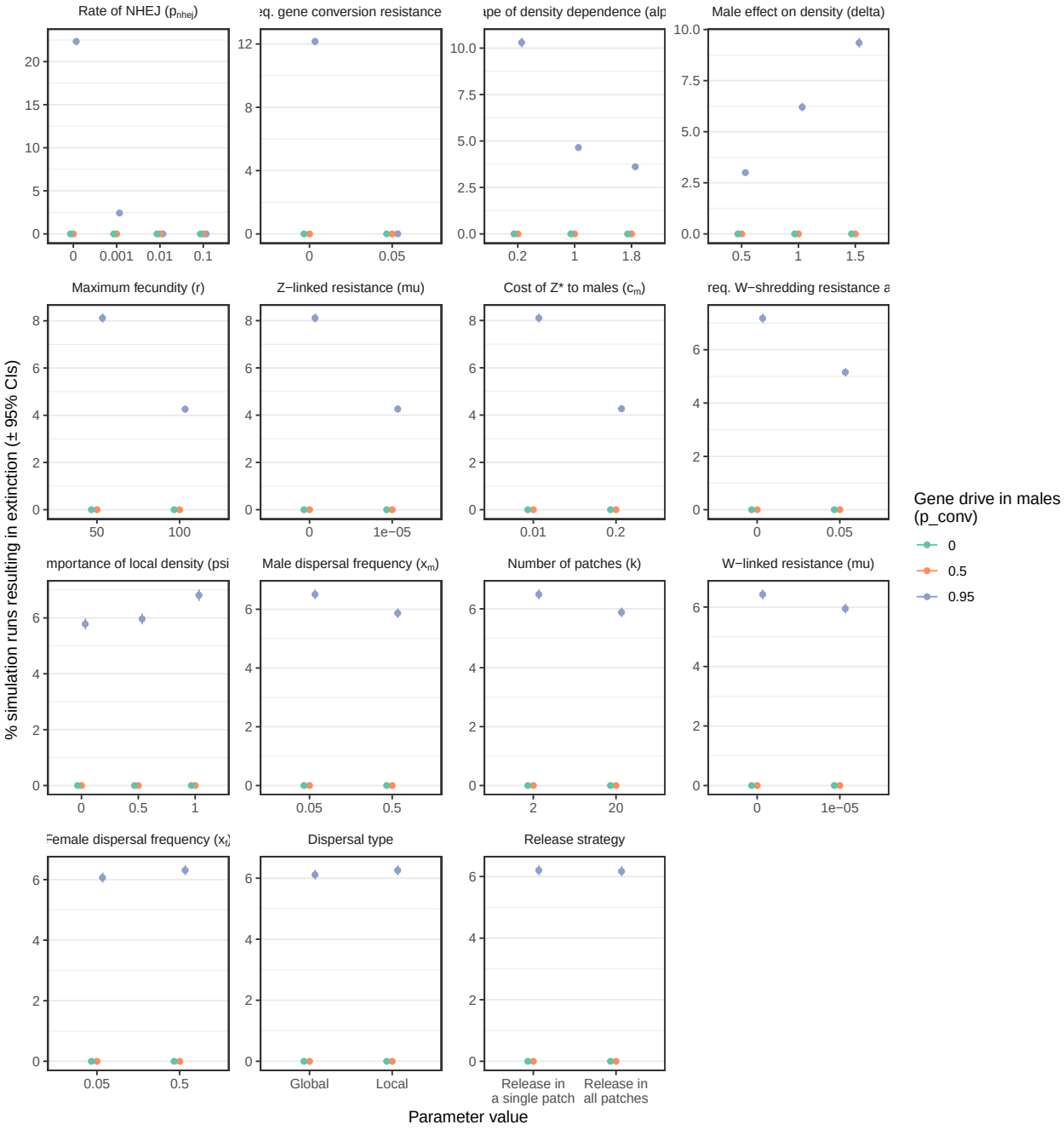

Figure S3: Analagous information to Figure 3, but showing the results for a female-sterilising Z^* allele instead of a W-shredder (n = 629,954 simulation runs).

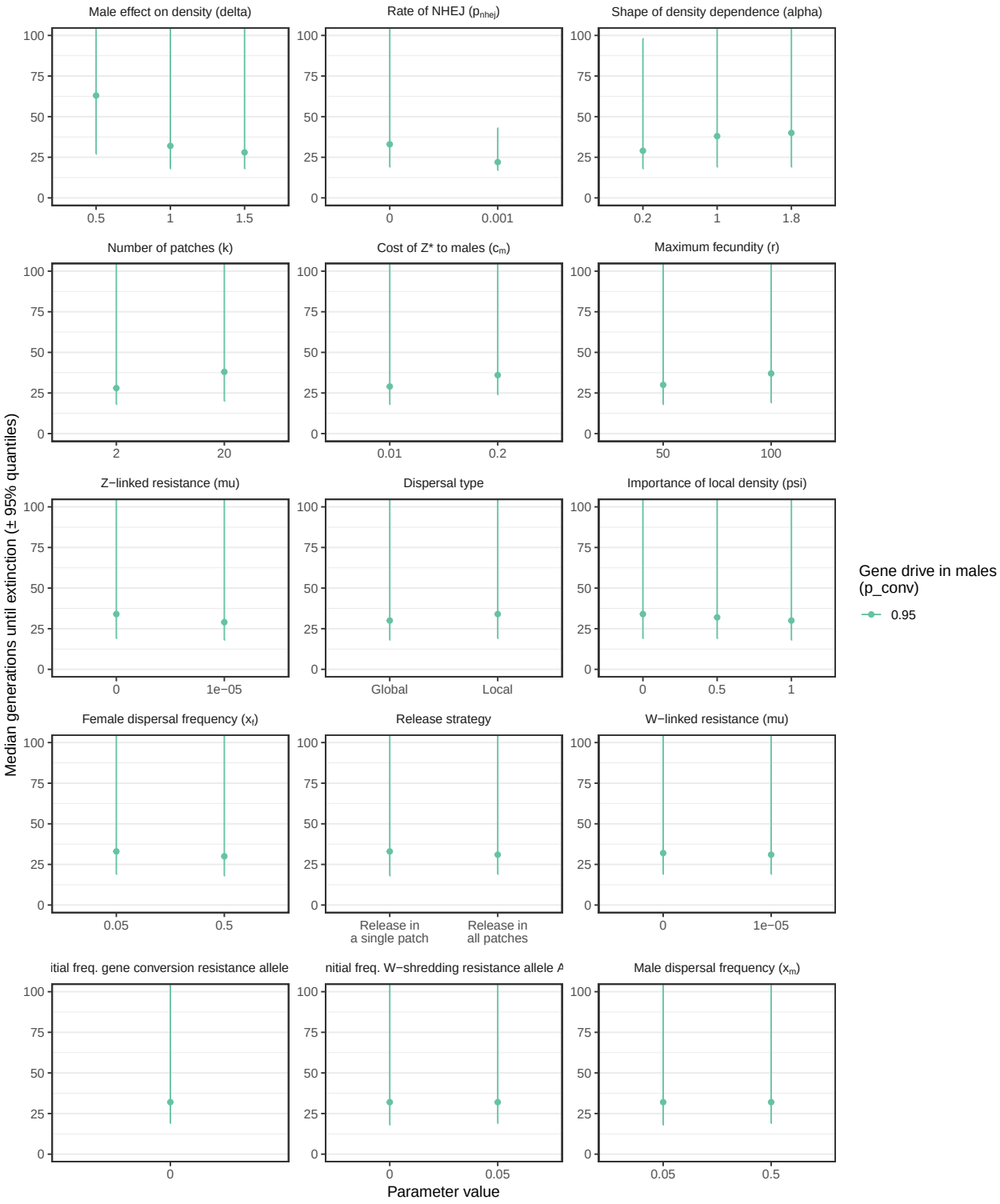

Figure S4: Analagous information to Figure S1, but showing the time to extinction for a female-sterilising Z^* allele instead of a W-shredder. Note that a median was only calculated if at least 40 simulation runs reached extinction (n = 12,992 simulation runs), and extinction only occurred when gene drive in males was strong ($p_{conv} = 0.95$).

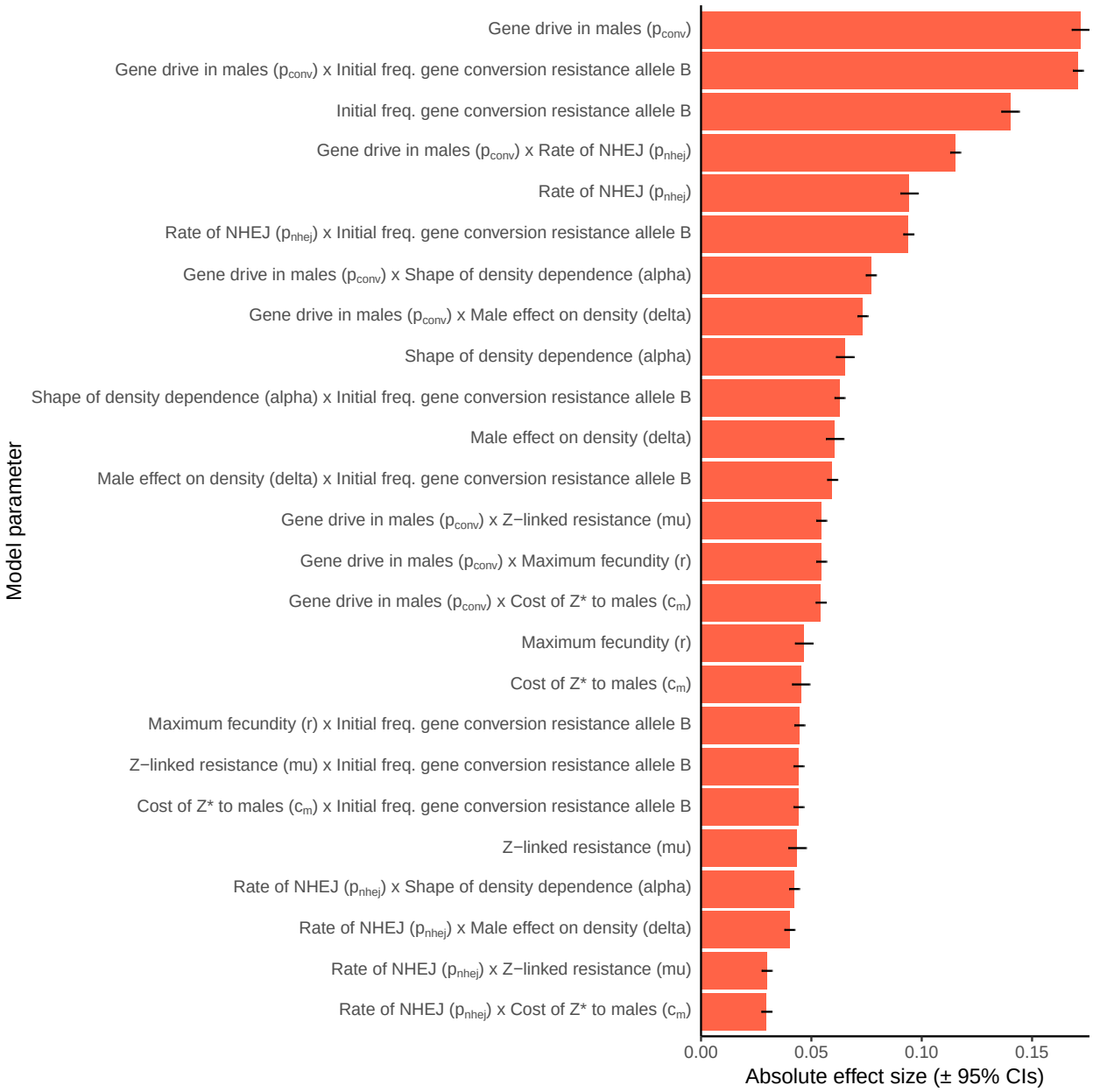

Figure S5: Relative parameter importance in the simulations of *Z*-linked female-sterilising gene drives, for the top 25 most important main effects or two-way interactions (from a binomial GLM that included all the main effects and all their two-way interactions; $n = 629{,}954$ simulation runs). Each predictor variable was scaled before running the model, meaning that the absolute effect size indicates how important each parameter is to the extinction probability, given the range of values plotted in Figure S3.

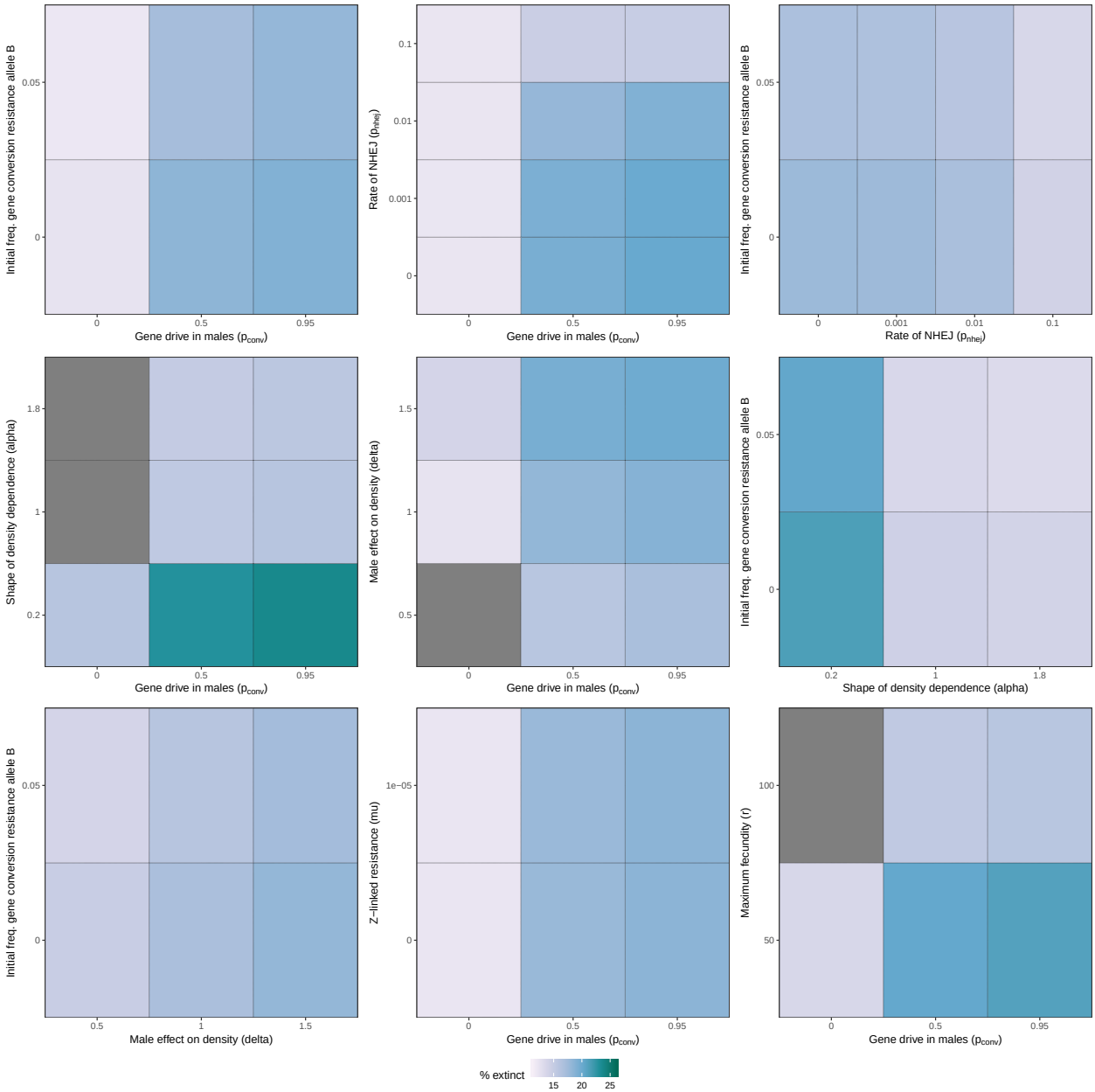

Figure S6: Heatmap illustrating the twelve strongest two-way interactions for simulations of a female-sterilising gene drive, as determined by the effect sizes from the GLM plotted in Figure S5 (n = 629,954 simulation runs).