Journal of Theoretical Biology 258 (2009) 250–265

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The effect of gene drive on containment of transgenic mosquitoes

John M. Marshall *

Department of Biomathematics, UCLA School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095-1766, USA

ARTICLE INFO

Article history:
Received 30 July 2008
Received in revised form
27 January 2009
Accepted 30 January 2009
Available online 7 February 2009

Keywords:
Branching process
Extinction probability
Field cage
Malaria
Dengue fever

ABSTRACT

Mosquito-borne diseases such as malaria and dengue fever continue to be a major health problem through much of the world. Several new potential approaches to disease control utilize gene drive to spread anti-pathogen genes into the mosquito population. Prior to a release, these projects will require trials in outdoor cages from which transgenic mosquitoes may escape, albeit in small numbers. Most genes introduced in small numbers are very likely to be lost from the environment; however, gene drive mechanisms enhance the invasiveness of introduced genes. Consequently, introduced transgenes may be more likely to persist than ordinary genes following an accidental release. Here, we develop stochastic models to analyze the loss probabilities for several gene drive mechanisms, including homing endonuclease genes, transposable elements, *Medea* elements, the intracellular bacterium *Wolbachia*, engineered underdominance genes, and meiotic drive. We find that *Medea* and *Wolbachia* present the best compromise between invasiveness and containment for the six gene drive systems currently being considered for the control of mosquito-borne disease.

Published by Elsevier Ltd.

1. Introduction

Mosquito-borne diseases such as malaria and dengue fever continue to pose a major health problem through much of the world. In the absence of any single effective disease control strategy, much interest has been directed at the use of gene drive mechanisms to spread anti-pathogen genes through mosquito populations (Craig, 1963; Curtis, 1968; Alphey et al., 2002). Several gene drive systems exist in nature, and it is hoped that refractory genes will be associated with these systems and driven into mosquito populations within a timeframe acceptable to public health goals (James, 2005). Some of the most promising gene drive systems currently being investigated include homing endonuclease genes (HEGs), transposable elements (TEs), *Medea* elements, the intracellular bacterium *Wolbachia*, engineered underdominance genes, and meiotic drive (Sinkins and Gould, 2006).

Any transgenic mosquito project is expected to involve several stages of testing—first in the laboratory, then in indoor cages, and then in outdoor cages exposed to the ambient environment in a region where transgenic mosquitoes might eventually be released (Alphey et al., 2002; Scott et al., 2002). Laboratory studies will investigate the efficacy of the transgene at preventing disease as well as testing for unintentional adverse effects. The Core Working Group on Guidance for Contained Field Trials (Benedict et al., 2008) has identified several potential adverse effects of transgenic mosquitoes that must be assessed prior to a release. These include an enhanced vectorial capacity for nontarget pathogens, increased mosquito longevity or reproductive capacity, behavioral changes that lead to a higher biting rate, and a decreased susceptibility to other control measures such as insecticides. The Working Group also expressed the need to investigate the rate of horizontal DNA transfer between mosquitoes and nontarget organisms, since other species may also acquire an increased capacity to transmit disease or disrupt an essential ecological function.

Much can be studied in the laboratory; however, there are some potential adverse effects of transgenic mosquitoes that can only be assessed in outdoor cages (Benedict et al., 2008). A realistic assessment of mosquito longevity and reproductive capacity must be carried out under more natural conditions of climate and light variation. Ambient cages are also necessary to assess the population growth rate and the carrying capacity of the environment. Furthermore, a female bias in the wild sex ratio is problematic since only female mosquitoes transmit disease. Realistic assessments of such a bias can only be studied in an ambient cage.

By segregating transgenic organisms from the field, ambient cages provide a useful intermediate research stage between the laboratory and the environment; however, complete physical containment can never be guaranteed. The Working Group (Benedict et al., 2008) has outlined several possible breaches of containment—some of which can be avoided, but some of which are very difficult to protect against. These include unpredictable environmental damage due to earthquakes or lightning, leakage of water containing eggs or larvae, breaches of containment due to

* Tel.: +1 310 825 1602; fax: +1 310 825 8685.
E-mail address: johnmm@ucla.edu

0022-5193/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.jtbi.2009.01.031

sabotage or burglary, and just simple human error. Therefore, there is a possibility that transgenic mosquitoes will be accidentally released into an environment that is conducive to their survival before the effectiveness and safety of the gene drive strategy has been ascertained.

Most genes introduced in small numbers are very likely to be lost from the environment, even in the presence of a selective advantage (Fisher, 1922; Haldane, 1927; Wright, 1931). However, gene drive mechanisms enhance the invasiveness of introduced genes, and therefore introduced transgenes may be less likely to be lost than ordinary genes following an accidental release. Given that the organism currently being considered for genetic alteration is a vector of human disease, it is particularly important that the invasiveness of selfish DNA be accounted for in the risk management of ambient cage trials.

Here, we analyze the probability that transgenic DNA consisting of an anti-pathogen gene and drive system is lost from a mosquito population following an accidental release. Several gene drive mechanisms are currently being considered to spread anti-pathogen genes into mosquito populations, each having its own unique dynamics. We therefore analyze the loss probability associated with each system separately.

For an initial comparison of gene drive strategies, we calculate the asymptotic extinction probability for each system. A major feature of the demography of *Anopheles gambiae*, the main vector of malaria in tropical Africa, is the existence of population size changes within and between years (Taylor et al., 2001; Manoukis, 2006). Given the influence of population size changes on gene loss, we also calculate the extinction probabilities under conditions of population growth and decline. Although there are several other factors that will influence the loss or persistence of transgenic DNA following an accidental release, it is hoped that these calculations will inform the risk management of planned ambient cage trials involving selfish DNA.

2. Homing endonuclease genes

HEGs are a class of highly specific DNA endonucleases found in some viruses, bacteria and eukaryotes (Windbichler et al., 2007). HEGs are able to spread through a population despite a fitness cost due to their overrepresentation in the gametes of a heterozygote. They achieve this by expressing an endonuclease which creates a double-stranded break at a highly specific site that lacks the HEG. Homologous DNA repair then copies the HEG to the cut chromosome (Rong and Golic, 2003).

To calculate the conditions under which a HEG allele can spread following an accidental release, we consider a two-type continuous-time branching process in which type-1 particles are heterozygous and type-2 particles are homozygous for the HEG. The branching process framework is favored due to its suitability for calculating extinction probabilities. Normally, in a branching process, each particle produces a burst of offspring at the moment of its death. However, in order to separate the birth and death rates, we consider a budding model in which each particle can die in one of two ways—it can truly die at a rate equal to the mosquito death rate (set to $\mu = 1\,\text{gen}^{-1}$); or it can disappear at a reproduction event to be replaced by itself and its offspring (Lange, 2002; Dorman et al., 2004). Reproduction events continue until the mosquito actually dies, and the mosquito can have several offspring each time.

To avoid confusion, we will henceforth only refer to the mosquito dying as "death". In this sense, reproduction and death are essentially different kinds of branching events. Fitness differences can then be accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have a fecundity of θ, where

$$\theta = 2(1 + r), \tag{1}$$

and r represents the population growth rate.

In the early stages of spread, almost all matings involve at least one wild-type mosquito. All matings between wild-type mosquitoes and mosquitoes homozygous for the HEG will produce heterozygotes. Matings between wild-type mosquitoes and heterozygotes for the HEG will produce heterozygotes with probability $(1+t)/2$ and wild-types with probability $(1-t)/2$. Here, the homing rate, t, represents the fraction by which the HEG allele is overrepresented in the gametes of a heterozygote.

Since homozygotes will be lost from the population after one generation, we are interested in the conditions under which mosquitoes that are heterozygous for the HEG are able to spread. We will calculate these conditions using a "reproductive threshold" defined as "the average number of offspring of the same type that a particle eventually generates, either directly or indirectly via other particle-type intermediates". Counting of offspring, in this case, stops at each point in the descendent tree when the offspring is of the same type as the original particle. The reproductive threshold is greater than one when spread is possible, and less than or equal to one when the particle is certain to be lost from the population. This quantity should not be confused with the basic reproductive number, which represents the average number of offspring produced over the lifetime of a single individual. For a heterozygote, the reproductive threshold is simply given by the average number of heterozygotes that it produces at a branching event, $f_{1,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with the HEG allele. We assume that a female homozygous for the HEG has a reduced fecundity of $\theta(1 - s)$. Male mosquitoes that mate with wild-type females do not suffer from reduced fecundity, and so the mean fecundity of a homozygote is $\theta(1 - s/2)$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0, 1]$, where $h = 1$ represents a dominant fitness cost, $h = 0$ represents a recessive fitness cost, and $h = 1/2$ represents an additive fitness cost. The mean fecundity of a heterozygote is therefore $\theta(1 - hs/2)$.

According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 1. The rate of branching events (death and reproduction) for heterozygotes is equal to

$$\lambda_1 = \mu + \theta(1 - hs/2), \tag{2}$$

while the rate for homozygotes is equal to

$$\lambda_2 = \mu + \theta(1 - s/2). \tag{3}$$

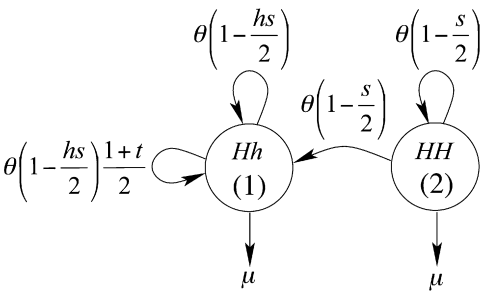

Fig. 1. Schematic for the early spread of a homing endonuclease gene (HEG) through a randomly mating mosquito population. Mosquitoes homozygous for the HEG (*HH*) have a reduced fecundity of $\theta(1 - s/2)$ and mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for the HEG (*Hh*). Heterozygotes have a reduced fecundity of $\theta(1 - hs/2)$ and mate with wild-type mosquitoes to give birth to heterozygotes with probability $(1+t)/2$ and wild-types with probability $(1-t)/2$. All mosquitoes have a death rate of μ.

The reproductive threshold for a mosquito heterozygous for the HEG is then given by

$$R_{1,1} = f_{1,1} = \frac{\theta}{\lambda_1}\left(1-\frac{hs}{2}\right)\left(1+\frac{1+t}{2}\right). \quad (4)$$

When the threshold is greater than one, the HEG has a nonzero probability of spreading through the mosquito population. Simplifying Eq. (4), this means that an accidentally released HEG has some chance of spreading into a wild population when

$$t > \frac{hs-(2-hs)r}{(2-hs)(1+r)}. \quad (5)$$

When the population size is constant, then the condition for HEG spread simplifies to

$$t > \frac{hs}{2-hs}. \quad (6)$$

One implication of this result is that only the fitness cost to the heterozygote determines whether the HEG has a chance of spreading or not. The magnitude of fitness cost that is tolerable is given by Eqs. (5) and (6).

When the reproductive threshold is greater than one, HEG spread is possible; however, the probability of HEG loss may still be very high. To determine the extinction probability of the HEG, we first define the probability generating function for the HEG branching process. This is a function of the vector $\mathbf{z} = (z_1, z_2)$ and is defined as

$$P_i(\mathbf{z}) = \sum_{j=1}^{2} p_{ij} z^j, \quad (7)$$

where p_{ij} is the probability that a type-i particle gives rise to j_1 type-1 particles and j_2 type-2 particles, and j is defined as the vector $j = (j_1 j_2)$. Substituting the p_{ij} terms from the HEG branching process into this equation, we have the probability generating function,

$$P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1}\left(1-\frac{hs}{2}\right)\frac{1-t}{2}z_1 + \frac{\theta}{\lambda_1}\left(1-\frac{hs}{2}\right)\frac{1+t}{2}z_1^2, \quad (8)$$

$$P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2}\left(1-\frac{s}{2}\right)z_1 z_2. \quad (9)$$

The probability that the HEG allele is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations (Harris, 1963),

$$P_i(\mathbf{e}) = e_i \forall i, \quad (10)$$

where e_1 is the loss probability beginning with a heterozygote for the HEG, and e_2 is the loss probability beginning with a homozygote for the HEG. For the HEG branching process, this system of equations has the solution,

$$e_1 = \min\left\{\frac{2}{(1+r)(2-hs)(1+t)^{-1}}\right\}, \quad (11)$$

$$e_2 = \frac{1}{1+(1+r)(2-s)(1-e_1)}. \quad (12)$$

The extinction probability beginning with a heterozygote, e_1, is a decreasing function of homing rate and population growth rate and an increasing function of heterozygote fitness cost. The extinction probability beginning with a homozygote, e_2, is a function of e_1 and is less than one when e_1 is less than one.

Note that we have calculated the asymptotic loss probabilities here, although the model is only valid in the early stages of spread when the reservoir of wild-types is particularly large. This paradox is averted by considering that, when mating events begin to occur between two mosquitoes having the HEG allele, then the HEG allele has already reached a high enough presence in the population that it is very unlikely to go extinct. This assumption similarly applies to each of the drive systems discussed in this paper.

To make sense of these equations, it helps to have some idea of the parameter values and ranges that exist in nature. For HEGs, the most important parameter is the homing rate, t, which describes the fractional increase in representation of HEGs in the gametes of a heterozygote. An estimate of homing rate has been hinted at for the transfer of a *Tet*-resistance transcription unit between two plasmids (Windbichler et al., 2007). The act of homing requires both DNA cleavage and repair using the HEG-carrying site as a template. In this case, the complete HEG allele was transferred to ~10% of cleaved sites, suggesting a range of homing rates between 0 and 10% since not all sites will be cleaved. This estimate may not be particularly insightful since it is likely that HEGs will behave differently in the mosquito germ line (Austin Burt, personal communication).

Rong and Golic (2003) have studied the dynamics of HEG alleles in *Drosophila melanogaster*; however, their studies have been directed more at molecular dynamics and less at the rates at which these processes occur. In the absence of good parameter estimates, Deredec et al. (2007) have investigated the full range of homing rates, $t \in [0, 1]$.

The fitness cost associated with the HEG allele depends on the disease control strategy being employed. A HEG designed to drive a refractory gene into the mosquito population will have a relatively small fitness cost equal to the fitness cost of the refractory gene. This is difficult to estimate due to the lack of reliable comparative fitness measurements in vector populations; however, fitness costs have been documented in several insect species due to both mounting an immune response (Moret and Schmid-Hempel, 2000; Ahmed et al., 2002) and maintaining the physiological machinery necessary to do so (Kraaijeveld and Godfray, 1997; Koella and Boëte, 2002).

A recent encouraging result is that transgenic mosquitoes have been engineered that exhibit no measurable fitness cost when fed on *Plasmodium*-free blood (Moreira et al., 2004) and exhibit a fitness benefit when fed on *Plasmodium*-infected blood (Marelli et al., 2007). Accounting for the proportion of mosquitoes infected with malaria parasites (Beier et al., 1999), these results correspond to a mean homozygote fitness cost in the range $s \in [-0.05, 0.04]$ (Marshall, 2008a). There is little data on the homozygosity of a fitness cost associated with a refractory allele and so this is best assumed as being additive ($h = 0.5$).

A HEG designed to induce a genetic load or bias the sex ratio in order to reduce the size of the vector population will have a much larger fitness cost. For this strategy, it is hoped that females homozygous for the HEG will be sterile and males will suffer minimal fitness cost (Austin Burt, personal communication). This suggests an average homozygous fitness cost of $s \leqslant 0.5$. The best estimate for the homozygosity of this fitness cost comes from classic data for *D. melanogaster* (Austin Burt, personal communication) in which recessive lethal alleles have a homozygosity of $h = 0.02$.

A third strategy of disease control is to use a HEG that disrupts a gene regulating the ability of mosquitoes to function as efficient vectors for the malaria parasite. This strategy is likely to confer a very small fitness cost on the mosquito, and may in fact confer a fitness benefit similar to that observed by Marelli et al. (2007) for the refractory gene approach.

The population dynamics of the wild-type mosquito population are equally relevant to an assessment of each of the gene drive strategies. The *An. gambiae* population of Banambani, Mali, serves as a well-studied example (Taylor et al., 2001; Lanzaro et al., 1998; Tripet et al., 2005; Touré et al., 1994), with recent

collections suggesting that peak population densities during the wet season are at least 10 times those during the dry season (Taylor et al., 2001). Assuming a mosquito generation time of ~16 days (Mahamadou Touré, personal communication) and a population change over a period of six months, this yields a population growth rate of $r = 0.2$ per generation, and a corresponding rate of decline of $r = -0.2$ per generation.

HEGs are one of the most invasive gene drive systems available; and hence are one of the most likely to spread following an accidental release. For both the refractory gene and genetic load HEG strategies, a single homozygous mosquito having a HEG with a relatively modest homing rate of $t = 0.1$ has a loss probability of less than 90% over the entire range of likely fitness costs. This is also likely the case for the gene disruption HEG strategy, which is expected to have similar fitness consequences to the refractory gene strategy. This leads to a persistence probability of more than 10% for all strategies and realistic parameterizations, which is very high for a single escapee (see supplemental Fig. 1A–C). These loss probabilities are noticeably reduced during periods of population growth.

The probability that the HEG becomes established in the population is most influenced by the number of escapees during an accidental release (Fig. 2). For both the refractory gene and genetic load HEG strategies, the loss probability steadily declines with the power of the number of escapees. Homozygotes are slightly more likely to persist in the population essentially because they bring two HEGs into the wild population; while heterozygotes only bring one. According to model predictions for the refractory gene strategy under default parameters, an escape of five homozygotes reduces the loss probability to 43%, while an escape of 25 homozygotes reduces the loss probability to 1.5% (Fig. 2A).

For the genetic load strategy, HEGs are able to persist despite a large fitness cost due to the recessiveness of the lethal allele. According to model predictions for a fitness cost of $s = 0.5$, an escape of five homozygotes reduces the loss probability to 54%, while an escape of 25 homozygotes reduces the loss probability to 5.4% (Fig. 2B). For both strategies, HEGs are able to spread under the full range of realistic conditions, and are more likely to persist than not for escape sizes greater than five.

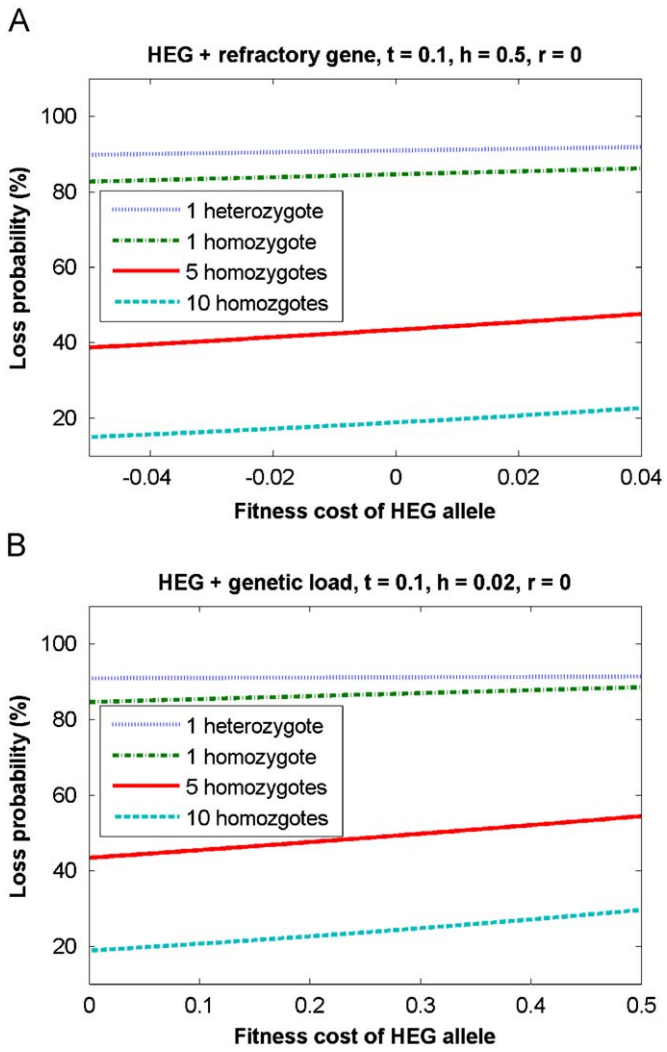

Fig. 2. Asymptotic probabilities of HEG loss as a function of fitness cost for: (A) the refractory gene HEG strategy and (B) the genetic load HEG strategy. The escapee genotype is varied and the homozygote release size is increased to 10.

3. Transposable elements

TEs are particularly interesting genomic components due to their ability to transpose replicatively and hence spread throughout a population despite a fitness cost (Charlesworth et al., 1994). The observation that *P* elements spread through most of the wild-type *D. melanogaster* population within a few decades (Engels, 1989) has inspired the idea of using TEs as drive mechanisms for spreading anti-pathogen genes into mosquito populations (Craig, 1963; Curtis, 1968).

To model the early stages of TE spread following an accidental release, we consider a *T*-type continuous-time branching process analogous to that proposed by Marshall (2008b) in which a type-i particle has i copies of the TE, where $i \in \{1, 2, \ldots, T\}$ and $T \geqslant 1$. Here, T can be approximated as the number of sites that will be occupied in the early stages of TE spread. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1\,\text{gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity θ as described by Eq. (1).

Transposition and deletion are modeled by assuming that a proportion α_i of gametes are derived from cells in which a replicative transposition event has occurred, while a proportion β_i of gametes are derived from cells in which an element deletion event has occurred. The replicative transposition rate for a type-i host, α_i, is equal to the replicative transposition rate per TE in a type-i host, u_i, multiplied by the number of TEs in the host genome, i (i.e. $\alpha_i = iu_i$). Here, u_i is a decreasing function of i to account for suppression of transposition with increasing copy number (Weinreich et al., 1994; Wu and Morris, 1999; Townsend and Hartl, 2000). We model transposition rate as an exponentially decreasing function of copy number

$$u_i = u_1 2^{-c(i-1)}, \tag{13}$$

where c determines the rate at which the replicative transposition rate falls off with additional element copies (Marshall, 2008a). Similarly, the deletion rate for a type-i host, β_i, is equal to the deletion rate per element, v, multiplied by the number of elements in the host genome, i (i.e. $\beta_i = iv$), where v is generally considered a constant.

The fitness cost associated with additional TE copies is modeled by assuming that a female having i TE copies has a reduced fecundity of $\theta(1-s_i)$, where s_i is an increasing function of i. For the early stages of TE spread, we describe fitness cost as a linear function of i (i.e. $s_i = id$), where d represents the fractional decrease in female fecundity associated with each additional TE copy in the genome. Male mosquitoes that mate with wild-type

females do not suffer from reduced fecundity, and so the mean fecundity of a type-i host is $\theta(1-s_i/2)$.

The number of TEs in the haploid gamete is determined by the number of TEs in the diploid cell that are passed on during meiosis. For a diploid cell with i copies of the TE we assume, to a first approximation, that all of these TEs are far enough apart from each other that they segregate independently. Under this assumption, the probability of having j copies in a gamete is proportional to the number of ways of choosing j elements from a total of i. Similarly, if a replicative transposition event has occurred in the diploid cell, then the probability of having j copies in the gamete is proportional to the number of ways of choosing j elements from i+1, or from $i-1$ if a deletion event has occurred.

In the early stages of spread, almost all matings involve at least one wild-type mosquito, and so the number of TEs in the offspring's genome is equal to the number of TEs in the gamete contributed by the parent infected with the TE. According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 3. In this process, a type-i host undergoes branching events (death and reproduction) at a rate:

$$\lambda_i = \mu + \theta(1-s_i/2), \tag{14}$$

and produces on average f_{ij} hosts having j element copies per branching event according to the equation:

$$f_{ij} = \frac{\theta}{\lambda_i}\left(1-\frac{s_i}{2}\right)\left[1_{\{i=j\}} + \frac{1}{2^{i-1}}\binom{i-1}{j}\beta_i + \frac{1}{2^i}\binom{i}{j}(1-\alpha_i-\beta_i) + \frac{1}{2^{i+1}}\binom{i+1}{j}\alpha_i\right], \tag{15}$$

where $i,j \notin \{1,2,\ldots,T\}$.

The reproductive threshold of a TE following an accidental release can by studied by assuming that, in the early stages of TE spread, the vast majority of infected hosts have only one or two TE copies. The reproductive threshold for a mosquito having a single TE copy, $R_{1,1}$, can then be defined recursively as

$$R_{1,1} = f_{1,1} + f_{1,2}R_{2,1}. \tag{16}$$

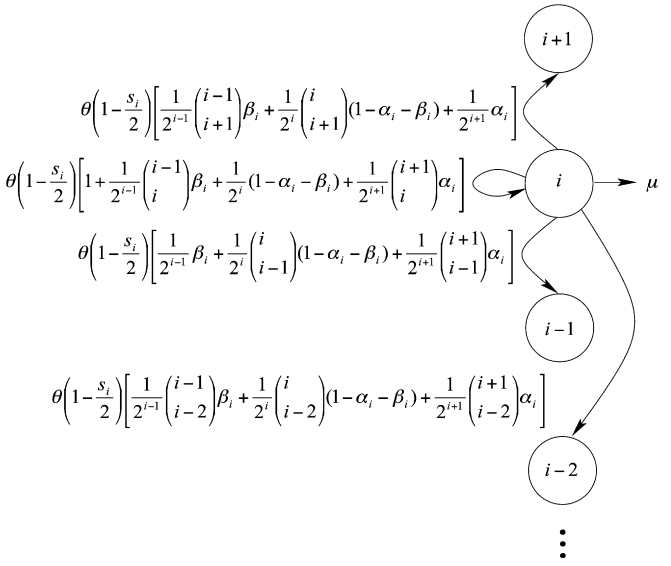

Fig. 3. Schematic for the early spread of a transposable element (TE) through a randomly mating mosquito population. Mosquitoes having i copies of the TE have a reduced fecundity of $\theta(1-s_i/2)$ and mate with wild-type mosquitoes to give birth to mosquitoes having $j \in \{0,\ldots,i+1\}$ element copies as described in Eq. (14). All mosquitoes have a death rate of μ.

Here, $f_{1,1}$ is the average number of type-1 offspring produced by a type-1 host per branching event, $f_{1,2}$ is the average number of type-2 offspring produced by a type-1 host per branching event, and $R_{2,1}$ is the average number of type-1 offspring that a type-2 host eventually generates, either directly or indirectly via type-2 particle intermediates. In the latter case, counting of offspring stops at each point in the descendent tree when the offspring is of type-1.

Following similar reasoning, $R_{2,1}$ can be defined recursively as

$$R_{2,1} = f_{2,1} + f_{2,2}R_{2,1}, \tag{17}$$

where $f_{2,1}$ and $f_{2,2}$ are similarly defined. Eq. (17) can then be rearranged and substituted into Eq. (16) to obtain the reproductive threshold for a type-1 host,

$$R_{1,1} = f_{1,1} + \frac{f_{1,2}f_{2,1}}{1-f_{2,2}}. \tag{18}$$

There is no simple solution to this equation; however, if we consider the simplified scenario in which transpositional regulation and element deletion are negligible ($c=\nu=0$), this leads to the reproductive threshold,

$$R_{1,1} = \frac{(d-2)(r+1)(6-u_1(u_1+4)-r(6+u_1(4+u_1))+d(r+1)(54+u_1(u_1+4)))}{4(d+3+r(d+2))(r-1+(r+1)(u_1-d(9+u_1)))}. \tag{19}$$

When the reproductive threshold is greater than one, TE spread is supercritical, and the TE has a nonzero probability of spreading through the mosquito population. When transpositional regulation and element deletion are not negligible, we must use Eq. (18) to determine the conditions for TE spread to be supercritical.

When TE spread is supercritical the eventual extinction of the TE is uncertain, however, the probability that the TE is lost from the population may still be very high. To determine the extinction probability of the TE, we first define the probability generating function for the TE branching process. This is a function of the vector $\mathbf{z} = (z_1,\ldots,z_T)$ and is defined as

$$P_i(\mathbf{z}) = \sum_j p_{ij}z^j = \sum_j p_{ij}z_1^{j_1}\ldots z_T^{j_T}, \tag{20}$$

where p_{ij} is the probability that a type-i particle gives rise to j_1 type-1 particles, j_2 type-2 particles, and so on, and j is defined as the vector $j = (j_1,\ldots,j_T)$. Substituting the p_{ij} terms from the TE branching process into this equation, we have the probability generating function,

$$P_i(\mathbf{z}) = \frac{\mu}{\lambda_i} + (f_{i,0}+f_{i,T-1})z_i\sum_{j=1}^{T} f_{ij}z_iz_j - \frac{\theta}{\lambda_i}\left(1-\frac{s_i}{2}\right)z_i^2. \tag{21}$$

The probability that the TE is eventually lost from the population is then given by the smallest solution of the system of T simultaneous equations described by Eq. (10). There is no simple analytic solution in this case, and so the system is best solved by numerically iterating:

$$e_i = \frac{\mu}{\lambda_i} + (f_{i,0}+f_{i,T+1})e_i + \sum_{j=1}^{T} f_{ij}e_ie_j - \frac{\theta}{\lambda_i}\left(1-\frac{s_i}{2}\right)e_i^2 \tag{22}$$

for $i \in \{1,\ldots,T\}$, where e_1 is the loss probability beginning with an individual having a single TE copy, e_2 is the loss probability beginning with an individual having two TE copies, and so on.

There is currently little or no data regarding the behavior of the candidate TEs in human disease vectors such as *An. gambiae* and *Aedes aegypti*, so most of these estimates have been taken from measurements in *D. melanogaster*.

We consider a baseline replicative transposition rate of $u_1 = 0.1$ per element per generation ($\mathrm{TE}^{-1}\,\mathrm{gen}^{-1}$). Although this is a fairly high transposition rate, it is realistic (Seleme et al., 1999;

Vasilyeva et al., 1999) and several modeling approaches have recommended it as a minimum requirement for gene drive to occur in a timeframe acceptable to public health goals (Rasgon and Gould, 2005; Le Rouzic and Capy, 2006). Following Marshall (2008a), we consider a transpositional regulation parameter of $c = 2.9$ per element (TE^{-1}). This value was chosen since it produces an equilibrium TE copy number consistent with the *Herves* element in *An. gambiae* (Subramanian et al., 2007). We also consider a deletion rate of $\nu = 4 \times 10^{-6}\,\mathrm{TE}^{-1}\,\mathrm{gen}^{-1}$ as suggested by pooled estimates from several laboratory line experiments (Nuzhdin et al., 1997; Maside et al., 2000).

The fitness cost associated with a TE can originate from both the mutagenic nature of a new genomic insertion and the effects of being associated with a refractory gene. Following Mackay et al. (1992), we estimate the fitness cost of a new genomic insertion by the average fitness cost of a spontaneous mutation. This has been estimated as ~0.02 per element (Mukai et al., 1972; Ohnishi, 1977; Crow and Simmons, 1983) and is a reasonable estimate in the early stages of TE spread when insertional mutagenesis is the dominant fitness cost and selection has not yet eliminated TEs with higher fitness costs (Charlesworth, 1991). The fitness cost of being associated with a refractory gene has been discussed for the case of HEGs, and has been estimated by the range [−0.05,0.04]. Combining these two estimates suggests a range for fitness costs on the order of $d \in [-0.03, 0.06]$ per element.

While TEs are not as invasive as HEGs, they are very able to spread through a mosquito population following an accidental release. A TE having a transposition rate greater than $u_1 = 0.05$ $\mathrm{TE}^{-1}\,\mathrm{gen}^{-1}$ has a small chance of persisting in the population over the entire range of likely fitness costs. For an ambitious yet feasible transposition rate of $u_1 = 0.1\,\mathrm{TE}^{-1}\,\mathrm{gen}^{-1}$, the persistence probability of a single escapee is between 4% and 9% for all realistic parameterizations. The prospects for a TE to colonize a wild population are dramatically improved during conditions of population growth (see supplemental Fig. 2A and B).

The probability that the TE is lost from the population is decreased by the number of TEs that are present in the escapee (see supplemental Fig. 2C); however, it is decreased even more by the number of escapees during an accidental release (Fig. 4). According to model predictions, an escape of five mosquitoes each infected with a single TE decreases the loss probability to 72%. For 10 escapees, the loss probability decreases to 51%; while for 25 escapees, the loss probability decreases to 19%. TEs are therefore less invasive than HEGs; however, they are still very likely to establish themselves in a population following an accidental release. Under default conditions, a TE is more likely to persist than not for escape sizes greater than 10.

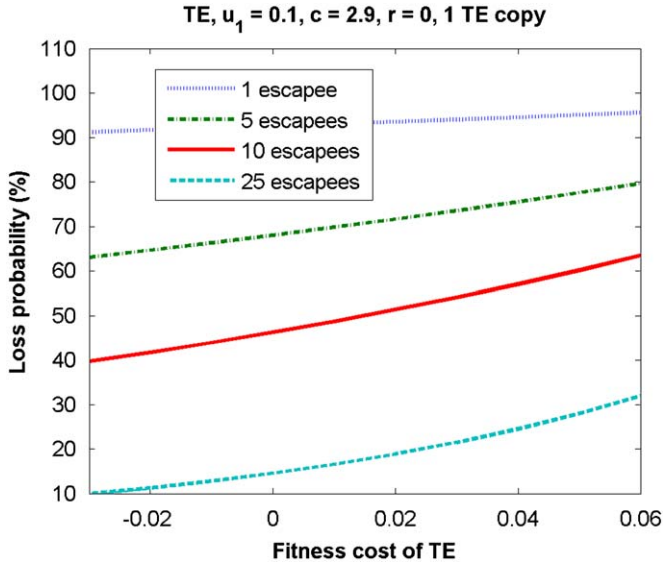

Fig. 4. Asymptotic probabilities of TE loss as a function of fitness cost. Release size is increased to 25.

4. Meiotic drive

Meiotic drive refers to any mechanism by which a heterozygous locus segregates at a greater-than-Mendelian frequency by destroying or disabling the homologous chromosome (Little, 1991). Various mechanisms are known to result in meiotic drive (Hickey and Craig, 1966; Lyttle, 1977). Some of these reduce the quantity of functional sperm; however, they do not necessarily result in reduced fertility (Sinkins and Gould, 2006). Alleles that promote meiotic drive are able to spread through a population despite a fitness cost as a consequence of their increased inheritance.

We consider a meiotic drive strategy which has been recently modeled (Huang et al., 2007) utilizing a Y-linked meiotic drive gene (Y^D) to drive an X-linked drive-insensitive response allele (X^{it}) into the mosquito population. The Y-linked meiotic drive gene is able to spread into the population by virtue of its overrepresentation in the gametes of $X^{sn}Y^D$ males, where X^{sn} is a wild-type drive-sensitive allele at the same locus as the X^{it} allele. As the meiotic drive gene increases its prevalence in the population, it selects for the drive-insensitive X^{it} allele. This selection occurs because, while the X^{sn} allele is underrepresented in the gametes of $X^{sn}Y^D$ males, the X^{it} allele is represented at normal Mendelian frequencies. If an anti-pathogen gene is linked to the X^{it} allele, it will therefore be driven into the population.

We consider a 5-type continuous-time branching process in which type-1 particles are males having the X-linked drive-sensitive allele and Y-linked meiotic drive gene ($X^{sn}Y^D$), type-2 particles are males having the X-linked drive-insensitive allele and Y-linked meiotic drive gene ($X^{it}Y^D$), type-3 particles are males having the X-linked drive-insensitive allele ($X^{it}Y^d$), type-4 particles are females homozygous for the drive-insensitive allele ($X^{it}X^{it}$), and type-5 particles are heterozygous females ($X^{it}X^{sn}$). We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1\,\mathrm{gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity θ as described by Eq. (1).

In the early stages of spread, almost all transgenic males will mate with wild-type females ($X^{sn}X^{sn}$) while almost all transgenic females will mate with wild-type males ($X^{sn}Y^d$). This means that, one generation following an accidental release, there will be no mosquitoes having both transgenic alleles because all offspring will inherit a wild-type allele from their wild-type parent. Matings involving type-1 males ($X^{sn}Y^D$) will produce more type-1 males with probability $(1+t)/2$ and will produce wild-type females with probability $(1-t)/2$. Here, the meiotic drive parameter, t, represents the fraction by which the Y^D allele is overrepresented in the gametes of $X^{sn}Y^D$ males.

Since the two transgenic alleles, Y^D and X^{it}, become separated in the early stages of spread, we are actually interested in calculating the reproductive thresholds of the two alleles separately. The Y^D allele will spread when the reproductive threshold of type-1 ($X^{sn}Y^D$) individuals is greater than one; and the X^{it} allele will spread when the reproductive thresholds of both type-3 ($X^{it}Y^d$) and type-5 ($X^{sn}X^{it}$) individuals are greater than one.

In order to calculate these quantities, we first need to consider the fitness costs associated with each transgenic allele. We assume that a female homozygous for the X^{it} allele has a reduced

fecundity of $\theta(1-s_{it})$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0,1]$ and hence male ($X^{it}Y^{d}$) and female ($X^{sn}X^{it}$) mosquitoes that only have one copy of the X^{it} allele have a reduced fecundity of $\theta(1-hs_{it})$. Male mosquitoes that have the Y^{D} allele ($X^{sn}Y^{D}$) are assumed to have a reduced fecundity of $\theta(1-s_D)$; and finally, male mosquitoes that have both the X^{it} allele and the Y^{D} allele ($X^{it}Y^{D}$) are assumed to have a reduced fecundity of $\theta(1-hs_{it}-s_D)$.

According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 5. The rates of branching events (death and reproduction) for each particle type are given by

$$\lambda_1 = \mu + \theta(1-s_D), \tag{23}$$

$$\lambda_2 = \mu + \theta(1-hs_{it}-s_D), \tag{24}$$

$$\lambda_3 = \lambda_5 = \mu + \theta(1-hs_{it}), \tag{25}$$

$$\lambda_4 = \mu + \theta(1-s_{it}). \tag{26}$$

The process is also characterized by the following offspring counts per branching event:

$$f_{1,1} = \frac{\theta(1-s_D)}{\lambda_1}\left(1+\frac{1+t}{2}\right), \tag{27}$$

$$f_{3,3} = \frac{\theta(1-hs_{it})}{\lambda_3}, \tag{28}$$

$$f_{3,5} = \frac{\theta(1-hs_{it})}{\lambda_3}\left(\frac{1}{2}\right), \tag{29}$$

$$f_{5,3} = \frac{\theta(1-hs_{it})}{\lambda_5}\left(\frac{1}{4}\right), \tag{30}$$

$$f_{5,5} = \frac{\theta(1-hs_{it})}{\lambda_5}\left(1+\frac{1}{4}\right). \tag{31}$$

Upon mating with wild-type females, type-1 males ($X^{sn}Y^{D}$) can only give rise to type-1 males or wild-type females. Their reproductive threshold is therefore given by

$$R_{1,1} = f_{1,1} = \frac{\theta(1-s_D)(3+t)}{2\mu + 2\theta(1-s_D)}. \tag{32}$$

When this threshold is greater than one, the meiotic drive gene has a nonzero probability of spreading through the mosquito population. Simplifying Eq. (32), this means that an accidentally released meiotic drive gene has a chance of spreading into a wild population when

$$t > \frac{s_D - r(1-s_D)}{1-s_D+r(1-s_D)}. \tag{33}$$

When the population size is constant, then the condition for spread of the meiotic drive gene simplifies to

$$t > \frac{s_D}{1-s_D}. \tag{34}$$

This result describes the maximum fitness cost that a meiotic drive gene of given strength, t, can tolerate when the population size is constant. Eq. (33) shows that these conditions are relaxed during periods of population growth.

Calculating the reproductive thresholds for type-3 ($X^{it}Y^{d}$) and type-5 ($X^{it}X^{sn}$) mosquitoes is a little more complicated since type-3 males can mate with wild-type females to produce type-5 females. Additionally, type-5 females can mate with wild-type males to produce type-3 males and more type-5 females. The reproductive thresholds of type-3 and type-5 mosquitoes are therefore defined recursively analogous to Eq. (18). The reproductive threshold for a type-3 male is

$$R_{3,3} = f_{3,3} + \frac{f_{3,5}f_{5,3}}{1-f_{5,5}} = \frac{1}{2} + \frac{4}{5}\frac{1}{hs_{it}(1+r)-r+1} + \frac{9}{10}\frac{1}{2hs_{it}(1+r)-2r-3}, \tag{35}$$

and the reproductive threshold for a type-5 female is

$$R_{5,5} = f_{5,5} + \frac{f_{5,3}f_{3,5}}{1-f_{3,3}} = \frac{(1+r)(1-hs_{it})(6+r-hs_{it}(1+r))}{6-4hs_{it}+4r(1-hs_{it})}. \tag{36}$$

When these thresholds are greater than one, the X-linked drive-insensitive allele has a nonzero probability of spreading through the mosquito population. Simplifying Eqs. (35) and (36), this means that an accidentally released drive-insensitive allele has a chance of spreading into a wild population when:

$$hs_{it} < \frac{r}{1+r}. \tag{37}$$

When the population size is constant, then the condition for spread of the drive-insensitive allele becomes

$$s_{it} < 0. \tag{38}$$

This result suggests that, if the initial release size is very small, then the drive-insensitive allele will only spread through the population if it confers a fitness benefit to the mosquito. Under conditions of population growth, the allele will be able to tolerate a fitness cost as described by Eq. (37); however, it is only the fitness cost of having a single X^{it} allele that matters.

Interestingly, the strength of the meiotic drive gene, t, does not relax the conditions for spread of the X^{it} allele. This is because, in the early stages of spread, almost all matings are with the reservoir of wild-types, and so the increase in the proportion of individuals having the X^{it} allele is insignificant. The condition for the spread of the meiotic drive gene is more relaxed than the condition for the spread of the drive-insensitive allele. This suggests that, following an accidental release, a very feasible possibility is that the X^{it} allele will be lost while the Y^{D} allele will continue to spread through the mosquito population.

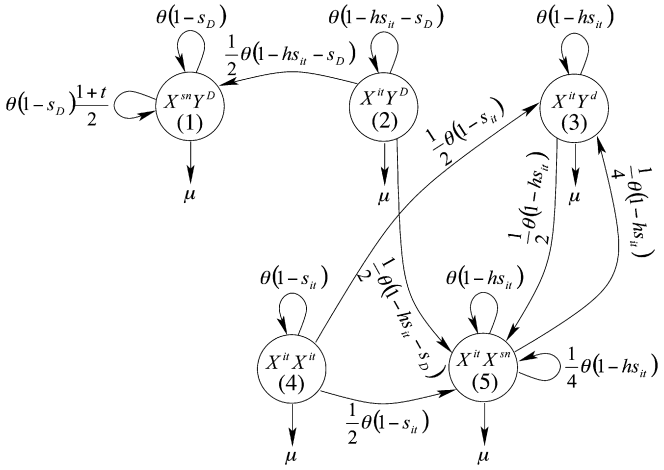

Fig. 5. Schematic for the early spread of a Y-linked meiotic drive gene (Y^{D}) and X-linked response allele (X^{it}) through a randomly mating mosquito population. Mosquitoes having both transgenic alleles ($X^{it}Y^{D}$) have a reduced fecundity of $\theta(1-hs_{it}-s_D)$ and mate with wild-type mosquitoes to give birth to $X^{it}X^{sn}$ females with probability 1/2 and $X^{sn}Y^{D}$ males with probability 1/2. $X^{sn}Y^{D}$ males have a reduced fecundity of $\theta(1-s_D)$ and mate with wild-type mosquitoes to give birth to $X^{sn}Y^{D}$ males with probability $(1+t)/2$ and wild-type females with probability $(1-t)/2$. The remaining transitions are described in the text. All mosquitoes have a death rate of μ.

When meiotic drive is supercritical, eventual extinction is uncertain; however, the extinction probability may still be very high. To determine this probability when the population size is not constant, we first define the probability generating function for the meiotic drive branching process. This is a function of the vector $\mathbf{z} = (z_1, \ldots, z_5)$ and is defined as

$$P_i(\mathbf{z}) = \sum_{j=1}^{5} p_{ij} z^j = \sum_{j=1}^{5} p_{ij} z_1^{j_1} z_2^{j_2} z_3^{j_3} z_4^{j_4} z_5^{j_5}, \tag{39}$$

where p_{ij} is the probability that a type-i particle gives rise to j_1 type-1 particles, j_2 type-2 particles and so on, and j is defined as the vector $j = (j_1, \ldots, j_5)$. Substituting the p_{ij} terms from the branching process into this equation, we have the probability generating functions,

$$P_1(\mathbf{z}) = \frac{\mu}{\lambda_1} + \frac{\theta(1-s_D)}{\lambda_1}\frac{1}{2}(1-t)z_1 + \frac{\theta(1-s_D)}{\lambda_1}\frac{1}{2}(1+t)z_1^2, \tag{40}$$

$$P_2(\mathbf{z}) = \frac{\mu}{\lambda_2} + \frac{\theta(1-hs_{it}-s_D)}{\lambda_2}\frac{1}{2}z_2 z_1 + \frac{\theta(1-hs_{it}-s_D)}{\lambda_2}\frac{1}{2}z_2 z_3, \tag{41}$$

$$P_3(\mathbf{z}) = \frac{\mu}{\lambda_3} + \frac{\theta(1-hs_{it})}{\lambda_3}\frac{1}{2}z_3 + \frac{\theta(1-hs_{it})}{\lambda_3}\frac{1}{2}z_3 z_5, \tag{42}$$

$$P_4(\mathbf{z}) = \frac{\mu}{\lambda_4} + \frac{\theta(1-s_{it})}{\lambda_4}\frac{1}{2}z_3 z_4 + \frac{\theta(1-s_{it})}{\lambda_4}\frac{1}{2}z_4 z_5, \tag{43}$$

$$P_5(\mathbf{z}) = \frac{\mu}{\lambda_5} + \frac{\theta(1-hs_{it})}{\lambda_5}\frac{1}{2}z_5 + \frac{\theta(1-hs_{it})}{\lambda_5}\frac{1}{4}z_3 z_5 + \frac{\theta(1-hs_{it})}{\lambda_5}\frac{1}{4}z_5^2. \tag{44}$$

The probability that the meiotic drive gene and drive-insensitive allele are both eventually lost from the population is then given by the smallest solution of the system of five simultaneous equations described by Eq. (10). For the meiotic drive branching process, this system of equations has the solution:

$$e_1 = \min\left\{\frac{1}{(1+r)(1-s_D)(1+t)}, 1\right\}, \tag{45}$$

$$e_3 = e_5 = \min\left\{\frac{1}{(1+r)(1-hs_{it})}, 1\right\}, \tag{46}$$

$$e_2 = \frac{1}{1+(2-e_1-e_3)(1+r)(1-s_D-hs_{it})}, \tag{47}$$

$$e_4 = \frac{1}{1+2(1-e_3)(1+r)(1-s_{it})}. \tag{48}$$

Here, e_1 is the loss probability beginning with a single type-1 male, e_2 is the loss probability beginning with a single type-2 male, and so on. When a type-1 male is released, these equations calculate the extinction probability of the Y^D allele; when a type-3, type-4 or type-5 female is released, the equations calculate the extinction probability of the X^{it} allele; and when a type-2 male is released, the equations calculate the probability that both alleles are eventually lost from the wild mosquito population.

The extinction probability beginning with a type-1 male is a decreasing function of meiotic drive strength and population growth rate and an increasing function of fitness cost due to the meiotic drive gene. Its form is similar to the extinction probability of a heterozygote for the HEG allele. The extinction probability beginning with a type-3 or -5 female is a decreasing function of population growth rate and an increasing function of the fitness cost due to a single drive-insensitive allele. As explained earlier, the meiotic drive parameter does not influence the extinction probability of the X^{it} allele. The extinction probability beginning with a type-4 female is a function of e_3 and is less than one when e_3 is less than one. Finally, the extinction probability beginning with a type-2 male is a function of both e_1 and e_3 and is less than one when both e_1 and e_3 are less than one.

One of the most important parameters for meiotic drive is the parameter t which represents the fraction by which the Y^D allele is overrepresented in the gametes of $X^{sn}Y^D$ males. There is currently very little data available for this parameter, and so we will rely somewhat on values used in the modeling literature. Although Cha et al. (2006a, b) have studied the effects of the meiotic drive gene M^D on the population dynamics of *Aedes aegypti*, these studies have not yielded any estimates for the parameter t. Researchers with experience modeling meiotic drive suggest that the degree of sex distortion should be as high as $t = 0.9$ (Yunxin Huang, personal communication); while the goal of some molecular biologists is to have no females produced at all (Fred Gould, personal communication). Huang et al. (2007) have investigated the full range of meiotic drive parameter values, $t \in [0, 1]$, with particular emphasis on the value $t = 0.8$.

The fitness cost associated with the meiotic drive system can originate from both the meiotic drive gene and the X-linked drive-insensitive allele. As for the meiotic drive parameter, t, very little data is available for these fitness costs, and so we will rely somewhat on values used in the modeling literature. Researchers with expertise in modeling meiotic drive suggest that a fitness cost on the order of $s_D \approx s_{it} \approx 0.05$ is desirable (Yunxin Huang, personal communication). Huang et al. (2007) have modeled the spread of meiotic drive for fitness costs of 0.05 and 0.2, and have found that a fitness cost on the order of $s_D = s_{it} = 0.05$ is necessary for gene drive to occur over default conditions. We will investigate a fitness cost in the range $s_D, s_{it} \in [0, 0.2]$.

For the meiotic drive strategy modeled by Huang et al. (2007), there is an additional fitness cost due to the X^{it} allele being associated with a refractory gene. This fitness cost has been discussed for the case of HEGs, and has been estimated by the range [−0.05,0.04]. Combining these two fitness costs for the X^{it} allele suggests a fitness cost in the range $s_{it} \in [-0.05, 0.24]$. We assume that the fitness cost of the X^{it} allele is additive and hence choose $h = 0.5$.

The meiotic drive strategy by which a Y-linked meiotic drive gene (Y^D) is used to drive an X-linked drive-insensitive allele (X^{it}) into the population is interesting in that, before the Y-linked drive gene has been driven into the population, the X-linked response allele is very likely to be lost from the population. In the absence of population growth, the X-linked response allele requires a fitness benefit ($s_{it} < 0$) in order to spread (Eq. (38)). The Y^D allele is able to persist in the population despite a fitness cost by virtue of its overrepresentation in the gametes of $Y^D X^{it}$ males. For a high yet feasible meiotic drive strength of $t = 0.8$, the persistence probability of a single escapee is between 19% and 30% for all realistic fitness costs, and the Y^D allele has a chance of persisting even during a dramatic population decline (see supplemental Fig. 3A–C).

The probability of transgenes persisting following an accidental release depends largely on the genotype of the escaped mosquito. If the escapee only carries the transgenic X^{it} allele (as is the case for genotypes $Y^d X^{it}$, $X^{sn}X^{it}$ and $X^{it}X^{it}$), then this allele must confer a fitness benefit in order to spread (Fig. 6). If the escapee only carries the transgenic Y^D allele (as is the case for genotype $Y^D X^{it}$), then the escapee has a persistence probability greater than 27% over the entire range of realistic fitness costs (Fig. 6). Interestingly, an escapee having both transgenic alleles ($Y^D X^{it}$) has a slightly smaller persistence probability than an escapee only having the transgenic Y^D allele. This is because it is usually only the Y^D allele that persists following an accidental release, and the transgenic X^{it} allele confers a fitness cost to the escapee.

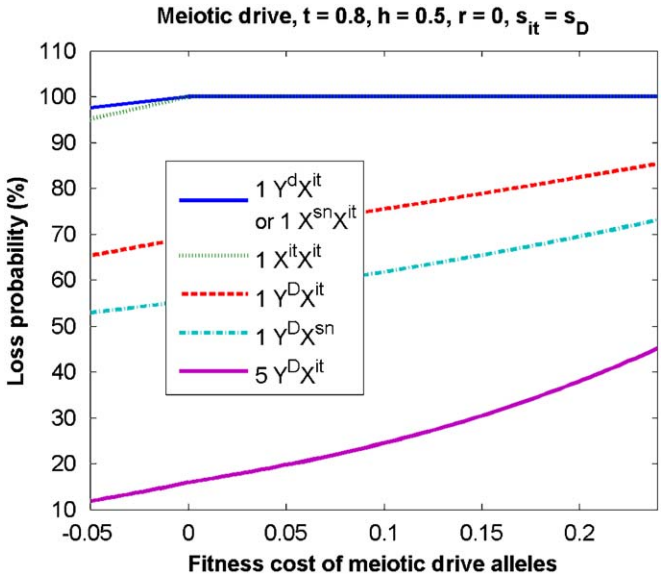

Fig. 6. Asymptotic probabilities of meiotic drive loss as a function of fitness cost. The escapee genotype is varied and the $X^{it}Y^{D}$ release size is increase to five.

Under default conditions, for five escapees having both transgenic alleles the loss probability decreases to 20%; while for 10 escapees the loss probability decreases to 4%. The meiotic drive allele, Y^D, is therefore more invasive than a HEG or TE; however, the X-linked response allele, X^{it}, is less invasive than either. The meiotic drive allele is more likely to persist than not for escapes consisting of more than two mosquitoes having the Y^D allele; while the drive-insensitive response allele requires a fitness benefit in order to have any chance of spreading, and even then is very likely to go extinct.

5. *Medea*

Medea, also known as maternal-effect-dominant embryonic arrest, is another form of meiotic drive. Its dynamics have been studied in *Tribolium* beetles (Wade and Beeman, 1994); however, it has attracted much recent attention since an engineered *Medea* element has been observed to rapidly spread through *Drosophila* populations in the laboratory (Chen et al., 2007). *Medea* is able to spread through a population despite a fitness cost through its ability to cause the death of all offspring of heterozygous females that do not inherit the *Medea* allele. This distorts the offspring ratio in favor of the *Medea* allele.

In the early stages of spread, the vast majority of mosquitoes belong to what may be thought of as a reservoir of wild-types. One implication of this is that the death of wild-type offspring of female heterozygotes will have very little influence on the spread of the *Medea* allele following an accidental release because the vast majority of mosquitoes will still be wild-types. The early spread of the *Medea* allele will therefore be very similar to the spread of a new mutation.

Taking this consideration into account, we model the spread of the *Medea* allele following an accidental release using a two-type continuous-time branching process in which type-1 particles are heterozygous and type-2 particles are homozygous for the *Medea* allele. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1\,\text{gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity θ as described by Eq. (1).

Matings between wild-type mosquitoes and mosquitoes homozygous for the *Medea* allele will produce heterozygotes; while matings between wild-type mosquitoes and heterozygotes for the *Medea* allele will produce offspring that are half heterozygotes and half wild-types (wild-types will die if their heterozygote parent is female). Since homozygotes will be lost from the population after one generation, we are interested in the reproductive threshold of mosquitoes that are heterozygous for the *Medea* allele. For a heterozygote, this is simply given by the average number of heterozygotes that it produces at a branching event, $f_{1,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with the *Medea* allele. We assume that a female homozygous for *Medea* has a reduced fecundity of $\theta(1-s)$. Male mosquitoes that mate with wild-type females do not suffer from reduced fecundity, and so the mean fecundity of a homozygote is $\theta(1-s/2)$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0,1]$, and hence the mean fecundity of a heterozygote is $\theta(1-hs/2)$.

According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 7. The rate of branching events (death and reproduction) for a heterozygote is equal to

$$\lambda_1 = \mu + \theta(1-hs/2), \tag{49}$$

and the rate for a homozygote is equal to

$$\lambda_2 = \mu + \theta(1-s/2). \tag{50}$$

The reproductive threshold for a mosquito heterozygous for the *Medea* allele is then given by

$$R_{1,1} = f_{1,1} = \frac{\theta}{\lambda_1}\left(1-\frac{hs}{2}\right)\left(1+\frac{1}{2}\right). \tag{51}$$

When the reproductive threshold is greater than one, *Medea* spread is supercritical, and the *Medea* allele has a nonzero probability of spreading through the mosquito population. Simplifying Eq. (51), this means that an accidentally released *Medea* element has some chance of spreading into a wild population when

$$hs < \frac{2r}{1+r}. \tag{52}$$

When the population size is constant, then the condition for *Medea* spread simplifies to

$$hs < 0. \tag{53}$$

One implication of this result is that only the fitness cost to the heterozygote determines whether the *Medea* allele has a chance of

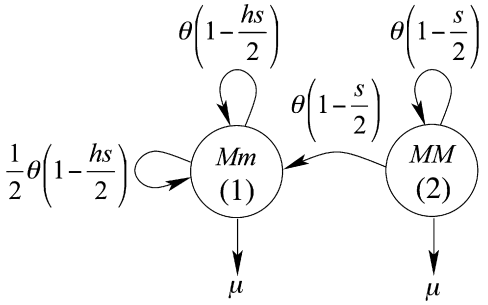

Fig. 7. Schematic for the early spread of a *Medea* allele through a randomly mating mosquito population. Mosquitoes homozygous for *Medea* (*MM*) have a reduced fecundity of $\theta(1-s/2)$ and mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for *Medea* (*Mm*). Heterozygotes have a reduced fecundity of $\theta(1-hs/2)$ and mate with wild-type mosquitoes to give birth to heterozygotes with probability 1/2 and wild-types with probability 1/2. All mosquitoes have a death rate of μ.

spreading or not. If the population size is constant, the *Medea* allele must confer a fitness benefit in order to have some chance of spreading. During periods of population growth, the magnitude of fitness cost that is tolerable is given by Eq. (52).

When the reproductive threshold is greater than one, *Medea* spread is possible; however, the probability of *Medea* loss may still be very high. To determine the extinction probability of the *Medea* allele, we first define the probability generating function for the *Medea* branching process. This is a function of the vector $\mathbf{z} = (z_1, z_2)$ and is defined in Eq. (7). Substituting the p_{ij} terms from the *Medea* branching process into this equation, we have the probability generating function,

$$P_1(\mathbf{z}) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1}\left(1 - \frac{hs}{2}\right)\frac{1}{2}z_1 + \frac{\theta}{\lambda_1}\left(1 - \frac{hs}{2}\right)\frac{1}{2}z_1^2, \tag{54}$$

$$P_2(\mathbf{z}) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2}\left(1 - \frac{s}{2}\right)z_1 z_2. \tag{55}$$

The probability that the *Medea* allele is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Eq. (10). For the *Medea* branching process, this system of equations has the solution,

$$e_1 = \min\left\{\frac{2}{(1+r)(2-hs)}, 1\right\}, \tag{56}$$

$$e_2 = \frac{1}{1 + (1+r)(2-s)(1-e_1)}. \tag{57}$$

The extinction probability beginning with a heterozygote, e_1, is a decreasing function of population growth rate and an increasing function of heterozygote fitness cost. The extinction probability beginning with a homozygote, e_2, is a function of e_1 and is less than one when e_1 is less than one.

To make sense of these equations, we need to have some idea of the fitness consequences of the *Medea* allele. While very few measurements exist for this parameter, Chen et al. (2007) have fitted data from their synthetic *Medea* element in *Drosophila* to a model of *Medea* spread. They found that their data is most consistent with an element that has no fitness cost; however, their confidence interval for this parameter, $s \in [-0.23, 0.1]$, is large. The fitness cost of being associated with a refractory gene has been discussed for the case of HEGs, and has been estimated by the range $[-0.05, 0.04]$. Combining these two estimates suggests a range for the fitness cost of a *Medea* allele on the order of $s \in [-0.28, 0.14]$.

Following Chen et al. (2007), we assume that the fitness cost of the *Medea* allele is additive and hence choose a degree of homozygosity of $h = 0.5$. The proportion of wild-type embryos that die because they do not produce an antidote to the toxin produced by a female *Medea*-infected parent, t, although important, is not relevant to the early spread of the *Medea* allele.

Medea is an ideal drive system for the needs of transgene containment since it will spread very quickly following an intentional release (Chen et al., 2007); however, it requires either a fitness benefit or period of population growth in order to spread following an accidental release (Fig. 8). During a period of population decline, the *Medea* allele has no chance of spreading; however, for a population growth rate of $r = 0.1\,\text{gen}^{-1}$, the persistence probability for a single homozygous escapee is between 11% and 17% over the full range of realistic parameterizations (see supplemental Fig. 4).

The probability that a beneficial *Medea* allele is lost from the population following an accidental release is most affected by the number of escapees during an accidental release (Fig. 8). For a realistic fitness benefit of $s = -0.05$, an escape of 10 mosquitoes homozygous for the *Medea* allele decreases the loss probability to

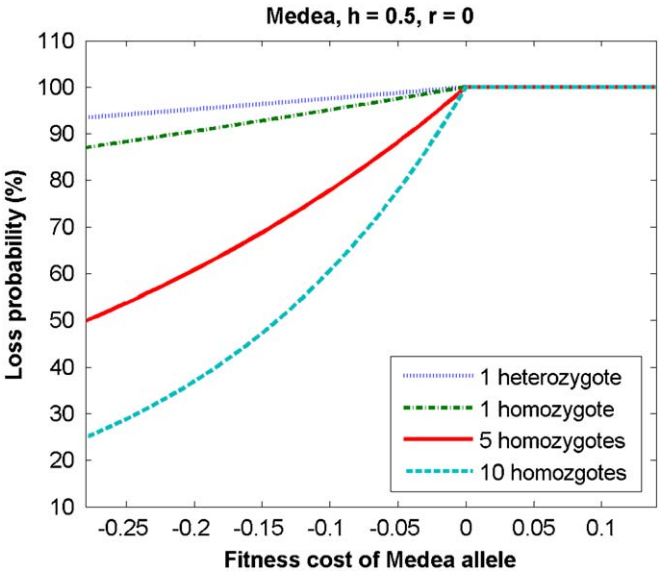

Fig. 8. Asymptotic probabilities of *Medea* loss as a function of fitness cost. The escapee genotype is varied and the homozygote release size is increased to 10.

78%. For 25 escapees, the loss probability decreases to 54%. Under default conditions, the *Medea* allele is more likely to persist than not for homozygous escape sizes greater than 27.

Medea is much less prone to spread following an accidental release than HEGs, TEs or meiotic drive genes; however, it should still be kept in mind that an escape of ~30 transgenic mosquitoes is entirely feasible. If these mosquitoes are refractory to malaria, it is also feasible that they may possess a slight fitness advantage over wild mosquitoes. Additionally, the observation of *Medea* alleles in nature suggests that naturally occurring *Medea* alleles may actually confer an innate selective advantage to their host; as hinted at by the confidence interval for fitness effects of the synthetic *Medea* allele studied by Chen et al. (2007). Therefore, while *Medea* is significantly safer than HEGs, TEs and meiotic drive genes; we should still be weary of its ability to spread following an accidental release.

6. *Wolbachia*

Wolbachia is a maternally inherited, intracellular bacterium found in a wide variety of invertebrate taxa. *Wolbachia* infections are associated with several host reproductive alterations including cytoplasmic incompatibility (Stouthammer et al., 1999), in which offspring of matings between infected males and uninfected females are completely or partially sterilized, while matings involving infected females always produce infected offspring. This favors the offspring ratio in favor of the *Wolbachia* infection, and *Wolbachia* is therefore able to spread rapidly through a population despite a fitness cost (Turelli and Hoffmann, 1999).

To calculate the reproductive threshold of a *Wolbachia* bacterium following an accidental release, we consider a two-type continuous-time branching process in which type-1 particles are females infected with the *Wolbachia* bacterium and type-2 particles are *Wolbachia*-infected males. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1\,\text{gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity θ as described by Eq. (1).

In the early stages of spread, almost all matings involve at least one mosquito uninfected by the *Wolbachia* bacterium. Crosses between *Wolbachia*-infected females and uninfected males tend to

produce *Wolbachia*-infected males and females; however, maternal transmission is imperfect such that a proportion, u, of their offspring are uninfected and a proportion, $1-u$, of their offspring are infected by *Wolbachia*. Crosses between *Wolbachia*-infected males and uninfected females tend to be infertile with a proportion, t, of their offspring suffering from CI-induced sterility and a proportion, $1-t$, of their offspring surviving. The viable offspring of this cross, however, are uninfected and may be thought of as belonging to the reservoir of uninfected hosts.

Since *Wolbachia* is maternally inherited, we are only interested in the reproductive threshold of *Wolbachia*-infected females. This is equal to the average number of *Wolbachia*-infected females that it produces at a branching event, $f_{1,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with a *Wolbachia* infection. We assume that a female infected with the *Wolbachia* bacterium has a reduced fecundity of $\theta(1-s)$. Infected males that mate with uninfected females do not suffer from reduced fecundity, and so their fecundity is θ.

According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 9. The rate of branching events (death and reproduction) for a *Wolbachia*-infected female is equal to

$$\lambda_1 = \mu + \theta(1-s), \tag{58}$$

and the rate for a *Wolbachia*-infected male is equal to

$$\lambda_2 = \mu + \theta. \tag{59}$$

The reproductive threshold for a *Wolbachia*-infected female is then given by

$$R_{1,1} = f_{1,1} = \frac{\theta(1-s)}{\lambda_1}\left(1+\frac{1-u}{2}\right). \tag{60}$$

When the reproductive threshold is greater than one, *Wolbachia* spread is supercritical, and the *Wolbachia* bacterium has a nonzero probability of spreading through the mosquito population. Simplifying Eq. (60), this means that an accidentally released *Wolbachia*-infected female has some chance of spreading into a wild population when

$$(1+r)(1-s)(1-u) > 1. \tag{61}$$

This equation implies that population growth can somewhat compensate for imperfect maternal transmission and a fitness cost associated with the *Wolbachia* bacterium. When the population size is constant ($r=0$), this equation is consistent with the condition derived by Turelli et al. (1992) for *Wolbachia* spread. According to Turelli et al. (1992), since maternal transmission is always imperfect ($u>0$) then, when the population size is constant, the *Wolbachia* bacterium must confer a fitness benefit to the mosquito in order for it to spread from a low prevalence. The magnitude of the fitness benefit, $-s$, required for *Wolbachia* spread is then given by

$$-s > \frac{1}{1-u} - 1. \tag{62}$$

When the reproductive threshold is greater than one, *Wolbachia* spread is possible; however, the probability of *Wolbachia* loss may still be very high. To determine the extinction probability of the *Wolbachia* bacterium, we first define the probability generating function for the *Wolbachia* branching process. This is a function of the vector $\mathbf{z} = (z_1, z_2)$ and is defined in Eq. (7). Substituting the p_{ij} terms from the *Wolbachia* branching process into this equation, we have the probability generating function,

$$P_1(\mathbf{z}) = \frac{\mu}{\lambda_1} + \frac{\theta(1-s)}{\lambda_1}uz_1 + \frac{\theta(1-s)}{\lambda_1}\frac{1-u}{2}z_1^2 + \frac{\theta(1-s)}{\lambda_1}\frac{1-u}{2}z_1z_2, \tag{63}$$

$$P_2(\mathbf{z}) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2}z_2. \tag{64}$$

The probability that *Wolbachia* is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Eq. (10). For the *Wolbachia* branching process, this system of equations has the solution,

$$e_1 = \min\left\{\frac{1}{(1+r)(1-s)(1-u)}, 1\right\}, \tag{65}$$

$$e_2 = 1. \tag{66}$$

The extinction probability beginning with a *Wolbachia*-infected female, e_1, is a decreasing function of population growth rate and an increasing function of fitness cost. It is also a decreasing function of the reduction in maternal transmission. An accidental release beginning with a *Wolbachia*-infected male is certain to go extinct because *Wolbachia* is maternally inherited.

One of the most important parameters for the spread of a *Wolbachia* infection is the proportion, t, of uninfected embryos from a cross involving an uninfected female and an infected male that die due to cytoplasmic incompatibility. This parameter is not so relevant in the early stages of spread, however, since surviving offspring belong to what may be thought of as a reservoir of uninfected hosts.

Another important parameter is the proportion, u, of offspring from *Wolbachia*-infected females that are uninfected by *Wolbachia*. Maternal transmission is high, so this parameter tends to be relatively small. Charlat et al. (2004) have estimated a 95% confidence interval for this parameter for *D. melanogaster* in the range $u \in [0.01, 0.17]$. Point estimates and ranges estimated for other species are within this range—for example, $u = 0.01$ for *Culex pipiens* in California (Rasgon and Scott, 2003), $u = 0.025$ for *D. melanogaster* in Australia (Hoffmann et al., 1994, 1998), and $u \in [0.03, 0.15]$ for *D. simulans* in California (Weeks et al., 2007).

The fitness consequences of a *Wolbachia* infection are particularly relevant to its ability to spread. Recent measurements suggest that a *Wolbachia* infection can induce either a fitness benefit or cost, depending on the host species, genetic background (Dean, 2006), and the amount of time that the infection has been present in the host species (Weeks et al., 2007). In terms of fitness costs, Hoffmann et al. (1990) have measured a reduction in fecundity of $s \in [0.1, 0.2]$ for *Wolbachia*-infected *D. melanogaster*, and Stevens and Wade (1990) have measured a reduction in overall fitness of $s = 0.37$ for *Wolbachia*-infected *Tribolium* beetles. Neutral *Wolbachia* infections ($s = 0$) have been observed in *Drosophila yakuba* (Charlat et al., 2004), *C. pipiens* (Rasgon and Scott, 2003), and in natural populations of *D. simulans* (Turelli and Hoffmann, 1995, 1999).

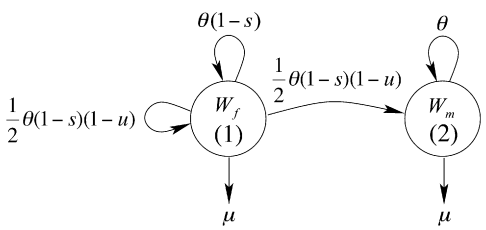

Fig. 9. Schematic for the early spread of a *Wolbachia* bacterium through a randomly mating mosquito population. *Wolbachia*-infected females (W_f) have a reduced fecundity of $\theta(1-s)$ and mate with wild-type mosquitoes to give birth to *Wolbachia*-infected males and females, each with probability $(1-u)/2$, and wild-types with probability u. All mosquitoes have a death rate of μ.

It is becoming increasingly recognized that *Wolbachia* may not just be a reproductive parasite but may also be a mutualist, and hence confer a selective advantage to its host (Steve Sinkins, personal communication). Weeks et al. (2007) have noted that, although 20 years ago California *D. simulans* were shown to have a reduced fecundity due to *Wolbachia* infection in the laboratory, *Wolbachia*-infected females now exhibit an increased fecundity of $s = -0.1$ in the laboratory. Dobson et al. (2002) have also observed a fecundity advantage due to *Wolbachia* superinfection in the mosquito *Aedes albopictus*; however, this is yet to be validated in field populations (Jason Rasgon, personal communication). Combining these measurements and accounting for the potential effects of a refractory gene suggests a range of fecundity effects due to *Wolbachia* infection on the order of $s \in [-0.1, 0.37]$.

Like *Medea*, *Wolbachia* is an ideal drive system for the needs of transgenic containment since it will spread following an intentional release (Turelli and Hoffmann, 1999); however, it requires either a fitness benefit or period of population growth in order to spread following an accidental release (Fig. 10). The conditions for spread are a little more restrictive for *Wolbachia* than they are for *Medea*, primarily because maternal transmission of *Wolbachia* is imperfect. As the efficiency of maternal transmission decreases, then the fitness benefit required for *Wolbachia* spread becomes larger (see supplemental Fig. 5A). Population growth can also compensate for a fitness cost and imperfect maternal transmission. For a population growth rate of $r = 0.1\ \text{gen}^{-1}$ and a maternal transmission rate of 97%, the *Wolbachia* bacterium will spread for fitness costs less than 6.5% (see supplemental Fig. 5B).

For a *Wolbachia* bacterium that is able to spread following an accidental release, its loss probability is most affected by the number of escapees during an accidental release (Fig. 10). For a realistic fitness benefit of $s = -0.05$, an escape of 10 *Wolbachia*-infected females decreases the loss probability to 83%. For 25 infected females, the loss probability decreases to 63%. Under default conditions, the *Wolbachia* bacterium is more likely to persist than not for infected female escape sizes greater than 37. It is likely that less than half of the escapees from an accidental release will be infected females, and so this corresponds to a required escape size greater than 74 in order for it to be more likely than not that the *Wolbachia* bacterium will spread following an accidental release.

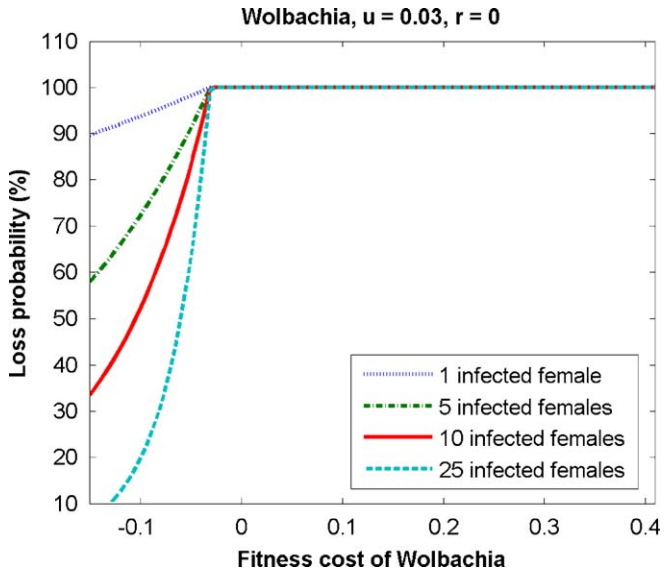

Fig. 10. Asymptotic probabilities of *Wolbachia* loss as a function of fitness cost. Release size is increased to 25.

Wolbachia is less prone to spread than *Medea* following an accidental release; however, as for *Medea*, it should be kept in mind that an escape of ~75 infected mosquitoes is possible. Recent measurements suggest that, in addition to a fitness benefit due to malaria refractoriness, the *Wolbachia* infection may in fact confer a selective advantage to the escapees over wild mosquitoes (Weeks et al., 2007). Additionally, the observation that *Wolbachia* bacteria exist in nature suggests that *Wolbachia* infections have been able to spread from very low prevalences, in contradiction to the prediction by Turelli and Hoffmann (1999) that the initial frequency of *Wolbachia* must exceed ~10% in order to spread. The fact that real populations are locally structured may assist in the emergence of a new *Wolbachia* infection (Steve Sinkins and Jason Rasgon, personal communication). Therefore, while *Wolbachia* is one of the safest drive systems; there is still a possibility that a transgenic *Wolbachia* strain will persist in a wild mosquito population following an accidental release.

It should also be noted that the extinction probabilities calculated here are for the *Wolbachia* drive strategy in which the refractory gene is engineered directly into the *Wolbachia* genome. Other drive strategies utilizing *Wolbachia* have been proposed by Turelli and Hoffmann (1999) and Sinkins and Godfray (2004) and may be associated with slightly different loss probabilities.

7. Engineered underdominance

The simplest case of underdominance is when a trait is determined by two alleles at a single locus and the fitness of a heterozygote is less than that of either homozygote (Hartl and Clark, 1989). The dynamics of underdominant traits are generally unstable and, depending on the initial frequency of the two alleles, one will tend to be lost while the other will reach fixation in the population (Crow and Kimura, 1970; Spiess, 1977). However, the problem with single-locus underdominance as a form of gene drive is that, in order for an introduced allele to reach fixation, it must be introduced at a very high frequency in the population which is often not feasible for gene drive strategies (Curtis, 1968).

A novel form of engineered underdominance has been suggested by Davis et al. (2001) that does not require such a high release size in order to spread into a naïve population. In this system there are two transgenic constructs, α and β, each of which possesses a lethal gene and a suppressor gene that down-regulates the expression of the lethal gene on the other construct. This system is most efficient when the two transgenic constructs are inserted at loci on nonhomologous chromosomes, and when there are anti-pathogen genes associated with each construct.

Since both loci can be homozygous for the wild-type allele, homozygous for the transgenic allele, or heterozygous, then there are nine possible genotypes for the engineered underdominance system. However, individuals that only possess one transgenic construct are unviable because they express a lethal gene while lacking its suppressor. This means that there are only five viable genotypes—four of which possess both transgenic constructs, and one which belongs to the reservoir of wild-types.

Following an accidental release, we are interested in tracking the four viable genotypes possessing both transgenic constructs. Let us denote the transgenic allele at the first locus as α, the wild-type allele at the first locus as A, the transgenic allele at the second locus as β, and the wild-type allele at the second locus as B. Utilizing these symbols, we consider a four-type continuous-time branching process in which type-1 particles are heterozygous at both loci ($\alpha A \beta B$), type-2 particles are heterozygous at the first locus ($\alpha A \beta\beta$), type-3 particles are heterozygous at the second locus ($\alpha\alpha \beta B$), and type-4 particles are homozygous at both loci ($\alpha\alpha\beta\beta$).

We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1\,\text{gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity θ as described by Eq. (1).

In the early stages following an accidental release, almost all matings involve at least one wild-type mosquito. This means that the only viable transgenic offspring will be heterozygous at both loci, since all offspring will inherit a wild-type allele from their wild-type parent at both loci. In calculating the reproductive threshold of the engineered underdominance constructs, we are therefore interested in the average number of type-1 offspring that a type-1 mosquito produces at a branching event, $f_{1,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with the engineered underdominance constructs. It is likely that the two constructs will each confer a fitness cost to the host, since binding of a suppressor to its corresponding lethal gene may be imperfect. We assume that a female homozygous for the α allele has their fecundity reduced by a fraction, s. Male mosquitoes that mate with wild-type mosquitoes do not suffer from reduced fecundity, and so the mean reduction in fecundity due to being homozygous for the α allele is $s/2$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0,1]$, and hence the mean reduction in fecundity due to being heterozygous for the α allele is $hs/2$. The β allele is expected to cause a similar reduction in fecundity, and so type-1 particles are expected to have a fecundity of $\theta(1-hs)$, type-2 and type-3 particles have an expected fecundity of $\theta(1-(s+hs)/2)$, and type-4 particles have an expected fecundity of $\theta(1-s)$.

According to standard branching process theory (Lange, 2002; Dorman et al., 2004), these assumptions lead to the branching process shown in Fig. 11. The rates of branching events (death and reproduction) for each particle type are equal to

$$\lambda_1 = \mu + \theta(1-hs), \tag{67}$$

$$\lambda_2 = \lambda_3 = \mu + \theta\left(1 - \frac{s+hs}{2}\right), \tag{68}$$

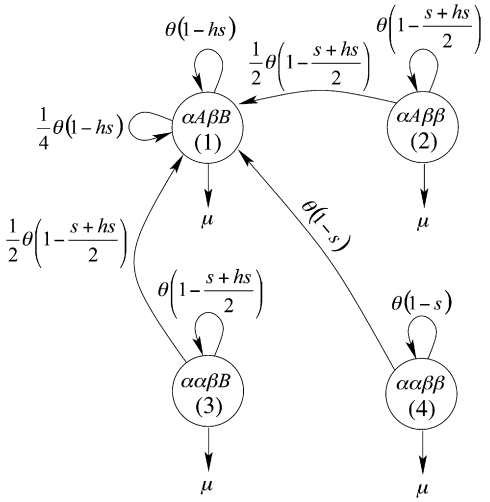

Fig. 11. Schematic for the early spread of a pair of engineered underdominance alleles through a randomly mating mosquito population. Mosquitoes homozygous for one or both alleles ($\alpha A\beta\beta$, $\alpha\alpha\beta B$ and $\alpha\alpha\beta\beta$) mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for both alleles ($\alpha A\beta B$). Heterozygotes for both alleles have a reduced fecundity of $\theta(1-hs)$ and mate with wild-type mosquitoes to give birth to heterozygotes for both alleles with probability 1/4, wild-types with probability 1/4 and unviable offspring with probability 1/2. All mosquitoes have a death rate of μ.

$$\lambda_4 = \mu + \theta(1-s). \tag{69}$$

The reproductive threshold for a mosquito heterozygous at both loci is then given by

$$R_{1,1} = f_{1,1} = \frac{\theta(1-hs)}{\lambda_1}\left(1+\frac{1}{4}\right). \tag{70}$$

When the reproductive threshold is greater than one, spread of the engineered underdominance alleles is supercritical, and the engineered constructs have a nonzero probability of spreading through the mosquito population. Simplifying Eq. (70), this means that an accidentally released mosquito having both underdominance constructs has some chance of spreading into a wild population when

$$-hs > \frac{1-r}{1+r}. \tag{71}$$

When the population size is constant, then the condition for engineered underdominance spread simplifies to

$$-hs > 1. \tag{72}$$

Noting that the fecundity of a type-1 particle is $\theta(1-hs)$, this condition means that, when the population size is constant, the underdominance alleles have a chance of spreading when the fecundity of a type-1 particle is greater than 2θ. The fecundity of a wild-type mosquito is equal to θ, and so this implies that a heterozygote for both underdominance alleles must be twice as fecund as a wild-type mosquito in order for the underdominance alleles to have some chance of spreading into a population of constant size. This condition is moderately relaxed during periods of population growth, as described by Eq. (71).

When the reproductive threshold is greater than one, spread of the underdominance alleles is possible; however, the probability of loss of these alleles may still be very high. To determine the extinction probability of the underdominance alleles, we first define the probability generating function for the engineered underdominance branching process. This is a function of the vector $\mathbf{z} = (z_1, z_2, z_3, z_4)$ and is defined as

$$P_i(\mathbf{z}) = \sum_{j=1}^{4} p_{ij} z^j = \sum_{j=1}^{4} p_{ij} z_1^{j_1} z_2^{j_2} z_3^{j_3} z_4^{j_4}, \tag{73}$$

where p_{ij} is the probability that a type-i particle gives rise to j_1 type-1 particles, j_2 type-2 particles, and so on, and j is defined as the vector $j = (j_1, \ldots, j_4)$. Substituting the p_{ij} terms from the engineered underdominance branching process into this equation, we have the probability generating function,

$$P_1(\mathbf{z}) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1}(1-hs)\frac{3}{4}z_1 + \frac{\theta}{\lambda_1}(1-hs)\frac{1}{4}z_1^2, \tag{74}$$

$$P_2(\mathbf{z}) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2}\left(1-\frac{s+hs}{2}\right)\frac{1}{2}z_2 + \frac{\theta}{\lambda_2}\left(1-\frac{s+hs}{2}\right)\frac{1}{2}z_2 z_1, \tag{75}$$

$$P_3(\mathbf{z}) = \frac{\mu}{\lambda_3} + \frac{\theta}{\lambda_3}\left(1-\frac{s+hs}{2}\right)\frac{1}{2}z_3 + \frac{\theta}{\lambda_3}\left(1-\frac{s+hs}{2}\right)\frac{1}{2}z_3 z_1, \tag{76}$$

$$P_4(\mathbf{z}) = \frac{\mu}{\lambda_4} + \frac{\theta}{\lambda_4}(1-s)z_1 z_4, \tag{77}$$

The probability that engineered underdominance alleles are eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Eq. (10). For the engineered underdominance branching process, this system of equations has the solution,

$$e_1 = \min\left\{\frac{2}{(1+r)(1-hs)}, 1\right\}, \tag{78}$$

$$e_2 = e_3 = \frac{2}{2(2 - e_1 + r(1 - e_1)) - (1 - e_1)(1 + h)s(1 + r)}, \quad (79)$$

$$e_4 = \frac{1}{3 + 2r - 2e_1(1 - s)(1 + r) - 2s(1 + r)}. \quad (80)$$

The extinction probability beginning with a heterozygote for both alleles, e_1, is a decreasing function of population growth rate and an increasing function of heterozygote fitness cost. The extinction probabilities beginning with any other viable genotype are all functions of e_1, and are less than one when e_1 is less than one.

The two allele, two locus strategy for engineered underdominance has only recently been proposed as a mechanism of gene drive and so there are no experimental measurements of its effects on female fecundity. We will therefore rely on fitness costs used in the modeling literature.

Magori and Gould (2006) have investigated a large range of fitness costs due to underdominance constructs, $s \in [0, 0.5]$. While they note that a neutral underdominance construct is unrealistic; their model predicts that a refractory gene could be driven into a population in the presence of a fitness cost of $s = 0.05$ provided a sufficient release ratio is achieved. Assuming that a beneficial refractory gene may be associated with each engineered construct, we explore an expanded range of fecundity effects on the order of $s \in [-0.05, 0.5]$. For completeness, we also consider a full range for the degree of homozygosity of $h \in [0, 1]$.

Engineered underdominance constructs are the safest gene drive system for the needs of transgenic containment. Their safety in the event of an accidental release primarily arises from the high release ratio required for them to spread. Davis et al. (2001) showed that, even under ideal conditions in which there are no fitness costs associated with the constructs, transgenic insects must exceed a frequency of ~27% in order to have some chance of spreading through the population. The required release ratio increases as fitness costs are accounted for (Magori and Gould, 2006); and fecundity effects due to underdominance constructs are unlikely to be positive given the constructs are engineered to express lethal genes down-regulated by suppressors.

According to Eq. (71), engineered underdominance constructs face certain extinction following an accidental release for the entire range of realistic parameters. Even for the most generous estimate of population growth rate ($r = 0.2\ \text{gen}^{-1}$), a heterozygote for one of the constructs must have a fitness benefit of at least 67% in order to have some chance of spreading from a low prevalence. This is unachievable for even the most generous estimates of the parameters h and s. In order to spread following an escape from a field cage, the escape size must represent a significant fraction of the wild population, as predicted by Davis et al. (2001) and Magori and Gould (2006).

It should be noted that the two allele, two locus strategy for engineered underdominance is not the only possible strategy for spreading underdominance constructs into a wild population; however, for the time being it is reasonable to model this since more complicated strategies are further down the road (Fred Gould, personal communication). Variants of this strategy have been proposed by Magori and Gould (2006); and although they are also very likely to be lost following an accidental release, their extinction probabilities and required release ratios may be smaller than for the two allele, two locus strategy.

8. Conclusions

In the event of an actual escape of transgenic mosquitoes from an ambient field cage, there will be many factors influencing the loss or persistence of transgenic DNA. Escapees from field cages will likely suffer from some degree of inbreeding depression and may be slightly maladapted to conditions in the wild. Control measures may also be put in place to reduce the spread of transgenic DNA—for example, the use of vegetation-free zones or "trap crops" surrounding the cage to restrict mosquito dispersal. Consequently, the loss probabilities calculated here contain systematic errors and are mainly of comparative interest for studying the possible outcomes under a variety of gene drive strategies.

Comparison of the predictions of the six branching processes in this paper suggests that engineered underdominance constructs provide the safest gene drive strategy for the needs of transgenic containment. According to model predictions, this system will face certain extinction for the entire range of realistic parameters provided that the release proportion is less than ~27% of the wild population (Davis et al., 2001). The release proportion required for transgenic spread is even higher if there is a fitness cost associated with the constructs.

The catch-22 of this result is that engineered underdominance constructs are also not very invasive following an intentional release. Any gene drive strategy with the aim of controlling disease over a large geographic area must have the ability to spread between several partially isolated subpopulations of disease vectors. The problem with engineered underdominance constructs is that, even under the most generous parameterizations, they require a migration rate consistently above 3% per generation in order to spread into neighboring populations (Davis et al., 2001). This required migration rate is higher than those observed between subpopulations of *An. gambiae* mosquitoes in Mali, West Africa (Taylor et al., 2001).

TEs and HEGs, on the other hand, are very capable of spreading between subpopulations of disease vectors following an intentional release (Taylor and Manoukis, 2003; Deredec et al., 2007); but are also very capable of spreading following an accidental release. According to model predictions with default parameters, a TE is more than 50% likely to persist for escape sizes greater than 10; while a HEG is more than 50% likely to persist for homozygote escape sizes greater than five. These escape sizes are entirely within the realm of possibility; and hence the invasiveness of TEs and HEGs is offset by the risk of establishment following an accidental release.

Meiotic drive systems are also of concern following an accidental release. For the meiotic drive strategy analyzed in which a Y-linked meiotic drive gene is used to drive an X-linked response allele into the population, the Y-linked drive gene is very capable of spreading following an accidental release, while the X-linked response allele requires population growth or a fitness benefit in order to spread. If we are concerned about the unknown effects of an attached refractory allele, then the configuration of this meiotic drive strategy is good since the refractory allele is to be attached to the X-linked response allele, and hence is likely to be lost following an accidental release. However, if we are also concerned about the unknown effects of the drive gene on the mosquito population, then this strategy is worrying since the drive gene is more than 50% likely to persist for escapes of more than two mosquitoes carrying the drive gene.

Medea and *Wolbachia* provide a good compromise between invasiveness following an intentional release and containment following an accidental release. Both systems will spread very quickly following an intentional release (Chen et al., 2007; Turelli and Hoffmann, 1999); however, they require either a fitness benefit or population growth in order to spread following an accidental release. The conditions for spread are a little more restrictive for *Wolbachia*, primarily because maternal transmission of *Wolbachia* is imperfect.

Despite this, we should still be wary of the ability of *Medea* and *Wolbachia* to persist in a wild mosquito population following an

accidental release. It is feasible that *Medea* or *Wolbachia*-infected mosquitoes could possess a slight fitness advantage over wild-type mosquitoes. Under such conditions, an advantageous *Medea* allele is more than 50% likely to persist for homozygote escape sizes greater than 27, while an advantageous *Wolbachia* bacterium is more than 50% likely to persist for escape sizes greater than 74. The concern that we should be wary of is that the existence of a fitness advantage under natural conditions will not be known until experiments have been carried out in ambient field cages. During these experiments, an escape of this size is entirely within the realm of possibility.

An additional concern is that *Medea* and *Wolbachia* may be more invasive following an accidental release than predicted by simple models of population dynamics. The observation of *Medea* alleles and *Wolbachia* bacteria in nature (Wade and Beeman, 1994; Stevens and Wade, 1990) suggests that both have been able to spread from very low prevalences. The existence of local population structure has not been accounted for in the branching processes in this paper; however, it may help to enable the establishment of a new *Wolbachia* infection, *Medea* allele or drive-insensitive response allele for a meiotic drive system from a very low initial prevalence.

Despite these concerns, *Medea* and *Wolbachia* still represent the best compromise between invasiveness and containment for the six gene drive systems currently being considered (Sinkins and Gould, 2006). The mechanism of gene drive should be thought of as a form of biological containment in the design and implementation of ambient cage trials. This is only one of many considerations in the design of such experiments; however, it is an important additional consideration to the comprehensive list compiled by Benedict et al. (2008). Given the lack of knowledge of the outcomes of such technology, all efforts should be taken to prevent the release of transgenic strains before their efficacy and side-effects have been adequately studied.

Acknowledgments

I am particularly grateful to Prof. Charles Taylor, Prof. Ken Lange and Dr. Mahamadou Touré for discussions and comments. I am also grateful to Prof. Jason Rasgon, Prof. Fred Gould, Prof. Steve Sinkins, Prof. Michael Turelli, Prof. Austin Burt and Dr. Yunxin Huang for advice on specific gene drive strategies; and to a panel of anonymous reviewers whose constructive comments have improved the manuscript. This research was supported by Grant number 1R56AI072549 from the National Institutes of Health.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at 10.1016/j.jtbi.2009.01.031.

References

Ahmed, A.M., Baggott, S.L., Maingon, R., Hurd, H., 2002. The costs of mounting an immune response are reflected in the reproductive fitness of the mosquito *Anopheles gambiae*. Oikos 97, 371–377.

Alphey, L., Beard, C.B., Billingsley, P., Coetze, M., Crisanti, A., et al., 2002. Malaria control with genetically manipulated insect vectors. Science 298, 119–121.

Beier, J.C., Killeen, G.F., Githure, J.I., 1999. Short report: entomologic inoculation rates and *Plasmodium falciparum* malaria prevalence in Africa. Am. J. Trop. Med. Hyg. 61, 109–113.

Benedict, M., D'Abbs, P., Dobson, S., Gottlieb, M., Harrington, L., et al., 2008. Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group. Vector-Borne Zoonotic Dis. 8, 127–166.

Cha, S.J., Mori, A., Chadee, D.D., Severson, D.W., 2006a. Cage trials using an endogenous meiotic drive gene in the mosquito *Aedes aegypti* to promote population replacement. Am. J. Trop. Med. Hyg. 74, 62–68.

Cha, S.J., Chadee, D.D., Severson, D.W., 2006b. Population dynamics of an endogenous meiotic drive system in *Aedes aegypti* in Trinidad. Am. J. Trop. Med. Hyg. 75, 70–77.

Charlat, S., Ballard, J.W.O., Mercot, H., 2004. What maintains noncytoplasmic incompatibility inducing *Wolbachia* in their hosts: a case study from a natural *Drosophila yakuba* population. J. Evol. Biol. 17, 322–330.

Charlesworth, B., 1991. Transposable elements in natural populations with a mixture of selected and neutral sites. Genet. Res. 57, 127–134.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215–220.

Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., et al., 2007. A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. Science 316, 597–600.

Craig, G.B., 1963. Prospects for vector control through manipulation of populations. Bull. World Health Organ. 29, 89–97.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. Harper and Row, New York.

Crow, J.F., Simmons, M.J., 1983. The mutation load in *Drosophila*. In: Ashburner, M., Carson, H.L., Thompson, J.N. (Eds.), The Genetics and Biology of Drosophila, vol. 3c. Academic Press, London, pp. 1–35.

Curtis, C.F., 1968. Possible use of translocations to fix desirable genes in insect pest populations. Nature 218, 368–369.

Davis, S.A., Bax, N., Grewe, P., 2001. Engineered underdominance allows efficient and economical introgression of traits into pest populations. J. Theor. Biol. 212, 83–98.

Dean, M.D., 2006. A *Wolbachia*-associated fitness benefit depends on genetic background in *Drosophila simulans*. Proc. R. Soc. B 273, 1415–1420.

Deredec, A., Burt, A., Godfray, C., 2007. Homing endonuclease genes: new tools for population engineering and control. Selfish DNA Meeting, Durham, North Carolina.

Dobson, S.L., Marsland, E.J., Rattanadechakul, W., 2002. Mutualistic *Wolbachia* infection in *Aedes albopictus*: accelerating cytoplasmic drive. Genetics 160, 1087–1094.

Dorman, K.S., Sinsheimer, J.S., Lange, K., 2004. In the garden of branching processes. SIAM Rev. 46, 202–229.

Engels, W.R., 1989. *P* elements in *Drosophila melanogaster*. In: Berg, D.E., How, M.M. (Eds.), Mobile DNA. ASM Press, Washington, DC, pp. 439–484.

Fisher, R.A., 1922. On the dominance ratio. Proc. R. Soc. Edinburgh 42, 321–341.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection. Part V: selection and mutation. Proc. Cambridge Philos. Soc. 23, 838–844.

Harris, T., 1963. The theory of branching processes. Springer, Berlin.

Hartl, D.L., Clark, A.G., 1989. Principles of Population Genetics. Sinauer Associates, Sunderland, Massachusettes.

Hickey, W.A., Craig, G.B., 1966. Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. Genetics 53, 1177–1196.

Hoffmann, A.A., Turelli, M., Harshman, L.G., 1990. Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. Genetics 126, 933–948.

Hoffmann, A.A., Clancy, D.J., Merton, E., 1994. Cytoplasmic incompatibility in Australian populations of *Drosophila melanogaster*. Genetics 136, 993–999.

Hoffmann, A.A., Hercus, M., Dagher, H., 1998. Population dynamics of the *Wolbachia* infection causing cytoplasmic incompatibility in *Drosophila melanogaster*. Genetics 136, 993–999.

Huang, Y., Magori, K., Lloyd, A.L., Gould, F., 2007. Introducing desirable transgenes into insect populations using Y-linked meiotic drive—a theoretical assessment. Evolution 61, 717–726.

James, A.A., 2005. Gene drive systems in mosquitoes: rules of the road. Trends Parasitol. 21, 64–67.

Koella, J.C., Boëte, C., 2002. A genetic correlation between age at pupation and melanization immune response of the yellow fever mosquito *Aedes aegypti*. Evolution 56, 1074–1079.

Kraaijeveld, A.R., Godfray, H.C., 1997. Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. Nature 389, 278–280.

Lange, K., 2002. Applied Probability. Springer, New York.

Lanzaro, G., Touré, Y., Carnahan, J., Zheng, L., Dolo, G., et al., 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in West Africa as revealed by microsatellite DNA analysis. Proc. Natl. Acad. Sci. USA 95, 14260–14265.

Le Rouzic, A., Capy, P., 2006. Reversible introduction of transgenes in natural populations of insects. Insect Mol. Biol. 15, 227–234.

Little, T.W., 1991. Segregation distorters. Annu. Rev. Genet. 25, 511–557.

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

Mackay, T.F., Lyman, R.F., Jackson, M.S., 1992. Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics 130, 315–332.

Magori, K., Gould, F., 2006. Genetically engineered underdominance for manipulation of pest populations: a deterministic model. Genetics 172, 2613–2620.

Manoukis, N.C., 2006. Studies on the ecology and adaptation of *Anopholes gambiae* in Mali and their impacts on malaria transmission and control. Ph.D. Thesis, University of California, Los Angeles.

Marelli, M.T., Li, C., Rasgon, J.L., Jacobs-Lorena, M., 2007. Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. Proc. Natl. Acad. Sci. USA 104, 5580–5583.

Marshall, J.M., 2008a. The impact of dissociation on transposon-mediated disease control strategies. Genetics 178, 1673–1682.
Marshall, J.M., 2008b. A branching process model for the early spread of a transposable element in a diploid population. J. Math. Biol. 57, 811–840.
Maside, X., Assimacopoulos, S., Charlesworth, B., 2000. Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. Genet. Res. 75, 275–284.
Moreira, L.A., Wang, J., Collins, F.H., Jacobs-Lorena, M., 2004. Fitness of anopheline mosquitoes expressing transgenes that inhibit *Plasmodium* development. Genetics 166, 1337–1341.
Moret, Y., Schmid-Hempel, P., 2000. Survival for immunity: the price of immune system activation for bumblebee workers. Science 290, 1166–1168.
Mukai, T., Chigusa, S.I., Mettler, L.E., Crow, J.F., 1972. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics 72, 335–355.
Nuzhdin, S.V., Pasyukova, E.G., Mackay, T.F.C., 1997. Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. Genetica 100, 167–175.
Ohnishi, O., 1977. Spontaneous and ethyl methane-sulfonate-induced mutations controlling variability in *Drosophila melanogaster* I: recessive lethal mutations. Genetics 87, 335–348.
Rasgon, J.L., Gould, F., 2005. Transposable element insertion location bias and the dynamics of gene drive in mosquito populations. Insect Mol. Biol. 14, 493–500.
Rasgon, J.L., Scott, T.W., 2003. *Wolbachia* and cytoplasmic incompatibility in the California *Culex pipiens* mosquito species complex: parameter estimates and infection dynamics in natural populations. Genetics 165, 2029–2038.
Rong, Y.S., Golic, K.G., 2003. The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in *Drosophila*. Genetics 165, 1831–1842.
Scott, T.W., Takken, W., Knols, B.G., Boëte, C., 2002. The ecology of genetically modified mosquitoes. Science 298, 117–119.
Seleme, M., Busseau, I., Malinsky, S., Bucheton, A., Teninges, D., 1999. High-frequency retrotransposition of a marked *I* factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. Genetics 151, 761–771.
Sinkins, S.P., Godfray, H.C., 2004. Use of *Wolbachia* to drive nuclear transgenes through insect populations. Proc. Biol. Sci. 271, 1421–1426.
Sinkins, S.P., Gould, F., 2006. Gene drive systems for insect disease vectors. Nat. Rev. Genet. 7, 427–435.
Spiess, E.B., 1977. Genes in Populations. Wiley, New York.
Stevens, L., Wade, M.J., 1990. Cytoplasmically inherited reproductive incompatibility in *Tribolium* flour beetles: the rate of spread and effect on population size. Genetics 124, 367–372.
Stouthammer, R., Breeuwer, J.A., Hurst, G.D., 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. 53, 71–102.
Subramanian, R.A., Arensburger, P., Atkinson, P.W., O'Brochta, D.A., 2007. Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae* s.s. Genetics 176, 2477–2487.
Taylor, C.E., Manoukis, N.C., 2003. Effective population size in relation to genetic modification of Anopholes gambiae sensu stricto. In: Takken, W., Scott, T.W. (Eds.), Ecological Aspects for Application of Genetically Modified Mosquitoes. Wageningen, The Netherlands, pp. 133–146.
Taylor, C., Touré, Y.T., Carnahan, J., Norris, D.E., Dolo, G., et al., 2001. Gene flow among populations of the Malaria vector *Anopholes gambiae* in Mali, West Africa. Genetics 157, 743–750.
Touré, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., et al., 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopholes gambiae s. str.* in Mali, West Africa. Genetica 94, 213–223.
Townsend, J.P., Hartl, D.L., 2000. The kinetics of transposable element autoregulation. Genetica 108, 229–237.
Tripet, F., Dolo, G., Lanzaro, G.C., 2005. Multilevel analyses of genetic differentiation in *Anopholes gambiae s.s.* reveal patterns of gene flow important for malaria-fighting mosquito projects. Genetics 169, 313–324.
Turelli, M., Hoffmann, A.A., McKechnie, S.W., 1992. Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics 132, 713–723.
Turelli, M., Hoffmann, A.A., 1995. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics 140, 1319–1338.
Turelli, M., Hoffmann, A.A., 1999. Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. Insect Mol. Biol. 8, 243–255.
Vasilyeva, L.A., Bubenshchikova, E.V., Ratner, V.A., 1999. Heavy heat shock induced retrotransposon transposition in *Drosophila*. Genet. Res. 74, 111–119.
Wade, M.J., Beeman, R.W., 1994. The population dynamics of maternal-effect selfish genes. Genetics 138, 1309–1314.
Weeks, A.R., Turelli, M., Harcombe, W.R., Reynolds, K.T., Hoffmann, A.A., 2007. From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. PLoS Biol. 5, e114.
Weinreich, M.D., Gasch, A., Reznikoff, W.S., 1994. Evidence that the *cis* preference of the *Tn*5 transposase is caused by nonproductive multimerization. Genet. Dev. 8, 2363–2374.
Windbichler, N., Papathanos, P.A., Catteruccia, F., Ranson, H., Burt, A., et al., 2007. Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. Nucleic Acids Res. 35, 5922–5933.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.
Wu, C.T., Morris, J.R., 1999. Transvection and other homology effects. Curr. Opin. Genet. Dev. 9, 237–246.