Modeling the mutation and reversal of engineered underdominance gene drives

Matthew P. Edgington*, Luke S. Alphey

The Pirbright Institute, Ash Road, Woking, Surrey GU24 0NF, UK

ABSTRACT

A range of gene drive systems have been proposed that are predicted to increase their frequency and that of associated desirable genetic material even if they confer a fitness cost on individuals carrying them. Engineered underdominance (UD) is such a system and, in one version, is based on the introduction of two independently segregating transgenic constructs each carrying a lethal gene, a suppressor for the lethal at the other locus and a desirable genetic "cargo". Under this system individuals carrying at least one copy of each construct (or no copies of either) are viable whilst those that possess just one of the transgenic constructs are non-viable. Previous theoretical work has explored various properties of these systems, concluding that they should persist indefinitely in absence of resistance or mutation. Here we study a population genetics model of UD gene drive that relaxes past assumptions by allowing for loss-of-function mutations in each introduced gene. We demonstrate that mutations are likely to cause UD systems to break down, eventually resulting in the elimination of introduced transgenes. We then go on to investigate the potential of releasing "free suppressor" carrying individuals as a new method for reversing UD gene drives and compare this to the release of wild-types; the only previously proposed reversal strategy for UD. This reveals that while free suppressor carrying individuals may represent an inexpensive reversal strategy due to extremely small release requirements, they are not able to return a fully wild-type population as rapidly as the release of wild-types.

1. Introduction

Gene drive systems have gained much attention in recent years for their predicted ability to increase the frequency of desirable genetic material within a population. These systems have been proposed to have a number of important applications (Gould, 2008). Our interest in these systems relates to their potential use in preventing the spread of mosquito-borne viruses such as dengue (Alphey, 2014). In this context, refractory genes have been developed that are capable of significantly reducing the ability of *Aedes aegypti* mosquitoes to transmit the virus (Franz et al., 2006). However, the incorporation of these genes into the genome resulted in individuals of lower fitness than their wild-type counterparts (Franz et al., 2014). It is thus necessary to develop gene drives that are capable of increasing the frequency of such desirable genes in spite of their expected fitness costs.

Engineered underdominance (UD) is one class of gene drive that has been proposed for driving desirable genetic traits into a pop-

ulation (Davis et al., 2001). This technique is based on the introduction of two transgenic constructs (*A* and *B*) at unlinked genetic loci (see Fig. 1). These constructs consist of three component genes, namely a lethal effector, a suppressor for the lethal at the other locus and a "cargo" gene conferring a desirable phenotype. Due to the cross suppressing transgenic constructs, individuals carrying one or more copies of either construct will be non-viable if they do not also possess at least one copy of the other construct. Individuals carrying one or more copies of both constructs are viable since each of the lethal effectors are inactivated by suppressors carried at the other locus. These effects combine to create a selection pressure for individuals to either carry both transgenic constructs or neither.

Theoretical work on UD has demonstrated that these systems, when introduced above a threshold frequency, are capable of spreading desirable genes thus replacing wild populations with those carrying UD transgenic constructs (Davis et al., 2001; Magori and Gould, 2006; Huang et al., 2009; 2011; Edgington and Alphey, 2017; 2018; Dhole et al., 2018). Whilst the existence of this threshold frequency means that UD systems may be relatively expensive to deploy in the field relative to other classes of gene drive, it

Construct A   Construct B
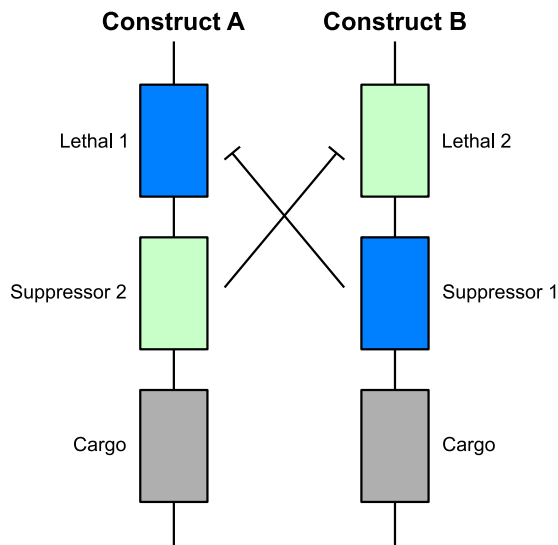
Fig. 1. A schematic diagram of the engineered underdominance gene drive system. Each transgenic construct possesses three genes; a lethal, a suppressor for the lethal on the other construct and a desirable genetic cargo. Genotypes possessing one or more copies of both constructs (or none of either) are viable since lethals will be deactivated by the suppressor(s) on the other construct. Those genotypes carrying one or more copies of either construct but none of the other are non-viable since they have non-suppressed lethal genes.

should restrict the invasion of transgenes into neighboring populations (Marshall, 2009; Marshall and Hay, 2012).

Previous modeling work has mostly assumed a population genetics framework neglecting the possibility of mutations forming within the introduced transgenes (Davis et al., 2001; Magori and Gould, 2006; Marshall and Hay, 2012; Marshall, 2009; Edgington and Alphey, 2017; Huang et al., 2011; 2009). Under this assumption it has been predicted that UD should persist indefinitely when introduced above the threshold frequency.

This threshold-dependent nature of UD systems has lead many to suggest that they may be reversed via the introduction of wild-type individuals (Champer et al., 2016; Marshall and Akbari, 2016). This should lower the transgene frequency to sub-threshold levels and result in the elimination of introduced transgenes. To our knowledge this is the only reversal strategy proposed for UD gene drives and has yet to be investigated in detail.

Here we extend upon results in the previous literature by formulating a population genetics model of two-locus UD (as proposed in Davis et al., 2001) that incorporates the effects of loss-of-function mutations in the introduced transgenes. In particular, we investigate the predicted dynamics of UD systems in the presence of a constant rate of mutation for each introduced transgene. We then go on to propose that the release of individuals carrying "free suppressors" could function as a genetics-based reversal strategy and compare this to the introduction of wild-types.

2. Mathematical modeling

We consider here a discrete generation deterministic population genetics model of two-locus UD gene drive in a panmictic (randomly mating), isolated (closed) population of infinite size. Mutation is allowed to occur in each transgenic construct at a constant rate (m per gene) and we assume that m is low enough for multiple mutations in a single generation to be neglected (Fig. 2 and Table 1). This gives a total of 18 alleles, with nine at each of the two independently segregating genetic loci (i.e. a, A, A_L, A_S, A_C, A_{LS}, A_{LC}, A_{SC} and A_{LSC} at one locus and b, B, B_L, B_S, B_C, B_{LS}, B_{LC}, B_{SC} and B_{LSC} at the other) where lower case denotes wild-type alleles, up-
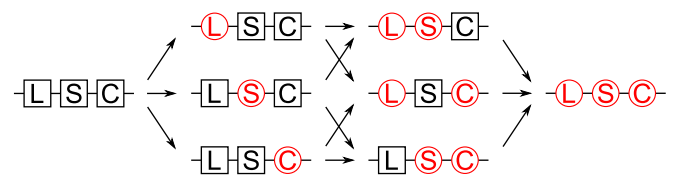
Fig. 2. Transgenic constructs are assumed to mutate at a rate m per gene. This is assumed low enough that multiple mutations per generation may be neglected. For example the initial transgenic construct (say, A) mutates at a rate of $3m$, producing mutations in the lethal (giving A_L), suppressor (A_S) and the cargo (A_C) gene each at a rate of m. Then, transgenic constructs possessing one mutated gene (e.g. A_L) mutate at a rate of $2m$ (giving A_{LS} and A_{LC} each at rate m). Transgenic constructs with two mutated genes (e.g. A_{LS}) then mutate at rate m producing constructs with all three genes mutated (i.e. A_{LSC}). Here non-mutated genes are represented by black squares whereas genes with loss-of-function mutations are shown in red circles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Table of parameter and variable definitions used throughout this study.

| Symbol	Definition
G_i	Frequency of genotype (i)
G_i^e	Proportional (i.e. non-normalized) genotype frequency
m	Rate of mutation per gene
α	Release ratio (α=introduced/wild)
$\bar{\Omega}$	Average fitness of the overall population
Ω_i	Relative fitness of genotype (i)
ε_j	Relative fitness conferred by transgenic construct (j)
β	Number of non-mutated copies of transgenic construct A
ϕ	Number of mutated copies of transgenic construct A (i.e. A_M)
μ	Number of non-mutated copies of transgenic construct B
ψ	Number of mutated copies of transgenic construct B (i.e. B_M)
γ_i	Lethality of genotype (i)

per case represents transgenic alleles and subscripts indicate loss-of-function mutations in a given genetic component (L = lethal, S = suppressor and C = cargo). This results in a total of 2025 possible genotypes, 819 of which are non-viable.

The fitness of each genotype is expressed relative to wild-type and represents a reduction in survival due to the carrying of transgenic constructs. Many genotypes also suffer a lethal effect from non-suppressed lethal genes. These factors are combined to give the relative fitness of each genotype:

$$\Omega_i = \varepsilon_A^{\beta} \varepsilon_{A_M}^{\phi} \varepsilon_B^{\mu} \varepsilon_{B_M}^{\psi} (1 - \gamma_i), \tag{1}$$

where ε denotes the relative fitness per construct conferred by non-mutated (A, B) or mutated (A_M, B_M where $M = L, S, C, LS, LC, SC, LSC$) transgenic constructs. We assume relative fitnesses act multiplicatively such that the overall relative fitness of an individual resulting from all transgenic constructs carried is restricted to the biologically feasible range $0 \leq \Omega \leq 1$. We also assume that all mutated transgenic constructs confer the same relative fitness regardless of the type or number of mutations (although resulting genotypes may confer a separate lethal effect (γ_i)). Exponents β, ϕ, μ and ψ denote the number (0, 1 or 2) of each transgenic construct type carried by a given genotype while γ_i represents the lethality of that genotype ($\gamma_i = 1$ if non-viable or $\gamma_i = 0$ if viable). Finally, we assume all lethal effectors are 100% effective when unsuppressed and that any number (i.e. one or two) of lethal effector copies are fully suppressed by any number (one or more) of the relevant suppressor (i.e. the strong suppression case in Edgington and Alphey, 2017).

For numerical simulations we used a set of MATLAB (MATLAB R2014b, The MathWorks Inc., Natick, MA) scripts run in parallel using the MATLAB Parallel Computing Toolbox through a MATLAB Distributed Computing Server. These allow simulation of the system without a manual formulation of 2025 difference equations
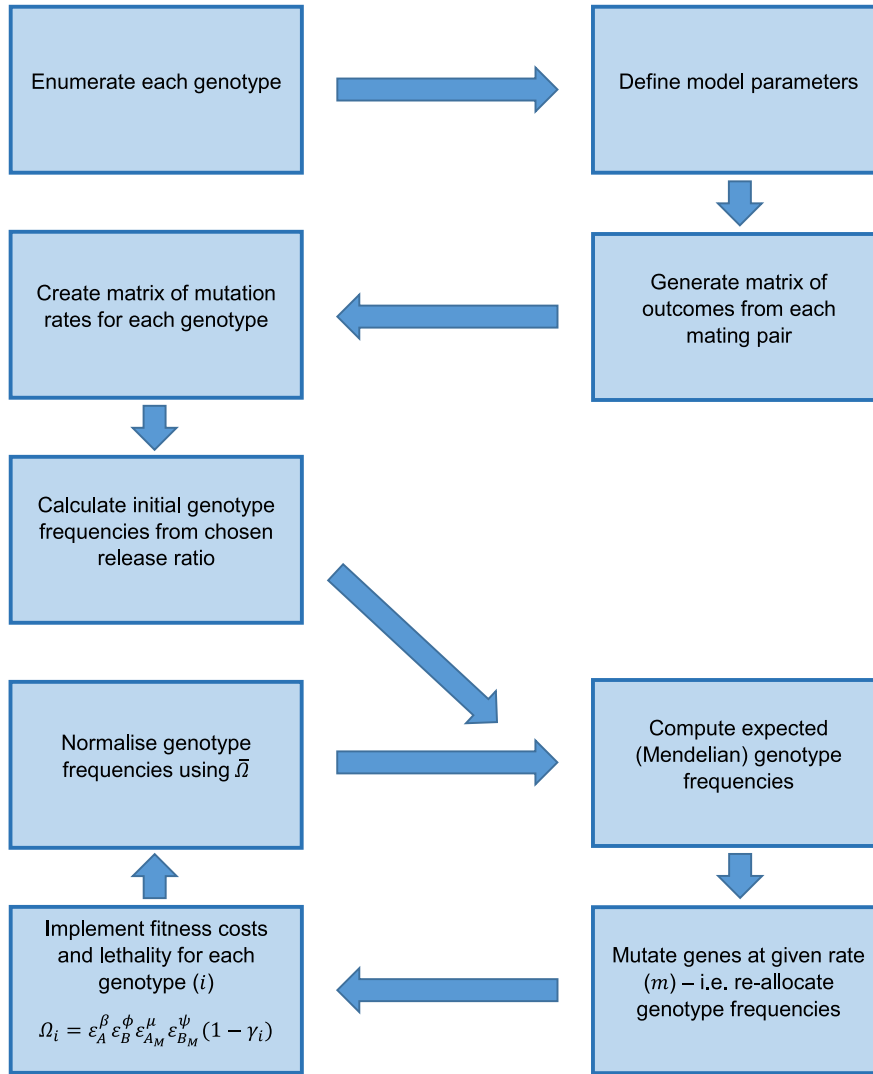
Fig. 3. Diagram showing the simulation procedure for an engineered underdominance system with a constant rate of mutation (m) per gene. Parameter symbols used here are: ε, the relative fitness conferred by a given allele (denoted in the subscripts where M represents a mutated allele); β and ϕ, the numbers of non-mutated copies of construct A and B, respectively; μ and ψ, the number of mutated copies of transgenic constructs A and B, respectively; γ_i, the lethality conferred by a given genotype (i); Ω_i, the overall relative fitness of individuals of genotype i; and $\bar{\Omega}$, the average relative fitness of the entire population.

(see Fig. 3). Briefly, the process begins by converting each possible genotype into a numerical form and computing relative fitnesses for each (using Eq. (1)). Initial conditions are then calculated according to:

$$G_{WT}^{t=0} = \frac{1}{1+\alpha}, \quad G_{AABB}^{t=0} = \frac{\alpha}{1+\alpha}, \quad (2)$$

where G_i^t is the i-th genotype frequency t generations after the initial release; α is the release ratio (introduced/wild) of non-mutated transgene double homozygotes ($AABB$) into a wild-type ($aabb$) population; and all other genotypes have an initial frequency of zero (i.e. $G_i^{t=0} = 0$ for $i \neq WT, AABB$). We assume the absence of mutations in released individuals is feasible due to quality controls at the rearing facility. Genotype frequencies in subsequent generations are computed iteratively. Expected (Mendelian) frequencies are first calculated in absence of any mutation using a matrix of outcomes from every genotype mating pair and then multiplied by a matrix of mutation rates (giving proportional frequencies G_i^e). As such, it is assumed that mutation occurs during gamete production, meaning fitness effects are applied based on an individuals genotype after mutations occur. Finally, proportional frequencies
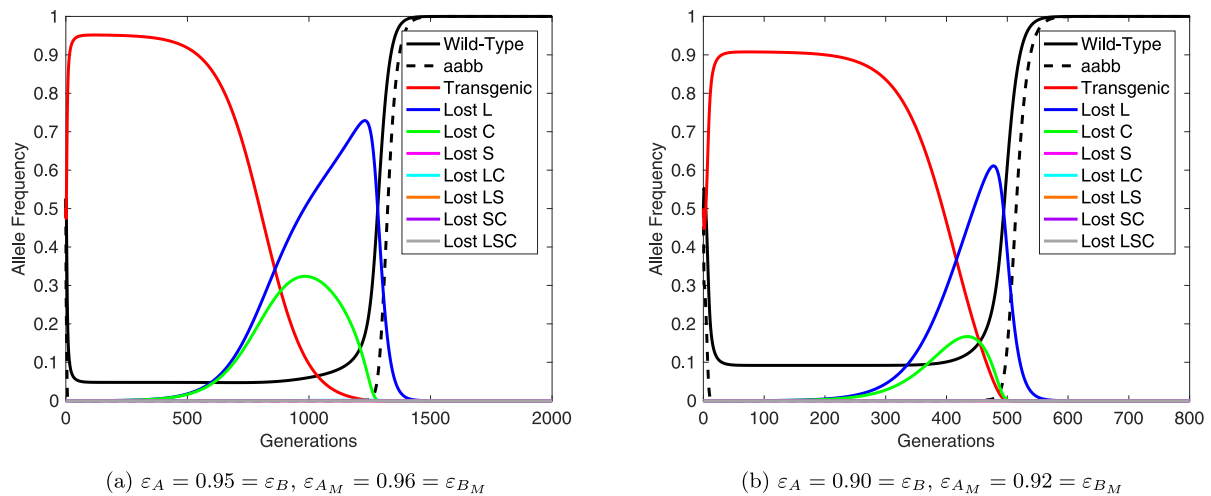
(G_i^e) are normalized using the average fitness of the overall population ($\bar{\Omega} = \sum \Omega_i G_i^e$) such that genotype frequencies (G_i^t) sum to one.

Simpler haplotype-based population genetics models of UD in absence of mutation are also considered here to allow stability analyses of various equilibrium states since they are more analytically tractable. Details of these models and their respective stability analyses are given in the Supplementary Information.

3. Results

3.1. Example numerical simulations with various rates of mutation per gene

To our knowledge, experimentally measured rates of mutation in key organisms considered as likely targets for UD gene drives are not well known. Thus, here we conduct numerical simulations for rates of mutation (per gene) spanning five orders of magnitude, namely $m = 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$ and 10^{-8}. Given previous estimates of the mutation rate in *Drosophila melanogaster* (2.8×10^{-9} Keightley et al., 2014 and 8.4×10^{-9} Haag-Liautard et al., 2007 per

(a) $\varepsilon_A = 0.95 = \varepsilon_B$, $\varepsilon_{A_M} = 0.96 = \varepsilon_{B_M}$

(b) $\varepsilon_A = 0.90 = \varepsilon_B$, $\varepsilon_{A_M} = 0.92 = \varepsilon_{B_M}$

Fig. 4. Mutation of transgenic constructs is predicted to return a fully wild-type population. Results here are presented for 1:1 (introduced:wild) introductions of non-mutated double homozygote (*AABB*) individuals into a wild-type (*aabb*) population. Here, transgenic constructs are assumed to mutate at a rate $m = 10^{-6}$ per gene, neglecting the possibility of multiple genes mutating within a single generation. Solid lines represent the allele frequency of wild-type, transgenic and each variety of mutated allele with details given in the figure legend whilst black dashed lines denote the genotype frequency of double wild-type homozygotes. Here a number of alleles only reach very low maximum frequencies (see Fig. 5(a)) and thus they appear to overlie one another along the horizontal axis. Panel (a) shows results for $\varepsilon_A = 0.95 = \varepsilon_B$ and $\varepsilon_{A_M} = 0.96 = \varepsilon_{B_M}$ whereas (b) is for $\varepsilon_A = 0.90 = \varepsilon_B$ and $\varepsilon_{A_M} = 0.92 = \varepsilon_{B_M}$. Note the difference in time-scales between these panels.

nucleotide per generation), the size of gene drive components used in previous studies (1–10 kb e.g. Reeves et al., 2014; Windbichler et al., 2011; Champer et al., 2017b) and an approximation that 1–10% of nucleotides in these sequences are essential (i.e. causes a loss of gene function when mutated) we anticipate that this range should include rates relevant to a range of proposed target species.

Fig. 4 shows some sample results considering a 1:1 (introduced:wild) release of double homozygote (i.e. *AABB*) individuals. In these examples, we assume each non-mutated transgenic construct confers a fitness load of either 5% or 10% (i.e. $\varepsilon_A = \varepsilon_B = 0.95$ or 0.90) and mutated constructs 4% or 8% (i.e. $\varepsilon_{A_M} = \varepsilon_{B_M} = 0.96$ or 0.92), respectively. Thus, mutated transgenic constructs have a small fitness advantage over non-mutated ones, yet still a deficit relative to wild-type. In examples conducted with mutated transgenic alleles conferring a greater fitness cost than non-mutated alleles, the UD system progressed without significant accumulation of mutated alleles, thus we do not consider these cases any further.

In Fig. 4 the non-mutated UD system initially reaches a high frequency, as previously modeled (Davis et al., 2001; Magori and Gould, 2006; Marshall and Hay, 2012; Marshall, 2009; Edgington and Alphey, 2017; Huang et al., 2011; 2009). However, each type of mutated construct begins to accumulate in the population, with large frequencies reached by constructs carrying a single mutation in either the lethal or cargo gene. This results in a concurrent decrease in the frequency of non-mutated constructs. Once these mutated constructs have reached a high frequency in the population, they begin to be replaced by the remaining wild-type alleles due to the relative fitness advantage of wild-type alleles over mutated transgenic alleles. This eventually returns the population to a fully wild-type state. Note that we did not observe the stable coexistence of non-mutated and mutated transgenes under any parameter combination considered within this study.

For the full range of mutation rates considered here we observe similar dynamics to those in Fig. 4 except that increasing (decreasing) mutation rates lead to faster (slower) accumulation and higher (lower) maximum frequencies of most types of mutated transgene allele (see Fig. 5(a)). This inevitably means that higher rates of mutation reduce the period over which the UD system persists (see
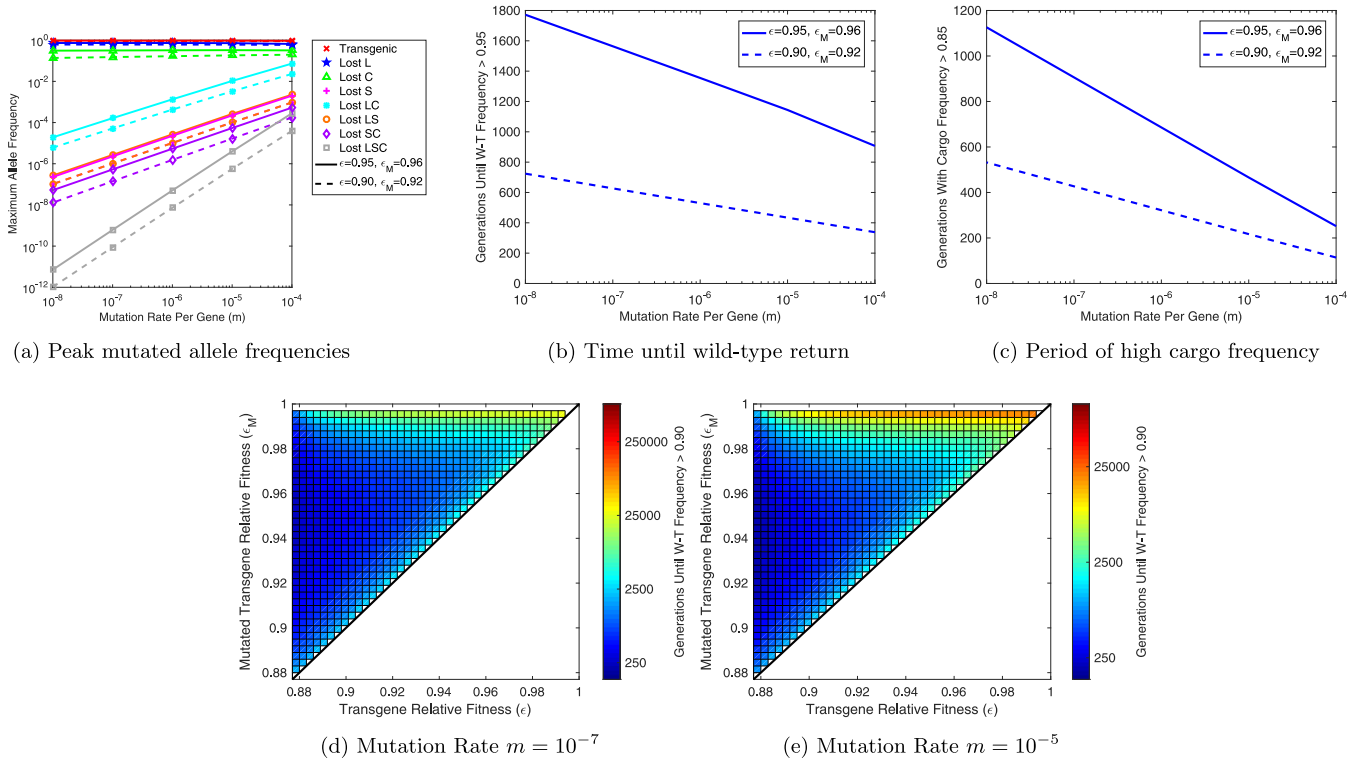
Fig. 5(b)) and also the period for which functional cargo genes are present at high frequency (see Fig. 5(c)). This should form a reasonable proxy for the efficacious period of a released UD system. Importantly, while mutation may eliminate the UD system, it is still likely to be maintained at high frequency for hundreds of generations - likely long enough for introduced cargo genes to have produced their desired effect.

3.2. Reversal strategies

To our knowledge, the only reversal strategy previously proposed for UD systems is the release of wild-type individuals in sufficient numbers to push the transgene frequency to sub-threshold levels (Champer et al., 2016; Marshall and Akbari, 2016). When successful, wild-type alleles should recover and eliminate transgenes from the population. Whilst this mechanism has been mentioned a number of times previously it has yet to be explored in any detail.

Based on results presented in Fig. 4 here we propose an alternative, genetics-based, reversal strategy for UD systems. In this strategy transgenic individuals carrying only the suppressor genes of the original UD system would be released. These "free suppressor" carrying individuals should, under certain conditions, be able to increase in frequency at the expense of the initial UD system. Once at high frequency, assuming they suffer even a small fitness load, these free-suppressor carriers should begin to decline in frequency as they are replaced with fitter wild-type individuals.

To provide theoretical support for this reversal mechanism we conduct mathematical stability analyses of equilibria resulting from deterministic haplotype-based population genetics models of UD systems both in absence and in the presence of free-suppressor constructs since this is more analytically tractable than the genotype-based model used elsewhere in this study (see Supplementary Information for details). The models used in these stability analyses are formulated in absence of mutation in other transgenic components. This ensures that the proposed reversal strategy is capable of reversing the initial UD system on its own (i.e. without requiring additional mutation in other transgenic components). Note also that the haplotype-based and genotype-based models considered in this study produce equivalent allele

(a) Peak mutated allele frequencies

(b) Time until wild-type return

(c) Period of high cargo frequency

(d) Mutation Rate $m = 10^{-7}$

(e) Mutation Rate $m = 10^{-5}$

Fig. 5. Rates of mutation in introduced engineered underdominance (UD) transgenes affect mutated allele frequencies and the efficacious period of the system. (a) Maximum frequencies attained by each of the non-mutated and mutated transgene alleles. Colors represent each different type of allele with details given in the figure legend. (b) The number of generations from the time of initial transgenic release until wild-type alleles return to a frequency greater than 0.95. (c) The number of generations that the UD system is able maintain a high frequency (> 0.85) of transgenes with a functional copy of the cargo gene. In all panels solid lines represent cases with relative fitness parameters $\varepsilon_A = 0.95 = \varepsilon_B$ and $\varepsilon_{A_M} = 0.96 = \varepsilon_{B_M}$ whilst dashed lines are for cases with $\varepsilon_A = 0.90 = \varepsilon_B$ and $\varepsilon_{A_M} = 0.92 = \varepsilon_{B_M}$. These timings should represent a reasonable proxy for the period over which the desired phenotype conferred by the UD system would be effective. (d) and (e) show the number of generations taken for a system to return to a high wild-type frequency (> 0.90) following the release of a two-locus UD system at a 1:1 introduced to wild-type ratio. Two different mutation rates are considered, namely $m = 10^{-7}$ in panel (d) and $m = 10^{-5}$ in panel (e).

frequencies for all parameter configurations tested and thus results obtained here should apply to both model formulations. In the models both with and without free-suppressor constructs we find that the transgene free (i.e. fully wild-type) equilibrium state is stable for all relative fitness parameters considered (i.e. $0 \leq \varepsilon \leq 1$ and where applicable $0 \leq \varepsilon_S \leq 1$). In the model without free suppressors, the transgene introgression equilibrium is stable if $\varepsilon^* \leq \varepsilon \leq 1$, where ε^* is the relative fitness threshold for a given transgene release ratio. However, under certain conditions, when free-suppressor constructs are introduced into the model this transgene introgression equilibrium is found to be unstable which is indicative of the ability of free-suppressor constructs to reverse an initially introduced UD system. The introduction of free-suppressor constructs with $\varepsilon_S \geq \varepsilon$ at any release ratio appears to result in a return to a fully wild-type population. Perhaps more surprisingly, even where free suppressors confer a slight relative fitness disadvantage compared to the initial transgenic constructs, they may still lead to a fully wild-type population so long as the release ratio is large enough. Some examples of this are given in Figures S3-S6. Interestingly, in spite of the large difference in release ratio of free-suppressor constructs and the differences in the maximum frequencies achieved (see Figures S4-S6), there does not appear to be a major difference in the time taken for a given ε and ε_S combination to return a fully wild-type population (See Figure S7).

We conduct numerical simulations in order to compare the different reversal strategies discussed here. Similar to the examples above, we consider the relative fitness conferred by the initial UD system to be $\varepsilon_A = 0.95 = \varepsilon_B$ or $\varepsilon_A = 0.90 = \varepsilon_B$ and the free

suppressor elements to confer a relative fitness of $\varepsilon_{\text{Rev}} = 0.96$ or $\varepsilon_{\text{Rev}} = 0.92$, giving them a small fitness advantage over the original UD constructs and a deficit relative to wild-type. Results are shown in Fig. 6 for a 2:1 (introduced:wild (all genotypes)) release of wild-type individuals and individuals carrying free suppressors at both loci. These releases are assumed to occur 100 generations after an initial 1:1 (introduced:wild) UD release so that the original system is at high frequency and mutations would not have accumulated to significant frequencies. As such we initially assume that no mutation occurs (i.e. $m = 0$) to ensure that free suppressors work in absence of other types of allele.
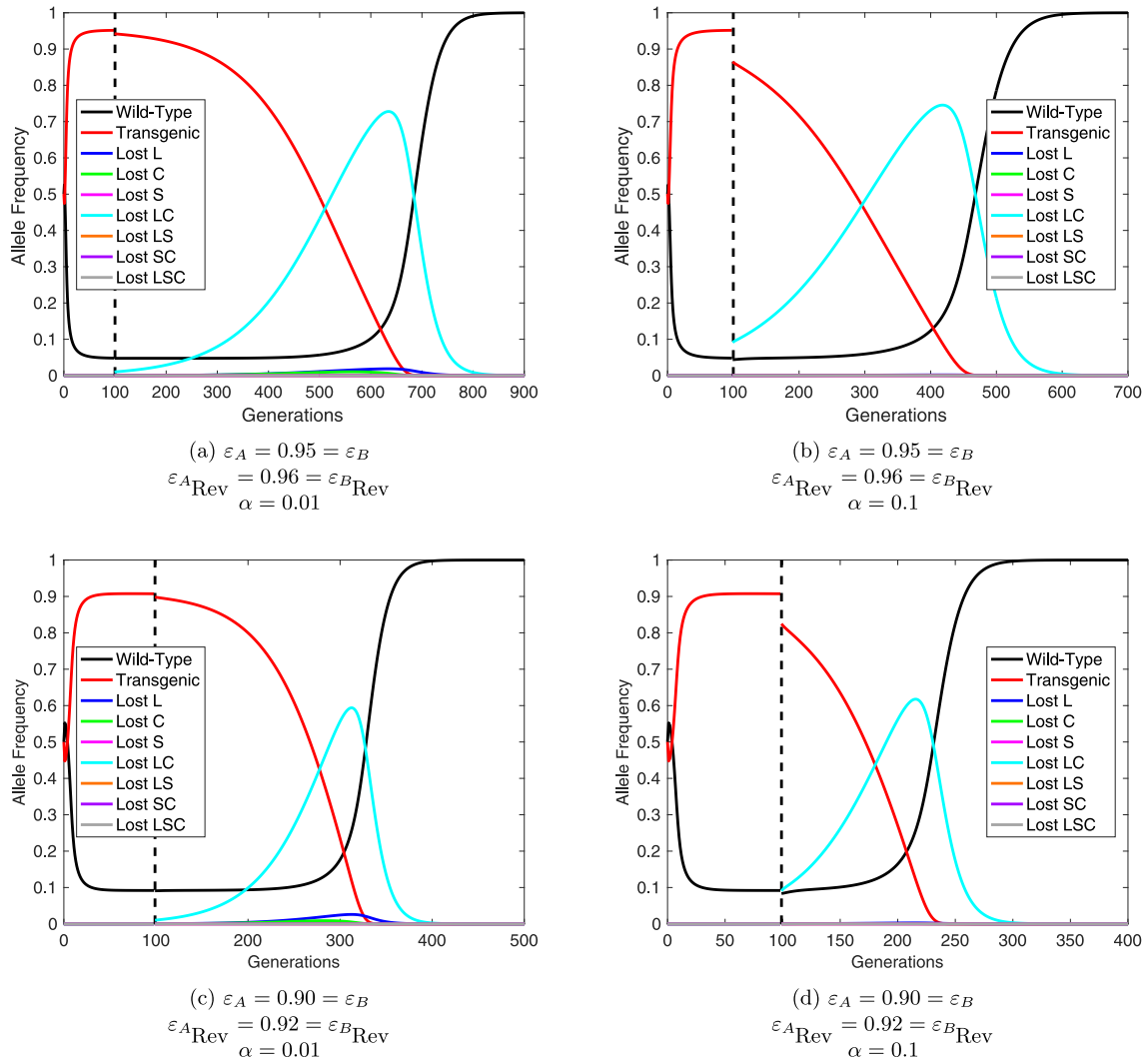
Fig. 6 (b) demonstrates the feasibility of reversing of UD systems via the introduction of free suppressors in the absence of mutated alleles. We then relax this assumption to consider the behavior of a reversal drive in the presence of mutation (here we assume $A_{\text{Rev}} = A_{LC}$ and $B_{\text{Rev}} = B_{LC}$). These numerical simulations demonstrate that this reversal strategy can function even with small releases (i.e. $\alpha = 0.01$ and $\alpha = 0.1$; see Fig. 7). However, one would need to be wary of stochastic effects when making such small releases. These results imply that the release of free suppressor carrying individuals could represent a very cost effective reversal strategy for UD systems.

It is clear that release of wild-type individuals provides a much faster option for returning a fully wild-type population (Fig. 6). However, releases of wild-type individuals must be significantly larger (for example ~ 1.75 times the wild population when $\varepsilon_A = 0.95 = \varepsilon_B$) in order to achieve reversal of the initial UD system. This likely makes the direct cost of releasing wild-type individuals far higher than for the use of free-suppressor carrying individuals.

(a) Release of *aabb* (wild-type)

(b) Release of $A_{\mathrm{Rev}}A_{\mathrm{Rev}}B_{\mathrm{Rev}}B_{\mathrm{Rev}}$
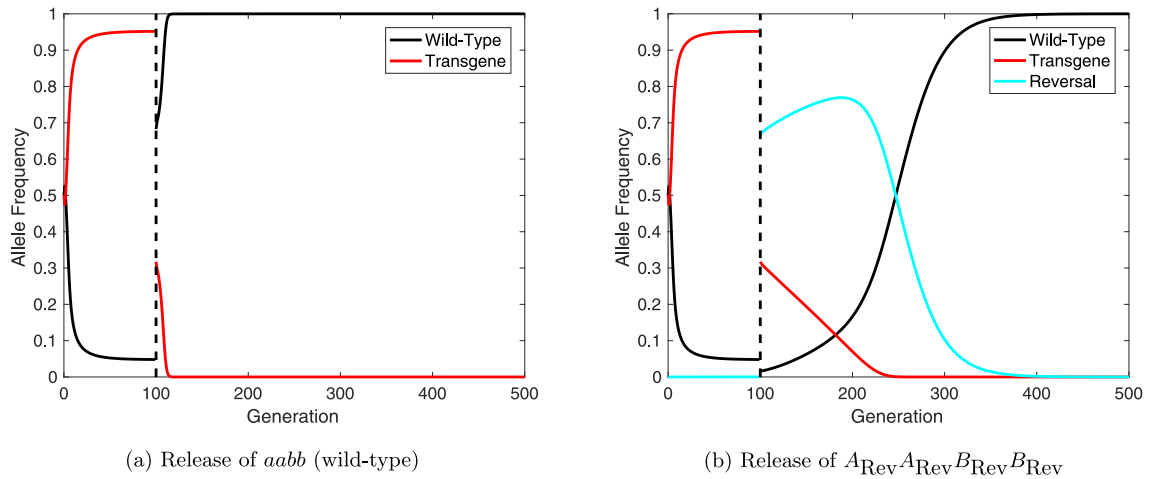
Fig. 6. Comparison of two different reversal strategies for engineered underdominance (UD) gene drive systems. (a) An example of reversal through the release of wild-type (*aabb*) individuals. (b) Release of individuals carrying free suppressors at both loci (i.e. $A_{\mathrm{Rev}}A_{\mathrm{Rev}}B_{\mathrm{Rev}}B_{\mathrm{Rev}}$). Here line colors denote each type of allele with black representing wild-type (*a*, *b*); red denoting non-mutated transgenes (*A*, *B*); and cyan showing suppressor-only alleles (A_{Rev}, B_{Rev}). In each example the initial UD release is made at a 1:1 (introduced:wild) ratio with relative fitness parameters $\varepsilon_A = 0.95 = \varepsilon_B$. After 100 generations one of the reversal strategies is released at a ratio of 2:1 (introduced:wild). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) $\varepsilon_A = 0.95 = \varepsilon_B$
$\varepsilon_{A_{\mathrm{Rev}}} = 0.96 = \varepsilon_{B_{\mathrm{Rev}}}$
$\alpha = 0.01$

(b) $\varepsilon_A = 0.95 = \varepsilon_B$
$\varepsilon_{A_{\mathrm{Rev}}} = 0.96 = \varepsilon_{B_{\mathrm{Rev}}}$
$\alpha = 0.1$

(c) $\varepsilon_A = 0.90 = \varepsilon_B$
$\varepsilon_{A_{\mathrm{Rev}}} = 0.92 = \varepsilon_{B_{\mathrm{Rev}}}$
$\alpha = 0.01$

(d) $\varepsilon_A = 0.90 = \varepsilon_B$
$\varepsilon_{A_{\mathrm{Rev}}} = 0.92 = \varepsilon_{B_{\mathrm{Rev}}}$
$\alpha = 0.1$

Fig. 7. Free suppressor constructs can reverse engineered underdominance (UD) gene drives with small releases. Here, releases at generation 100 of free suppressor constructs (assuming $A_{\mathrm{Rev}} = A_{lC}$ and $B_{\mathrm{Rev}} = B_{lC}$) are numerically simulated for release ratios of $\alpha = 0.01$ (panels (a) and (c)) and $\alpha = 0.1$ (panels (b) and (d)) - far smaller than those required for reversal with wild-type. Lines represent the possible alleles with colors denoting the nature of each (details given in figure legends).

4. Discussion

Since gene drive systems were first proposed they have gained much attention for their potential to help fight a number of important global issues (e.g. Alphey (2014); Gould (2008); Prowse et al. (2017); Thresher et al. (2014)). However, various genetic control measures have been shown to be limited by the potential generation of mutation or resistance (Alphey et al., 2011; Watkinson-Powell and Alphey, 2017; Hammond et al., 2017; Unckless et al., 2017; Champer et al., 2017b; 2017a) the dissociation of gene drive components (e.g. Marshall (2008)) and loss-of function mutation similar to the type studied here (Beaghton et al., 2017). A number of studies have also begun to investigate mechanisms capable of reversing gene drives in case they produce unexpected consequences (Vella et al., 2017; DiCarlo et al., 2017; Wu et al., 2016). Here we showed that mutations will likely lead to the break down of two-locus UD systems and the eventual elimination of introduced transgenes (a feature similar to that seen previously for homing-based gene drives Beaghton et al., 2017) although under the assumptions used, the UD system is likely to persist at high allele frequency for hundreds of generations. We then went on to demonstrate that reversal of UD systems is feasible via the release of either wild-type individuals or individuals carrying free suppressors.

As with all mathematical models, the work presented here relies on a number of simplifying assumptions, the majority of which are common in this type of study (Edgington and Alphey, 2017; Davis et al., 2001; Magori and Gould, 2006; Marshall and Hay, 2012). Since these assumptions have been discussed previously, we do not consider them any further here. There are however a few areas specific to this study where more detailed modeling would be useful to further elucidate the effects of mutation in UD systems. For example, here mutations are assumed to fully eliminate gene function whereas in reality they may produce only a partial loss of function. We also assumed that all mutated transgenic constructs confer the same fitness cost regardless of the number and type of mutations carried. These assumptions could be relaxed in future models by expressing gene efficacy as a function of mutations carried. Another assumption is that the consideration of a population with discrete generations is realistic. For some target species, such as mosquitoes, this may be a reasonable approximation where populations are synchronized either by climate (e.g. wet and dry seasons) or laboratory conditions. However, it is unlikely that this assumption will hold in other areas where populations are thought to reproduce continuously - although we would not anticipate vastly different conclusions emerging from overlapping generation and/or population dynamics models. Finally, it is possible that both natural genetic polymorphism and mutation outside of the introduced transgenic constructs within a target population could mean that certain individuals would be, at least partially, resistant to the effects of introduced genes. In such cases, the UD system may place resistance alleles under a selective pressure causing them to increase in frequency and reduce the efficacy of UD systems. In the worst case scenario, such polymorphisms/mutations would be fully effective and dominant suppressors of one or both lethal transgenic components. We would anticipate that this would produce similar dynamics to the introduction of a free-suppressor construct albeit with slight differences in inheritance patterns since such polymorphisms/mutations would not be allelic to either of the UD constructs. Screening and sampling of the target population would likely ensure that any pre-existing troublesome polymorphisms were not present at high frequency within the population.

Results presented here assume that individuals carrying mutated transgenes are fitter than those carrying non-mutated transgenic constructs. This is a common assumption when modeling the formation of resistance/mutations in gene drive systems (e.g. Unckless et al., 2017; Alphey et al., 2011). The fitness advantage of individuals carrying specific mutations in each transgene (or combination of transgenes) is likely to determine the exact dynamics and time scales that would be observed in experimental work. Results presented here are intended to give a guide as to the type of behavior we would expect to emerge rather than giving definitive predictions for specific applications.

Previous literature has estimated the mutation rate in *Drosophila melanogaster* to be $\sim 5.6 \times 10^{-9}$ per nucleotide per generation (mean of estimates in Keightley et al. (2014) and Haag-Liautard et al. (2007)). Given the size of gene drive components used in previous experimental studies (1–10 kb e.g. Reeves et al., 2014; Windbichler et al., 2011; Champer et al., 2017b) and an approximation that 1–10% of nucleotides in these sequences are essential (i.e. causes a loss of gene function when mutated), we believe the mutation rates explored here to be feasible for a range of proposed target organisms. Using these mutation rates, we assumed only one mutation was able to emerge in a single generation. In practice, deletions could remove two adjacent components simultaneously. However, assuming the ordering of components shown in Fig. 1 such deletions would leave individuals without suppressor elements, thus making them non-viable in the presence of the construct at the other locus. Coupling this with the extremely low probability of multiple mutations occurring simultaneously, we do not anticipate that relaxing this assumption would lead to major differences from the results presented here. Within this study we have also neglected the possibility of recombination leading to loss of transgenic components. However, since the components of each transgenic construct are located extremely close together, We would expect this to occur at an extremely low rate. As such, we would not expect any major differences to emerge if recombination were incorporated into the model.

A number of recent studies have discussed the need to explore reversal strategies for gene drives in case they have unexpected consequences. While the majority of work has focused on the reversal of CRISPR-Cas9 gene drive systems (e.g. Vella et al., 2017; DiCarlo et al., 2017; Oye et al., 2014; Wu et al., 2016), it has been suggested that the release of wild-type individuals could reverse UD systems (Champer et al., 2016; Marshall and Akbari, 2016). The release of wild-type individuals represents a threshold dependent reversal strategy with the precise number of wild-types required to reverse a released UD system depending on the fitness of UD carrying individuals and the size of the wild population, both of which are difficult to measure. What is known however, is that the size of these wild-type releases would need to be very large. For example, when transgenic constructs each confer a 5% fitness cost (i.e. $\varepsilon_A = 0.95 = \varepsilon_B$), the wild-type release would need to be ~ 1.75 times the size of the entire wild population. Importantly, insufficiently sized releases of wild-type individuals would fail to reverse a given UD system. While it would be feasible to make further wild-type releases, this would have obvious cost implications. To help overcome these issues, here we proposed a genetics-based alternative (release of free suppressor carriers) that appears to be threshold independent (i.e. it is predicted to be effective from very small releases, leaving aside stochastic effects) if free-suppressor constructs confer an equal or smaller fitness load than the originally introduced transgenic constructs. This approach may also be effective where free-suppressor constructs confer a slightly larger fitness load than the original transgenic constructs, however much larger releases may be required for this to work. One of the main issues associated with employing free-suppressor carrying individuals as a reversal strategy is that it will take longer to return a fully wild-type population than release of wild-types. It may also be difficult to convince the public that releasing further transgenic individuals is an acceptable method of eliminating a gene drive that

produced unexpected consequences. Here we do not wish to draw a firm conclusion on which reversal strategy should be pursued, since this may depend on case-specific social, economic and technical factors that are beyond the scope of this study.

In spite of demonstrating that UD gene drives will likely break down over time with the emergence of mutations, this study provides reason to be optimistic about the prospect of using UD gene drives to spread desirable genes through a target population. In particular, even though introduced transgenes are likely to be eliminated, the desirable genetic cargo is expected to reach and maintain a high frequency in the population for hundreds of generations. This is likely long enough for them to have produced their desired effect. Here we have given an initial theoretical examination of this UD break down due to mutation. We anticipate that future modeling studies will be able to produce more application specific models, thus refining the predictions presented here.

Acknowledgements

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2019.06.024.

References

Alphey, L., 2014. Genetic control of mosquitoes. Annu. Rev. Entomol. 59, 205–224.
Alphey, N., Bonsall, M., Alphey, L., 2011. Modeling resistance to genetic control of insects. J. Theoret. Biol. 270 (1), 42–55.
Beaghton, A., Hammond, A., Nolan, T., Crisanti, A., Godfray, H., Burt, A., 2017. Requirements for driving antipathogen effector genes into populations of disease vectors by homing. Genetics 205 (4), 1587–1596.
Champer, J., Buchman, A., Akbari, O., 2016. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nat. Rev. Genet. 17 (3), 146–159.
Champer, J., Liu, J., Oh, S., Reeves, R., Luthra, A., Oakes, N., Clark, A., Messer, P., 2017a. Reducing resistance allele formation in CRISPR gene drives. bioRxiv 150276.
Champer, J., Reeves, R., Oh, S., Liu, C., Liu, J., Clark, A., Messer, P., 2017b. 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.
Davis, S., Bax, N., Grewe, P., 2001. Engineered underdominance allows efficient and economical introgression of traits into pest populations. J. Theoret. Biol. 212 (1), 83–98.
Dhole, S., Vella, M., Lloyd, A., Gould, F., 2018. Invasion and migration of spatially self-limiting gene drives: a comparative analysis. Evolut. Appl. 11 (5), 794–808.
DiCarlo, J., Chavez, A., Dietz, S., Esvelt, K., Church, G., 2017. Safeguarding CRISPR-Cas9 gene drives in yeast. Nat. Biotechnol. 33, 1250–1255.
Edgington, M., Alphey, L., 2017. Conditions for success of engineered underdominance gene drive systems. J. Theoret. Biol. 430, 128–140.
Edgington, M., Alphey, L., 2018. Population dynamics of engineered underdominance and killer-rescue gene drives in the control of disease vectors. PLOS Comput. Biol. 14 (3), e1006059.
Franz, A., Sanchez-Vargas, I., Adelman, Z., Blair, C., Beaty, B., James, A., Olson, K., 2006. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified aedes aegypti. Proc. Natl. Acad. Sci. 103 (11), 4198–4203.
Franz, A., Sanchez-Vargas, I., Raban, R., Black IV, W., James, A., Olson, K., 2014. Fitness impact and stability of a transgene conferring resistance to dengue-2 virus following introgression into a genetically diverse aedes aegypti strain. PLoS Negl. Trop. Dis. 8 (5), e2833.
Gould, F., 2008. Broadening the application of evolutionarily based genetic pest management. Evolution 62 (2), 500–510.
Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D., Charlesworth, B., Keightley, P., 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in drosophila. Nature 445 (7123), 82–85.
Hammond, A., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F., D'Aurizio, R., Crisanti, A., Nolan, T., 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. PLOS Genet. 13 (10), 1–16.
Huang, Y., Lloyd, A., Legros, M., Gould, F., 2009. Gene-drive in age-structured insect populations. Evolut. Appl. 2 (2), 143–159.
Huang, Y., Lloyd, A., Legros, M., Gould, F., 2011. Gene-drive into insect populations with age and spatial structure: A theoretical assessment. Evolut. Appl. 4 (3), 415–428.
Keightley, P., Ness, R., Halligan, D., Haddrill, P., 2014. Estimation of the spontaneous mutation rate per nucleotide site in a drosophila melanogaster full-sib family. Genetics 196 (1), 313–320.
Magori, K., Gould, F., 2006. Genetically engineered underdominance for manipulation of pest populations: A deterministic model. Genetics 172 (4), 2613–2620.
Marshall, J., 2008. The impact of dissociation on transposon-mediated disease control strategies. Genetics 178 (3), 1673–1682.
Marshall, J., 2009. The effect of gene drive on containment of transgenic mosquitoes. J. Theoret. Biol. 258 (2), 250–265.
Marshall, J., Akbari, O., 2016. Gene drive strategies for population replacement. In: Genetic Control of Malaria and Dengue. Academic Press, Boston, pp. 169–200.
Marshall, J., Hay, B., 2012. Confinement of gene drive systems to local populations: a comparative analysis. J. Theoret. Biol. 294, 153–171.
Oye, K., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., Lightfoot, S.-Y., McNamara, J., Smidler, A., Collins, J., 2014. Regulating gene drives. Science 345 (6197), 626–628.
Prowse, T., Cassey, P., Ross, J., Pfitzner, C., Wittmann, T., Thomas, P., 2017. Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. Proc. R. Soc. B 284 (1860).
Reeves, R., Bryk, J., Altrock, P., Denton, J., Reed, F., 2014. First steps towards underdominant genetic transformation of insect populations. PLoS One 9 (5), e97557.
Thresher, R., Hayes, K., Bax, N., Teem, J., Benfey, T., Gould, F., 2014. Genetic control of invasive fish: technological options and its role in integrated pest management. Biol. Invasions 16 (6), 1201–1216.
Unckless, R., Clark, A., Messer, P., 2017. Evolution of resistance against CRISPR/Cas9 gene drive. Genetics 205 (2), 827–841.
Vella, M., Gunning, C., Lloyd, A., Gould, F., 2017. Evaluating strategies for reversing CRISPR-Cas9 gene drives. Scientif. Rep. 7 (1), 11038.
Watkinson-Powell, B., Alphey, N., 2017. Resistance to genetic insect control: modelling the effects of space. J. Theoret. Biol. 413, 72–85.
Windbichler, N., Menichelli, M., Papathanos, P., Thyme, S., Li, H., Ulge, U., Hovde, B., Baker, D., Monnat, R., Burt, A., Crisanti, A., 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature 473 (7346), 212–215.
Wu, B., Luo, L., Gao, X., 2016. Cas9-triggered chain ablation of Cas9 as a gene drive brake. Nat. Biotechnol. 34, 137–138.