Evading evolution of resistance to gene drives

Richard Gomulkiewicz*, **Micki L. Thies***, **and James J. Bull**[1]

***School of Biological Sciences, Washington State University, Pullman, WA 99164**
[1]**Dept of Biological Sciences, University of Idaho, Moscow, ID 83843**

Corresponding author:
R. Gomulkiewicz

Email address: gomulki@wsu.edu

ABSTRACT

Gene drives offer the possibility of altering and even suppressing wild populations of countless plant and animal species, and CRISPR technology now provides the technical feasibility of engineering them. However, population-suppression gene drives are prone to select resistance, should it arise. Here we develop mathematical and computational models to identify conditions under which suppression drives will evade resistance, even if resistance is present initially. We show that linkage between the resistance and drive loci is critical to the evolution, that evolution of resistance requires (negative) linkage disequilibrium between the two loci. When the two loci are unlinked or only partially so, a suppression drive that causes limited inviability can evolve to fixation while causing only a minor increase in resistance frequency. Once fixed, the drive allele no longer selects resistance. Single drives of this type would achieve only partial population suppression, but multiple drives (perhaps delivered sequentially) would allow arbitrary levels of suppression. Given that it is now possible to engineer CRISPR-based gene drives capable of circumventing allelic resistance, this design may allow for the engineering of suppression gene drives that are effectively resistance-proof.

INTRODUCTION

The ability to engineer gene drives has recently become feasible for countless species, especially insects, many of which carry disease or eat crops (Gantz and Bier, 2015; Champer et al., 2016; Kyrou et al., 2018; Oberhofer et al., 2019). Gene drives are genetic elements that use biased segregation or directed killing of alternative alleles to give themselves an 'unfair' evolutionary advantage, possibly to the detriment of their carriers. One possible use of engineered drives is to suppress populations ('suppression' drives), to reduce or even eradicate an entire species (Hamilton, 1967; Lyttle, 1977; Burt, 2003). Perhaps the major remaining biological hurdle in successfully implementing suppression drives is avoiding the evolution of resistance. In the early history of gene drive research, of both natural and experimental examples, resistance was the outstanding cause of the persistence of gene drives that would otherwise have extinguished populations or suppressed them to far greater levels (Lewontin and Dunn, 1960; Sandler et al., 1959; Sandler and Hiraizumi, 1959; Lyttle, 1979, 1981). Of course, examples from nature are necessarily devoid of past extinctions, so only experimental work can truly reflect on the *a priori* likelihood of resistance evolving to suppress a drive that will cause extinction.

Gene drive engineering has been profoundly facilitated by CRISPR technology (Gantz and Bier, 2015; Champer et al., 2016; Oberhofer et al., 2019). CRISPR uses an RNA-directed nuclease to cut specific sites in DNA, enabling a variety of gene drive designs from homing endonucleases that operate by segregation distortion to killer-rescue systems that destroy sensitive alleles. The most obvious form of resistance to CRISPR-based drives is mutation in the target sequence at which the nuclease cuts (Burt, 2003; Unckless et al., 2015, 2017; Champer et al., 2017). However, contrary to early suspicions that all possible target sites would be prone to such mutations (e.g., Drury et al., 2017), genomic sites have been identified that are apparently immutable, and the ability of single gene drives to direct CRISPR cutting at multiple sites in the same gene (any of which suffice to allow the drive to function) appear to ameliorate concerns about the inevitability of target-site resistance (Oberhofer et al., 2018; Champer et al., 2018; Kyrou et al., 2018). Other forms of resistance to CRISPR-based gene drives are theoretically possible, but it is too early to

tell whether they will prove to interfere with implementations: inbreeding, proteins that block CRISPR complex activity, and suppressors of CRISPR gene expression (Stanley and Maxwell, 2018; Bull et al., 2019). Experimental work by Lyttle 1979; 1981 using a natural gene drive mechanism raise the spectre of diverse and unpredictable forms of resistance to strongly suppressing gene drives, so resistance evolution is intrinsically plausible from our limited evidence.

All forms of resistance are not necessarily equal. Target-site resistance differs from other potential types of resistance in one important respect: linkage to the drive. For gene drives that operate as homing endonucleases, target-site resistance is allelic, segregating opposite the drive itself (Unckless et al., 2015, 2017; Drury et al., 2017). Many of the possible alternative types of resistance will not be allelic to the drive, they may even be completely unlinked (Lyttle, 1981; Champer et al., 2019). The purpose here is to understand the evolution of resistance to a population-suppressing gene drive when the resistance is potentially unlinked to the drive. Analytical models are developed for simple cases to understand the processes affecting the evolution. Loose linkage between the resistance locus and the drive locus works against the evolution of resistance for a wide range of conditions. This realization allows us identify strategies that may avoid the evolution of resistance to gene drives and still – by the sequential introduction of multiple drives – allow eventual population suppression to arbitrary levels.

RESULTS

Population genetics with non-allelic resistance

A two-locus model with selection in haploids (gametes) illustrates some key properties of selection on and evolution of gene drives. The setup for this model is extreme for purposes of illustration: segregation distortion is assumed to be 100%, the two loci are fully unlinked, and resistance is dominant, complete, and has no fitness cost; analyses in the supplement are more permissive (e.g., partial distortion or linkage). Segregation distortion operates at the A/a locus, resistance to segregation distortion operates at the R/r locus:

Diploid genotype	Gametes produced
Aa rr	Ar
Aa Rr	ar, aR, Ar, AR
Aa RR	AR, aR

Table 1. Segregation ratios for the three possible genotypes that are heterozygous at the drive locus. Only the genotype of the top row experiences segregation distortion, as any genome with allele *R* suppresses distortion. All of the gamete types listed for a diploid are produced in equal abundance; the double heterozygote produces all 4 in equal abundance because of the absence of linkage between the two loci. Segregation is Mendelian for the 6 other possible diploid genotypes not shown here – those involving *AA* and *aa* homozygotes at the drive locus.

Fitness effects are expressed in the gamete stage, with only the gene drive allele having an effect: The

Gamete type	Fitness
ar	1
Ar	$1-s_A$
aR	1
AR	$1-s_A$

Table 2. Haplotype fitnesses. s_A is the coefficient of selection agains the drive allele.

following sections explain in stepwise fashion how unlinked resistance is selected.

The non-drive allele is depressed by segregation distortion but bolstered by its fitness advantage

In the absence of resistance alleles, the net effect of each life cycle is to further depress the a allele frequency. But this net effect is the outcome of the opposing effects of segregation distortion and gametic fitness. Assuming transmission occurs prior to selection against the drive, the frequency of allele a after transmission but before selection (p_a^*) as a function of its frequency at the start of the generation (p_a) is

$$p_a^* = p_a - H/2 \tag{1}$$

where H is the frequency of diploid genotypes that are resistance-free and heterozygous for the drive ($Aarr$). The term $H/2$ describes the depression of allele a from segregation distortion. Note that resistance does nothing to blunt this negative effect.

Selection against the drive subsequently counters this. The frequency of a is increased from p_a^* by the factor $1/\overline{w}$,

$$p_a' = p_a^* \left(\frac{1}{\overline{w}} \right), \tag{2}$$

where p_a' denotes the frequency of a at the start of the next generation and $\overline{w} = 1 - p_A^* s_A$ is mean fitness, necessarily less than 1 when the drive allele is present. The term in parens represents the gain of allele a from its fitness advantage over A. The loss in (1) is always larger than the gain (2) when the drive allele is spreading. However, it is the separation of these effects that is critical to understanding the evolution of resistance.

Resistance has no advantage in the transmission stage

This point is perhaps the most surprising, but is easy to understand. Considering the resistance allele R before and after the segregation distortion stage, there is no change in its frequency:

$$p_R^* = p_R \tag{3}$$

(Supplement eq. 11; see also eq. 18b). This result may be understood from the fact that every genotype with R segregates normally. Even in the case of the $AaRr$ double heterozygote, the lack of segregation distortion (recall R is dominant) means that R will on average occur in half the gametes (regardless of the haplotypes that formed the double heterozygote), the same frequency as in the parent. This holds even if the loci are linked. Segregation distortion occurs only in the $Aarr$ dipoid, and all gametes carry r, so again there is no change in its frequency.

Given that neither R nor r has an effect on gametic fitness, there is no intrinsic advantage that one allele has over the other during the transmission stage. However, the relative benefit of resistance changes when we consider associations among loci, as next.

Synergy between resistance and fitness of the non-drive allele

The foregoing two sections establish that the non-drive allele a gains in survival over A each generation but loses potentially more from segregation distortion. By itself, the resistance allele has no fitness advantage and also does not benefit by suppressing distortion at the other locus. Both outcomes change when we consider resistance combined with the fitness advantage of the non-drive allele. The two alleles synergize when they occur in the same genotype. When associated with R, the a allele is emancipated from its loss in segregation distortion because segregation distortion is abolished; now its only deviation from neutrality is the gain due to its intrinsic fitness advantage in gametes. In turn, allele R gains by segregating with a, experiencing the fitness advantage a has in gametes. Thus, both R and a benefit from their co-occurrence, each allele providing a benefit to the other.

For this association to provide a net benefit across the population, the alleles of the different loci must be statistically associated. In population genetics terms, there must be (negative) linkage disequilibrium, which can be defined as

$$D = p_a p_R - x_{aR} \tag{4}$$

where x_{aR} is the frequency of gametes that contain both a and R; $D < 0$ indicates that a and R occur together more frequently than at random. Note that D is bounded below by the larger of $-(1-p_a)p_R$ and $-p_a(1-p_R)$ (e.g., Crow and Kimura, 1970). This lower bound to the strength of the association between a and R approaches zero in particular with the frequency of the drive allele A.

During transmission, free recombination halves the magnitude of D but, seperately, segregation distortion forces the association between a and R closer:

$$D^* = D/2 - p_R H/2. \tag{5}$$

The subtracted term shows that the amount of association created by meiotic drive increases with the frequencies of resistance-free drive heterozygotes (H) and the resistance allele itself. If D is initially

positive, recombination and segregation distortion work in concert to reduce it. Once D is negative, the two forces work in opposition with segregation distortion exaggerating and recombination eroding the association between a and R.

The statistical association between a and R has consequences for R when a is selected. After selection its frequency becomes

$$p'_R = p_R - D^* \left(\frac{s_A}{\overline{w}} \right). \tag{6}$$

This result shows that negative linkage disequilibrium ($D^* < 0$) is required for R to rise in frequency – that a and R must exist together in the same gamete more often than random. Thus, non-allelic resistance increases by genetic hitchhiking. (Maynard Smith and Haigh, 1974).

Versions of these dynamics can be derived that account for incomplete segregation distortion, partial linkage between drive and resistance loci, costly resistance, and diploid selection, including lethal recessive drives (see Supplement). These models show that the scope for resistance evolution increases with more effective segregation distortion, closer linkage, and less costly resistance. Despite their quantitative influences on the evolution of resistance, those features do not affect the three main features emphasized above: (i) Segregation distortion has no direct effect on resistance (unless allelic to the drive). (ii) Segregation distortion directly promotes a tighter statistical association between a and R. (iii) Alleles a and R must co-occur in gametes more often than at random (i.e., negative linkage disequilibrium is required) for resistance to rise in frequency via hitchhiking.

Evading resistance evolution is possible and practical

The preceding results facilitate developing protocols for suppression drives that diminish the evolution of resistance. Prior to exploring whether such protocols are theoretically feasible, we offer several points to serve as intuitive guiding principles.

1. Allelic resistance should be avoided. Resistance is least prone to evolve when weakly linked to the drive. Some designs using CRISPR homing endonucleases can greatly limit the opportunity for allelic resistance.

2. Resistance may impose a cost, independent of the drive itself. Any cost will work against resistance evolution.

3. The drive should be designed to evolve to fixation. Even if resistance arises, it cannot maintain the required linkage disequilibrium once the drive allele fixes, so it will cease to function as the drive sweeps to high levels. Persistent heterozygosity of a drive allele provides an indefinite opportunity for resistance to reverse the drive at any time prior to species extinction. Drive fixation precludes future reversals.

4. The drive homozygote should be viable and fertile, with a substantial but not extreme reduction ($\sim$10%–20%) in fitness. A modest fitness effect not only facilitates drive fixation (point 3, above) but also avoids the spread of resistance alleles with a high fitness cost.

We offer numerical analyses of a model that fulfils many of the preceding points (supplement 2?). Unlike the model used above for mathematical analysis, this numerical model assigns fitness effects to diploids – a design easily implemented with current technology. Additional details include the following. (i) The drive is expressed in one sex only (males, for the sake of illustration). A drive that operates in both sexes may be equally useful, but perhaps harder to implement and is not necessary in many practical contexts. (ii) The drive operates with 100% segregation distortion in males, has no effect on heterozygote fitness, but imposes a fitness effect of s_{AA} on the homozygote. A value of s_{AA} substantially less than 1 reflects the drive's mild population suppression effect. In some implementations, the fitness effect accrues to homozygotes of both sexes; in others just to one sex. (iii) Resistance (allele R) is dominant and fully suppresses the segregation distortion of the drive. Total suppression represents the most favorable case of its evolution. Any diploid carrying R realizes a fitness-reducing cost s_R, equally in Rr heterozygotes and RR homozygotes; fitness is $1 - s_R$. (iv) Both the drive allele and resistance allele are introduced at low frequency into the population. The two loci are unlinked in most trials. (A version of this model with a recessive lethal gene drive, $s_{AA} = 1$, that manifests in both sexes is analyzed in the Supplement.)

Figure 1 shows long term (equilibrium) results of trials in which the homozygote fitness ($1 - s_{AA}$) is varied along with the cost of resistance, s_R. The drive evolves to fixation (loss of a) provided $s_{AA} \leq 0.5$). It is immediately apparent that drives with even substantial homozygous fitness effects (up to $s_{AA} = 0.35$) evolve to fixation, even when resistance imposed no fitness cost, but the outcome of evolution depends on the starting frequency of resistance (compare top and bottom panels).

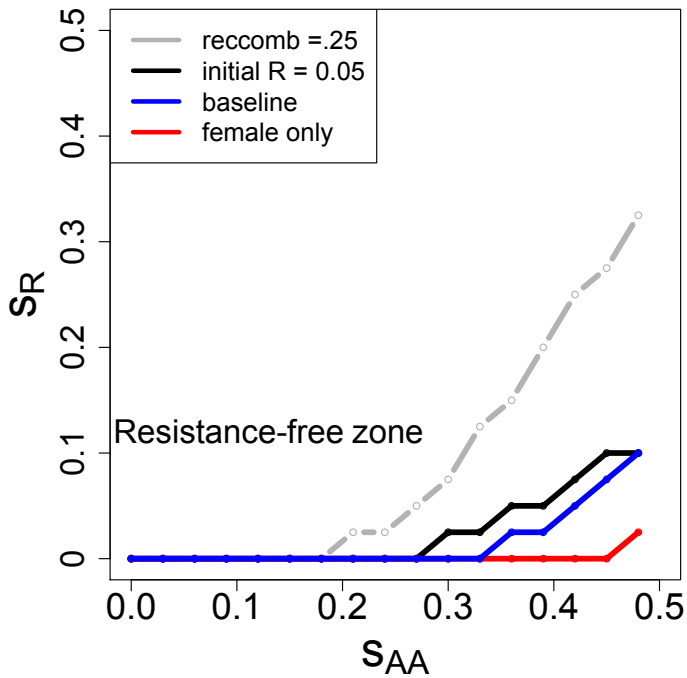

Figure 1. Based on numerical trials, some kinds of suppression gene drives can evolve to fixation despite the presence of resistance alleles in the population. Gene drives evolve to fixation above and to the left of the curves, in the 'resistance-free' zone. The baseline trial (blue) assumes drive is 100% efficient in males (females do not experience drive), the drive allele starts at frequency 0.005, resistance at 0.015. The drive allele impairs fitness of homozygotes of both sexes, their fitness being $1 - s_{AA}$. Resistance to the drive's segregation distortion is unlinked, dominant, complete, and impairs fitness of Rr and RR gentypes as $1 - s_R$. The pattern is that even cost-free resistance does not evolve until $s_{AA} > 0.3$, and at higher values of s_{AA}, resistance evolves and blocks fixation of the drive allele only if its cost (s_R) is sufficiently low. Other curves deviate from the baseline case in single respects: (red) the fitness effect of drive homozygotes s_{AA} is experienced only in females; (black) the initial frequency of resistance is increased to 0.05; (gray) the drive and resistance loci are linked with a recombination rate of 0.25. The red line is flat out to $s_{AA} = 0.54$ if the fitness effect is assigned only to males. All trials were run at least 700 generations.

Figure 2 shows the frequency dynamics of resistance, drive and disequilibrium over time for a single trial. Resistance does experience a modest gain during gene drive evolution, reflecting the dynamic nature of linkage disequilibrium, but the increase is not enough to halt drive fixation at least for these initial conditions. Resistance will 'ratchet' itself up during the process, even when it does not ultimately block fixation of the drive. This level of resistance will remain in the population unless resistance imposes a fitness cost.

DISCUSSION

In theory, gene drives offer an ability to quickly alter the genetics of populations, with two important uses: modification and suppression (Sandler and Novitski, 1957; Hamilton, 1967; Lyttle, 1977; Burt, 2003;

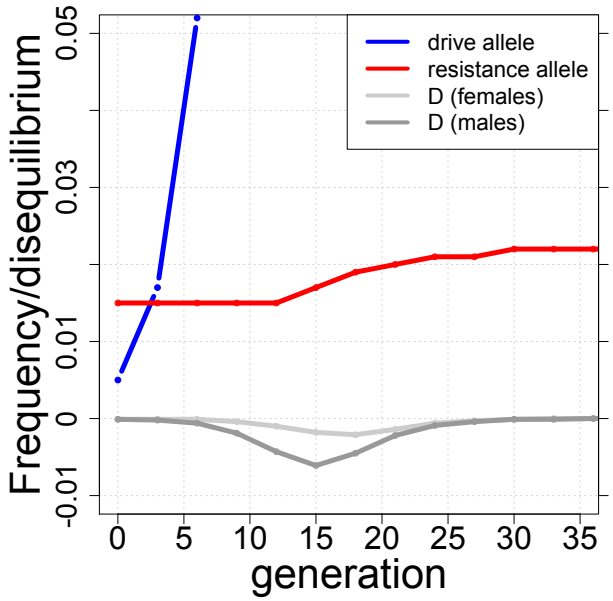

Figure 2. Dynamics of gene drive, resistance, and linkage disequilibrium. Linkage disequilibrium (D) is more pronounced in males (the sex experiencing the drive). The resistance allele (red) rises via indirect response to selection against the drive allele A during the transient submersion of disequilibrium below 0. For this trial, $s_{AA} = 0.2$, the resistance allele started at frequency 0.015 and entailed no cost ($s_R = 0$). This trial otherwise corresponded to the baseline case in Fig. 1.

Gould et al., 2006, 2008; Gould, 2008). Modification drives facilitate the spread of genes without greatly affecting the population in other ways. Suppression drives, our focus here, can be used to permanently depress population densities, even to the point of extinction. Suppression is simply a matter of creating a gene drive that has deleterious effects in drive homozygotes (Burt, 2003). CRISPR endonucleases now trivially enable the engineering of a type of drive known as a homing endonucleases (HEGs) that can be used for suppression. The drive is merely engineered to disrupt a gene of arbitrary importance to fitness, such that heterozygotes are nearly normal but drive homozygotes are impaired (Burt, 2003).

Homing endonuclease genes work on the principle that the enzyme cuts a specific and unique DNA target sequence in the chromosome (Burt, 2003). Homology-based DNA repair mechanisms of the cell recognize the cut ends and fill in the gap by copying from the homologous chromosome. Thus in a heterozygote for the homing endonuclease gene, if the endonuclease is encoded opposite the cut site, the cell will repair the cut by copying the endonuclease gene into the other chromosome, in essence turning the heterozygote into a homozygote. Given this simple mechanism, it initially seemed that any suppression homing endonuclease gene would quickly select resistance in the form of a mutation in the target sequence (Burt, 2003; Unckless et al., 2015, 2017; Drury et al., 2017). Resistance would then be allelic to the drive. However, this type of resistance now seems generally avoidable by designing CRISPR endonucleases to several sites within the target gene, only one of which needs to be wild-type for the drive to function (Champer et al., 2018; Oberhofer et al., 2018). Other types of resistance to suppression drives are possible in principle, such as micro RNAS that block expression and function of CRISPR, inbreeding, and proteins that interfere with the CRISPR complex (Bull, 2016; Stanley and Maxwell, 2018). These various mechanisms will not typically be allelic to the drive and may even be fully unlinked (Lyttle, 1981; Champer et al., 2019). They may also entail a cost to fitness (Lyttle, 1981).

As shown here, the evolution of unlinked resistance can be avoided with suppression drives that impose only a modest fitness cost on the organism. Non-allelic resistance evolves by genetic hitchhiking from (negative) linkage disequilibrium with the drive locus. If the drive allele can evolve quickly to fixation then D will decay rapidly to zero, precluding further increases in resistance. Given that any single drive would provide only modest levels of population suppression, arbitrary levels of suppression

could then be achieved by careful choice of suppression drives introduced in succession, with no threat of resistance evolving to block it.

How feasible is population suppression by this approach? The most assured design for any single drive is to target it to destroy a gene that is haplo-sufficent but compromises fitness moderately (not severely) in null homozygotes. Population extinction would likely require a succession of such gene drives, each targeting different genes, although simultaneous release multiple drives may offer a solution. Choose a target gene with too strong of a fitness effect, and the drive will not fix and may well select resistance – thwarting all future gene drives that work by the same mechanism. Choice of an appropriate target gene may be relatively easy for the first drive release but may become increasingly challenging (without sufficient background work) as the organism's fitness declines from previous drive fixations. Nonetheless, various tricks might be employed to ensure success, such as choosing a target gene with a delayed effect on population fitness (e.g., important in defense against an initially uncommon predator or important in surviving an environmental factor that can be controlled). But another drawback of sequential releases is the increased time to attain a combined effect.

Our study has explored the evolution of resistance against drives that operate by transmission distortion, and then only for a subset of possible conditions, such as dominant resistance. We conjecture that similar outcomes can be obtained for other systems and conditions, perhaps drives that operate as killer-rescue systems (Gould et al., 2008; Marshall and Hay, 2011; Oberhofer et al., 2019). Given concerns about the potential of gene drives to do unintended harm, especially transmission drives with highly deleterious effects on the homozygote (Collins, 2018; Report, 2016), the approach identified here may alleviate some of these concerns. Thus, any single suppression drive with mild fitness consequences may be introduced to reduce the economic or public health burden of a pest without causing pest extinction. Escape of of a 'mild' drive into other species is also expected to be less consequential than is escape of an extremely deleterious drive. And drives that have only modest fitness effects may be feasible by targeting genes that are poorly conserved between species, reducing their likelihood of escape. However, whatever genomic target is chosen, it must be sufficiently invariant that the drive can fix throughout the entire species.

Fixation of a gene drive does not necessarily end the evolutionary response to a gene drive. Loss of gene function throughout a species may select compensatory evolution in other genes, ameliorating the impact (Burt 03). Furthermore, any off-target activity of CRISPR may select suppression of CRISPR activity (Jeong et al., 2020), thereby limiting opportunities to introduce additional drives to further suppress the population. Implementations of resistance-free drives will thus need to be monitored for long term effects.

As is true of most studies of gene drive resistance, our models assumed fully mixed populations, a state especially conducive to gene drive spread. Population structure, a key component of many natural populations, will retard gene drive spread and may even qualitatively alter outcomes. It will be interesting to extend the present analyses to population structure to see how selection of resistance is altered (North et al., 2013; Beaghton et al., 2016; North et al., 2019; Champer et al., 2020). Minimally, we expect structure to alter the dynamics of linkage disequilibrium that is key to the evolution of resistance. Inbreeding, as one form of structure, is known to hamper the spread of suppression drives (refs) directly but also impacts the process of genetic hitchhiking (Hedrick, 1980) which could indirectly favor the rise of non-allelic resistance. It may also affect the evolution of other types of resistance.

ACKNOWLEDGMENTS

We thank B. Hay for comments. RG was supported by a WSU Honors College Distinguished Professorship, MLT by a WSU College of Arts and Sciences Undergraduate Research grant, and JJB by NIH grant GM 122079.

SUPPLEMENT

Model of costly resistance to a linked gene drive: haploid selection

Consider the model described in the main text but assume the gene drive and resistance loci are linked with recombination rate c ($0 \leq c \leq \frac{1}{2}$) and that the fitness of any R-bearing haplotypes is reduced by a multiplicative factor $1 - s_R$. We also allow for partial segregation distortion such that each resistance-free drive locus heterozygote produces $\frac{1}{2} + \delta$ A and $\frac{1}{2} - \delta$ a alleles, where δ is the distortion ($0 \leq \delta \leq \frac{1}{2}$). Complete distortion occurs if $\delta = \frac{1}{2}$.

At the beginning of a generation, the frequency of gametes with i drive alleles A and j resistance alleles R is denoted x_{ij} for $i, j, = 0, 1$. The four frequencies sum to 1 can be used to compute allele frequencies and disequlibrium: $p_A = 1 - p_a = x_{10} + x_{11}$, $p_R = 1 - p_r = x_{01} + x_{11}$, and $D = x_{00}x_{11} - x_{01}x_{10}$.

Diploids are formed by random union of these gametes. Let Z_{ij} be frequency of diploids with i copies of A and j copies of R, then

$$Z_{ij} = \sum_{i'+i''=i} \sum_{j'+j''=j} x_{i'j'} x_{i''j''}. \tag{7}$$

Note that the frequency of resistance-free drive heterozygotes is

$$H = Z_{10} = 2x_{00}x_{10}. \tag{8}$$

After transmission, the frequencies of gametes produced by the diploids are

$$\begin{aligned}
x^*_{00} &= Z_{00} + \tfrac{1}{2}Z_{01} + (\tfrac{1}{2} - \delta)Z_{10} + (1-c)x_{00}x_{11} + cx_{01}x_{10} \\
x^*_{01} &= Z_{02} + \tfrac{1}{2}Z_{12} + \tfrac{1}{2}Z_{01} + cx_{00}x_{11} + (1-c)x_{01}x_{10} \\
x^*_{10} &= Z_{20} + \tfrac{1}{2}Z_{21} + (\tfrac{1}{2} + \delta)Z_{10} + cx_{00}x_{11} + (1-c)x_{01}x_{10} \\
x^*_{11} &= Z_{22} + \tfrac{1}{2}Z_{21} + \tfrac{1}{2}Z_{12} + (1-c)x_{00}x_{11} + cx_{01}x_{10}
\end{aligned}$$

These are equivalent to

$$x^*_{00} = x_{00} - cD - \delta H \tag{9a}$$
$$x^*_{01} = x_{01} + cD \tag{9b}$$
$$x^*_{10} = x_{10} + cD + \delta H \tag{9c}$$
$$x^*_{11} = x_{11} - cD \tag{9d}$$

One can see directly from equations (9) that, for any distortion δ or recombination rate c,

$$p^*_a = x^*_{00} + x^*_{01} = p_a - \delta H \tag{10}$$

and

$$p^*_R = x^*_{01} + x^*_{11} = p_R. \tag{11}$$

These equations can also be used to show that disequilibrium after transmission is

$$D^* = (1-c)D - \delta H p_R. \tag{12}$$

Selection on the haploid genotypes after transmission results in the following frequencies in the next generation

$$\begin{aligned}
x'_{00} &= x^*_{00}/\bar{w} \\
x'_{01} &= x^*_{01}(1-s_R)/\bar{w} \\
x'_{10} &= x^*_{10}(1-s_A)/\bar{w} \\
x'_{11} &= x^*_{11}(1-s_A)(1-s_R)/\bar{w}
\end{aligned}$$

where

$$\begin{aligned}
\overline{w} &= x^*_{00} + x^*_{10}(1-s_A) + x^*_{01}(1-s_R) + x^*_{11}(1-s_A)(1-s_R) \\
&= (1-p^*_A s_A)(1-p^*_R s_R) + D^* s_A s_R
\end{aligned} \tag{13}$$

These expressions can be used to show that the corresponding frequencies of the drive and resistance alleles are

$$p'_A = p^*_A \frac{(1-s_A)(1-p^*_R s_R)}{\bar{w}} - D^* \frac{s_R(1-s_A)}{\bar{w}} \tag{14a}$$
$$p'_R = p^*_R \frac{(1-s_R)(1-p^*_A s_A)}{\bar{w}} - D^* \frac{s_A(1-s_R)}{\bar{w}} \tag{14b}$$

If resistance has no cost, $s_R = 0$, equations (14) are equivalent to equations (2) and (6). However, selection against resistance, $s_R > 0$, could indirectly promote the drive allele A if the loci have a negative association (second term in eq. 14a). For $s_A > \frac{1}{2}$, this model has an unstable equilibrium with no resistance ($\hat{p}_R = 0$) and polymorphism for the drive

$$\hat{p}_A = \frac{s_A - 2\delta(1 - s_A)}{s_A^2}. \tag{15}$$

Model of a lethal gene drive with linked, costly resistance

In this section, we describe a model similar to the above except that the gene drive is recessive lethal. Assuming death occurs after diploids are formed as above (Eqs. 7) but before transmission, the post-death frequencies (indicated by the symbol #) are

$$Z_{20}^{\#} = Z_{21}^{\#} = Z_{22}^{\#} = 0$$

and, for $i = 0, 1$,

$$Z_{ij}^{\#} = Z_{ij}/\bar{w} \tag{16}$$

where

$$\overline{w} = \sum_{i=0}^{1}\sum_{j=0}^{2} Z_{ij} = 1 - p_A^2, \tag{17}$$

the fraction surviving. From this it can be shown that

$$p_A^{\#} = p_A - \frac{p_A^2 q_A}{\bar{w}} \tag{18a}$$

$$p_R^{\#} = p_R - D\frac{p_A}{\bar{w}}. \tag{18b}$$

Equation (18b) shows that when $D < 0$ the benefit to resistance from lethality increases with the frequency of the drive.

Transmission from the surviving genotypes produces the following gamete frequencies:

$$\begin{aligned}
x_{00}^* &= (x_{00} - cD - \delta H)/\overline{w} \\
x_{01}^* &= (x_{01} + cD)/\overline{w} \\
x_{10}^* &= (x_{10}p_a + cD + \delta H)/\overline{w} \\
x_{11}^* &= (x_{11} - cD)/\overline{w}
\end{aligned}$$

Note that

$$p_a^* = p_a^{\#} - \delta H^{\#} \tag{19}$$

and

$$p_R^* = p_R^{\#} \tag{20}$$

where $H^{\#} = Z_{10}^{\#} = Z_{10}/\overline{w}$ is the frequency of resistance-free heterozygotes among survivors.

Finally, haploid selection on resistance results in the following gamete frequencies at the start of the next generation

$$\begin{aligned}
x'_{00} &= x_{00}^*/\overline{w}^* \\
x'_{01} &= x_{01}^*(1 - s_R)/\overline{w}^* \\
x'_{10} &= x_{10}^*/\overline{w}^* \\
x'_{11} &= x_{11}^*(1 - s_R)/\overline{w}^*
\end{aligned}$$

where

$$\overline{w}^* = x_{00}^* + x_{10}^* + x_{01}^*(1-s_R) + x_{11}^*(1-s_R) = 1 - p_R^* s_R = 1 - p_R^\# s_R. \tag{21}$$

The corresponding frequencies of the drive and resistance alleles are

$$p_A' = p_A^* - D^* \frac{s_R}{\overline{w}^*} \tag{22a}$$

$$p_R' = p_R^* \left(\frac{1-s_R}{\overline{w}^*} \right) = p_R^\# - \frac{p_R^\#(1-p_R^\#)s_R}{1-p_R^\# s_R} \tag{22b}$$

REFERENCES

Beaghton, A., Beaghton, P. J., and Burt, A. (2016). Gene drive through a landscape: Reaction-diffusion models of population suppression and elimination by a sex ratio distorter. *Theoretical Population Biology*, 108:51–69.

Bull, J. J. (2016). Lethal gene drive selects inbreeding. *Evolution, Medicine, and Public Health*, 2017(1):1–16.

Bull, J. J., Remien, C. H., and Krone, S. M. (2019). Gene-drive-mediated extinction is thwarted by population structure and evolution of sib mating. *Evolution, Medicine, and Public Health*, 2019(1):66–81.

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings. Biological Sciences / The Royal Society*, 270(1518):921–928.

Champer, J., Buchman, A., and Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews. Genetics*, 17(3):146–159.

Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., Clark, A. G., and Messer, P. W. (2018). Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):5522–5527.

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., and Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS genetics*, 13(7):e1006796.

Champer, J., Wen, Z., Luthra, A., Reeves, R., Chung, J., Liu, C., Lee, Y. L., Liu, J., Yang, E., Messer, P. W., and Clark, A. G. (2019). Crispr gene drive efficiency and resistance rate is highly heritable with no common genetic loci of large effect. *Genetics*, 212(1):333–341.

Champer, J., Zhao, J., Champer, S. E., Liu, J., and Messer, P. W. (2020). Population dynamics of underdominance gene drive systemss in continuous space. *ACS synthetic biology*, 9(4):779–792.

Collins, J. P. (2018). Gene drives in our future: challenges of and opportunities for using a self-sustaining technology in pest and vector management. *BMC proceedings*, 12(Suppl 8):9.

Crow, J. F. and Kimura, M. (1970). *An introduction to population genetics theory*. Harper & Row, Publishers, New York, NY.

Drury, D. W., Dapper, A. L., Siniard, D. J., Zentner, G. E., and Wade, M. J. (2017). CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science Advances*, 3:e1601910.

Gantz, V. M. and Bier, E. (2015). Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science (New York, N.Y.)*, 348(6233):442–444.

Gould, F. (2008). Broadening the application of evolutionarily based genetic pest management. *Evolution*, 62(2):500–510.

Gould, F., Huang, Y., Legros, M., and Lloyd, A. L. (2008). A killer-rescue system for self-limiting gene drive of anti-pathogen constructs. *Proceedings. Biological Sciences / The Royal Society*, 275(1653):2823–2829.

Gould, F., Magori, K., and Huang, Y. (2006). Genetic strategies for controlling mosquito-borne diseases. *American Scientist*, 94:238 – 346.

Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science (New York, N.Y.)*, 156(3774):477–488.

Hedrick, P. W. (1980). Hitchhiking: a comparison of linkage and partial selfing. *Genetics*, 94(3):791–808.

Jeong, Y. K., Song, B., and Bae, S. (2020). Current status and challenges of dna base editing tools. *Molecular Therapy: The Journal of the American Society of Gene Therapy*.

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T., and Crisanti, A. (2018). A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*.

Lewontin, R. C. and Dunn, L. C. (1960). The evolutionary dynamics of a polymorphism in the house mouse. *Genetics*, 45(6):705–722.

Lyttle, T. W. (1977). Experimental population genetics of meiotic drive systems. I. Pseudo-Y chromosomal drive as a means of eliminating cage populations of Drosophila melanogaster. *Genetics*, 86(2 Pt. 1):413–445.

Lyttle, T. W. (1979). Experimental population genetics of meiotic drive systems II. Accumulation of genetic modifiers of segregation distorter (SD) in laboratory populations. *Genetics*, 91(2):339–357.

Lyttle, T. W. (1981). Experimental population genetics of meiotic drive systems. III. Neutralization of sex-ratio distortion in Drosophila through sex-chromosome aneuploidy. *Genetics*, 98(2):317–334.

Marshall, J. M. and Hay, B. A. (2011). Inverse Medea as a novel gene drive system for local population replacement: a theoretical analysis. *The Journal of Heredity*, 102(3):336–341.

Maynard Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35.

North, A., Burt, A., Godfray, H. C. J., and Buckley, Y. (2013). Modelling the spatial spread of a homing endonuclease gene in a mosquito population. *The Journal of Applied Ecology*, 50(5):1216–1225.

North, A. R., Burt, A., and Godfray, H. C. J. (2019). Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC biology*, 17(1):26.

Oberhofer, G., Ivy, T., and Hay, B. A. (2018). Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40):E9343–E9352.

Oberhofer, G., Ivy, T., and Hay, B. A. (2019). Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences of the United States of America*, 116(13):6250–6259.

Report, N. A. (2016). Gene drive research in non-human organisms: Recommendations for responsible conduct.

Sandler, L. and Hiraizumi, Y. (1959). Meiotic drive in natural populations of Drosophila melanogaster. II. Genetic variation at the segregation-distorter locus. *Proceedings of the National Academy of Sciences of the United States of America*, 45(9):1412–1422.

Sandler, L., Hiraizumi, Y., and Sandler, I. (1959). Meiotic drive in natural populations of Drosophila melanogaster. I. The cytogenetic basis of segregation-distortion. *Genetics*, 44(2):233–250.

Sandler, L. and Novitski, E. (1957). Meotic drive as an evolutionary force. *The American Naturalist*, 91(857):105–110.

Stanley, S. Y. and Maxwell, K. L. (2018). Phage-3ncoded anti-CRISPR defenses. *Annual Review of Genetics*.

Unckless, R. L., Clark, A. G., and Messer, P. W. (2017). Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics*, 205(2):827–841.

Unckless, R. L., Messer, P. W., Connallon, T., and Clark, A. G. (2015). Modeling the manipulation of natural populations by the mutagenic chain reaction. *Genetics*, 201(2):425–431.