The effect of mating complexity on gene drive dynamics

Prateek Verma[1,*]

R. Guy Reeves[2]

Samson Simon [3]

Mathias Otto [3]

Chaitanya S. Gokhale[1]

1. Research Group for Theoretical Models of Eco-evolutionary Dynamics, Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany;

2. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany;

3. Federal Agency for Nature Conservation (BfN), I 2.6 Biosafety, Konstantinstrasse 110, D-53179 Bonn, Germany.

∗ Corresponding author; e-mail: verma@evolbio.mpg.de.

Keywords: gene drive, mating complexity, mate choice, mating system, mating network, mathematical modelling, risk assessment.

Abstract

Gene drive technology is being presented as a means to deliver on some of the global challenges humanity faces today in healthcare, agriculture and conservation. However, there is a limited understanding of the consequences of releasing self-perpetuating transgenic organisms into the wild populations under complex ecological conditions. In this study, we analyze the impact of three factors, mate-choice, mating systems and spatial mating network, on the population dynamics for two distinct classes of modification gene drive systems; distortion and viability-based ones. All three factors had a high impact on the modelling outcome. First,

we demonstrate that distortion based gene drives appear to be more robust against the mate-choice than viability-based gene drives. Second, we find that gene drive spread is much faster for higher degrees of polygamy. With fitness cost, speed is the highest for intermediate levels of polygamy. Finally, the spread of gene drive is faster and more effective when the individuals have fewer connections in a spatial mating network. Our results highlight the need to include mating complexities while modelling the population-level spread of gene drives. This will enable a more confident prediction of release thresholds, timescales and consequences of gene drive in populations.

Introduction

Gene drive technology is being developed to potentially deliver on some of the critical challenges in human health, agriculture or biodiversity conservation (Brossard et al., 2019; Buchman et al., 2018; Johnson et al., 2016; Prowse et al., 2017; Windbichler et al., 2011). A prominent example of gene drive is its proposed use to push transgenes into wild mosquito populations that make them resistant to malaria parasites (Carballar-Lejarazú et al., 2020; Gantz et al., 2015). For biodiversity conservation, the potential of gene drives to control the spread of invasive species or implementing disease resistance in endangered species is being discussed (Godwin et al., 2019; Johnson et al., 2016; Prowse et al., 2017). In agriculture, gene drive could, it is argued, control pest populations like fruit flies (Buchman et al., 2018) in cherry plantations or transform the pest population to make them more susceptible to pesticides (Barrett et al., 2019). Though to date, no gene drive organisms have been released into the wild populations. All gene drive constructs are necessarily transgenic in nature and require the release of genetically modified organisms into wild populations. The possibility exists that not all the unintended consequences of gene drive releases can be reversed, consequently modeling is key to evaluating this technology.

Theoretical and laboratory studies indicate that some driving transgenic-constructs could spread through wild populations in a relatively small number of generations (Burt, 2003; Deredec et al., 2008; Simoni et al., 2020; Windbichler et al., 2011). However, such results may only

be valid under ideal conditions, such as random mating and other simplified ecological interaction. Such estimates, therefore, may in some circumstances not provide robust predictions of the drive's behavior under field conditions. Several studies related to the risk assessment of gene drives have highlighted the relevance of ecological and technological bottlenecks like resistance evolution, mate-choice, mating system, and spatial interaction in successfully deploying gene drive organisms (Collins, 2018; Giese et al., 2019; Moro et al., 2018; National Academies of Sciences Engineering and Medicine, 2016; Oye et al., 2014). Thus, assessing model assumption's validity is an essential task that any gene drive technology needs to overcome to become an option for a field release. While numerous assumptions made in the laboratory may be violated in the wild, we choose to focus on aspects relating to the ecological complexity of mating. We demonstrate the effect of mate-choice, aspects of the mating system, and spatial aspects of finding mating partners may change the course of eco-evolutionary trajectories of gene drive systems and thus model predictions.

Gene drive leverages sexual reproduction by biasing the inheritance of a specific gene from one generation to the next. Hence, it becomes imperative to account for the target species' reproductive biology and mating pattern to predict key parameters for a release such as the threshold of gene-drive organisms (GDOs) needed in order to be successful (Moro et al., 2018; National Academies of Sciences Engineering and Medicine, 2016). While the theoretical explorations and laboratory experiments of gene drive techniques often assume simplified mating conditions based on random mating, important other factors also influence mating success in the wild. Non-random mating may result from a range of factors and processes, such as inbreeding, mate-choice, and multiple matings, which are often a part of complex mating systems. These aspects have already been recognized in gene drive research (Deredec et al., 2008; Noble et al., 2017; Qureshi et al., 2019; Unckless et al., 2015). Inbreeding could diminish the frequency of heterozygotes in the population, slowing the spread of gene drive. In natural meiotic drive, females of some species can discriminate against males carrying drive when the region containing the drive gene is linked to mate-choice signals (Price and Wedell, 2008; Wedell and Price, 2015). For

example, the naturally occurring selfish genetic element (*t*-complex) in *Mus domesticus* exhibits mate preference whereby both sexes appear to avoid heterozygous mate using olfactory cues (Lenington, 1983; Lenington, Sarah, 1991; Lindholm et al., 2013).

A newly evolved natural distorter system may remain at low frequency due to reduced fertility of drive carrying individuals, with the resulting potential to selection for mating bias (Charlesworth and Charlesworth, 2010; Wedell and Price, 2015). Though it remains unclear if bias in mate preference can quickly evolve for laboratory-engineered gene drives. A study by Drury et al. (2017) showed that non-random mating caused by inbreeding could render the CRISPR based gene drive inefficient against standing genetic variation resulting in cleavage resistance for Cas9 target sites in the flour beetle *Tribolium castaneum*. Bull (2017) suggested that mild levels of initial inbreeding can lead to the evolution of selfing in hermaphrodites (plants) in response to a homing endonuclease gene drive. Suppression gene drives, aimed at the local eradication of target species, can lead to the evolution of sib-mating, significantly hampering the spread of the driven gene (Bull et al., 2019*b*).

The mating system of target species will also play an essential role in determining the population dynamics of the spread of gene drives. For example, even in the absence of pre-copulatory mate-choice the *t*-haplotype meiotic drive in mice can be limited by their polyandrous mating system where females mate with multiple males in a breeding cycle (Lindholm et al., 2016; Manser et al., 2017). The *t*-haplotype carrying males have reduced fertility, so when a female mates with multiple males, the fertilization of non-drive carrying male due to sperm competition is more likely (Manser et al., 2020, 2017). A sex-linked gene drive based on utilizing *t*-haplotypes has been proposed to suppress the rodent populations (Godwin et al., 2019; Leitschuh et al., 2018). The impact of polyandry on the population-level dynamics of one such proposed gene drive construct (*t-Sry*) has been studied by Manser et al. (2019). *t-Sry* has two components: *t*-haplotypes and sex-determining *Sry* gene and polyandry negatively effect its spread Manser et al. (2019). Focusing on an age-structured population, Huang et al. (2009) showed that the mating system for Medea and engineered under-dominance gene drives can significantly change the predicted

threshold number of released transgenic individuals for successful population transformation. They also found that low polyandry levels can hamper gene drive spread if only males are released. When the gene drive causes male scarcity (Y-shredder), in polygamous systems where males mate with multiple females the efficacy of spread is hampered (Prowse et al., 2019).

Most wild populations do not exist in a single panmictic population but multiple heterogeneous communities across rugged, disconnected landscapes. In a spatially segregated population, individuals are more likely to interact with others in their vicinity than randomly with everyone in the population. Some mathematical models of gene drive use reaction-diffusion models to account for spatial interaction (Beaghton et al., 2017; Girardin et al., 2019; Tanaka et al., 2017). In these systems, the time required for a gene to spread depends on the interaction zone where the wildtype meets the transgenics. This zone is the wave's leading edge in the reaction-diffusion models (Beaghton et al., 2017; Girardin et al., 2019; Tanaka et al., 2017). In the case of suppression drives, the wave sweeps through the wild population, leaving empty space (Barton and Turelli, 2011; Bull et al., 2019*a*; North et al., 2013). Compared to the panmictic models, the suppression drive can be less effective and slow in spatial models (Champer et al., 2021, 2020; North et al., 2013). When considering long-range dispersal, the wildtypes could occupy the empty space created by the suppression drive resulting in local cycles of drive eradication and reoccupation by the wildtype (Champer et al., 2021). Similar cyclical dynamics is possible for reversal drives released to convert the previously established homing drives (Girardin et al., 2019). A question primarily ignored in some of these spatial models concerns the effect of heterogeneous interaction among individuals during mating. For example, the interactions in mathematical models using reaction-diffusion equations are assumed to be homogeneous. The spread of the gene drives relies on sexual reproduction, which is most likely not spatially or temporally uniform for all individuals in a population. A population structured on a network can help account for the natural heterogeneity in mating success. We use concepts from network theory and build a model to investigate how spatial mating networks could affect the gene drive's spread.

Risk assessors face fundamental challenges when using models in their assessments. First, understanding modelling approaches and the underlying assumptions for complex applications like synthetic gene drives is far from trivial. Second, evaluating the effects of ecological factors on gene drive efficacy is not intuitive. Hence, in general, risk assessment of GDOs will be complex and include more uncertainties than current GM crops designed for release into the environment (Simon et al., 2018). Analogous to other risk-assessments (EFSA document on good modelling practice), modelling can be a valuable tool for risk assessment of GMOs, acknowledging that modelling is complex even for presumably simple questions like the impact of Bt Toxins from transgenic maize (Dolezel et al., 2020; EFSA Panel on Plant Protection Products and their Residues, 2014; Fahse et al., 2018). While modelling ecological effects with respects to gene drives is still in its infancy (Dhole et al., 2020), much research focuses on efficacy modelling. However, the view of risk assessors needs to be much broader than only efficacy.

The population-dynamic consequences of mate-choice, mating systems, and mating structure on gene drives are crucial in predicting the transgenic constructs' probability and time to fixation and the release threshold for invading wild population. The effects of mate-choice and mating systems are studied here using deterministic ordinary differential equations. In contrast, the spatial mating structure uses a network model. Although we use different modelling frameworks for different mating complexities, the underlying gene drive model extends from a population genetic perspective. Gene drive systems have previously categorized based on standard terminology; distortion, fertility selection and viability selection (Verma et al., 2021). Here, we extend this approach by adding a generalizable understanding of the effect of some aspects of mating complexity on gene drive dynamics.

Model and Results

As is typical for a functioning gene drive, we assume a diploid organism whose life cycle consists of three stages: zygote, adults and gametes. An adult produces gametes that combine to form

a zygote. The zygote grows up to become an adult, and the cycle continues. We also assume that the organisms are diploid with two alleles for the gene of interest, the wild type allele (W) and the modified allele aimed to be driven (D). Hence, an individual can be either of the three genotypes: WW, DD and WD. Previous work has shown that the gene drive can arise if a drive carrying genotype undergoes distortion, viability or fertility selection that acts during the different life stages of an organism (Verma et al., 2021). Hence, one can categorize various gene drive systems based on pre-existing standard population-genetic terminology (distortion, fertility selection and viability selection). Manipulating the strength of these forces via the engineered construct influences the probability of inheritance, giving rise to gene drive (Verma et al., 2021). Gene drives can also be classified into two types based on the purpose of the release: modification and suppression drives. Suppression drives are aimed to reduce or completely eradicate the wild population, while modification drives are intended to replace the wild population with organisms carrying the gene drives. In this article, we will discuss the modifications that result from distortion and viability selection.

At the gamete level, distortion favors the transmission of the drive allele in the heterozygote. It can give rise to meiotic drive (Lindholm et al., 2016; Sandler and Novitski, 1957) and CRISPR based homing endonuclease gene drive (Noble et al., 2018, 2017). Gametes combine to form zygotes, but specific genotypes may become non-viable. The engineered constructs that work principally by manipulating viability selection are those using zygotic toxin-antidote mechanisms as Medea (Beeman et al., 1992; Gokhale et al., 2014; Ward et al., 2011), Inverse Medea Marshall and Hay (2011) and Semele (Marshall et al., 2011). Fertility selection acts at the adult stage. Empirical studies have shown that selfish genetic elements can reduce the fertility of drive allele carrying organisms (offspring production) (Dyer and Hall, 2019; Larner et al., 2019). These evolutionary forces can become the source or the by-product of the gene drive mechanism. The population dynamics of these systems have been studied independently in (Verma et al., 2021). Here, we subject the target population to three additional factors relevant for field populations: mate-choice, mating structure and mating systems to understand their effect on gene drive population

dynamics (figure 1).

Mate-choice

We first consider the null case where there is no gene drive and understand how mate-choice bias of wildtype against transgenic will affect the population dynamics. The mating rate among the wildtypes is set to one. Similarly, the mating rate among the drive types is also one. Mate choice bias in our model is captured by the parameter h (figure 1). The mating rate among the wildtypes (WW) and the drive types (WD or DD) is $(1-h)$. If $h=0$, the wildtype (WW) are equally likely to mate with the drive carrying genotype (WD and DD). While if $h=1$, the wildtype (WW) and the drive type (WD or DD) do not mate at all. During the exploration of parameter space (h), we work under the assumption that the wildtype genotypes are less likely to mate with individuals carrying the drive allele (WD and DD); therefore, $0 \leq h \leq 1$. The above assumption can be justified with the observation that for natural gene drives or even in sterile insect technique (SIT), when a female choice of mates is "active," i.e. females choose among males, wild females preferred wild males over a drive carrying males or mass-reared sterile males (Price and Wedell, 2008; Robinson and Hendrichs, 2005; Wedell and Price, 2015). For simplicity, in our model, both sexes (male and female) of WW have an equal bias against mating with WD or DD. The rate of the production for the three genotypes assuming an infinitely large population and random segregation of alleles during meiosis is given by,

$$\begin{aligned} F_{WW} &= x_{WW}^2 + (1-h)x_{WW}x_{WD} + \frac{x_{WD}^2}{4} \\ F_{WD} &= (1-h)x_{WW}x_{WD} + x_{WD}x_{DD} + 2(1-h)x_{WW}x_{DD} + \frac{x_{WD}^2}{2} \\ F_{DD} &= x_{DD}^2 + x_{WD}x_{DD} + \frac{x_{WD}^2}{4} \end{aligned} \tag{1}$$

where x_α and F_α are the frequency and rate of genotype production respectively, and $\alpha \in (WW, WD, DD)$. The following set of differential equations governs the population dynamics

of the genotypes in continuous time:

$$\dot{x}_\alpha = F_\alpha - x_\alpha \bar{F}. \tag{2}$$

where $\bar{F}$ is the average fitness of the three genotypes:

$$\bar{F} = \sum_\alpha F_\alpha. \tag{3}$$

The frequencies of all genotypes is normalised to one.

$$x_{WW} + x_{WD} + x_{DD} = 1. \tag{4}$$

The above constraints on frequencies allow us to represent the dynamics of equation (2) on a de Finetti diagram. The frequency of the three genotypes (WW, WD and DD) without mate-choice ($h = 0$) converge to Hardy Weinberg equilibrium (Gokhale et al., 2014; Verma et al., 2021). When we introduce the mate-choice parameter into the rate equations (1), the dynamics deviate from Hardy Weinberg equilibrium and is governed by the fixed points that appear in the interior of the de Finetti diagram. In this context, a fixed point is a specific composition of the population $(x^*_{WW}, x^*_{WD}, x^*_{DD})$ where the proportion of all the genotypes does not change. Specifically, where $\dot{x}_\alpha = 0 \; \forall \; \alpha \in (WW, WD, DD)$. Primarily, there are two types of fixed points: stable and unstable. If the population is at the stable fixed point, a slight change in the population composition will bring the population to the stable fixed point. While in unstable fixed points, a small change will diverge the population composition away from an unstable fixed point. The position of these fixed points governs the overall population dynamics of a specific case. For example, population dynamics for a particular case of $h = 0.9$ is shown in the inset of figure 2A. The position of an unstable interior fixed point decides the evolutionary fate of the population.

In figure 2, we plot the positions and trajectories of these interior fixed points for different mate-choice (h) values under scenarios such as null case, viability selection, distortion, fertility selection. The null case is when only the effect of mate-choice is considered without any gene drive arising from viability selection, distortion, and fertility selection (figure 2A). Even under slight mate-choice bias ($h = 0.01$), the dynamics quickly deviates from the Hardy Weinberg

equilibrium. An unstable fixed point (saddle point) appears in the interior of the de Finetti diagram. The threshold frequency of transgenic genotype (DD or WD) required for population transformation is closely related to the position of these unstable fixed points. The area to the left of the unstable fixed point is the basin of attraction of wild-type genotype. The trajectories of the initial conditions in this area lead to the extinction of the modified allele. In contrast, the area on the right is the basin of attraction of drive homozygotes (DD), leading to population transformation. Increasing the mate-choice bias (or as h increases from 0.01 to approximately 1), the position of the interior fixed point moves towards the middle of WW and DD line (figure 2A). It implies that when the mate-choice bias increases, the threshold amount of transgenics (DD and WD) required to transform the wildtype population increases even without the gene drive.

Mate-choice with Viability Selection (Medea)

Many toxin-antidote gene drive designs, including Medea, Inverse Medea, Semele, and designed under-dominance drive, exhibit viability selection (Beeman et al., 1992; Marshall and Hay, 2011; Marshall et al., 2011). In such systems, specific offsprings become non-viable during the zygote stage of the life cycle. We have focused on the Medea gene drive system in our analysis, where d measures the drive efficiency. In Medea gene drive, wildtype homozygous offspring of heterozygous mothers become non-viable (Akbari et al., 2014; Buchman et al., 2018; Gokhale et al., 2014; Ward et al., 2011). The rate of production of genotypes in the for Medea gene drive with the incorporation of mate-choice bias can be written as:

$$\begin{aligned} F_{WW} &= x_{WW}^2 + (1-h)(1-0.5d)x_{WW}x_{WD} + (1-d)\frac{x_{WD}^2}{4} \\ F_{WD} &= (1-h)\frac{x_{WW}x_{WD}}{2} + x_{WD}x_{DD} + 2(1-h)x_{WW}x_{DD} + \frac{x_{WD}^2}{2} \\ F_{DD} &= x_{DD}^2 + x_{WD}x_{DD} + \frac{x_{WD}^2}{4} \end{aligned} \tag{5}$$

Figure 2B shows the position and trajectory of the unstable fixed point for viability selection based Medea gene drive with 100% efficiency, i.e. $d = 1$. The population dynamics equation can been derived using equation (2) and (5). When the mating rate between transgenic and wildtype

decreases via h, the unstable fixed point moves towards DD vertex in the de Finetti diagram following a projectile trajectory (figure 2B). Hence here, mate-choice bias increases the threshold release of transgenics. For $h \approx 1$, the number of transgenics released needs to be almost half the target population size for achieving total population replacement. These results are also consistent with the invasion condition of equation (A3) derived in appendix A for Medea gene drive.

Mate-choice with Distortion

Here we will consider the case of distorted allele transmission in addition to mate-choice bias introduced by h. There are several distortion based gene drives, but here we will focus on a meiotic drive where the distortion efficiency is p. More specifically, p is the probability of transmission of drive allele from heterozygous parent to offspring. If $p = 1$, the gene drive system mimics CRISPR/Cas-9 based homing endonuclease drive with 100% efficiency (Noble et al., 2017). If a drive allele is transmitted from heterozygous parents with probability p, the rate of genotype production then changes to,

$$\begin{aligned} F_{WW} &= x_{WW}^2 + 2(1-h)(1-p)x_{WW}x_{WD} + (1-p)^2 x_{WD}^2 \\ F_{WD} &= 2(1-h)px_{WW}x_{WD} + 2(1-p)x_{WD}x_{DD} + 2(1-h)x_{WW}x_{DD} + 2p(1-p)x_{WD}^2 \\ F_{DD} &= x_{DD}^2 + 2px_{WD}x_{DD} + p^2 x_{WD}^2 \end{aligned} \tag{6}$$

Again the population dynamics for the distorted case is given by equation (2), but the effective genotype production rate changes according to equation (6). In figure 2C we focus on the scenario when the distortion based gene drive such as meiotic drive or CRISPR drive with 100% efficiency (refer equation (6) for $p = 1$). We observe that the interior unstable fixed point only appears after the mate-choice bias becomes greater than 50% or $h > 0.5$, unlike viability based gene drive Medea (figure 2B & C). For $h < 0.5$, a small transgenic release is enough for population transformation to drive homozygotes (DD). Hence, the distortion based gene drives appear to be more robust against the mate-choice than viability-based gene drive Medea. These results

are also consistent with the condition of invasion derived in appendix A for the distortion based gene drive (see equation (A6)).

Mate-choice with Fertility Selection

The relative number of offspring produced may differ because of the variation in the mating pair's fertility resulting from their genotypes. The fitness component due to differential fertilities is included in the parameters f_α where $\alpha \in$ (WW, WD, DD). The rate of the offspring production for the three genotypes because of fertility selection changes to,

$$\begin{aligned} F_{WW} &= f_{WW}^2 x_{WW}^2 + (1-h) f_{WW} f_{WD} x_{WW} x_{WD} + f_{WD}^2 \frac{x_{WD}^2}{4} \\ F_{WD} &= (1-h) f_{WW} f_{WD} x_{WW} x_{WD} + f_{WD} f_{DD} x_{WD} x_{DD} + 2(1-h) f_{WW} f_{DD} x_{WW} x_{DD} + f_{WD}^2 \frac{x_{WD}^2}{2} \\ F_{DD} &= f_{DD}^2 x_{DD}^2 + f_{WD} f_{DD} x_{WD} x_{DD} + f_{WD}^2 \frac{x_{WD}^2}{4}. \end{aligned} \tag{7}$$

To observe the effect of fitness cost on fertility, we consider a scenario where $f_{WW} = 1$, $f_{WD} = (1-c)$, $f_{DD} = (1-c)^2$ for the dynamical equations derived using equation (7). Here, we assume multiplicative fitness cost and c denotes the fertility-fitness cost of the drive allele. The two internal fixed points appear only after substantial mate-choice bias $h \approx 0.656$ (figure 2D). One of the fixed points is unstable, and the other is stable. Therefore, with multiplicative fitness cost on the fertility of the transgenic organism, due to drive-allele payload, mate-choice can result in the coexistence of all the three genotypes. When $h < 0.656$, the global stable fixed point lies at the vertex of wildtype population (WW); hence no amount of drive release can replace the wild population; however, complete fixation may not be a necessary aim in all applied scenarios.

Besides understanding the impact of mate-choice on the population dynamics, we also indirectly probe the threshold fraction of transgenic organisms needed to be released for complete population replacement relative to the target population size. In figure 3, we numerically calculate the threshold frequency of drive homozygotes (DD) necessary to invade a population consisting of wildtypes (WW). We evaluate the impact of mate-choice bias (h), gene drive efficiency and fertility-fitness cost for two gene drive systems, namely meiotic drive and Medea.

Figure 3A shows that the mate-choice bias increases the invasion threshold frequency of DD required for complete population replacement for Medea drive. The threshold frequency of DD also slightly increases with decreasing drive efficiency. The change in threshold frequency due to drive-efficiency reduces for a higher bias in mate-choice. The release threshold is close to zero for lower mate-choice bias, represented by the heatmap's light colour. The position of fixed point for the case of 100% drive efficiency ($p = 1$ and $d = 1$) for both figure 3A & B corresponds to the scenario studied in 2B & C respectively. For the distortion-based drive, lower mate-choice and sufficiently high distortion probability do not change the threshold frequency. The region in the heatmap where a minimal transgenic release can transform the population is significantly high for the distortion-based drive than Medea drive. When the mate-choice bias is high enough $(h > 0.5)$, an increase in distortion probability only slightly decreases the invasion threshold of DD. In this regime $(h > 0.5)$, a substantial frequency of DD is required for the population of wildtype to be invaded even for a very high distortion probability.

In figure 3C & D corresponds to the case when there is a cost on the fertility fitness of the drive carrying organism ($c = 0.1$ hence $f_{WD} = 0.9$ and $f_{DD} = 0.81$). Fitness cost leads to an increase in the invasion threshold frequency for both the gene drive systems overall. Moreover, any DD release is insufficient to invade the wildtype population for inefficient drives under low mate-choice bias. The dark colour represents this region in the heatmap. Interestingly, increasing the mate-choice bias can facilitate the invasion by DD even for less efficient drives. The distortion based gene drive appears to be more robust against the ecological stress of mate-choice bias even when considering the fitness costs.

Mating systems

Gene drive technology relies on sexual reproduction between the mating pairs for its transmission in the population. Most of the target species of interest have a polygamous mating system instead of the commonly assumed monogamous mating system (Moro et al., 2018; Rode et al., 2019). As introduced in the previous section of mate-choice, the model is modified here to incorporate

this aspect of the mating system. In this model, we will consider two separate populations of the two sexes. We assume that the offspring of both sexes are produced in equal proportion. The frequency of male and female's genotypes are denoted using x_i and y_j. There are three possible genotypes: wildtype (WW), drive heterozygotes (WD) and drive homozygotes (DD). Let us consider the mating system when one male mates with r females. Hence $r = 1$ represents the monogamous mating system while $r > 1$ corresponds to the polygynous mating system. The following set of equations gives the frequencies of the genotypes produced with the polygamous mating system, as the equation holds for both males and females (with a change in variable x_i and y_j):

$$F_k(r) \quad = \quad \sum_{\alpha,\beta_1,\beta_2,\ldots\beta_r} \sum_{j=1}^{r} M_k\left(\alpha,\beta_j\right) f_\alpha x_\alpha \prod_{i=1}^{r} f_{\beta_i} y_{\beta_i} \tag{8}$$

Here, $M_k\left(\alpha,\beta_j\right)$ is the proportion of genotype k produced from the mating between a male of genotype α and a female of genotype β_j. α and β_j are dummy indexes for any of the three genotypes WW, WD or DD. The elements of the matrix $M_k\left(\alpha,\beta_j\right)$ will depend on the gene drive type as well. Matrix M_k for Medea (equation (A7)-(A9)) and distortion based gene drive system (equation (A10)-(A12)) is given in appendix A. The summation over α and β_j is carried out over the set of all genotypes (WW, WD, DD). We have also assumed a polygamous mating system of mating ratio r, i.e. one male mates with r female or vice-versa. Equation (8) may be interrupted in parts as selecting a male of genotype α and selecting r females of genotype $\beta_1, \beta_2, \ldots, \beta_r$. Finally, the contribution of all possible matings in producing genotype k is summed up.

Simplifying equation (8) by expansion formula for multinomial expression yields,

$$F_k(r) = rF_k(1)(f_{WW}y_{WW} + f_{WD}y_{WD} + f_{DD}y_{DD})^{r-1} \tag{9}$$

The following set of differential equations governs the population dynamics of the genotypes in continuous time:

$$\begin{aligned} \dot{x}_\alpha &= \frac{1}{2}F_\alpha(r) - x_\alpha \bar{F}(r) \\ \dot{y}_\alpha &= \frac{1}{2}F_\alpha(r) - y_\alpha \bar{F}(r) \end{aligned} \tag{10}$$

where $\bar{F}$ is the sum of rates of genotype production:

$$\bar{F}(r) = \sum_{\alpha} F_{\alpha}(r) \tag{11}$$

The total population of both males and females remains constant and sum up to unity.

$$x_{WW} + x_{WD} + x_{DD} = 1 \tag{12}$$

$$y_{WW} + y_{WD} + y_{DD} = 1 \tag{13}$$

In equation (9), $F_k(r = 1)$ and $F_k(r > 1)$ is the production rate of genotype k for monogamous ($r = 1$) and polygamous ($r > 1$) mating system respectively. It implies that the equilibrium population dynamics for both monogamous ($r = 1$) and polygamous ($r > 1$) mating systems, even with gene drives, are equivalent. In other words, the final population composition of the genotypes remains the same for both polygamous and monogamous mating systems. Previous studies without any gene drive also support that the equilibrium dynamics for monogamy and polygamy remain the same (Karlin, 1978; O'Donald, 1980). However, the difference lies in the relative time to reach population equilibrium. It can be shown that after simplifying the equation (10) obtained for $r > 1$, the rate of increase of different genotypes is equivalent to the case of monogamy ($r = 1$) with rescaled time. The expectation is that the gene drive will spread faster in polygamous mating species compared to monogamy (Moro et al., 2018). Hence, the time required for the drive allele to spread through the population should increase for the monogamous mating system. Our result also supports the expected outcome. Here we quantify the same.

If we first look at the case where there is no fitness cost of the gene drive, and only the efficiency of the two gene drive system based on distortion and viability selection are varied. Figure 4A, B & C shows that gene drive will spread faster for species with a high degree of polygamy (r). It can also be seen by comparing figures 4A & B that the distortion-based gene drive will spread faster than the viability-based Medea drive. The time for the gene drive to reach 99% frequency is an order of magnitude higher for Medea drive compared to CRISPR homing drive or meiotic drive. A higher degree of polygamy (r) reduces the time required to

reach critical drive frequency (99%) for both the gene drive system. This reduction in absolute time value becomes more pronounced when the gene drive is less efficient (figure 4A & B).

Figure 4C clearly shows that the relative time required for the drive allele to reach 99% frequency is rescaled exactly by a factor of $1/r$ for the polygamy relative to the monogamous mating system. This is in line with the relation obtained in equation (9). When $f_{WW} = f_{WD} = f_{DD}$, the production rate of offspring for polygamy is r times that for the monogamous mating system. But, when we have a fitness cost c for carrying a drive allele, the relation between the time to reach 99% frequency and degree of polygamy becomes more complex (figure 4D). An increase in the degree of polygamy first decreases the relative time to reach the drive allele's critical frequency ($r = 2$ and $r = 4$), but a further increase in the degree of polygamy ($r = 6, 8, 10$) elevates it. In figure 4, it can also be noted that when the distortion probability is low ($p < 0.625$), the drive allele is not able to invade the wildtype population. This is in congruence with the condition of invasion derived for the monogamous case in equation (A6) in the appendix.

The above result can be understood from the equation (9) where the fitness cost makes the factor $(f_{WW}y_{WW} + f_{WD}y_{WD} + f_{DD}y_{DD})^{r-1}$ less than one. The factor $(f_{WW}y_{WW} + f_{WD}y_{WD} + f_{DD}y_{DD})^{r-1}$ decreases exponentially with increasing level of polygamy r. Hence the time is rescaled by the factor of $\frac{1}{r(f_{WW}y_{WW}+f_{WD}y_{WD}+f_{DD}y_{DD})^{r-1}}$ effectively. The time first decreases when dominated by $1/r$ with an increase in r but later on decreases when dominated by $1/(f_{WW}y_{WW} + f_{WD}y_{WD} + f_{DD}y_{DD})^{r-1}$. When the fitness cost is $c = 0.2$, the relative time until the drive allele reaches 99% frequency with respect to monogamy decreases for $r = 2$ and $r = 4$, but then it starts to increase for $r = 6$. For $r = 8$ and $r = 10$ spread of gene-drive becomes slower compared to monogamy. Another way to understand the results is that the rate of production genotype DD first increases up to a point for increasing level of polygamy r but later decreases for moderate fitness cost (figure B1). Hence the time in spreading gene drive is lowest for intermediate levels of polygamy. Further increase in the degree of polygamy reduces the production of DD and therefore increases the time to spread the drive allele.

Spatial network interaction

The population dynamics of CRISPR based homing endonuclease gene drive have been extensively studied for well-mixed infinitely large (Noble et al., 2017) and finite populations (Noble et al., 2018). But most species occur in a partially heterogeneous landscape where they interact and mate with other individuals in their vicinity. Hence, a network-based population is an appropriate framework to model dynamics in such structured populations.

We considered a structured population of n individuals. The individuals live on a random network with an average degree of k; thus, each individual has k connections on average. Here k controls the number of mating opportunities and the level of competition for an individual. The population is updated via a death-birth process (figure 5) described as follows: First, an individual is chosen randomly for death. Then two parents are selected, who are neighbours of the dead individual with probability proportional to their fertility fitness. According to their genetic archetype, the selected parents contribute their gametes, where other genetic effects like distortion can come into play. The combination of these contributed gametes forms the offspring that replaces the dead individual in the network. The population is updated until it fixes to all WW or all DD states.

In figure 6, we exhibit the stochastic network model by running the simulation several times and plotting fixation probability and conditional fixation time with variation in the average number of interacting individuals per site (represented by k). We also studied the impact of increasing the number of released transgenic (WD and DD) and different genotypes (WD and DD). Here, k controls the number of mating opportunities and competition during the birth process. When k increases, the fixation probability of DD decreases mainly due to higher competition during the birth update per site (figure 6A). As expected, distortion probability has a positive impact on the fixation probability of DD. The effect is more pronounced for lower values of an average degree since the heterogeneity in the number of connected individuals is also high for this case. Fixation probability also increases as the number of released DD increases (figure 6A). Unexpect-

edly, DD transgenic release has a lower chance of getting fixed than a WD release (figure 6B). This observation is mainly because the fitness cost of DD is relatively high compared to WD ($f_{WD} = 0.50, f_{DD} = 0.25$). If the fitness cost is negligible and the drive efficiency is high, the release of the DD genotype is expected to fix the gene drive with a higher probability. The effect on fixation probability by the release of WD compared to DD becomes more pronounced with the increase in average degree k (figure 6B). It increases first with an increase in the release number of transgenic, attains a maximum and decreases later. We also plotted conditional fixation time with variation in the number of releases and the network degree (figure 6C). The conditional fixation time is lower for a high number of releases and lower values of average degree k. The difference in the value of conditional fixation time is increased when the release number hence the loss of drive due to stochasticity is high. In all of our simulations for the release nodes of transgenic are chosen at random.

Discussion

Gene drive is one of the tools of synthetic biology that has the potential to transform whole wild populations. The transformation uses and modulates one of the foundational tenets of evolution - the inheritance of traits through sexual reproduction. Thus, variation in the reproductive biology and the mating behaviour of the target species can affect the eventual spread of the gene drive. While previous studies have emphasized the evolution of resistance to gene drives, we examined some of the ecological assumptions related to mating systems and their effect on the potential outcome of a drive release (Champer et al., 2018; Noble et al., 2017; Unckless et al., 2017). Herein, we have examined some factors related to complex life histories and social interactions which, depending on the organism are relevant under field conditions, namely mate-choice, multiple matings, and spatial aspects of the mating network. We found that the above factors substantially influence predictions of the number of transgenic gene drive individuals to be released into a population for a successful invasion. The factors are also highly relevant to estimate better the

fixation time of the drive gene to plan any field release.

First, we considered a monogamous situation that deviates from panmixis when individuals actively choose partners (mate choice). Any drive linked to the ornaments of mate-choice, a mate-choice bias can develop with a significant effect on the release threshold of a gene drive, as shown in figure 3. Inefficient drive and fitness costs due to drive-payload aggravate the situation, and the predicted threshold release is drastically different compared to a situation with no mate-choice bias. This finding is not only important to estimate the drive efficacy but also highly relevant for the risk assessment of the drive. Comparing different drive approaches, we found that distortion based gene drives fare much better than drives based on viability selection under the ecological more realistic condition of mate choice. Hence for regulatory checks, gene drive constructs should be evaluated for their robustness against various ecological stressors. The findings may also be relevant for resistance evolution against the drive if the target species could evolve such mate-choice preferences. A fast evolution can be assumed as many drives target fast reproducing species. Experience from sterile insect techniques has taught that different rearing conditions in the lab and wild can also give rise to different behavioural and genetic traits leading to divergent mating preferences and eventual program failure (Eberhard, 1999; Lance et al., 1998; McInnis et al., 1996; Robinson and Hendrichs, 2005).

Next, we consider the potential for multiple matings for both females and males. Under these scenarios, the final evolutionary outcome of the spread of the gene drive (distortion or Medea drive) for a polygamous mating was the same in our simulations as that of the monogamous system. Even the species with a higher degree of polygamy will converge to the same evolutionary fate for a given gene drive system. However, the time needed for the spread of the drive gene will be affected by the level of multiple matings and the fitness cost linked to the drive. Time to fixation will be smaller for a higher degree of polygamy in the absence of any fitness cost. However, a moderate fitness cost under different polygamy levels will trigger a non-linear outcome for the time till drive establishment. This non-linearity is because the production rate of drive homozygote first increases but later decreases in line with the degree of polygamy for

moderate fitness cost (figure B1). Hence, the drive gene is expected to spread faster for species with intermediate levels of polygamy when there is an associated fitness cost of the drive allele.

Last, we examined the spatial implications of finding mating partners (mating opportunities) on the model outcome. To this end, the framework developed for the spatial mating interaction can be applied to any diploid population, regardless of the presence of gene drive. Considering a finite population on a network allows us to understand the probable outcomes of gene drive release. A finite population leads to stochastic fluctuations in the frequencies of the genotypes resulting in different outcomes for the same initial conditions. We found that the spread of transgenic release is lowered when individuals, on average, have more mating opportunities and intra-sexual competition. Thus the fixation time for the transgenic increases with an increase in the average degree of the mating network. Concerning the question of how the connectivity of mating networks are varied in wild populations , it is reported that selective pressures under which species evolve shape their network structure in the environment (Pinter-Wollmann et al., 2014). Hence, changes in environmental conditions such as resource availability, seasonal effects, selective pressure, and life-history traits all can impact on the network structure. Within a species itself, variation at the individual level can also lead to heterogeneous connectivity. Species with sparsely connected individuals on the mating network have a higher chance of fixing drive genes and in a shorter time. We also observe that the success in fixation of drive homozygotes can be mitigated by releasing more transgenic individuals. Furthermore, when there is a high fitness cost associated with carrying a drive allele payload, releasing drive heterozygotes instead of homozygotes would result in a higher chance of gene drive fixation.

In this study, we have decided to focus on three factors related to the mating complexities of target species. Still, many other ecological and environmental factors can impact the spread of gene drives. Known factors include the age structure of a population, spatial landscape and seasonality (Eckhoff et al., 2017; Huang et al., 2009, 2011; North et al., 2013, 2019; North, Ace R and Burt, Austin and Godfray, H Charles J, 2020). Gene drive behavior and the interactions of a drive released in a complex ecosystem over long time periods is highly complex. Navigating

this ecological complexity may seem insurmountable (Levin, 2003). However, for any technology aiming to intervene in complex systems, we will face a similar control problem. As such, it is not workable to address *in silico* all possible ecological and evolutionary pressures and scenarios that an engineered system will meet in the real world (Denton and Gokhale, 2019; Lindvall and Molin, 2020). Undoubtedly, modelling will play a key role in understanding drive spread. Our study emphasizes that modelling needs a more ecological reality to be really predictive of the drive behavior. Identifying and collecting necessary information on the effect of primary ecological and evolutionary pressures will be thus crucial to access the risk before any field deployment (James et al., 2018; Long et al., 2020; National Academies of Sciences Engineering and Medicine, 2016).

Conclusion

To date, most of the gene drive modelling exercises focus on drive spread under simplified conditions such as panmixis. In this study, we tested whether more complex assumptions, characteristic of many species and mating systems, justify the use of simplified assumptions. The results show that ecological factors related to mating can substantially change drive spread and, as supposedly many other ecological factors, may strongly impact the temporal and spatial dynamics of gene drive systems. Modelling may be used to predict gene drive spread and thus to assist the risk assessment. In this case, mating-related parameters, as all critical assumptions related to the ecology of the species, need to undergo a reality check. In a wider sense, the new modelling framework, including tools for analyzing spatial interactions or multiple matings, are generic and have the potential to be applied to any diploid population, independent of gene drive applications.

Acknowledgments

The work has been supported by funds from the Max Planck Society and is part of the R&D project "Risk assessment of synthetic gene-drive applications"(FKZ 3518 84 0500) funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety on behalf of the Federal Agency for Nature Conservation (BfN), Germany.

Data and Code Availability

All data and simulation codes for generating figures are available on Github (`https://github.com/tecoevo/genedrives_mating`).

Appendix A: Additional Methods

Invasion condition for Medea drive with Mate choice (h)

If we consider the case of Medea gene drive with fertility selection. The rate of production of the three genotype is given by the combination of equation (5) and (7),

$$\begin{aligned} F_{WW} &= f_{WW}^2 x_{WW}^2 + (1-h)(1-0.5d) f_{WW} f_{WD} x_{WW} x_{WD} + (1-d) f_{WD}^2 \frac{x_{WD}^2}{4} \\ F_{WD} &= (1-h) f_{WW} f_{WD} \frac{x_{WW} x_{WD}}{2} + f_{WD} f_{DD} x_{WD} x_{DD} + \\ &\quad 2(1-h) f_{WW} f_{DD} x_{WW} x_{DD} + f_{WD}^2 \frac{x_{WD}^2}{2} \\ F_{DD} &= f_{DD}^2 x_{DD}^2 + f_{WD} f_{DD} x_{WD} x_{DD} + f_{WD}^2 \frac{x_{WD}^2}{4} \end{aligned} \tag{A1}$$

The rate of change of frequencies of each genotype is still given by equation (2). We use the constraint on the frequencies of the three genotypes in equation (4) to reduce the population dynamics of the genotypes to two variables after replacing $x_{WD} = 1 - x_{WW} - x_{DD}$ in equation (2). The drive will not invade the wildtype population if both the eigenvalues of the dynamical system are negative. Eigenvalues can be deduced from the Jacobian matrix (J_d) of the system at $(x_{WW}, x_{WD}, x_{DD}) = (1, 0, 0)$,

$$J_d = \begin{pmatrix} f_{WD} f_{WW}(1-h) - f_{WW}^2 & f_{WD} f_{WW}(1-h) - 2 f_{DD} f_{WW}(1-h) \\ 0 & -f_{WW}^2 \end{pmatrix} \tag{A2}$$

Hence, Medea gene drive can invade a population of wildtype if

$$(1-h) f_{WD} > f_{WW} \tag{A3}$$

Note that the above invasion condition is independent of the efficiency of the Medea gene drive (d).

Invasion condition for Distortion drive with Mate choice (h)

Consider the scenario of distortion based gene drive with fertility selection. The rate of production of the three genotypes will then be governed by the combination of equation (6) and (7),

$$\begin{aligned} F_{WW} &= f_{WW}^2 x_{WW}^2 + 2(1-h)(1-p) f_{WW} f_{WD} x_{WW} x_{WD} + (1-p)^2 f_{WD}^2 x_{WD}^2 \\ F_{WD} &= 2(1-h) p f_{WW} f_{WD} x_{WW} x_{WD} + 2(1-p) f_{WD} f_{DD} x_{WD} x_{DD} \\ &\quad + 2(1-h) f_{WW} f_{DD} x_{WW} x_{DD} + 2p(1-p) f_{WD}^2 x_{WD}^2 \\ F_{DD} &= f_{DD}^2 x_{DD}^2 + 2p f_{WD} f_{DD} x_{WD} x_{DD} + p^2 f_{WD}^2 x_{WD}^2 \end{aligned} \tag{A4}$$

Similar to the Medea gene drive scenario, the population dynamics of the above system can be written in the form of two variables x_{WW} and x_{DD} using equation (4). The Jacobian matrix (J_m) of the system at $(x_{WW}, x_{WD}, x_{DD}) = (1, 0, 0)$ is given by

$$J_d = \begin{pmatrix} 2f_{WD} f_{WW}(1-h)p - f_{WW}^2 & 2f_{WD} f_{WW}(1-h)p - 2f_{DD} f_{WW}(1-h) \\ 0 & -f_{WW}^2 \end{pmatrix} \tag{A5}$$

From the condition on the eigenvalues, the gene drive can invade wildtype population if,

$$2(1-h)p f_{WD} > f_{WW} \tag{A6}$$

Note that when there is no mate choice ($h = 0$) the above condition reduces to the invasion condition derived by Noble et al. (2017) for CRISPR gene drive.

[See section figure legends for Figure A1]

$M_k\left(\alpha, \beta_j\right)$ *in equation* (8) *for Medea and Distortion Based Gene Drive*

Medea Gene Drive

$$M_{WW} = \begin{bmatrix} 1 & 0.5(1-d_m) & 0 \\ 0.5 & 0.25(1-d_m) & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{A7}$$

$$M_{WD} = \begin{bmatrix} 1 & 0.5(1-d_m) & 0 \\ 0.5 & 0.25(1-d_m) & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{A8}$$

$$M_{DD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.25 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \tag{A9}$$

Distortion Based Gene Drive

$$M_{WW} = \begin{bmatrix} 1 & (1-p) & 0 \\ (1-p) & (1-p)^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{A10}$$

$$M_{WD} = \begin{bmatrix} 0 & p & 1 \\ p & 2p(1-p)) & (1-p) \\ 1 & (1-p) & 0 \end{bmatrix} \tag{A11}$$

$$M_{DD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & p^2 & p \\ 0 & 0 & 1 \end{bmatrix} \tag{A12}$$

Appendix B: Supplementary Figures

See section figure legends for Figure B1.

Literature Cited

Akbari, O. S., C.-H. Chen, J. M. Marshall, H. Huang, I. Antoshechkin, and B. A. Hay. 2014. Novel synthetic medea selfish genetic elements drive population replacement in drosophila; a theoretical exploration of medea-dependent population suppression. ACS synthetic biology 3:915–928.

Barrett, L. G., M. Legros, N. Kumaran, D. Glassop, S. Raghu, and D. M. Gardiner. 2019. Gene drives in plants: opportunities and challenges for weed control and engineered resilience. Proceedings of the Royal Society B 286:20191515.

Barton, N. H., and M. Turelli. 2011. Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of allee effects. The American Naturalist 178:E48–E75.

Beaghton, A., P. J. Beaghton, and A. Burt. 2017. Vector control with driving y chromosomes: modelling the evolution of resistance. Malaria journal 16:286.

Beeman, R., K. Friesen, and R. Denell. 1992. Maternal-effect selfish genes in flour beetles. Science 256:89–92.

Brossard, D., P. Belluck, F. Gould, and C. D. Wirz. 2019. Promises and perils of gene drives: Navigating the communication of complex, post-normal science. Proceedings of the National Academy of Sciences of the United States of America 116:7692–7697.

Buchman, A., J. M. Marshall, D. Ostrovski, T. Yang, and O. S. Akbari. 2018. Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. Proceedings of the National Academy of Sciences of the United States of America 115:4725–4730.

Bull, J. J. 2017. Lethal gene drive selects inbreeding. Evolution, medicine, and public health 2017:1–16.

Bull, J. J., C. H. Remien, R. Gomulkiewicz, and S. M. Krone. 2019*a*. Spatial structure undermines parasite suppression by gene drive cargo. PeerJ 7:e7921.

Bull, J. J., C. H. Remien, and S. M. Krone. 2019*b*. Gene-drive-mediated extinction is thwarted by population structure and evolution of sib mating. Evolution, medicine, and public health 2019:66–81.

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

Carballar-Lejarazú, R., C. Ogaugwu, T. Tushar, A. Kelsey, T. B. Pham, J. Murphy, H. Schmidt, Y. Lee, G. C. Lanzaro, and A. A. James. 2020. Next-generation gene drive for population modification of the malaria vector mosquito, anopheles gambiae. Proceedings of the National Academy of Sciences 117:22805–22814.

Champer, J., I. K. Kim, S. E. Champer, A. G. Clark, and P. W. Messer. 2021. Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. Molecular Ecology 30:1086–1101.

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

Champer, J., J. Zhao, S. E. Champer, J. Liu, and P. W. Messer. 2020. Population dynamics of underdominance gene drive systems in continuous space. ACS Synthetic Biology 9:779–792.

Charlesworth, B., and D. Charlesworth. 2010. Elements of evolutionary genetics. Roberts and Company Publishers. Greenwood Village, CO.

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 12:9.

Denton, J. A., and C. S. Gokhale. 2019. Synthetic mutualism and the intervention dilemma. Life 9:15.

583 Deredec, A., A. Burt, and H. C. J. Godfray. 2008. The population genetics of using homing
584 endonuclease genes in vector and pest management. Genetics 179:2013–2026.

585 Dhole, S., A. L. Lloyd, and F. Gould. 2020. Gene drive dynamics in natural populations: The
586 importance of density dependence, space, and sex. Annual Review of Ecology, Evolution, and
587 Systematics 51:505–531.

588 Dolezel, M., C. Lüthi, and H. Gaugitsch. 2020. Beyond limits–the pitfalls of global gene drives
589 for environmental risk assessment in the european union. BioRisk 15:1.

590 Drury, D. W., A. L. Dapper, D. J. Siniard, G. E. Zentner, and M. J. Wade. 2017. Crispr/cas9 gene
591 drives in genetically variable and nonrandomly mating wild populations. Science advances
592 3:e1601910.

593 Dyer, K. A., and D. W. Hall. 2019. Fitness consequences of a non-recombining sex-ratio drive
594 chromosome can explain its prevalence in the wild. Proceedings of the Royal Society B
595 286:20192529.

596 Eberhard, W. G. 1999. Sexual behavior and sexual selection in the mediterranean fruit fly, ceratitis
597 capitata (dacinae: Ceratitidini). Pages 477–508 *in* Fruit Flies (Tephritidae). CRC Press.

598 Eckhoff, P. A., E. A. Wenger, H. C. J. Godfray, and A. Burt. 2017. Impact of mosquito gene drive
599 on malaria elimination in a computational model with explicit spatial and temporal dynamics.
600 Proceedings of the National Academy of Sciences 114:E255–E264.

601 EFSA Panel on Plant Protection Products and their Residues. 2014. Scientific opinion on good
602 modelling practice in the context of mechanistic effect models for risk assessment of plant
603 protection products. EFSA Journal 12:3589.

604 Fahse, L., P. Papastefanou, and M. Otto. 2018. Estimating acute mortality of lepidoptera caused
605 by the cultivation of insect-resistant bt maize–the lepix model. Ecological Modelling 371:50–59.

606 Gantz, V. M., N. Jasinskiene, O. Tatarenkova, A. Fazekas, V. M. Macias, E. Bier, and A. A.
607 James. 2015. Highly efficient cas9-mediated gene drive for population modification of the
608 malaria vector mosquito anopheles stephensi. Proceedings of the National Academy of Sci-
609 ences 112:E6736–E6743.

610 Giese, B., J. Frieß, N. Barton, P. Messer, F. Débarre, M. Schetelig, N. Windbichler, H. Meim-
611 berg, and C. Boëte. 2019. Gene drives: dynamics and regulatory matters—a report from the
612 workshop "evaluation of spatial and temporal control of gene drives," april 4–5, 2019, vienna.
613 BioEssays 10:1900151.

614 Girardin, L., V. Calvez, and F. Débarre. 2019. Catch Me If You Can: A Spatial Model for a
615 Brake-Driven Gene Drive Reversal. Bulletin of Mathematical Biology 81:5054–5088.

616 Godwin, J., M. Serr, S. K. Barnhill-Dilling, D. V. Blondel, P. R. Brown, K. Campbell, J. Delborne,
617 A. L. Lloyd, K. P. Oh, T. A. Prowse, et al. 2019. Rodent gene drives for conservation: opportu-
618 nities and data needs. Proceedings of the Royal Society B 286:20191606.

619 Gokhale, C. S., R. G. Reeves, and F. A. Reed. 2014. Dynamics of a combined medea-
620 underdominant population transformation system. BMC Evolutionary Biology 14:98.

621 Huang, Y., A. L. Lloyd, M. Legros, and F. Gould. 2009. Gene-drive in age-structured insect
622 populations. Evolutionary Applications 2:143–159.

623 ———. 2011. Gene-drive into insect populations with age and spatial structure: A theoretical
624 assessment. Evolutionary applications 4:415–428.

625 James, S., F. H. Collins, P. A. Welkhoff, C. Emerson, H. C. J. Godfray, M. Gottlieb, B. Greenwood,
626 S. W. Lindsay, C. M. Mbogo, F. O. Okumu, et al. 2018. Pathway to deployment of gene drive
627 mosquitoes as a potential biocontrol tool for elimination of malaria in sub-saharan africa: rec-
628 ommendations of a scientific working group. The American journal of tropical medicine and
629 hygiene 98:1–49.

Johnson, J. A., R. Altwegg, D. M. Evans, J. G. Ewen, I. J. Gordon, N. Pettorelli, and J. K. Young. 2016. Is there a future for genome-editing technologies in conservation? Animal Conservation 19:97–101.

Karlin, S. 1978. Comparisons of positive assortative mating and sexual selection models. Theoretical population biology 14:281–312.

Lance, D. R., D. O. McInnis, P. Rendon, and C. G. Jackson. 1998. Courtship among sterile and wild *Ceratitis capitata* (diptera: Tephritidae) in field cages in hawaii and guatemala. Annals of the Entomological Society of America 93.

Larner, W., T. Price, L. Holman, and N. Wedell. 2019. An x-linked meiotic drive allele has strong, recessive fitness costs in female drosophila pseudoobscura. Proceedings of the Royal Society B 286:20192038.

Leitschuh, C. M., D. Kanavy, G. A. Backus, R. X. Valdez, M. Serr, E. A. Pitts, D. Threadgill, and J. Godwin. 2018. Developing gene drive technologies to eradicate invasive rodents from islands. Journal of Responsible innovation 5:S121–S138.

Lenington, S. 1983. Social preferences for partners carrying 'good genes' in wild house mice. Animal Behaviour 31:325–333.

Lenington, Sarah. 1991. The t complex: a story of genes, behavior, and populations. Advances in the Study of Behavior 20:51–86.

Levin, S. A. 2003. Complex adaptive systems: Exploring the known, the unknown and the unknowable. Bulletin of the AMS - American Mathematical Society 40:3–19.

Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, L. Holman, H. Johannesson, U. Knief, H. Kokko, A. M. Larracuente, et al. 2016. The ecology and evolutionary dynamics of meiotic drive. Trends in ecology & evolution 31:315–326.

653 Lindholm, A. K., K. Musolf, A. Weidt, and B. König. 2013. Mate choice for genetic compatibility
654 in the house mouse. Ecology and evolution 3:1231–1247.

655 Lindvall, M., and J. Molin. 2020. Designing for the Long Tail of Machine Learning. arXiv .

656 Long, K. C., L. Alphey, G. J. Annas, C. S. Bloss, K. J. Campbell, J. Champer, C.-H. Chen, A. Choud-
657 hary, G. M. Church, J. P. Collins, et al. 2020. Core commitments for field trials of gene drive
658 organisms. Science 370:1417–1419.

659 Manser, A., S. J. Cornell, A. Sutter, D. V. Blondel, M. Serr, J. Godwin, and T. A. Price. 2019.
660 Controlling invasive rodents via synthetic gene drive and the role of polyandry. Proceedings
661 of the Royal Society B 286:20190852.

662 Manser, A., B. König, and A. K. Lindholm. 2020. Polyandry blocks gene drive in a wild house
663 mouse population. Nature communications 11:1–8.

664 Manser, A., A. K. Lindholm, L. W. Simmons, and R. C. Firman. 2017. Sperm competition sup-
665 presses gene drive among experimentally evolving populations of house mice. Molecular
666 ecology 26:5784–5792.

667 Marshall, J. M., and B. A. Hay. 2011. Inverse Medea as a Novel Gene Drive System for Local
668 Population Replacement A Theoretical Analysis. Journal of Heredity 103:336–341.

669 Marshall, J. M., G. W. Pittman, A. B. Buchman, and B. A. Hay. 2011. Semele: a killer-male, rescue-
670 female system for suppression and replacement of insect disease vector populations. Genetics
671 187:535–551.

672 McInnis, D. O., D. R. Lance, and C. G. Jackson. 1996. Behavioral resistance to the sterile insect
673 technique by mediterranean fruit fly (*diptera: tephritidae*) in hawaii. Annals of the Entomological
674 Society of America 89:739–744.

675 Moro, D., M. Byrne, M. Kennedy, S. Campbell, and M. Tizard. 2018. Identifying knowledge gaps

for gene drive research to control invasive animal species: the next crispr step. Global Ecology and Conservation 13:e00363.

National Academies of Sciences Engineering and Medicine. 2016. Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values. The National Academies Press, Washington, DC.

Noble, C., B. Adlam, G. M. Church, K. M. Esvelt, and M. A. Nowak. 2018. Current crispr gene drive systems are likely to be highly invasive in wild populations. Elife 7:e33423.

Noble, C., J. Olejarz, K. M. Esvelt, G. M. Church, and M. A. Nowak. 2017. Evolutionary dynamics of CRISPR gene drives. Science Advances 3.

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

North, A. R., A. Burt, and H. C. J. Godfray. 2019. Modelling the potential of genetic control of malaria mosquitoes at national scale. BMC biology 17:1–12.

North, Ace R and Burt, Austin and Godfray, H Charles J. 2020. Modelling the suppression of a malaria vector using a crispr-cas9 gene drive to reduce female fertility. BMC biology 18:1–14.

O'Donald, P. 1980. Genetic models of sexual and natural selection in monogamous organisms. Heredity 44:391–415.

Oye, K. A., K. Esvelt, E. Appleton, F. Catteruccia, G. Church, T. Kuiken, S. B.-Y. Lightfoot, J. McNamara, A. Smidler, and J. P. Collins. 2014. Regulating gene drives. Science 345:626–628.

Pinter-Wollmann, N., E. A. Hobson, J. E. Smith, A. J. Edelman, D. Shizuka, S. de Silva, J. S. Waters, S. D. Prager, T. Sasaki, G. Wittemyer, J. Fewell, and D. B. McDonald. 2014. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. Behavioral Ecology 25:242–255.

699 Price, T. A., and N. Wedell. 2008. Selfish genetic elements and sexual selection: their impact on
700 male fertility. Genetica 132:295.

701 Prowse, T. A., F. Adikusuma, P. Cassey, P. Thomas, and J. V. Ross. 2019. A y-chromosome
702 shredding gene drive for controlling pest vertebrate populations. Elife 8:e41873.

703 Prowse, T. A., P. Cassey, J. V. Ross, C. Pfitzner, T. A. Wittmann, and P. Thomas. 2017. Dodg-
704 ing silver bullets: good crispr gene-drive design is critical for eradicating exotic vertebrates.
705 Proceedings of the Royal Society B: Biological Sciences 284:20170799.

706 Qureshi, A., A. Aldersley, B. Hollis, A. Ponlawat, and L. J. Cator. 2019. Male competition and
707 the evolution of mating and life-history traits in experimental populations of aedes aegypti.
708 Proceedings of the Royal Society B 286:20190591.

709 Robinson, A. S., and J. Hendrichs. 2005. Prospects for the Future Development and Application
710 of the Sterile Insect Technique, pages 727–760. Springer Netherlands.

711 Rode, N. O., A. Estoup, D. Bourguet, V. Courtier-Orgogozo, and F. Débarre. 2019. Population
712 management using gene drive: molecular design, models of spread dynamics and assessment
713 of ecological risks. Conservation Genetics pages 1–20.

714 Sandler, L., and E. Novitski. 1957. Meiotic drive as an evolutionary force. The American Natu-
715 ralist 91:105–110.

716 Simon, S., M. Otto, and M. Engelhard. 2018. Synthetic gene drive: Between continuity and
717 novelty: Crucial differences between gene drive and genetically modified organisms require
718 an adapted risk assessment for their use. EMBO reports 19:e45760.

719 Simoni, A., A. M. Hammond, A. K. Beaghton, R. Galizi, C. Taxiarchi, K. Kyrou, D. Meacci,
720 M. Gribble, G. Morselli, A. Burt, et al. 2020. A male-biased sex-distorter gene drive for the
721 human malaria vector anopheles gambiae. Nature biotechnology 38:1054–1060.

Tanaka, H., H. A. Stone, and D. R. Nelson. 2017. Spatial gene drives and pushed genetic waves. Proceedings of the National Academy of Sciences 114:8452–8457.

Unckless, R. L., A. G. Clark, and P. W. Messer. 2017. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. Genetics 205:827–841.

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

Verma, P., R. G. Reeves, and C. S. Gokhale. 2021. A common gene drive language eases regulatory process and eco-evolutionary extensions. BMC Ecology and Evolution 21:1–21.

Ward, C. M., J. T. Su, Y. Huang, A. L. Lloyd, F. Gould, and B. A. Hay. 2011. Medea selfish genetic elements as tools for altering traits of wild populations: a theoretical analysis. Evolution 65:1149–1162.

Wedell, N., and T. A. R. Price. 2015. Selfish Genetic Elements and Sexual Selection, pages 165–190. Springer Netherlands.

Windbichler, N., M. Menichelli, P. A. Papathanos, S. B. Thyme, H. Li, U. Y. Ulge, B. T. Hovde, D. Baker, R. J. Monnat, A. Burt, and A. Crisanti. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature 473:212–215.

738

Figure legends

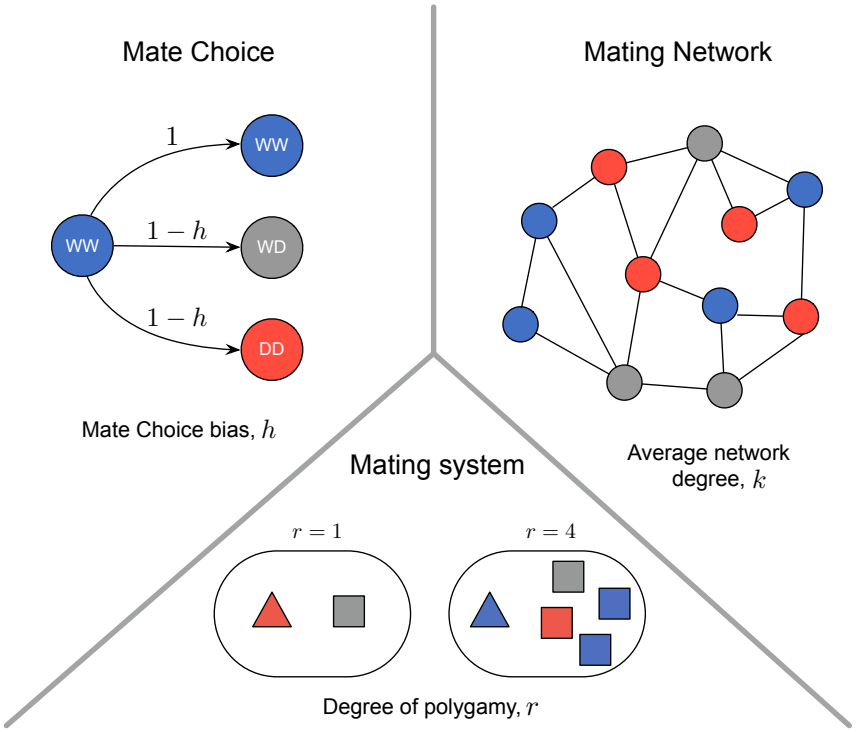

Figure 1: **Pictorial representation of the three mating complexities: mate-choice, mating network, and mating system that can affect gene drive's population dynamics.** Blue, gray and red colours represent individuals with genotype WW, WD and DD, respectively. When there is no distinction between the two sexes, individuals are represented by circles, while triangles and squares denote individuals belonging to different sexes. Under mate-choice bias, the wildtype genotype (WW) are less likely to mate with drive carrying genotype (DD and WD). Mate-choice bias is denoted by h in our model, where $(1-h)$ is the mating rate between the wildtypes (WW) and the transgenics (WD or DD). In structured mating, individuals mate and reproduce with other receptive individuals in their vicinity, and their likely interactions are modelled on a mating network of average degree k. The consequence of mating with one (monogamy $r = 1$) or multiple mating partners (polygamy, $r > 1$) on the gene drive dynamics is studied under the mating systems.

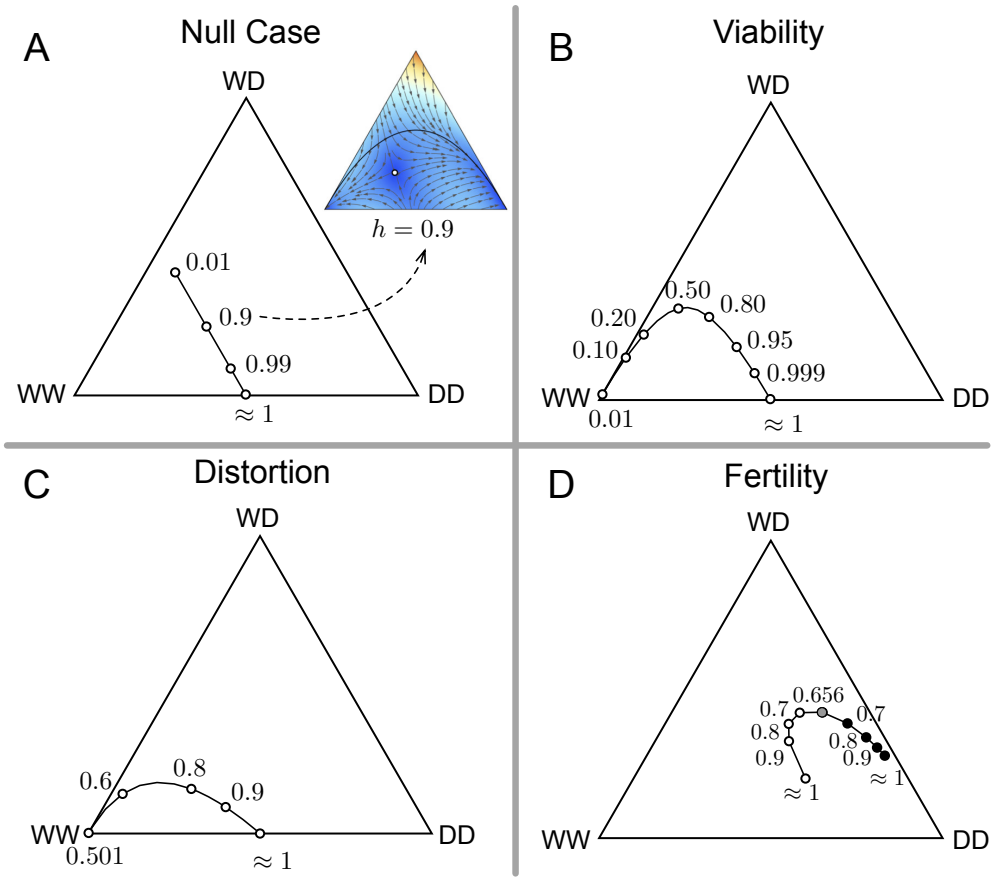

Figure 2: **Effect of mate-choice bias (h) on the internal fixed point of the population dynamics without (null case) and with gene drives system based on viability selection, distortion and fertility selection.** Fixed points appear in the interior of the de Finetti diagrams when the fitnesses of all the genotypes are the same. Open circles denote unstable fixed points of the dynamics, while closed black circles denote stable fixed points. The Gray circle denotes the bifurcation point where both unstable and stable points emerge. The position of these fixed points changes with mate-choice bias (h) and hence the overall population dynamics, including the release threshold. Solid black lines show in the trajectory of these fixed points for varying mate-choice parameter h. **(A)** Null case (without drive) considers the effect of mate-choice alone on the population dynamics. The population dynamics for a specific case of $h = 0.9$ is shown in the inset of figure 2A. The position of the fixed point is pointed out through a dashed line. **(B)** Medea drive efficiency is set to 100%, $d = 1.0$ **(C)** Distortion based drive is assumed to be fully efficient (probability $p = 1.0$) **(D)** Fertility fitness cost, $c = 0.2$. When other parameters are not changed their values are: $d = 0, p = 0.5, f_{WW} = 1, f_{WD} = 1, f_{DD} = 1$.

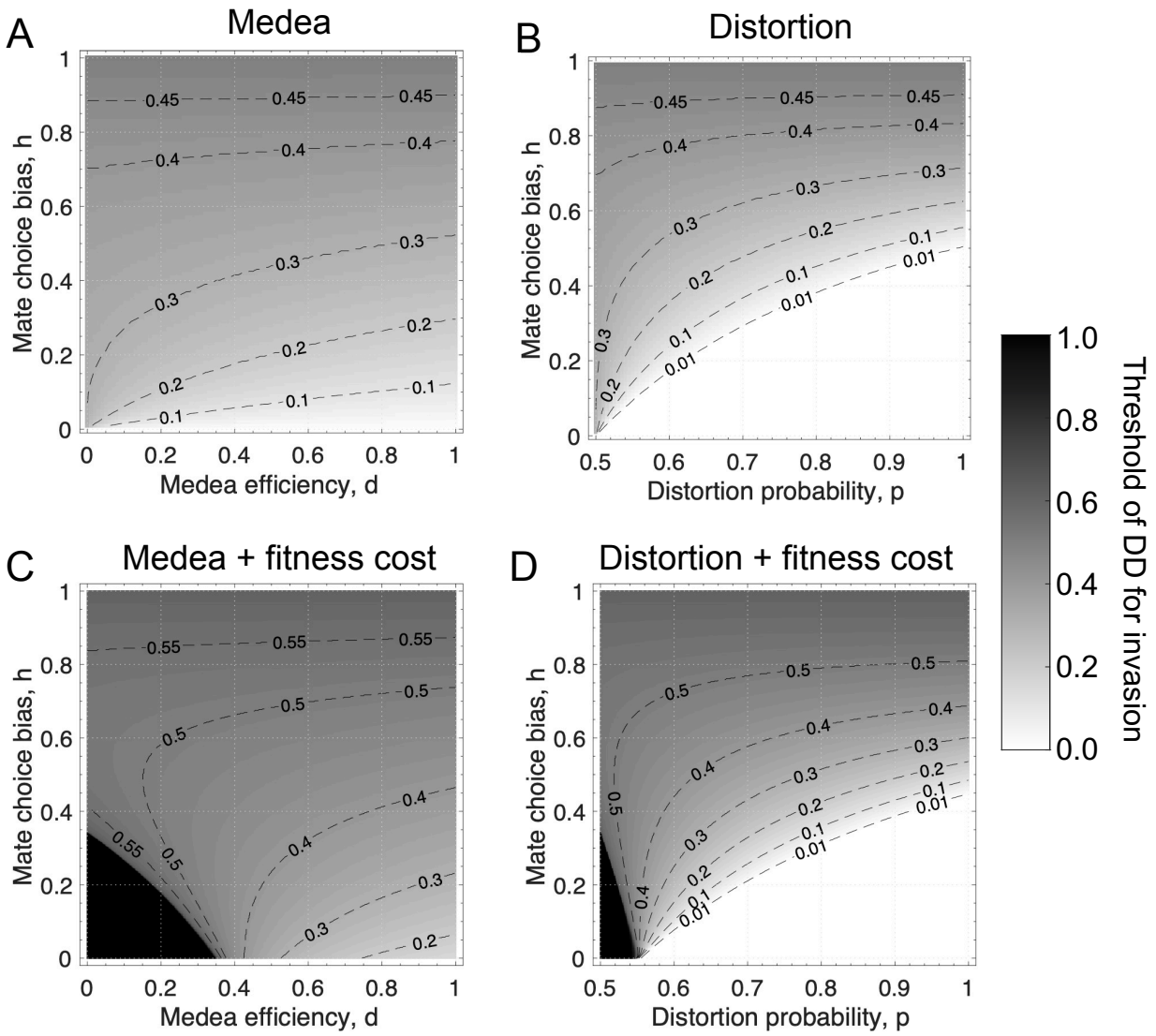

Figure 3: **Heatmap shows the threshold frequency of drive homozygotes (DD) required to invade a population of wildtype homozygotes (WW) with respect to variation in mate-choice bias (h) for the following gene drive systems: Medea and distortion based drive.** Black dashed lines correspond to the contour lines showing the threshold frequency of drive homozygotes (DD). (**A**) Medea gene drive with no fitness cost i.e. $c = 0$. (**B**) Distortion based gene drive with no fitness cost to drive i.e. $c = 0$. (**C**) Medea gene drive where the fitness cost due to drive alelle is $c = 0.1$ hence $f_{WD} = 0.9$ and $f_{DD} = 0.81$. (**D**) Distortion based gene drive where the fitness cost due to drive alelle is $c = 0.1$ hence $f_{WD} = 0.9$ and $f_{DD} = 0.81$.

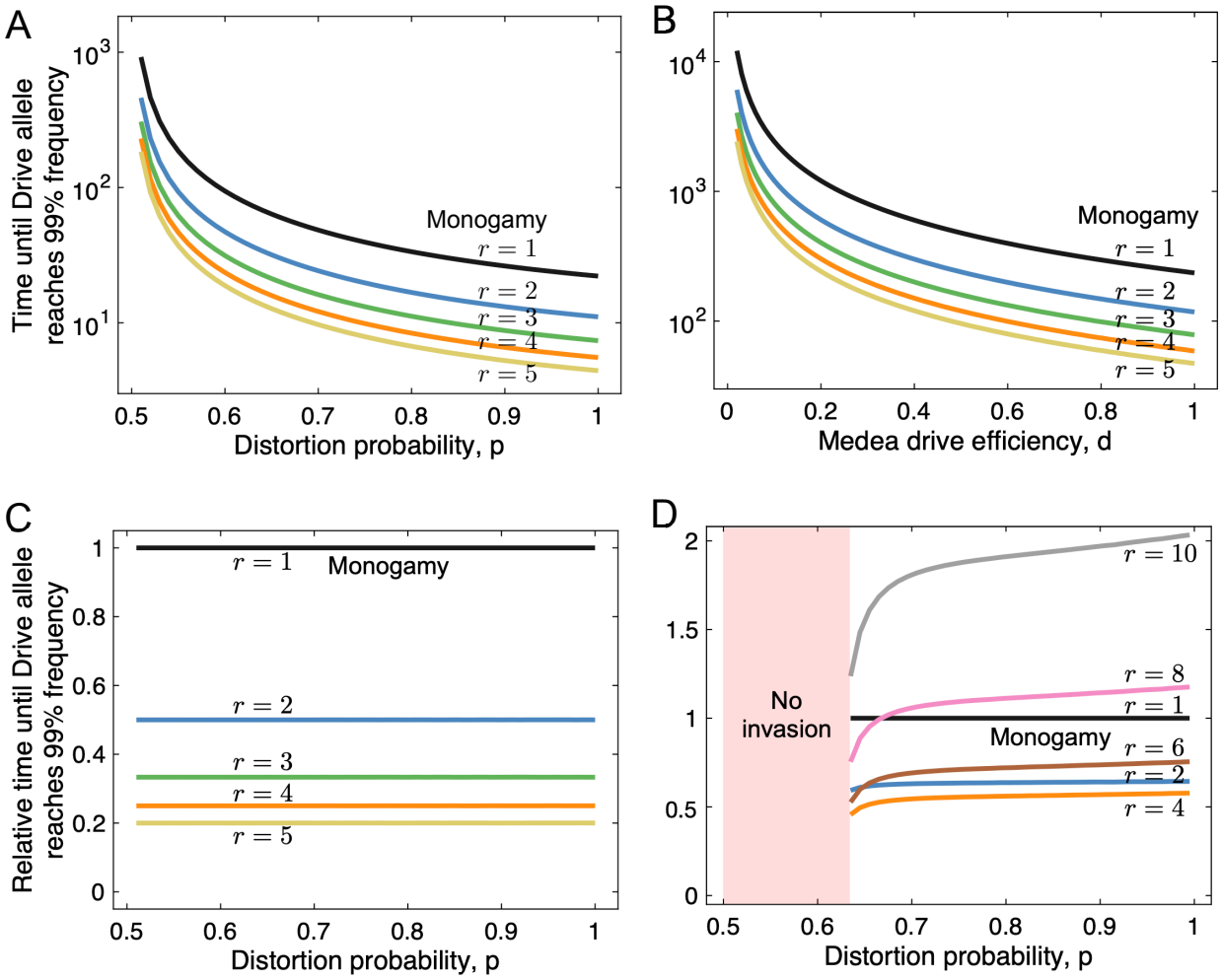

Figure 4: **Effect of mating system and drive efficiency on the time for the drive allele to reach 99% frequency.** We start from a population consisting of the 99% wildtypes (WW) and 1% the drive heterozygotes (DD) with 100% drive efficiency and varying fitness cost. The population is evolved until frequency of drive allele reaches 99%. **(A)** Absolute time is plotted for distortion based gene drive with no fitness cost, $c = 0$ and $p = 1$. **(B)** Absolute time is plotted for Medea gene drive with no fitness cost, $c = 0$ and $d = 1$. **(C)** Relative time with respect to monogamy ($r = 1$) case is plotted for distortion based gene drive without fitness cost, $c = 0$ and $p = 1$. **(D)** Relative time with respect to monogamy ($r = 1$) case is plotted for distortion based gene drive with fitness cost, $c = 0.2$ and $p = 1$. The red shaded area is the region where the drive hetrogygotes are not able to invade the wildtype population.

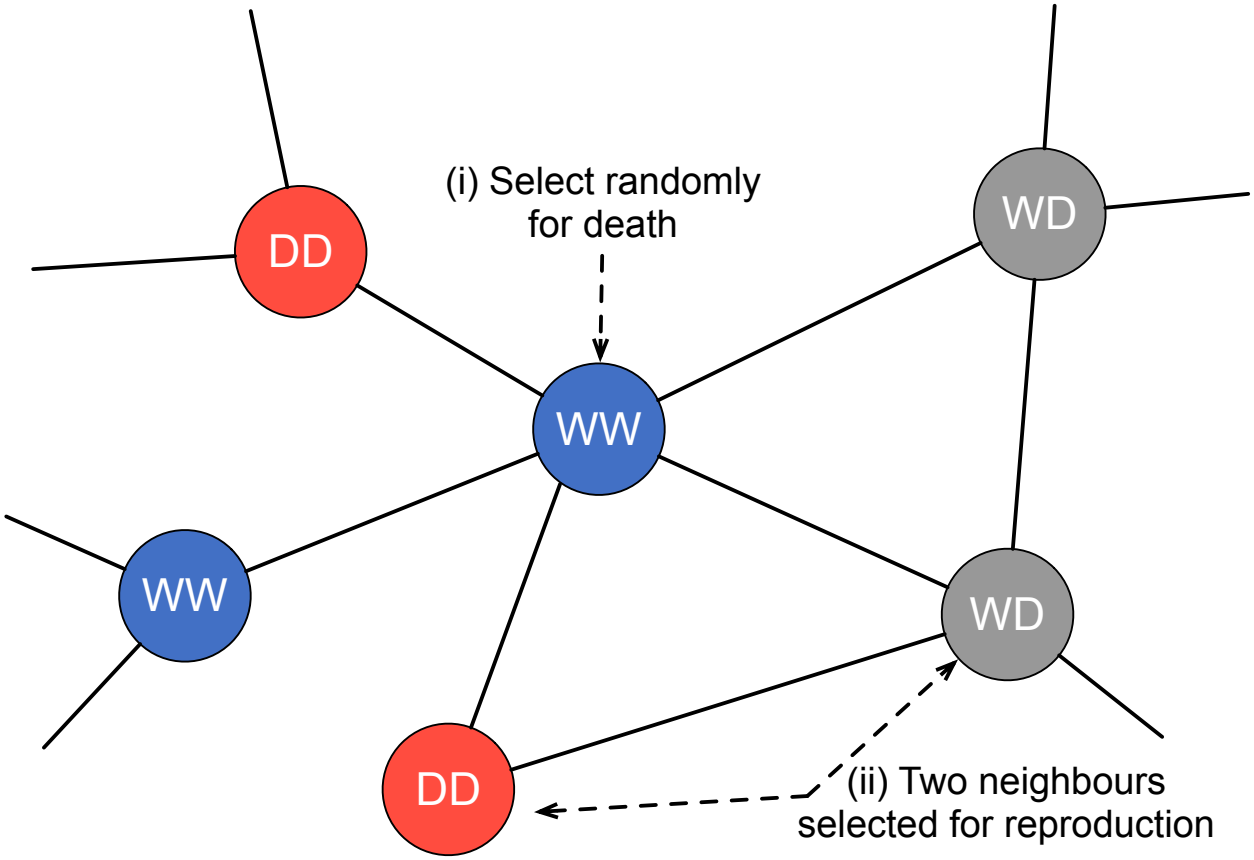

Figure 5: **Spatial model explaining the population update mechanism** The blue, gray and red colours represent individuals of WW, WD and DD genotype, respectively. Population update happens in 2 steps: firstly, a random individual is selected for death. This step creates space at that particular network position. Secondly, two random neighbours of the dead individuals are chosen as parents to produce offspring. The genotype of the offspring is determined from the parents, and it replaces the dead individual.

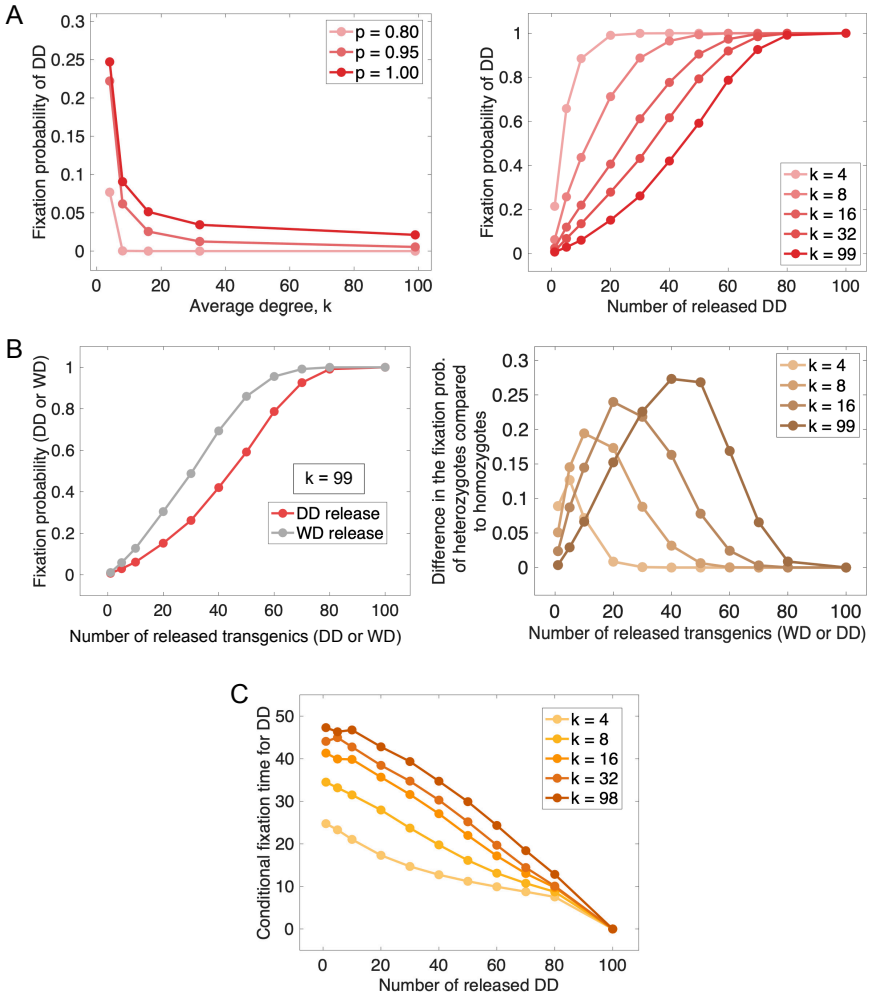

Figure 6: **Fixation probability and conditional fixation time of DD with variation in average degree k, distorsion probability p and initial number of released transgenic individuals WD or DD. (A)** Plots show the fixation probability of drive homozygotes against average degree k (left panel) and the number of released DD (right panel) for different values of p and k respectively. Left: one DD individual is initially released in the population consisting only of WW. **(B)** Left: Fixation probability is plotted against the number of released DD and WD for a complete graph ($k = 99$). Right: the difference between the fixation probability of WD and DD release is plotted against the number of released transgenic for varying average degree k. **(C)** Shows the average number of generations when the drive individuals get fixed in the population against an initial number of released DD with varying average degree k. A generation consists of n death-birth step. Hence in a generation, the whole population is updated on an average. All simulations were performed for a population size of $n = 100$ and 10,000 trials to estimate fixation probability and conditional fixation time. If not mentioned distortion probability and fitness cost are fixed to $p = 0.95$ and $c = 0.5$, respectively.

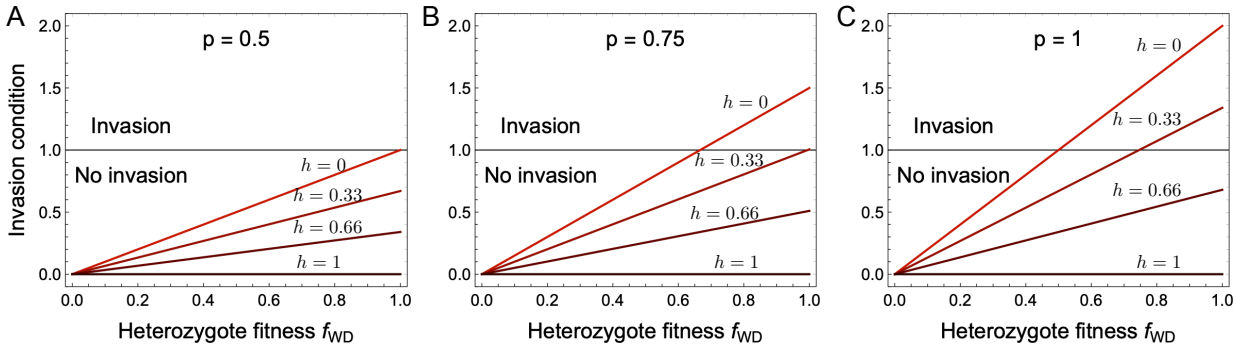

Figure A1: **Invasion condition with varying mate-choice bias (h) against heterozygotes fitness** f_{WD}. (**A**) Medea drive or no distortion, $p = 0.5$. Wildtype population cannot be invaded for any value of mate choice bias, h. (**B**) Distortion based gene drive with $p = 0.75$. Wildtype population can be invaded if there is no-mate choice bias $h = 0$ and $f_{WD} > 2/3$. (**C**) Distortion based gene drive with $p = 1$. Wildtype population can be invaded if mate choice bias not very high i.e. for $h = 0$ and $h = 0.33$.

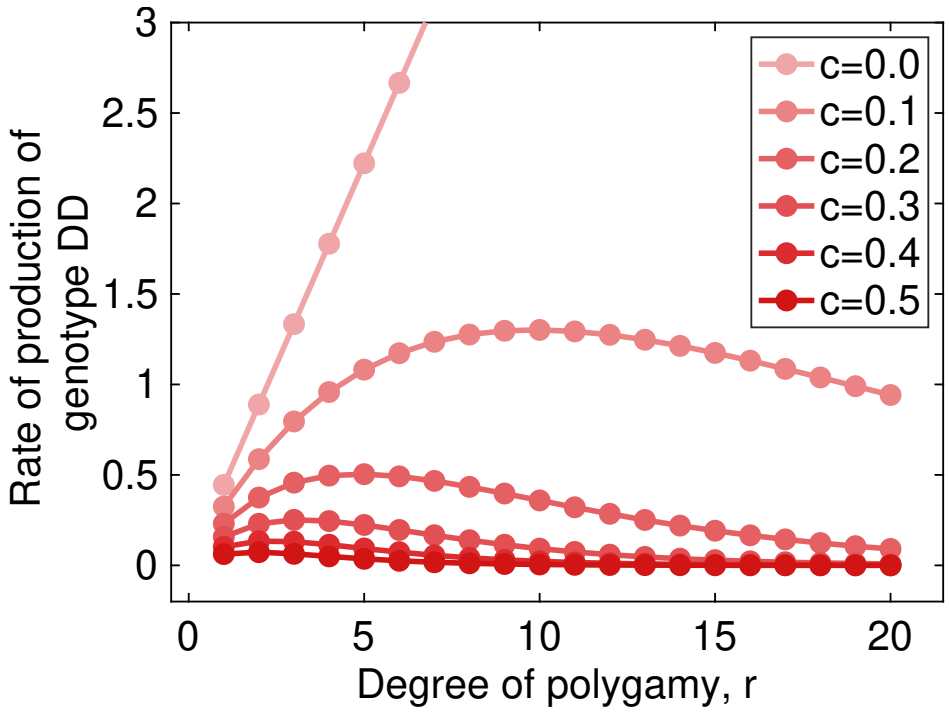

Figure B1: **Effect on the rate of production of DD genotype with increases in degree of polygamy (r) for different fitness cost (c).** We start from a population with an equal abundance of all three genotypes with 100% drive efficiency of distortion-based gene drive for different fitness costs. In essence, we plotted equation (9) for varying r and c keeping $x_{WW} = 1/3$, $x_{WD} = 1/3$, $x_{DD} = 1/3$ and $p = 1$. Increasing the fitness cost of the drive allele decreases the overall production of the DD genotype. For a moderate level of fitness cost, production of genotype DD first increases up to a point for species with a higher level of polygamy but then started to decrease.