DOI: 10.1534/genetics.110.124479

Semele: A Killer-Male, Rescue-Female System for Suppression and Replacement of Insect Disease Vector Populations

John M. Marshall,*,†,[1] Geoffrey W. Pittman,† Anna B. Buchman† and Bruce A. Hay†

**Department of Infectious Disease Epidemiology, Imperial College London, London, W2 1PG, United Kingdom and* †*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Manuscript received October 21, 2010
Accepted for publication November 10, 2010

ABSTRACT

Two strategies to control mosquito-borne diseases, such as malaria and dengue fever, are reducing mosquito population sizes or replacing populations with disease-refractory varieties. We propose a genetic system, *Semele*, which may be used for both. *Semele* consists of two components: a toxin expressed in transgenic males that either kills or renders infertile wild-type female recipients and an antidote expressed in females that protects them from the effects of the toxin. An all-male release results in population suppression because wild-type females that mate with transgenic males produce no offspring. A release that includes transgenic females results in gene drive since females carrying the allele are favored at high population frequencies. We use simple population genetic models to explore the utility of the *Semele* system. We find that *Semele* can spread under a wide range of conditions, all of which require a high introduction frequency. This feature is desirable since transgenic insects released accidentally are unlikely to persist, transgenic insects released intentionally can be spatially confined, and the element can be removed from a population through sustained release of wild-type insects. We examine potential barriers to *Semele* gene drive and suggest molecular tools that could be used to build the *Semele* system.

MOSQUITO-BORNE diseases such as malaria and dengue fever continue to pose a major health problem through much of the world. The goal of the Roll Back Malaria Initiative to halve malaria deaths by 2010 was not successful even in reducing malaria deaths (Shiff 2000; World Health Organization 2009), and a treatment for dengue fever still remains elusive. The failure of existing methods to control these diseases has renewed interest in approaches to disease prevention that involve the use of genetically modified mosquitoes (Braig and Yan 2001; Alphey *et al.* 2002; Sinkins and Gould 2006; Marshall and Taylor 2009).

There are two main strategies being considered to control vector-borne diseases using transgenic vectors. The first involves the release of genetically modified males that will mate with wild females and produce unviable offspring (Whitten and Foster 1975; Alphey *et al.* 2002; Dyck *et al.* 2005; Catteruccia *et al.* 2009). This is a genetic version of the sterile insect technique and is intended to dramatically reduce the vector population size and consequently reduce disease transmission. The technology for this strategy has already been developed for *Aedes aegypti*—the main vector of dengue fever—and preparations are currently being made for an environmental release (Vasan 2009).

The second strategy for disease prevention is to replace entire populations of mosquitoes with varieties that are refractory to disease transmission. A variety of genes conferring disease refractoriness have been identified in nature and engineered in the laboratory. For example, with respect to malaria, Ito *et al.* (2002) engineered a gene that saturates the receptor sites that the malaria parasite requires to pass through the mosquito gut following ingestion; de Lara Capurro *et al.* (2000) developed antibodies that kill malaria parasites; Riehle *et al.* (2006) discovered genes that govern refractoriness in natural populations; and Corby-Harris *et al.* (2010) activated a signaling pathway that dramatically reduces both parasite development and mosquito longevity. Expression of RNAs that induce RNA interference targeting dengue virus has also been shown to reduce dengue transmission (Franz *et al.* 2006). Mosquitoes carrying genes that mediate disease refractoriness are not expected to experience a fitness benefit in both the presence and the absence of infection (Lambrechts *et al.* 2008) and may well experience a cost (Schmid-Hempel 2005). Given that a very high fraction of mosquitoes must be disease refractory to achieve significant levels of disease protection (Boete and Koella 2002, 2003), it is generally thought that population replacement will require that genes mediating disease refractoriness be linked to a

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.110.124479/DC1.

Available freely online through the author-supported open access option.

[1]*Corresponding author:* Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London W2 1PG, United Kingdom. E-mail: john.marshall@imperial.ac.uk

genetic system capable of driving them into the population (Braig and Yan 2001; James 2005; Sinkins and Gould 2006).

A number of gene drive systems have been proposed, including naturally occurring selfish genetic elements such as transposons, B chromosomes, meiotic drive, *Medea* elements, homing endonuclease genes, and the intracellular bacterium Wolbachia. Another set of approaches to bringing about population replacement involves creating insects in which genes of interest are linked to engineered chromosomes: compound chromosomes or translocations (Curtis 1968; Foster *et al.* 1972; Gould and Schliekelman 2004) or pairs of unlinked lethal genes, each of which is associated with a repressor of the lethality induced by expression of the other lethal gene—a system known as engineered underdominance (Davis *et al.* 2001; Magori and Gould 2006).

A synthetic version of the *Medea* drive system was recently created and observed to spread rapidly through laboratory populations of *Drosophila melanogaster* (Chen *et al.* 2007). The ability of *Medea* to spread and the rate at which it spreads are a function of its fitness cost and introduction frequency. *Medea* elements with large fitness costs are expected to require high introduction frequencies to spread; but elements with small fitness costs are expected to spread from very low frequencies, particularly in real populations where population structure and stochastic effects become relevant. These features make *Medea* an interesting system for large-scale population replacement with a fully characterized arsenal of antipathogen genes.

In the early stages of testing, particularly in the field, it would be desirable to have gene drive systems that are either self-limiting or unlikely to spread following an accidental release (Benedict and Robinson 2003; Benedict *et al.* 2008; Marshall 2009). Candidates include a system known as killer rescue, which is designed to spread a linked transgene locally for a limited period of time before falling out of the population (Gould *et al.* 2008), and engineered underdominance, which requires a population frequency of >67% for a single-locus system and 27% for a two-locus system to spread (Davis *et al.* 2001). Neither of these systems has been implemented to date. Gene drive systems with high release thresholds are desirable since they may be confined to single populations or nearby populations exchanging large numbers of migrants with each other. This is an important property during open field trials, which must ultimately occur to test the efficacy of these systems. A high release threshold also creates a mechanism for removing the drive system and antipathogen genes from the population through large-scale release of wild-type insects, thus diluting the drive system to subthreshold frequencies.

Here, we describe a genetic system, *Semele*, which may be used for both suppression and replacement of

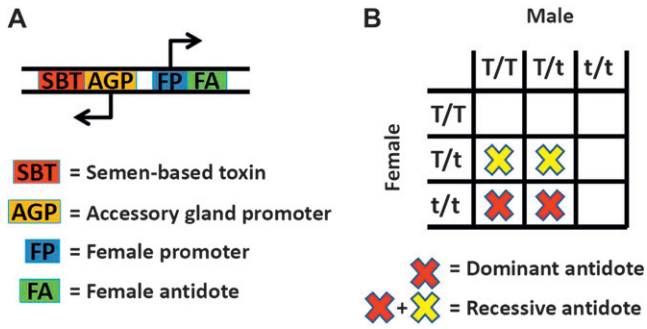

Figure 1.—Schematic diagram of a *Semele* element. (A) The element is composed of two genes: a toxin expressed in the accessory glands and an antidote expressed in female somatic tissues or a toxin expressed in the male germ line and an antidote expressed in the female germ line for deposition in the egg. (B) Crosses between transgenic males and wild-type females result in loss of progeny or adult female death because wild-type females do not express the antidote to the male's toxic ejaculate. If the antidote is recessive (requiring inheritance of two copies of the *Semele* element to function), then only homozygous females or their progeny are protected from the toxin.

disease vector populations. The system consists of two components—a toxin expressed in the reproductive system of transgenic males and an antidote expressed in females that protects them from being killed or rendered infertile following mating with a transgenic male (Figure 1). We name this system after the mortal female in Greek mythology, *Semele* (pronounced "Sem-uh-lee"), who was impregnated by Zeus but later died after witnessing his godliness because she was not herself a god. The name also stands for semen-based lethality.

A number of approaches may be taken for engineering the *Semele* system, none of which have been realized to date. The toxin could be expressed in the premeiotic germ line, with the product being shared by all haploid products of meiosis. This approach has been proposed in the context of utilizing genes that mediate cytoplasmic incompatibility (CI) in the intracellular bacterium Wolbachia (Sinkins *et al.* 1997; Turelli and Hoffmann 1999). Alternatively, the toxin could be expressed in somatic cells that synthesize components of semen transferred to the female upon mating (Gould 2007). Upon mating with a wild-type female, the ejaculate either renders the female infertile through the action of a toxin released by sperm that kills fertilized eggs or kills her through the action of a semen-associated toxin on, for example, her nervous system. The second component of this system is an antidote expressed in females that protects them from being rendered infertile or killed following mating with a transgenic male (Figure 1). A release of purely transgenic males results in population suppression, since wild females who mate with transgenic males produce no offspring. A release that includes transgenic females results in gene drive under permissive conditions since females having the *Semele*

allele are favored due to their immunity to the toxic ejaculate of transgenic males.

Importantly, the *Semele* drive system has a release threshold. At low population frequencies, transgenic males are disadvantaged when they mate with wild-type females because these crosses produce no offspring. However, this disadvantage diminishes at high population frequencies, when element-bearing females are common. In short, when the advantage conferred on transgenic females outweighs the disadvantage conferred on transgenic males, *Semele* is driven into the population.

Here, we present simple one- and two-locus genetic models that describe the population dynamics of the *Semele* drive system. We explore a range of parameters over which *Semele* can function as a gene drive system, including element-associated fitness costs and efficiency of toxin action. We also explore several system variants, X-linked alleles, and recessive antidotes. Finally, we explore the severity of potential barriers to spread such as the prior existence of an allele conferring toxin resistance in the population and assortative mating. In conclusion, we discuss ways in which the *Semele* drive mechanism could be engineered.

MODEL DEVELOPMENT

We use discrete-generation difference equations to model the population dynamics of the *Semele* system and its variants. We consider the *Semele* element as a single allele, which we denote as "*T*". This allele carries a minimum of two genes: one that encodes the toxin expressed in males and a second that encodes an antidote expressed in females. By placing the toxin gene within an intron of the antidote gene, these genes may be considered inextricably linked, since any product of breakage and rejoining will lack a functional antidote and will therefore be rendered unviable (Chen *et al.* 2007). We refer to the corresponding position on the wild-type chromosome as "*t*". In the simplest form of our model, we consider the case of a dominant toxin and a dominant antidote. In this case, only one copy of the toxin gene is sufficient to kill susceptible females and one copy of the antidote gene is sufficient to neutralize the toxin in a transgenic female, independent of whether her mating partner has one or two copies of the toxin gene.

A one-locus model can then be used for the discrete generation dynamics. For the deterministic version of this model, we assume random mating and an infinite population size. We account for a fitness cost due to having one or two copies of the allele and allow this cost to differ in males and females. This gender specificity is incorporated to account for the fact that a toxin expressed in males (which is designed to kill) is likely to be more costly than an antidote (whose only function is to inhibit the toxin) expressed in females. We also account for the possibility that toxin efficiency is <100% and allow for an unequal gender ratio at the time of release.

The mathematical formulation of this model is as follows: the proportions of the *k*th generation that are males of genotypes *tt*, *Tt*, and *TT* are denoted by $u_{\mathrm{m},k}$, $v_{\mathrm{m},k}$, and $w_{\mathrm{m},k}$, respectively. The corresponding proportions for females are $u_{\mathrm{f},k}$, $v_{\mathrm{f},k}$, and $w_{\mathrm{f},k}$. By considering all possible mating pairs, the genotypes of embryos in the next generation are described by the ratio $u_{\mathrm{e},k+1} : v_{\mathrm{e},k+1} : w_{\mathrm{e},k+1}$, where

$$u_{\mathrm{e},k+1} = u_{\mathrm{m},k}u_{\mathrm{f},k} + 0.5u_{\mathrm{m},k}v_{\mathrm{f},k} + 0.25v_{\mathrm{m},k}v_{\mathrm{f},k} + 0.5v_{\mathrm{m},k}u_{\mathrm{f},k}(1 - e_{\mathrm{T}}) \tag{1}$$

$$v_{\mathrm{e},k+1} = u_{\mathrm{m},k}w_{\mathrm{f},k} + 0.5u_{\mathrm{m},k}v_{\mathrm{f},k} + 0.5v_{\mathrm{m},k}v_{\mathrm{f},k} + 0.5v_{\mathrm{m},k}w_{\mathrm{f},k} + 0.5w_{\mathrm{m},k}v_{\mathrm{f},k} + 0.5v_{\mathrm{m},k}u_{\mathrm{f},k}(1 - e_{\mathrm{T}}) + w_{\mathrm{m},k}u_{\mathrm{f},k}(1 - e_{\mathrm{T}}) \tag{2}$$

$$w_{\mathrm{e},k+1} = w_{\mathrm{m},k}w_{\mathrm{f},k} + 0.5w_{\mathrm{m},k}v_{\mathrm{f},k} + 0.5v_{\mathrm{m},k}w_{\mathrm{f},k} + 0.25v_{\mathrm{m},k}v_{\mathrm{f},k}. \tag{3}$$

Here, e_{T} denotes toxin efficiency, which is equal to the probability that, when a transgenic male mates with a wild-type female, the female will either die or be rendered infertile due to the toxin. In our basic model, we assume that toxin efficiency is the same in heterozygous and homozygous transgenic males and that antidote efficiency is 100% in transgenic females. Later, we relax both of these assumptions. Genotype frequencies in the next generation are then given by

$$u_{\mathrm{m},k+1} = \frac{0.5u_{\mathrm{e},k+1}}{W_{k+1}} \tag{4}$$

$$v_{\mathrm{m},k+1} = \frac{0.5v_{\mathrm{e},k+1}(1 - h_{\mathrm{m}}s_{\mathrm{m}})}{W_{k+1}} \tag{5}$$

$$w_{\mathrm{m},k+1} = \frac{0.5w_{\mathrm{e},k+1}(1 - s_{\mathrm{m}})}{W_{k+1}} \tag{6}$$

$$u_{\mathrm{f},k+1} = \frac{0.5u_{\mathrm{e},k+1}}{W_{k+1}} \tag{7}$$

$$v_{\mathrm{f},k+1} = \frac{0.5v_{\mathrm{e},k+1}(1 - h_{\mathrm{f}}s_{\mathrm{f}})}{W_{k+1}} \tag{8}$$

$$w_{\mathrm{f},k+1} = \frac{0.5w_{\mathrm{e},k+1}(1 - s_{\mathrm{f}})}{W_{k+1}}. \tag{9}$$

Here, s_{m} and $h_{\mathrm{m}}s_{\mathrm{m}}$ represent the fitness costs for males that are homozygous and heterozygous for the *Semele* allele, respectively; and s_{f} and $h_{\mathrm{f}}s_{\mathrm{f}}$ represent the equivalent fitness costs for females. The normalizing term, W_{k+1}, is given by

$$W_{k+1} = u_{\mathrm{e},k+1} + v_{\mathrm{e},k+1}(1 - 0.5h_{\mathrm{m}}s_{\mathrm{m}} - 0.5h_{\mathrm{f}}s_{\mathrm{f}}) + w_{\mathrm{e},k+1}(1 - 0.5s_{\mathrm{m}} - 0.5s_{\mathrm{f}}). \tag{10}$$

We consider a release of *TT* individuals in which both the release size and the gender ratio may be varied.

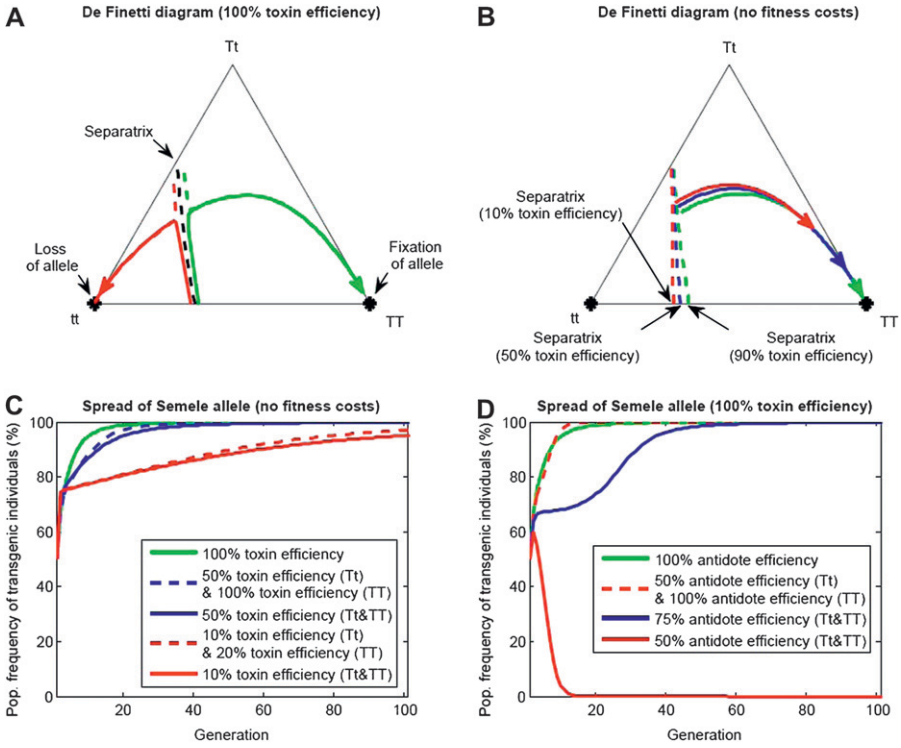

FIGURE 2.—Population dynamics of a *Semele* element with no fitness cost. (A) De Finetti diagram for the case of 100% toxin efficiency. A family of threshold points (separatrix) exists, above which the construct is fixed and below which the construct is lost. (B) De Finetti diagram showing separatrices for a variety of toxin efficiencies. (C) Time-series dynamics for the element in B incorporating additive and nonadditive toxin efficiencies. Reducing toxicity leads to much slower rates of allele spread. (D) Time-series dynamics for the element in B incorporating additive and nonadditive antidote efficiencies. For the nonadditive case, reducing antidote efficiency increases the release threshold and leads to slower rates of allele spread.

Considering a release at generation 0, the initial condition for the difference equations is given by

$$w_{m,0} = rw_0 \tag{11}$$

$$w_{f,0} = (1 - r)w_0 \tag{12}$$

$$u_{m,0} = u_{f,0} = 0.5(1 - w_0). \tag{13}$$

Here, the released individuals represent a proportion, w_0, of the total population, and a fraction, r, of the released individuals are male. Using this initial condition and the difference equations described above, the equilibria, thresholds, and time-series dynamics of the *Semele* system can be calculated.

All of the models in this article are based on this simple one-locus model. Using this model, we calculate the optimal gender ratio for a release, fitness cost effects, and the degree of toxin inefficiency that can be tolerated for the system to still drive. We develop a stochastic version of the model to estimate loss probabilities and expected times to loss or fixation in a finite population. Additionally, we adapt the model to account for a continuous release. This variant of the model has implications for spread of the allele into secondary populations. Two variants of the *Semele* system are also considered. First, we consider the case of a dominant toxin and a recessive antidote. In this case, two copies of the antidote gene are required to neutralize the toxin in a transgenic female. Second, we consider the case of a *Semele* element located on the X chromosome rather than on an autosome.

Finally, we explore some of the barriers that may prevent the *Semele* system from working. First, we develop a two-locus model for the prior existence of toxin resistance in the population. We consider a natural toxin-resistance allele, denoted by "*R*", which is unlinked to the *Semele* allele and has a prior equilibrium frequency in the population. Second, we develop a two-locus model to account for assortative mating and its implications for gene drive. In this model, an unlinked allele causes laboratory-reared mosquitoes to be less appealing to wild mosquitoes. Such an allele, which we denote as "*A*", could be considered the product of laboratory inbreeding.

All of these models apply to the gene drive application of the *Semele* system. We focused on gene drive because the genetic version of the sterile insect technique has already been modeled (PHUC *et al.* 2007) and the dynamics of a male-only release of *Semele* are expected to be very similar. That said, we use the continuous release model to investigate the effect a few stray transgenic females would have on an all-male transgenic release intended for population suppression.

RESULTS

No fitness costs: We begin by considering an autosomal *Semele* element with a dominant toxin produced by males and a dominant antidote produced by mature females. First, we consider the case where there is no fitness cost associated with the *Semele* allele. All mating pairs produce equal numbers of male and female

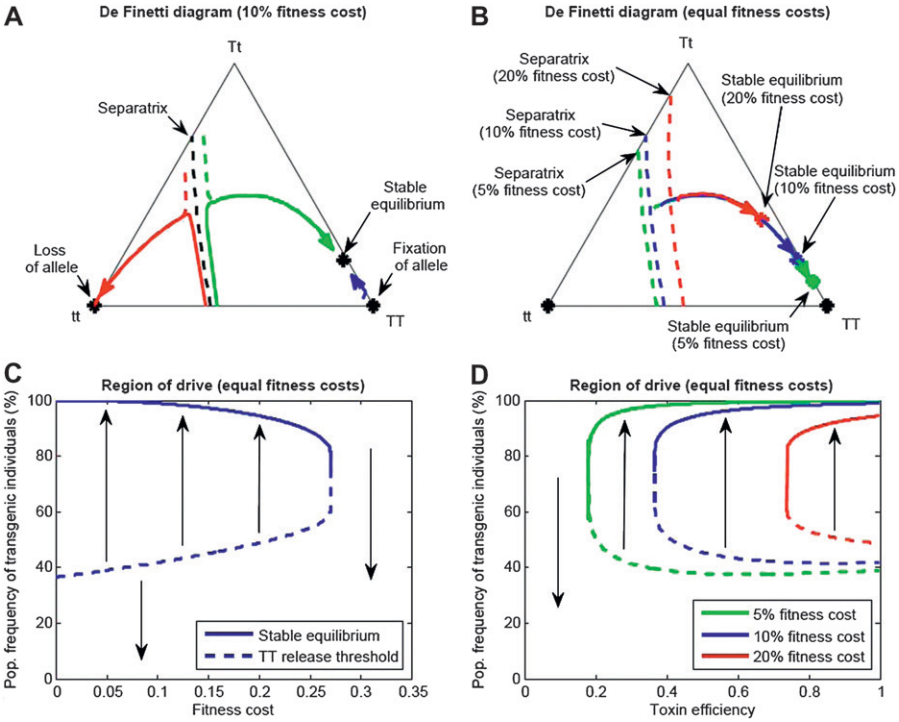

FIGURE 3.—Population dynamics of a *Semele* element with equal fitness costs in males and females. (A) De Finetti diagram for the case of a 10% fitness cost. Above the release threshold, the allele approaches a stable equilibrium consisting mostly of heterozygotes and homozygotes. (B) De Finetti diagram showing separatrices and stable equilibria for a variety of fitness costs. (C) Region of drive as a function of fitness cost. The region of drive is the set of population frequencies above the release threshold and below the stable transgenic equilibrium between which the *Semele* element will increase in frequency up to the stable equilibrium. (D) Region of drive as a function of toxin efficiency for a variety of fitness costs.

offspring and so, even if the gender ratio is initially unequal, the genotype distribution among males and females will be identical from the second generation on. The proportions of the kth generation that are individuals of genotypes *tt*, *Tt*, and *TT* may then be denoted by u_k, v_k, and w_k. The following simplified form of Equations 1–10 then applies,

$$u_{k+1} = \frac{u_k^2 + 0.25v_k^2 + u_k v_k(1 - 0.5e_T)}{W_{k+1}} \tag{14}$$

$$v_{k+1} = \frac{0.5v_k^2 + v_k w_k + u_k v_k(1 - 0.5e_T) + u_k w_k(2 - e_T)}{W_{k+1}} \tag{15}$$

$$w_{k+1} = \frac{w_k^2 + 0.25v_k^2 + v_k w_k}{W_{k+1}}, \tag{16}$$

where W_{k+1} is the normalizing term defined in Equation 10. This system has three biologically feasible equilibrium points:

$$(u_*, w_*) = (0,1), (1,0), \left(0.5, 0.5(3 - e_T - \sqrt{e_T^2 - 6e_T + 8})\right). \tag{17}$$

The first of these points represents allele fixation, the second represents absence or loss of the *Semele* allele, and the third represents coexistence of wild, heterozygous, and homozygous individuals in the population. We calculate the stabilities of these points in supporting information, File S1. Our analysis shows that both fixation and loss of the element are represented by stable equilibrium points and the intermediate point is unstable.

The location and stability of the three equilibrium points suggest the existence of a threshold, above which the element becomes fixed and below which the element is lost. Mapping genotype trajectories onto a De Finetti diagram, we see that there is a family of points that act as a threshold between loss and fixation (Figure 2A). This family of points includes the third equilibrium point in Equation 17 and is referred to as a separatrix. The fact that an unstable equilibrium point exists for all values of $e_T > 0$ suggests that, in the absence of fitness costs, even a minimally efficient toxin can facilitate gene drive (Figure 2B). The case where $e_T = 0$ corresponds to Hardy–Weinberg equilibrium.

Figure 2C depicts time-series dynamics for *Semele* alleles having no fitness costs and a variety of toxin efficiencies. Transgenic males that express a less-efficient toxin have more offspring and consequently require a smaller release frequency to spread. However, elements with reduced toxicity cause allele frequencies to change less quickly, leading to much weaker drive. For example, a release of *TT* males and females at a population frequency of 40% is expected to result in fixation of the *T* allele regardless of toxin efficiency. For a 100%-efficient toxin, wild-type individuals are predicted to fall below a population frequency of 1% within 22 generations; however, the same reduction is expected to take 230 generations if the toxin is 10% efficient.

In the above analysis, we assumed that toxin efficiency is the same in heterozygous and homozygous transgenic males; however, this assumption can be relaxed by replacing Equation 15 above with the following modified equation:

$$v_{k+1} = \frac{0.5v_k^2 + v_k w_k + u_k v_k(1 - 0.5e_T) + u_k w_k(2 - \min\{1,\ 2e_T\})}{W_{k+1}}. \tag{18}$$

Here, e_T represents toxin efficiency in heterozygous males; while toxin efficiency in homozygous transgenic males is either 100% or $2e_T$, whichever is smaller. Figure 2C also depicts the time-series dynamics for *Semele* alleles having additive toxin efficiencies. The dynamics are very similar to those for nonadditive toxin efficiencies; however, additive toxin efficiencies lead to slightly faster gene drive. For a 40% release and 10% toxin efficiency, wild-type individuals are predicted to fall below a population frequency of 1% within 160 generations, as opposed to 230 generations for the nonadditive case.

Figure 2D depicts time-series dynamics for *Semele* alleles having a variety of antidote efficiencies. For *Semele* alleles having additive antidote efficiencies, the following modified form of Equations 14–16 applies:

$$u_{k+1} = \frac{u_k^2 + 0.25v_k^2 e_A + u_k v_k}{W_{k+1}} \tag{19}$$

$$v_{k+1} = \frac{0.5v_k^2 e_A + 0.5v_k w_k(e_A + \min\{1, 2e_A\}) + u_k v_k + u_k w_k}{W_{k+1}} \tag{20}$$

$$w_{k+1} = \frac{w_k^2 \min\{1, 2e_A\} + 0.25v_k^2 e_A + 0.5v_k w_k(e_A + \min\{1,\ 2e_A\})}{W_{k+1}}. \tag{21}$$

Here, e_A represents antidote efficiency in heterozygous females, which is equal to the probability that, when a transgenic male mates with a heterozygous female, the female will survive and retain fertility. Antidote efficiency in homozygous transgenic females is either 100% or $2e_A$, whichever is smaller. For the case of nonadditive antidote efficiencies, this latter quantity is simply equal to e_A. For the additive case, changing antidote efficiency between 50 and 100% has very little effect on the spread of the *Semele* allele—in all cases, nearly all individuals are transgenic within ~15 generations. For the nonadditive case, decreasing antidote efficiency leads to higher release thresholds and slower gene drive. For 75% antidote efficiency, wild-type individuals fall below 1% within 50 generations of a 50% release, as opposed to 20 generations for the case of 100% antidote efficiency. For 50% antidote efficiency, a 50% release is insufficient to achieve gene drive.

Equal fitness costs: Next, we consider the case where there is an equal fitness cost associated with the element in both females and males ($s_f = s_m$). As in the previous case, all mating pairs produce equal numbers of male and female offspring and so the genotype distribution is identical among males and females from the second generation on. The following simplified form of Equations 1–10 applies:

$$u_{k+1} = \frac{u_k^2 + 0.25v_k^2 + u_k v_k(1 - 0.5e_T)}{W_{k+1}} \tag{22}$$

$$v_{k+1} = \frac{(0.5v_k^2 + v_k w_k + u_k v_k(1 - 0.5e_T) + u_k w_k(2 - e_T))(1 - hs)}{W_{k+1}} \tag{23}$$

$$w_{k+1} = \frac{(w_k^2 + 0.25v_k^2 + v_k w_k)(1 - s)}{W_{k+1}}. \tag{24}$$

Here, s and hs represent the fitness costs for individuals that are homozygous and heterozygous for the *Semele* allele, respectively. For nonzero fitness costs, this system has four biologically feasible equilibrium points. Two of these correspond to fixation and loss of the *Semele* allele. The other two represent the coexistence of wild, heterozygous, and homozygous individuals and have expressions too complex to be useful, even if simplifications are made such as 100% toxin efficiency ($e_T = 1$). We calculate the stabilities of fixation and loss of the *Semele* allele in File S1. Our analysis shows that fixation is unstable for a fitness cost that is recessive or shows any degree of heterozygosity, but is stable for a completely dominant fitness cost. Loss is stable under all scenarios.

Observation of time-series data together with the location and stability of the four equilibrium points suggests the existence of a separatrix—or family of threshold points—above which the element reaches a nontrivial equilibrium frequency in the population and below which the element is lost (Figure 3, A and B). Fixation is not a realistic equilibrium since any perturbation will bring the *Semele* allele back to the nontrivial equilibrium frequency and any real population is subject to perturbations.

As the fitness cost on the element increases, the threshold required for spread increases and the nontrivial equilibrium reached decreases, reducing the impact of drive from both sides (Figure 3C). To illustrate this, for an element having a 10% fitness cost in homozygotes, *TT* males and females must be released at a frequency >41.7% for transgenic individuals to spread to a population frequency of 98.9%. For a 20% fitness cost, the release threshold increases to 48.8%, above which transgenic individuals spread to a population frequency of 94.4%. The maximum tolerable fitness cost is 27%, at which point the release threshold and the nontrivial equilibrium are identical and drive does not occur. For fitness costs >27%, the *Semele* allele is lost for all cases other than initial fixation.

In the presence of a fitness cost, decreasing toxin efficiency also has the effect of reducing the impact of drive (Figure 3D). The major effect is that a less efficient toxin causes the allele to reach a lower equilibrium

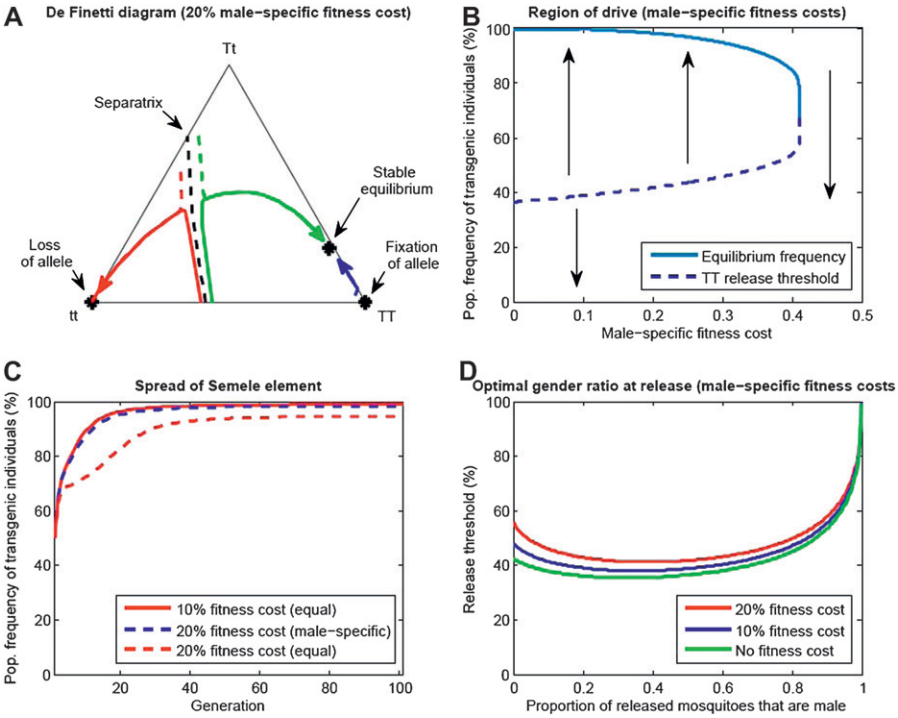

FIGURE 4.—Population dynamics of a *Semele* element with male-specific fitness costs. (A) De Finetti diagram for the case of a 20% male-specific fitness cost. (B) Region of drive as a function of male-specific fitness cost. (C) Time-series dynamics for constructs with 10% and 20% sex-independent and male-specific fitness costs. A construct with a 20% male-specific fitness cost behaves similarly to one with a 10% fitness cost expressed in males and females. (D) Release threshold as a function of release gender ratio. The threshold is minimized for a female-biased release; however, a release consisting of equal numbers of males and females is acceptable.

frequency in the population because the driving force is weakened relative to the fitness cost. The same effect causes the release threshold to increase as the toxin becomes less efficient; however, at the same time, more transgenic offspring survive at lower toxin efficiencies, causing the release threshold to decline. The result is a release threshold that declines as the toxin begins to become less efficient and then rises. The minimum toxin efficiency required for drive to occur depends on the fitness cost: for a 10% fitness cost, toxin efficiency must exceed 36.9%; and for a 20% fitness cost, toxin efficiency must exceed 74%. Below these efficiencies, the *Semele* allele is lost for all cases other than initial fixation. As before, reducing toxin efficiency greatly reduces the speed of spread.

Putting these results into perspective requires some idea of the real-world fitness costs that we could expect; however, since the *Semele* system has not yet been engineered, we are limited to estimates of fitness costs due to refractoriness and hypothetical considerations of the cost of a male-expressed toxin and a female-expressed antidote. Mounting an immune response is generally thought to be associated with an evolutionary cost in insects (KRAAIJEVELD and GODFRAY 1997; SCHMID-HEMPEL 2005). For example, AHMED *et al.* (2002) measured egg production to be reduced by 18.6% in *Anopheles gambiae* mosquitoes whose immune system was artificially stimulated with lipopolysaccharides. However, transgenic mosquitoes have also been engineered that have no noticeable fitness cost when fed on Plasmodium-free blood (MOREIRA *et al.* 2004) and have a 35–50% fitness advantage when fed on Plasmodium-infected blood (MARELLI *et al.* 2007). This fitness advantage would be much smaller in a real population in which only a fraction of mosquitoes are infected with malaria parasites. These observations give some idea of the large range of fitness costs that must be explored. As discussed earlier, a toxin expressed in the male accessory glands is likely to be much more costly for males than an antidote expressed specifically in females. For this reason, we explore the scenario of male-specific fitness costs in the following section.

Male-specific fitness costs: To model male-specific fitness costs, we use the basic model in Equations 1–13, with the one simplification that the *Semele* allele confers no fitness cost on females ($s_f = 0$). Symbolic analysis of these equations is too complex to be useful; however, the system can be easily numerically iterated (Figure 4).

Observation of time-series data reveals dynamics very similar to the case of equal fitness costs for males and females: a family of threshold points above which the allele reaches a nontrivial equilibrium frequency in the population and below which the allele is lost (Figure 4A). As the fitness cost on the element increases, the threshold required for spread increases and the nontrivial equilibrium reached decreases. The maximum tolerable fitness cost is 41%, at which point the release threshold and nontrivial equilibrium are identical (Figure 4B). An element with a 20% fitness cost on males spreads to a transgenic equilibrium frequency of 98.3% within 40 generations, which is very similar to an element conferring a 10% fitness cost on males and females (Figure 4C). An element with a 10% fitness cost on males spreads to a transgenic equilibrium frequency of 99.6% within 30 generations.

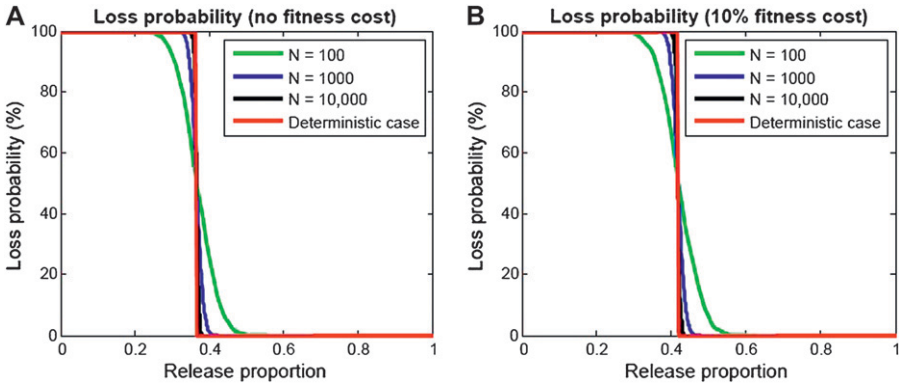

FIGURE 5.—Stochastic spread of *Semele* for population sizes of 100, 1000 and 10,000. (A) Loss probability as a function of release proportion for an element with no fitness cost. Release thresholds are not strict cutoff frequencies as suggested by the deterministic analysis. (B) Loss probability as a function of release proportion for an element with a 10% fitness cost.

Interestingly, male-specific fitness costs have little effect on the optimal gender ratio of a release. Gender ratio is clearly relevant since an all-male release cannot result in spread while, in the absence of fitness costs, a release of equal numbers of males and females has a threshold of 36.4%. This threshold is minimized when the release is 33% male. For a 10% fitness cost on males and females, the threshold is minimized for a 38% male release: the same as the optimal gender ratio for a 20% fitness cost on males only (Figure 4D). We generally consider a release of equal numbers of males and females since this is biologically convenient and is close to the optimal gender ratio in terms of release threshold.

Stochastic formulation: Real populations have a finite number of individuals and are subject to a multitude of chance events. For this reason, we consider a stochastic version of *Semele* dynamics. For simplicity, we consider the equal fitness cost model with 100% toxin efficiency. At generation k, the number of individuals with genotypes *tt* and *TT* is denoted by i_k and j_k, respectively, half being male and half being female. The total population size is denoted by N. Following from Equations 22–24, the genotypes of individuals in the next generation are described by the expected proportions

$$p_{tt,k+1} = \frac{i_k^2 + 0.5i_k(N - i_k - j_k) + 0.25(N - i_k - j_k)^2}{W_{k+1}} \tag{25}$$

$$p_{TT,k+1} = \frac{(j_k^2 + j_k(N - i_k - j_k) + 0.25(N - i_k - j_k)^2)(1 - s)}{W_{k+1}}. \tag{26}$$

The normalizing term, W_{k+1}, is analogous to that in Equation 10 and is given by

$$\begin{aligned} W_{k+1} = {} & i_k^2 + 0.5i_k(N - i_k - j_k) + 0.25(N - i_k - j_k)^2 \\ & + (j_k^2 + j_k(N - i_k - j_k) + 0.25(N - i_k - j_k)^2)(1 - s) \\ & + (i_k j_k + 0.5i_k(N - i_k - j_k) + 0.5(N - i_k - j_k)^2 \\ & + j_k(N - i_k - j_k))(1 - hs). \end{aligned} \tag{27}$$

At each generation, these expected proportions are reduced to a population of N adults consisting of i_{k+1} wild types and j_{k+1} homozygotes by sampling from the multinomial distribution,

$$\begin{aligned} & \Pr(i_{k+1}, j_{k+1} \mid i_k, j_k, N) \\ & \quad = \frac{N!}{i_{k+1}!\, j_{k+1}!(N - i_{k+1} - j_{k+1})!} \\ & \qquad \times p_{tt,k+1}^{i_{k+1}} p_{TT,k+1}^{j_{k+1}} (1 - p_{tt,k+1} - p_{TT,k+1})^{N - i_{k+1} - j_{k+1}}. \end{aligned} \tag{28}$$

We consider a release of j_0 homozygotes in a population of i_0 wild types at generation 0. We calculate the loss probability by iterating until the *Semele* allele is either fixed or lost and calculating the proportion of trials that reach the state $(i, j) = (N, 0)$. The distribution of extinction times can be calculated by recording the generation that this state is reached. The distribution of fixation times can be calculated by recording the generation in which the state $(i, j) = (0, N)$ is reached.

Results are shown in Figure 5 for population sizes of 100, 1000, and 10,000. The mosquito populations of interest for disease control are on the order of 1000–10,000 individuals per village. Malaria-transmitting *A. gambiae* populations show large seasonal variation in size, but have been estimated to behave like randomly mating populations of several thousand individuals (TAYLOR and MANOUKIS 2003). Dengue-transmitting *A. aegypti* mosquitoes are similarly numerous in villages, but tend to have more structured populations, suggesting smaller effective population sizes (REITER *et al.* 1995; TRIPS *et al.* 1995; HARRINGTON *et al.* 2005; JEFFERY *et al.* 2009).

The implication of these results is that the release thresholds mentioned earlier are not strict cutoff frequencies, above which the allele spreads and below which it does not. The deterministic thresholds are in fact the frequencies at which the allele is equally likely to spread or to go extinct. For a *Semele* allele with no fitness cost, the deterministic threshold is 36.4%; however, in a population of 1000, the allele has a 10% chance of spreading for a 34.6% release and a 1% chance of spreading for a 33.1% release (Figure 5A). A similar pattern is seen in the presence of a fitness cost (Figure 5B).

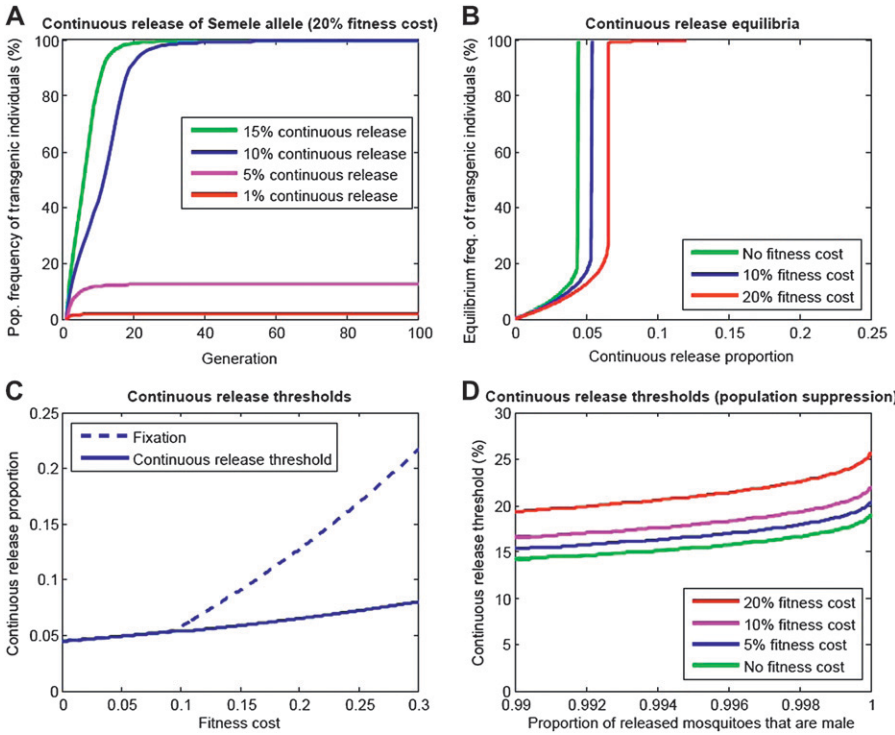

FIGURE 6.—Continuous release of a *Semele* allele. (A) Time-series dynamics for the case of a 20% fitness cost. Above the continuous release threshold, the allele either fixes or reaches a high equilibrium consisting mostly of transgenic individuals. (B) Transgenic equilibrium frequency as a function of continuous release proportion for a variety of fitness costs. This clearly shows the existence of a continuous release threshold. (C) Continuous release threshold as a function of fitness cost. Also shown is the continuous release proportion required for fixation to occur. (D) Continuous release threshold as a function of release gender ratio for release gender ratios between 99 and 100% male. Even a small number of transgenic females in an all-male transgenic release intended for population suppression can lead to gene drive occurring instead.

The results of the stochastic simulations also suggest that, in the event of an accidental release, a *Semele* allele is very unlikely to persist in the wild. The allele will almost certainly become extinct for a release frequency $<30\%$ in a population of ≥ 1000 (Figure 5). Accidental releases are expected to be much smaller than this and hence even more likely to be self-limiting.

Continuous release: For an intentional release, population replacement is most easily achieved by a sustained release of transgenic individuals. If the transgenic individuals are released at a high enough rate, they will accumulate in the population and eventually exceed the release threshold, at which point they are capable of spreading on their own. To model a

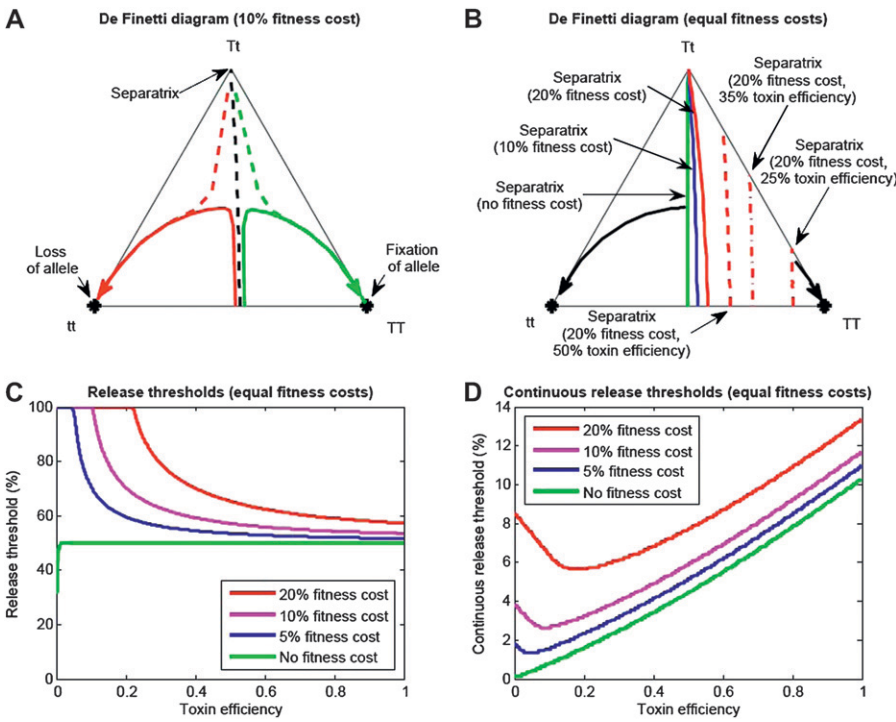

FIGURE 7.—Population dynamics of a *Semele* element with a recessive antidote. (A) De Finetti diagram for the case of a 10% fitness cost. Above the release threshold, the allele is completely fixed. (B) De Finetti diagram showing separatrices for a variety of fitness costs and toxin efficiencies. (C) Release threshold as a function of toxin efficiency for a variety of fitness costs. Above these thresholds, the element is always fixed. (D) Continuous release threshold as a function of toxin efficiency for a variety of fitness costs.

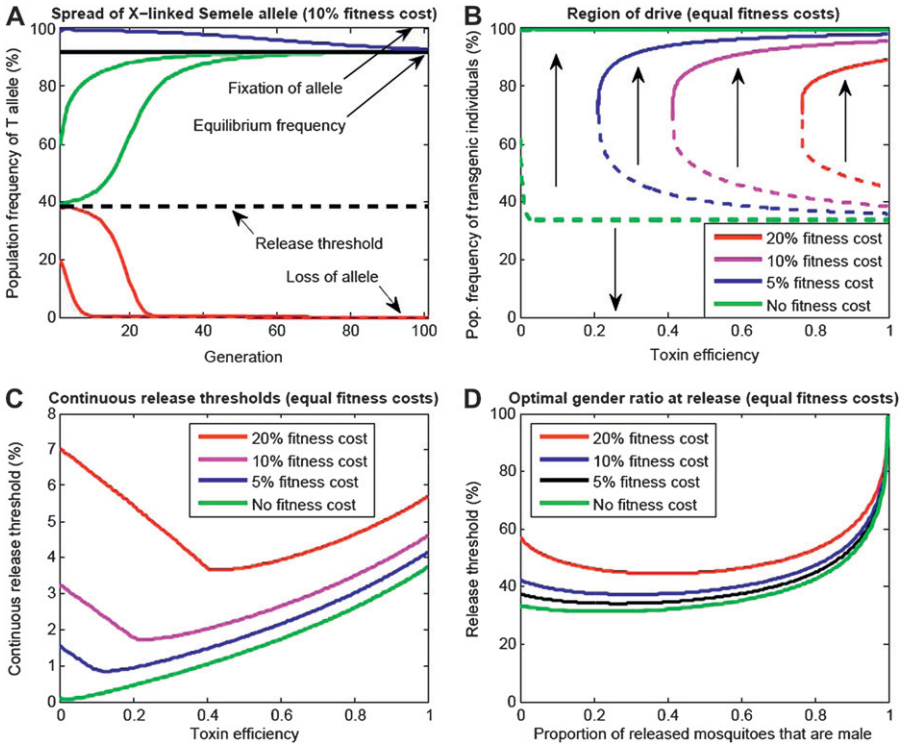

FIGURE 8.—Population dynamics of a *Semele* element located on the X chromosome. (A) Time-series dynamics for an element with a 10% fitness cost. Above the release threshold, the allele approaches a stable equilibrium consisting mostly of transgenic individuals. (B) Region of drive as a function of toxin efficiency for a variety of fitness costs. (C) Continuous release threshold as a function of toxin efficiency for a variety of fitness costs. (D) Release threshold as a function of release gender ratio. The threshold is minimized for a female-biased release.

sustained release, we add a proportion, μ, of *Semele* homozygotes to the mating pool at each generation. This proportion is measured relative to the total population prior to introduction. A fraction, r, of the released individuals are male. The relative proportion of homozygous males in the mating pool then becomes $w_{m,k} + r\mu$ and the relative proportion of homozygous females is $w_{f,k} + (1 - r)\mu$. Modifying Equations 1–10 accordingly, only Equations 2 and 3 are affected:

$$\begin{aligned} v_{e,k+1} = {} & u_{m,k}(w_{f,k} + (1 - r)\mu) + 0.5u_{m,k}v_{f,k} + 0.5v_{m,k}v_{f,k} \\ & + 0.5v_{m,k}(w_{f,k} + (1 - r)\mu) + 0.5(w_{m,k} + r\mu)v_{f,k} \\ & + 0.5v_{m,k}u_{f,k}(1 - e_T) + (w_{m,k} + r\mu)u_{f,k}(1 - e_T) \end{aligned} \quad (29)$$

$$\begin{aligned} w_{e,k+1} = {} & (w_{m,k} + r\mu)(w_{f,k} + (1 - r)\mu) + 0.5(w_{m,k} + r\mu)v_{f,k} \\ & + 0.5v_{m,k}(w_{f,k} + (1 - r)\mu) + 0.25v_{m,k}v_{f,k}. \end{aligned} \quad (30)$$

Since the release is incorporated into the difference equations, then the population is entirely wild type at generation 0 ($u_{m,0} = u_{f,0} = 0.5$). Using this initial condition and the difference equations described above, the continuous release thresholds, equilibria, and time-series dynamics can be calculated.

Figure 6A depicts the time-series dynamics for a sustained release of *Semele* homozygotes with a 20% fitness cost. For a continuous release of 5% homozygotes per generation, the allele persists at a low frequency and transgenic individuals persist at <12% in the population; however, for release rates of ≥10% per generation, the allele either fixes or spreads to a very high frequency. Figure 6B depicts the equilibria reached for a variety of release rates. These confirm the existence of a continuous release threshold that depends on the element fitness cost: in the absence of a fitness cost, the threshold is 4.4% per generation; while for a 20% fitness cost, the threshold is 6.5% per generation. For fitness costs <10%, the allele will fix in the population if this threshold is exceeded (Figure 6C). The population will return to the stable equilibria described earlier if continuous releases are terminated and occasional wild-type individuals enter the population.

These results have several implications for population replacement. First, if a transgenic release on the order of 37–50% homozygotes is considered unfeasible, then a sustained release of 5–7% homozygotes per generation would provide an achievable solution. Second, it is very likely that a release of mosquitoes with *Semele* alleles will be confined to the release population – a very desirable feature in the early stages of testing. If the allele fixes in the release population, this population will act as a source for neighboring populations; however, mosquito migration rates between villages in Africa tend to be <1% per generation (TAYLOR *et al.* 2001), suggesting that the allele will persist only at low levels in neighboring populations.

Finally, the results have implications for an all-male release intended for population suppression. If the toxin is 100% efficient, then transgenic males will have no offspring with wild females, leading to suppression; but uncertainty arises when a small number of transgenic females are included in the release. Figure 6D depicts the continuous release thresholds that must be

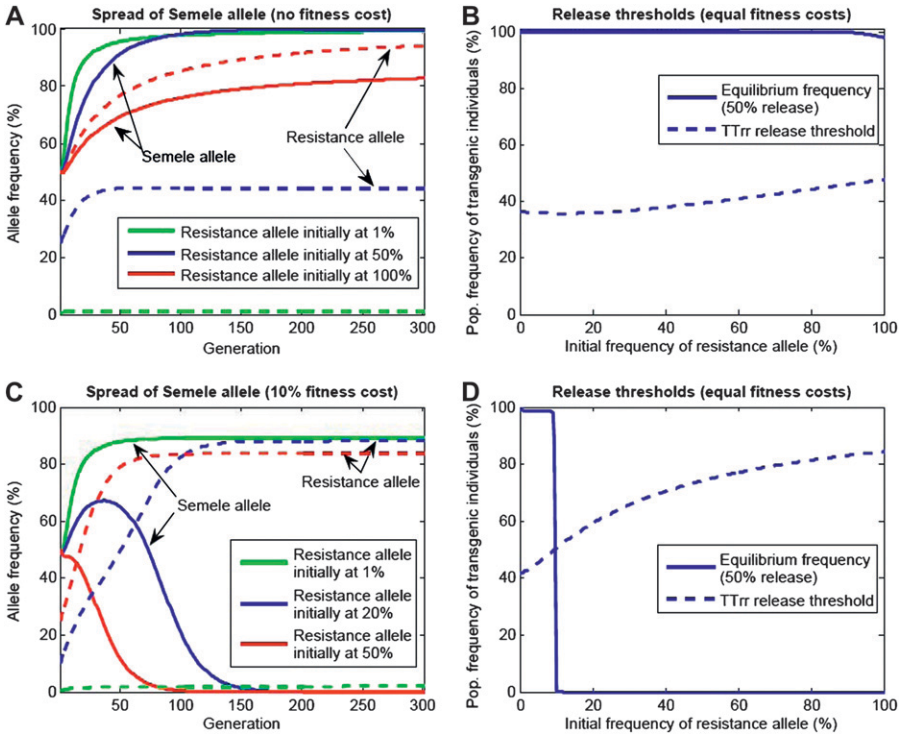

FIGURE 9.—Prior existence of a toxin-resistance allele, *R*, in a natural population. (A) Time-series dynamics for a *Semele* allele with no fitness cost released at 50% into a population having various prior frequencies of the *R* allele. (B) Release thresholds and stable equilibria following a 50% release as a function of prior *R*-allele population frequency. (C) Time-series dynamics for a construct with a 10% fitness cost released at 50% into a population having various prior frequencies of the *R* allele. (D) Release thresholds and stable equilibria following a 50% release as a function of prior *R*-allele population frequency.

exceeded for population replacement to occur when the transgenic male-to-female ratio is ≥99:1. For a 20% fitness cost, a 99:1 gender ratio will lead to population replacement for release rates >19.5% per generation. For a 999:1 gender ratio, the threshold is 23.5% per generation. These release rates are much less than those planned for population suppression (PHUC *et al.* 2007), suggesting that, if sexing is not perfect, a separate element lacking the antidote gene should be used for population suppression.

Recessive antidote: We consider two simple variants of the *Semele* system. First, we consider the case in which two copies of the antidote gene are required to neutralize the toxin instead of one. The dynamics of this system are the same as those for a dominant antidote with the exception that crosses between transgenic males and heterozygous females are also unviable. For the case of equal fitness costs in males and females, the following modified form of Equations 1–10 applies,

$$u_{k+1} = \frac{u_k^2 + 0.5u_kv_k + (0.25v_k^2 + 0.5u_kv_k)(1-e_T)}{W_{k+1}} \tag{31}$$

$$v_{k+1} = \frac{(u_kw_k + 0.5u_kv_k + 0.5v_kw_k + (0.5v_k^2 + 0.5v_kw_k + 0.5u_kv_k + u_kw_k)(1-e_T))(1-hs)}{W_{k+1}} \tag{32}$$

$$w_{k+1} = \frac{(w_k^2 + 0.5v_kw_k + (0.5v_kw_k + 0.25v_k^2)(1-e_T))(1-s)}{W_{k+1}}, \tag{33}$$

where W_{k+1} is the normalizing term defined in Equation 10. Assuming 100% toxin efficiency, this system has three biologically feasible equilibrium points:

$$(u_*, w_*) = (0,1), (1,0), (0,0). \tag{34}$$

The first of these points represents fixation, the second represents loss, and the third represents a completely heterozygous population. We calculate the stabilities of these points in File S1. In conjunction with time-series data (Figure 7, A and B), we see that fixation and loss are stable equilibria, while the case of an all-heterozygote population lies on a separatrix above which the allele is fixed and below which the allele is lost. These results are interesting because they imply that, above the release threshold, a *Semele* allele with a recessive antidote will be driven to fixation in a population regardless of its fitness cost. A fitness cost merely has the effect of increasing the release threshold.

The case of imperfect toxin efficiency in conjunction with a recessive antidote does not lend itself to analytic treatment; however, observation of time-series data suggests that the release threshold increases as the toxin becomes less efficient (Figure 7B). The minimum toxin efficiency required for drive to occur depends on the fitness cost: for a 10% fitness cost, toxin efficiency must exceed 9.6%; and for a 20% fitness cost, toxin efficiency must exceed 21.6% (Figure 7C). Above these efficiencies, a super-threshold release will result in element fixation regardless of its inefficiency and fitness cost. Below these efficiencies, the element will be lost for all cases other than initial fixation.

The release thresholds for *Semele* with a recessive antidote are relatively high: in the absence of fitness costs, *TT* males and females must be released at a frequency >50% to spread; and for a 20% fitness cost,

the threshold increases to 57.3% (Figure 7C). The same result can be achieved by a sustained release: in the absence of fitness costs, a continuous release rate >10.3% homozygotes per generation will lead to fixation; and for a 20% fitness cost, the threshold is 13.4% per generation (Figure 7D). These results are encouraging because they suggest that a *Semele* allele with a recessive antidote is even more confinable to a single population than an element with a dominant antidote.

X-linked *Semele*: Second, we consider the case in which the *Semele* allele is inserted at a location on the X chromosome. The dynamics of this system differ due to the fact that females carry two copies of the X chromosome, while males carry only one. The only unviable cross is between X^TY males and X^tX^t females. The proportions of the kth generation that are males of genotypes X^tY and X^TY are denoted by $u_{m,k}$ and $v_{m,k}$, respectively. The corresponding proportions for females are unchanged. For the case of equal fitness costs in males and females, the following modified form of Equations 1–10 applies,

$$u_{f,k+1} = \frac{0.5u_{m,k}u_{f,k} + 0.25u_{m,k}v_{f,k}}{W_{k+1}} \tag{35}$$

$$u_{m,k+1} = \frac{0.5u_{m,k}u_{f,k} + 0.25u_{m,k}v_{f,k} + 0.25v_{m,k}v_{f,k} + 0.5v_{m,k}u_{f,k}(1-e_T)}{W_{k+1}} \tag{36}$$

$$v_{f,k+1} = \frac{(0.5u_{m,k}w_{f,k} + 0.25u_{m,k}v_{f,k} + 0.25v_{m,k}v_{f,k} + 0.5v_{m,k}u_{f,k}(1-e_T))(1-hs)}{W_{k+1}} \tag{37}$$

$$v_{m,k+1} = \frac{(0.5u_{m,k}w_{f,k} + 0.25u_{m,k}v_{f,k} + 0.5v_{m,k}w_{f,k} + 0.25v_{m,k}v_{f,k})(1-hs)}{W_{k+1}} \tag{38}$$

$$w_{f,k+1} = \frac{(0.5v_{m,k}w_{f,k} + 0.25v_{m,k}v_{f,k})(1-s)}{W_{k+1}}, \tag{39}$$

where W_{k+1} is the normalizing term defined in Equation 10. We consider a release of X^TY males and X^TX^T females at generation 0,

$$v_{m,0} = rw_0 \tag{40}$$

$$w_{f,0} = (1-r)w_0 \tag{41}$$

$$u_{m,0} = u_{f,0} = 0.5(1-w_0), \tag{42}$$

where w_0 is the release proportion and r is the fraction of released individuals that are male. Using this initial condition and the difference equations described above, the dynamics of the system can be calculated.

Figure 8A depicts the time-series dynamics of an X-linked *Semele* allele with a 10% fitness cost. The dynamics are very similar to those of an autosomal allele: there is a release threshold of 38.5%, above which the allele spreads to an equilibrium frequency of 92%, and below which the allele is lost. Although the *Semele* allele reaches only a frequency of 92%, >96% of the population is transgenic at this allele frequency.

Figure 8B depicts the release thresholds and equilibria reached for a variety of fitness costs and toxin efficiencies. For 100% toxin efficiency, an X-linked allele with no fitness cost will fix in the population for releases >33.3%; an allele with a 5% fitness cost will spread to a transgenic frequency of 98% for a release >35.7%; and an allele with a 20% fitness cost will spread to a transgenic frequency of 89.2% for a release >45%. As for an autosomal allele (Figure 8D), the region of drive decreases from both sides as the fitness cost increases and the toxin becomes less efficient. X-linked alleles have a larger region of drive for high toxin efficiencies; but allele location makes less difference at low efficiencies.

Similarly, for the case of 100% toxin efficiency, an X-linked allele has a slightly lower continuous release threshold than an autosomal allele (Figure 8C). In the absence of fitness costs, the threshold is 3.7% per generation (compared to 4.4% per generation for an autosomal allele), and for a 20% fitness cost, the threshold is 5.7% per generation (compared to 6.5% per generation for an autosomal allele). Although X-linked alleles are less confinable than autosomal alleles at these toxin efficiencies, the amount of migration required to colonize a secondary population is still greater than that observed between typical African villages (Taylor *et al.* 2001) .

The gender ratio of a release is particularly relevant for X-linked alleles because released females have two copies of the allele while males have only one. In the absence of fitness costs, majority-female releases are favored for this reason; however, more equal ratios are favored in the presence of high fitness costs because released females have more copies of the allele, making them more vulnerable to fitness costs (Figure 8D).

Barriers to spread: Simple models can predict gene drive systems to spread in abstract populations; however, real populations are far more complex and may have complicating features that prevent spread. In this section, we explore two potential complicating features of real populations: first, the prior existence of toxin resistance in nature; and second, the tendency for wild females to be less attracted to transgenic males.

Prior toxin resistance: In both cases, we use a two-locus model to study the discrete generation dynamics. For prior toxin resistance in nature, we consider a toxin-resistance allele, denoted by *R*, which is unlinked to the *Semele* allele, *T*. We then use a series of 81 dihybrid crosses to keep track of the proportions of each generation that are males and females of genotypes *TTRR*, *TTRr*, *TTrr*, *TtRR*, *TtRr*, *Ttrr*, *ttRR*, *ttRr*, and *ttrr*. For simplicity, we assume equal fitness costs in both sexes due to the *Semele* allele and 100% toxin and antidote efficiency. Females that have either the *Semele* allele or

the natural toxin-resistance allele are protected against the *Semele* toxin. The only unviable crosses are therefore between males having the *Semele* allele and *ttrr* females. The Matlab code for this model is available from the authors upon request.

We assume that none of the released mosquitoes have the natural toxin-resistance allele and that, prior to a release, this allele exists in the population at Hardy–Weinberg equilibrium with frequency f_0. Considering a release of *TTrr* males and females at generation 0, the initial condition for the difference equations is given by

$$p_{TTrr,0} = w_0 \tag{43}$$

$$p_{ttRR,0} = (1 - w_0)f_0^2 \tag{44}$$

$$p_{ttRr,0} = (1 - w_0)2f_0(1 - f_0) \tag{45}$$

$$p_{ttrr,0} = (1 - w_0)(1 - f_0)^2. \tag{46}$$

Here, the released individuals represent a proportion, w_0, of the total population. Using this initial condition and the model described above, the dynamics of the *T* and *R* alleles can be calculated.

A *Semele* allele without fitness costs is relatively unaffected by the prior existence of toxin resistance in a population. Figure 9A depicts the time-series dynamics of a 50% transgenic release without fitness costs. Here, even if the *R* allele is initially at 50% in the population, the *T* allele will spread to fixation within 100 generations. Both the *T* and the *R* alleles are selected for following the release; however, the *T* allele is generally released at such a high frequency that it still reaches fixation. For a 50% transgenic release, it is only when the *R* allele is initially at a frequency >83% that it prevents the *T* allele from being fixed. In this case, the *R* allele is fixed first and the *T* allele plateaus at an equilibrium frequency >85%, corresponding to a frequency of transgenic individuals >98% (Figure 9B).

Prior toxin resistance is potentially debilitating when the *Semele* allele confers a fitness cost. Figure 9C depicts the time-series dynamics of a 50% transgenic release for an element having a 10% fitness cost. If the *R* allele has an initial population frequency <10%, the *T* allele will spread to an equilibrium frequency of ~90%, corresponding to a frequency of transgenic individuals of ~98.5%, within 100 generations (Figure 9D). However, if the *R* allele has an initial frequency >10%, the *T* allele will be lost from the population. This suggests that, for moderate to high fitness costs, only low prior levels of toxin resistance in the population can be tolerated. Experiments would be advised to test for toxin resistance in the environment and determine its frequency prior to a release.

Assortative mating: We model the impact of assortative mating on the spread of a *Semele* allele in File S1. Our model is based on the assumption that there is nothing in the biology of *Semele*, in and of itself, that would lead to an assortative mating phenotype; however, mosquitoes released with the *Semele* allele will have been raised in an industrial setting and will differ from wild mosquitoes due to laboratory selection, inbreeding, and strain differences. We model these differences in the form of a single unlinked allele, *A*, which is responsible for some degree of unattractiveness of transgenic males to females. Our analysis shows that, if only wild females are less attracted to *AA* males, then assortative mating has very minor effects on the spread of the *Semele* allele. If both transgenic and wild females are less attracted to *AA* males, then strong assortative mating tendencies may require increased introduction frequencies; however, provided that these frequencies are exceeded, the *Semele* allele will spread to the same equilibrium frequency in the population.

DISCUSSION

The failure of existing technology to control mosquito-borne diseases has renewed interest in the development of transgenic mosquitoes as a component of an integrated strategy for controlling insect disease vectors (Braig and Yan 2001; Alphey *et al.* 2002; Sinkins and Gould 2006; Marshall and Taylor 2009). Here, we have described a genetic system that may be used for both suppression of mosquito population sizes and replacement of mosquito populations with disease-refractory varieties. The system has several features that make it attractive in the early stages of testing and development, when it is essential that the spread of transgenes be limited in space and time.

As a population suppression system, *Semele* has the potential to control disease locally without persisting over time and spreading from one population to another. As a population replacement system, *Semele* is highly unlikely to spread following an accidental release and can be reasonably confined to a single population following an intentional release. *Semele* alleles can also be removed from a population through a combination of mosquito control measures and the introduction of large numbers of wild-type mosquitoes. Finally, as a chromosomally located toxin–antidote system, the original *Semele* allele can be bumped out of the population in favor of a new allele consisting of the old antidote in combination with a new toxin–antidote pair, provided that the new element is located at the same genomic location (Chen *et al.* 2007). The *Semele* system therefore satisfies many of the safety criteria required for release of transgenic mosquitoes.

An all-male release of mosquitoes with *Semele* results in population suppression because wild females that mate with transgenic males produce no offspring. Our modeling suggests, however, that a separate *Semele* allele lacking the antidote gene should be used for population suppression because even the tiniest contamination of transgenic females in an all-male release will result in gene drive rather than suppression.

It is therefore safer to create a sterile variant than to rely on perfect sexing prior to an all-male release.

For gene drive to occur, the element must include all components of the *Semele* system: the toxin, the antidote, and two promoters. One promoter must allow expression in the male germ line or accessory glands, and the other must allow expression in the female germ line or somatic tissues exposed to seminal fluid. The toxins used to build such an element are likely to be nearly 100% efficient; however, they may also confer a significant fitness cost on the males expressing them. To be conservative, let us consider a high fitness cost on males, approximated by a 10–20% fitness cost on both males and females. Our modeling then suggests that a 50% release should result in gene drive; however, the release could also be spread out over multiple generations by releasing 7% transgenic mosquitoes each generation. A release of equal numbers of males and females is adequate. We predict that such an element will spread to a transgenic frequency of 95% within 15 generations and to a transgenic frequency of 98% within 30 generations. The *Semele* system is therefore able to drive into a population quickly and efficiently.

Counterintuitively, a *Semele* allele with a recessive antidote is expected to drive into a population more quickly and efficiently (while having a higher release threshold) than an element with a dominant antidote. This is because, with a recessive antidote, the *Semele* allele distorts the offspring ratio even when there are no wild-type individuals in the population. This has the effect of driving the *Semele* allele to fixation, provided that the initial release exceeds a certain threshold. For comparable fitness costs, modeling suggests that a 55% release should result in gene drive and that the *Semele* allele will be fixed within 20 generations. Interestingly, the allele is expected to spread to fixation regardless of the fitness cost. The release threshold is slightly higher; but the flipside of this is that the element requires higher migration rates to spread into adjacent populations, suggesting it is more strongly confined to a single population.

These results inspire interest in the ability to engineer traits that are expressed only when an element is present in two copies (in this case at a common position on both homologs), but not one. Such a cellular counting system has not yet been engineered by humans; but it is a task that has been achieved by nature, in the context of sex determination in Drosophila (Sanchez 2008) and during X chromosome inactivation and allelic exclusion in mammals (Keverne 2009; Zakharova *et al.* 2009). One possible method for achieving this takes advantage of pairing-sensitive silencing, a phenomenon in which the presence of specific sequences near genes located at the same site on homologous chromosomes results in strong silencing of these genes in homozygotes, but much weaker silencing in heterozygotes (Kassis 2002). Perhaps a recessive antidote could be engineered by using pairing-sensitive silencing to repress a repressor of antidote activity in homozygotes, but not in heterozygotes.

The spread of a *Semele* allele is relatively immune to several potential complicating features of real populations, such as assortative mating and the prior existence of resistance alleles. This is largely due to the fact that the strategy is directed at single populations, involves releases at high proportions, and reaches an equilibrium frequency within a small number of generations. Assortative mating is a small hindrance that may lead to increased release thresholds. Prior toxin resistance in the population is a problem only when the *Semele* allele confers a moderate to high fitness cost and toxin resistance is present at a population frequency of $\sim \geq 10\%$. Mutational inactivation of the antidote gene is selected against and inconsequential. Mutational inactivation of the toxin and refractory genes are consequential, but are a feature of any toxin–antidote-based drive system, and can be forestalled to some extent through gene multimerization. There are likely other complicating factors that we have not considered.

As for any mathematical model, simplifications were made that may compromise the quality of the predictions. In using difference equations for the majority of our models, we considered an infinite, randomly mating population with discrete, nonoverlapping generations. We ignored the population structure of mosquito populations, for example, spatial structure, age structure, and mating structure. We also ignored density dependence and behavior effects. Despite this, discrete-generation difference equations have been successfully used to gain insight into several other gene drive systems (Wade and Beeman 1994; Davis *et al.* 2001; Deredec *et al.* 2008; Gould *et al.* 2008), including those generated using nuclear-encoded CI-causing factors (Turelli and Hoffmann 1999), which display analogous dynamics to the *Semele* system. We believe there is a mandate to use these models and that they capture the main features of *Semele* dynamics.

Engineering the *Semele* system: Finally, there is reason to believe that the *Semele* drive system can be constructed using existing reagents or reagents that could be created using existing technologies. Synthesis could be achieved in three ways: first, by manipulating gene expression in male and female somatic tissues; second, by manipulating gene expression in the germ line; and third, by isolating the genes that mediate CI and inserting these onto a nuclear chromosome under the control of promoters that recapitulate their patterns of expression in an infected insect. In the case where genes are expressed in somatic tissues, peptides or proteins can be expressed under the control of promoters that drive expression in male accessory glands (Sirot *et al.* 2009). If these proteins enter the female along with other seminal fluid components during mating, they may be able to disrupt essential functions, perhaps in the nervous system. Candidates might include insect-specific peptide neurotoxins

(Nicholson 2007). The activity of these toxins in female recipients could be blocked through the female-specific expression and secretion into the hemolymph of neutralizing antibodies or through female-specific expression of variants of the receptor targeted by the toxin that are insensitive to toxin function, perhaps in conjunction with the use of RNAi to silence expression of the transcript encoding the endogenous receptor.

In the case where genes are expressed in the germ line, sperm-based toxins could be used in conjunction with oocyte- or egg-based antidotes. Candidate toxins include DNAses that cleave zygotic DNA following fertilization, but do not cleave haploid sperm DNA. In this approach strategies are necessary to prevent toxin expression until late stages of spermatogenesis, by which time spermatid chromatin is in a highly condensed form, hopefully resistant to cleavage. UTR sequences that mediate translational repression during earlier stages of spermatogenesis provide one approach to achieving this goal (Schafer *et al.* 1995; Blumer *et al.* 2002). Sperm-specific expression of microRNAs that are designed to translationally silence the toxin-encoding transcript provide another approach. To protect against a sperm-based toxin, the antidote must be presynthesized in the newly fertilized egg, ready for immediate action following sperm entry and chromatin decondensation. Candidate antidotes might include a protease that cleaves a target site engineered into the toxin or a maternally expressed intrabody—cytoplasmic versions of antibodies (Lo *et al.* 2008)—that binds to the nuclease, neutralizing its toxic function.

A first step toward a sperm-based toxin has been taken with the demonstration that expression of the homing endonuclease I-PpoI under the control of a male germline-specific promoter results in zygote lethality in *A. gambiae*. Lethality is caused by cleavage of I-PpoI target sites in zygote DNA, found in multiple copies in the X chromosome-linked 28S ribosomal RNA gene cluster (Windbichler *et al.* 2008). I-PpoI expression in the male germ line also results in bias toward Y-bearing spermatozoa, probably due to I-PpoI-dependent cleavage of the ribosomal gene cluster found in X-bearing sperm. This damage presumably leads to their loss during spermatogenesis. This last fact highlights the necessity of being able to silence nuclease expression until stages of spermatogenesis in which nuclear DNA is hidden from such an enzyme.

Semele could also be created using molecules (presumably proteins) that mediate Wolbachia-induced CI. Unidirectional CI is seen in crosses between infected and uninfected individuals: matings between infected males and uninfected females result in death of some or all progeny, while matings between infected or uninfected males and infected females produce viable, infected progeny. As a result, infected females gain a reproductive advantage in the presence of Wolbachia, and since Wolbachia is transmitted through the female germ line, it benefits as well (Werren *et al.* 2008). Unidirectional CI behaves as though sperm produce a toxin, which is counteracted in the zygote by a maternally provided antidote. The idea then is to link genes that mediate toxin and rescue CI activities and express them from the nuclear genome in patterns that facilitate their CI-inducing function (Sinkins *et al.* 1997; Turelli and Hoffmann 1999). This approach to *Semele* generation is attractive because CI can be quite robust. In addition, Wolbachia strains exist that display CI but do not rescue each other, implying the existence of multiple, independent toxin–antidote functions. These positive points notwithstanding, the genes mediating CI remain to be identified. In addition, it is unclear whether the site and timing of expression of these proteins can be recapitulated from the nuclear genome. It is also important to note that *Semele* elements generated using components of the CI system could be used only in populations uninfected by Wolbachia bacteria expressing the same CI proteins.

Finally, it is interesting to consider the possibility that *Semele* elements could evolve in nature as a consequence of male–female conflict over reproduction. Male mating often results in a cost to females as a consequence of male traits designed to increase their paternity. These costs are sometimes mediated by seminal fluid components or modifications of sperm. In response, females sometimes evolve counteradaptations that decrease these costs (Chapman 2006; Parker 2006; Wolfner 2009). Typically, these effects are imagined to be the results of actions at unlinked loci. Here we note that linkage between a gene that mediates a male-derived cost to females and a locus expressed in females that counters this cost would create a *Semele*-like element. While *Semele* requires a threshold frequency to be exceeded for spread to occur, even when it carries no fitness cost, highly structured populations may provide an environment in which such elements could gain a foothold. Over evolutionary time, it is possible that such elements could sweep through a population, with the only hint that such an element existed being tight linkage between genes regulating fitness in response to mating in a reciprocal manner.

The authors thank Fred Gould for helpful insight into the *Semele* system, Catherine Ward for helpful discussions on model design, and two anonymous reviewers whose constructive comments improved the manuscript. John M. Marshall was supported by grant DP1 OD003878 to Bruce A. Hay from the National Institutes of Health.

LITERATURE CITED

Ahmed, A. M., S. L. Baggott, R. Maingon and H. Hurd, 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., C. B. Beard, P. Billingsley, M. Coetzee, A. Crisanti *et al.*, 2002 Malaria control with genetically manipulated insect vectors. Science **298:** 119–121.

Benedict, M. Q., and A. S. Robinson, 2003 The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends Parasitol. **19:** 349–355.

Benedict, M., P. D'Abbs, S. Dobson, M. Gottlieb, L. Harrington *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.

Blumer, N., K. Schreiter, L. Hempel, A. Santel, M. Hollmann *et al.*, 2002 A new translational repression element and unusual transcriptional control regulate expression of don juan during Drosophila spermatogenesis. Mech. Dev. **110:** 97–112.

Boete, C., and J. C. Koella, 2002 A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control. Malar. J. **1:** 3.

Boete, C., and J. C. Koella, 2003 Evolutionary ideas about genetically manipulated mosquitoes and malaria control. Trends Parasitol. **19:** 32–38.

Braig, H. R., and G. Yan, 2001 The spread of genetic constructs in natural insect populations, pp. 251–314 in *Genetically Engineered Organisms: Assessing Environmental and Human Health Effects*, edited by D. K. Letourneau and B. E. Burrows. CRC Press, Cleveland, OH/Boca Raton, FL.

Catteruccia, F., A. Crisanti and E. A. Wimmer, 2009 Transgenic technologies to induce sterility. Malar. J. **8**(Suppl. 2): S7.

Chapman, T., 2006 Evolutionary conflicts of interest between males and females. Curr. Biol. **16:** R744–R754.

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

Corby-Harris, V., A. Drexler, L. Watkins de Jong, Y. Antonova, N. Pakpour *et al.*, 2010 Activation of *Akt* signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. PLoS Pathog. **6:** e1001003.

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

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

de Lara Capurro, M., J. Coleman, B. T. Beerntsen, K. M. Myles, K. E. Olson *et al.*, 2000 Virus-expressed, recombinant single-chain antibody blocks sporozoite infection of salivary glands in Plasmodium gallinaceum-infected Aedes aegypti. Am. J. Trop. Med. Hyg. **62:** 427–433.

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

Dyck, V. A., J. Hendrichs and A. S. Robinson (Editors), 2005 *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management*. Springer, Dordrecht, The Netherlands.

Foster, G. G., M. J. Whitten, T. Prout and R. Gill, 1972 Chromosome rearrangements for the control of insect pests. Science **176:** 875–880.

Franz, A. W., I. Sanchez-Vargas, Z. N. Adelman, C. D. Blair, B. J. Beaty *et al.*, 2006 Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. Proc. Natl. Acad. Sci. USA **103:** 4198–4203.

Gould, F., 2007 Other mechanisms for strain replacement. Workshop on selfish DNA and the genetic control of vector-borne diseases, pp. 36–40. NESCent, Durham, NC.

Gould, F., and P. Schliekelman, 2004 Population genetics of autocidal control and strain replacement. Annu. Rev. Entomol. **49:** 193–217.

Gould, F., Y. Huang, M. Legros and A. L. Lloyd, 2008 A killer-rescue system for self-limiting gene drive of anti-pathogen constructs. Proc. Biol. Sci. **275:** 2823–2829.

Harrington, L. C., T. W. Scott, K. Lerdthusnee, R. C. Coleman, A. Costero *et al.*, 2005 Dispersal of the dengue vector Aedes aegypti within and between rural communities. Am. J. Trop. Med. Hyg. **72:** 209–220.

Ito, J., A. Ghosh, L. A. Moreira, E. A. Wimmer and M. Jacobs-Lorena, 2002 Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature **417:** 452–455.

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

Jeffery, J. A., N. Thi Yen, V. S. Nam, T. Nghia le, A. A. Hoffmann *et al.*, 2009 Characterizing the Aedes aegypti population in a Vietnamese village in preparation for a Wolbachia-based mosquito control strategy to eliminate dengue. PLoS Negl. Trop. Dis. **3:** e552.

Kassis, J. A., 2002 Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. Adv. Genet. **46:** 421–438.

Keverne, B., 2009 Monoallelic gene expression and mammalian evolution. BioEssays **31:** 1318–1326.

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

Lambrechts, L., J. C. Koella and C. Boete, 2008 Can transgenic mosquitoes afford the fitness cost? Trends Parasitol. **24:** 4–7.

Lo, A. S., Q. Zhu and W. A. Marasco, 2008 Intracellular antibodies (intrabodies) and their therapeutic potential. Handb. Exp. Pharmacol., **181:** 343–373.

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

Marelli, M. T., C. Li, J. L. Rasgon and M. Jacobs-Lorena, 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., 2009 The effect of gene drive on containment of transgenic mosquitoes. J. Theor. Biol. **258:** 250–265.

Marshall, J. M., and C. E. Taylor, 2009 Malaria control with transgenic mosquitoes. PLoS Med. **6:** e20.

Moreira, L. A., J. Wang, F. H. Collins and M. Jacobs-Lorena, 2004 Fitness of anopheline mosquitoes expressing transgenes that inhibit Plasmodium development. Genetics **166:** 1337–1341.

Nicholson, G. M., 2007 Fighting the global pest problem: preface to the special Toxicon issue on insecticidal toxins and their potential for insect pest control. Toxicon **49:** 413–422.

Parker, G. A., 2006 Sexual conflict over mating and fertilization: an overview. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361:** 235–259.

Phuc, H. K., M. H. Andreasen, R. S. Burton, C. Vass, M. J. Epton *et al.*, 2007 Late-acting dominant lethal genetic systems and mosquito control. BMC Biol. **5:** 11.

Reiter, P., M. A. Amador, R. A. Anderson and G. G. Clark, 1995 Short report: dispersal of Aedes aegypti in an urban area after blood feeding as demonstrated by rubidium-marked eggs. Am. J. Trop. Med. Hyg. **52:** 177–179.

Riehle, M. M., K. Markianos, O. Niare, J. Xu, J. Li *et al.*, 2006 Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. Science **312:** 577–579.

Sanchez, L., 2008 Sex-determining mechanisms in insects. Int. J. Dev. Biol. **52:** 837–856.

Schafer, M., K. Nayernia, W. Engel and U. Schafer, 1995 Translational control in spermatogenesis. Dev. Biol. **172:** 344–352.

Schmid-Hempel, P., 2005 Evolutionary ecology of insect immune defenses. Annu. Rev. Entomol. **50:** 529–551.

Shiff, C. J., 2000 Can roll back malaria achieve its goal? A challenge. Parasitol. Today **16:** 271–272.

Sinkins, S. P., and F. Gould, 2006 Gene drive systems for insect disease vectors. Nat. Rev. Genet. **7:** 427–435.

Sinkins, S. P., C. F. Curtis and S. L. O'Neill, 1997 The potential application of inherited symbiont systems to pest control, pp. 155–175 in *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*, edited by S. L. O'Neill, A. A. Hoffmann and J. H. Werren. Oxford University Press, Oxford.

Sirot, L. K., B. A. LaFlamme, J. L. Sitnik, C. D. Rubinstein, F. W. Avila *et al.*, 2009 Molecular social interactions: Drosophila melanogaster seminal fluid proteins as a case study. Adv. Genet. **68:** 23–56.

Taylor, C., Y. T. Toure, J. Carnahan, D. E. Norris, G. Dolo *et al.*, 2001 Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics **157:** 743–750.

Taylor, C. E., and N. C. Manoukis, 2003 Effective population size in relation to genetic modification of Anopheles gambiae sensu stricto, pp. 133–146 in *Ecological Aspects for Application of Genetically Modified Mosquitoes*, edited by W. Takken and T. W. Scott. Wageningen UR, Wageningen, The Netherlands.

Trips, M., W. Hausermann and G. B. Craig, Jr., 1995 Estimates of population size, dispersal, and longevity of domestic Aedes ae-

gypti aegypti (Diptera: Culicidae) by mark-release-recapture in the village of Shauri Moyo in eastern Kenya. J. Med. Entomol. **32:** 27–33.

Turelli, M., and A. A. Hoffmann, 1999 Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. Insect Mol. Biol. **8:** 243–255.

Vasan, S. S., 2009 Transgenic insects: from laboratory to field. AsPac. J. Mol. Biol. Biotechnol. **17:** 53–54.

Wade, M. J., and R. W. Beeman, 1994 The population dynamics of maternal-effect selfish genes. Genetics **138:** 1309–1314.

Werren, J. H., L. Baldo and M. E. Clark, 2008 Wolbachia: master manipulators of invertebrate biology. Nat. Rev. Microbiol. **6:** 741–751.

Whitten, M. J., and G. G. Foster, 1975 Genetical methods of pest control. Annu. Rev. Entomol. **20:** 461–476.

Windbichler, N., P. A. Papathanos and A. Crisanti, 2008 Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. PLoS Genet **4:** e1000291.

Wolfner, M. F., 2009 Battle and ballet: molecular interactions between the sexes in Drosophila. J. Hered. **100:** 399–410.

World Health Organization, 2009 *World Malaria Report 2009*. WHO Press, Geneva.

Zakharova, I. S., A. I. Shevchenko and S. M. Zakian, 2009 Monoallelic gene expression in mammals. Chromosoma **118:** 279–290.

Communicating editor: M. K. Uyenoyama

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.110.124479/DC1

Semele: A Killer-Male, Rescue-Female System for Suppression and Replacement of Insect Disease Vector Populations

John M. Marshall, Geoffrey W. Pittman, Anna B. Buchman and Bruce A. Hay

DOI: 10.1534/genetics.110.124479

FILE S1

I. Local stability analysis:

To calculate the local stability of an equilibrium point, we calculate the eigenvalues of the Jacobian matrix. The equilibrium is locally stable if all eigenvalues have modulus less than one. If one of more of the eigenvalues have modulus greater than one, the equilibrium is unstable. If the eigenvalue with the largest modulus has a modulus equal to one, then the linear stability analysis is inconclusive. In the case of an inconclusive analysis, we use numerical simulation to determine stability; however a nonlinear analysis can be used for more rigorous determination (Elaydi 1995).

No fitness costs: For a *Semele* element without fitness costs, the proportions of the k th generation that are individuals of genotypes *tt*, *Tt* and *TT* may then be denoted by u_k, v_k and w_k. The system of difference equations for this system is:

$$u_{k+1} = (u_k^2 + 0.25v_k^2 + u_k v_k(1 - 0.5e_T))/W_{k+1} \tag{14}$$

$$v_{k+1} = (0.5v_k^2 + v_k w_k + u_k v_k(1 - 0.5e_T) + u_k w_k(2 - e_T))/W_{k+1} \tag{15}$$

$$w_{k+1} = (w_k^2 + 0.25v_k^2 + v_k w_k)/W_{k+1} \tag{16}$$

where W_{k+1}, the normalizing term, is given by:

$$W_{k+1} = u_k^2 + u_k(v_k + w_k)(2 - e_T) + v_k^2 + 2v_k w_k + w_k^2 \tag{S1}$$

This system has three biologically-feasible equilibrium points:

$$(u_*, w_*) = (0,1),(1,0),\left(0.5,0.5(3 - e_T - \sqrt{e_T^2 - 6e_T + 8})\right) \tag{17}$$

The first of these points represents allele fixation, the second represents absence or loss of the *Semele* allele, and the third represents coexistence of wild, heterozygous and homozygous individuals in the population. We can calculate the stabilities of these points by calculating the eigenvalues of the Jacobian matrix:

$$\left(\begin{array}{cc} \frac{\partial u_{k+1}}{\partial u_k} & \frac{\partial u_{k+1}}{\partial w_k} \\ \frac{\partial w_{k+1}}{\partial u_k} & \frac{\partial w_{k+1}}{\partial w_k} \end{array} \right) \Bigg|_{(u_k, w_k)=(u_*, w_*)} \tag{S2}$$

The equilibrium $(u_*, w_*) = (0,1)$ has eigenvalues equal to 0 and 1, suggesting an inconclusive analysis; however numerical simulation indicates that fixation is stable over full range of $e_T \in (0,1]$. The equilibrium $(u_*, w_*) = (1,0)$ has eigenvalues equal to 0 and $1 - 0.5e_T$. Both of these eigenvalues are less than one for $e_T \in (0,1]$, suggesting that element loss is also a stable equilibrium. The third equilibrium point has eigenvalues equal to:

$$1 + e_T + z - \frac{4(2 - e_T)}{z} + \frac{\sqrt{(4 - e_T)^5 (4 - e_T(17 - 6z - 2e_T(6 - e_T - z)))}}{(4 - e_T)^3} \tag{S3}$$

and:

$$1 + e_T + z - \frac{4(2 - e_T)}{z} - \frac{\sqrt{(4 - e_T)^5 (4 - e_T(17 - 6z - 2e_T(6 - e_T - z)))}}{(4 - e_T)^3} \tag{S4}$$

where $z = \sqrt{e_T^2 - 6e_T + 8}$. The first of these eigenvalues is greater than 1 for $e_T \in (0,1]$, suggesting that the third equilibrium point is locally unstable. The stability analysis therefore shows that both fixation and loss of the element are represented by stable equilibrium points separated by an unstable intermediate equilibrium.

Equal fitness costs: For a *Semele* element with equal fitness costs in males and females, the system of difference equations is:

$$u_{k+1} = (u_k^2 + 0.25v_k^2 + u_k v_k (1 - 0.5e_T)) / W_{k+1} \tag{22}$$

$$v_{k+1} = (0.5v_k^2 + v_k w_k + u_k v_k (1 - 0.5e_T) + u_k w_k (2 - e_T))(1 - hs) / W_{k+1} \tag{23}$$

$$w_{k+1} = (w_k^2 + 0.25v_k^2 + v_k w_k)(1 - s) / W_{k+1} \tag{24}$$

where W_{k+1}, the normalizing term, is given by:

$$W_{k+1} = u_k^2 + 0.25v_k^2 + u_k v_k(1-0.5e_T) + (w_k^2 + 0.25v_k^2 + v_k w_k)(1-s) + ((u_k v_k + 2u_k w_k)(1-0.5e_T) + 0.5v_k^2 + v_k w_k)(1-hs) \quad \text{(S5)}$$

This system has four biologically-feasible equilibrium points; however only two of them are analytically tractable; namely:

$$(u_*, w_*) = (0,1),(1,0) \quad \text{(S6)}$$

The first of these points represents allele fixation, and the second represents absence or loss. We can calculate the stabilities of these points by calculating the eigenvalues of the Jacobian matrix (Equation S2).

The equilibrium $(u_*, w_*) = (0,1)$ has eigenvalues equal to 0 and

$$1 + \frac{s(1-h)}{1-s} \quad \text{(S7)}$$

The second eigenvalue is greater than one for $s \in (0,1)$ and $h \in [0,1)$, and is equal to one for two cases – the absence of fitness costs ($s = 0$), and the case of a completely dominant fitness cost ($s > 0$, $h = 1$). Numerical simulation suggests that fixation is a stable equilibrium in both these cases. The stability analysis therefore shows that fixation is unstable for a fitness cost that is recessive or shows any degree of heterozygosity, but is stable for a completely dominant fitness cost.

The equilibrium $(u_*, w_*) = (1,0)$ has eigenvalues equal to 0 and $(1-0.5e_T)(1-hs)$. Both of these eigenvalues are less than one for $e_T \in (0,1]$, $h \in [0,1]$ and $s \in [0,1)$, suggesting that allele loss is stable for all scenarios.

Recessive antidote: For a *Semele* element with a recessive antidote, the system of difference equations is:

$$u_{k+1} = (u_k^2 + 0.5u_k v_k + (0.25v_k^2 + 0.5u_k v_k)(1-e_T))/W_{k+1} \quad \text{(31)}$$

$$v_{k+1} = (u_k w_k + 0.5u_k v_k + 0.5v_k w_k + (0.5v_k^2 + 0.5v_k w_k + 0.5u_k v_k + u_k w_k)(1-e_T))(1-hs)/W_{k+1} \quad \text{(32)}$$

$$w_{k+1} = (w_k^2 + 0.5v_k w_k + (0.5v_k w_k + 0.25v_k^2)(1-e_T))(1-s)/W_{k+1} \quad \text{(33)}$$

where W_{k+1}, the normalizing term, is given by:

$$\begin{aligned} W_{k+1} &= u_k^2 + 0.5u_k v_k + (0.25v_k^2 + 0.5u_k v_k)(1 - e_T) \\ &+ (u_k w_k + 0.5u_k v_k + 0.5v_k w_k \\ &+ (0.5v_k^2 + 0.5v_k w_k + 0.5u_k v_k + u_k w_k)(1 - e_T))(1 - hs) \\ &+ (w_k^2 + 0.5v_k w_k + (0.5v_k w_k + 0.25v_k^2)(1 - e_T))(1 - s) \end{aligned} \quad \text{(S8)}$$

This system has three biologically-feasible equilibrium points; however, for the general case, only two are analytically tractable:

$$(u_*, w_*) = (0,1), (1,0) \quad \text{(S9)}$$

The first of these points represents allele fixation, and the second represents absence or loss. We can calculate the stabilities of these points by calculating the eigenvalues of the Jacobian matrix (Equation S2).

The equilibrium $(u_*, w_*) = (0,1)$ has eigenvalues equal to 0 and

$$\frac{(2 - e_T)(1 - hs)}{2(1 - s)} \quad \text{(S10)}$$

Both of these eigenvalues are less than one for $e_T \in (0,1]$, $h \in [0,1]$ and $s \in [0,1)$, suggesting that fixation is stable for all scenarios. The equilibrium $(u_*, w_*) = (1,0)$ has eigenvalues equal to 0 and $(1 - 0.5e_T)(1 - hs)$. Both of these eigenvalues are less than one for $e_T \in (0,1]$, $h \in [0,1]$ and $s \in [0,1)$, suggesting that element loss is also stable for all scenarios.

If we assume 100% toxin efficiency, then all three of the biologically-feasible equilibrium points are analytically tractable:

$$(u_*, w_*) = (0,1), (1,0), (0,0) \quad \text{(34)}$$

We have already shown that allele fixation and loss are stable equilibria for all scenarios. The eigenvalues of the Jacobian matrix for the third equilibrium point are indeterminate; however numerical simulation confirms that the intermediate equilibrium point is unstable for the full range of fitness costs ($h \in [0,1]$ and $s \in [0,1)$).

II. Assortative mating:

Several models have been proposed to describe the effects of assortative mating due to one allele on the distribution of another (Jennings 1917; Wright 1920). We consider two variations of a two-locus model to describe

the effect of assortative mating on the spread of a *Semele* allele. In both models, we account for the fact that male mosquitoes must compete in order to find a female mating partner. We consider an unlinked allele, "*A*", to be responsible for the degree of unattractiveness of transgenic males to females. Such an allele could be considered the product of laboratory inbreeding. We then use a series of 81 dihybrid crosses to keep track of the proportions of each generation that are males and females of genotypes *TTAA*, *TTAa*, *TTaa TtAA*, *TtAa*, *Ttaa*, *ttAA*, *ttAa* and *ttaa*.

For simplicity, we assume equal fitness costs due to the *T* allele and 100% toxin efficiency. All crosses between males having the *T* allele and females lacking the *T* allele are unviable. In model A, we assume that wild females are less attracted to transgenic males, while transgenic females are less discerning. Crosses between *AA* males and *aa* females have a reduced weighting of $1-a$, where a is the strength of assortative mating. We assume that the *A* allele is additive in its effect, and hence crosses between *AA* males and *Aa* females have a reduced weighting of $1-0.5a$, as do crosses between *Aa* males and *aa* females (Figure S1).

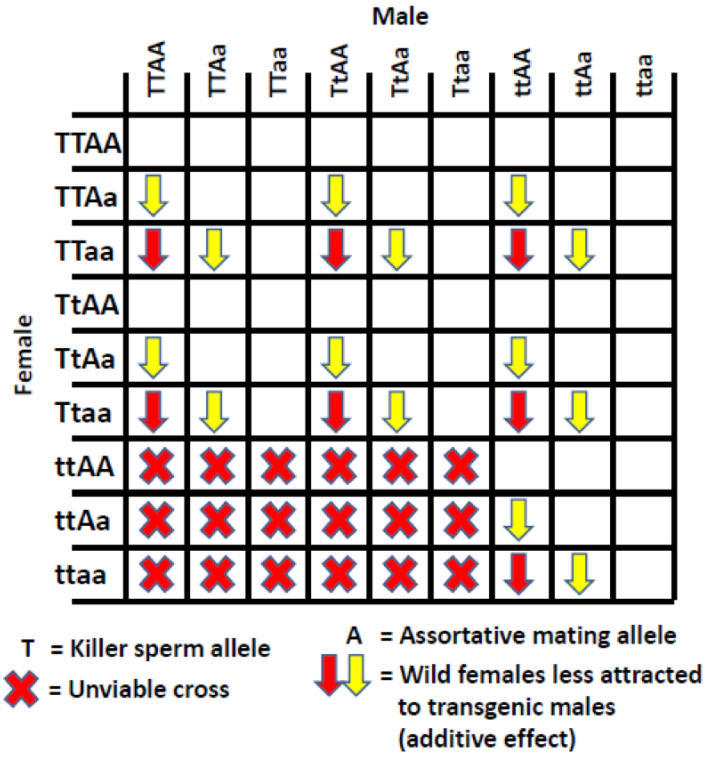

FIGURE S1.—Schematic diagram used to model the spread of a *Semele* allele, *T*, and an unlinked assortative mating allele, *A*, assumed to be the product of laboratory inbreeding. Each box represents a dihybrid cross used to keep track of the proportions of each genotype in discrete generations. All crosses between males having the *T* allele and females lacking the *T* allele are unviable. In this model (model A), *aa* females are less attracted to *AA* males, and so these crosses have a reduced weighting. The *A* allele is assumed to be additive in its effect and so crosses between *aa* females and *Aa* males and between *Aa* females and *AA* males also have reduced weight.

We assume that all released mosquitoes are inbred and are therefore homozygous for the *A* allele. Considering a release of *TTAA* males and females at generation 0, the initial condition for the system is given by:

$$p_{TTAA,0} = w_0 \tag{S11}$$

$$p_{ttaa,0} = 1 - w_0 \tag{S12}$$

Using this initial condition and the model described above, the dynamics of the *T* and *A* alleles can be calculated.

Assortative mating, modeled in this way, has only very minor effects on the spread of the *Semele* allele. Figure S2A depicts the time-series dynamics of a *Semele* allele having a 10% fitness cost for strengths of assortative mating between 10% and 90%. In all cases, for a 45% transgenic release, the *Semele* allele spreads to a transgenic frequency of ~99% within 40 generations. Increasing assortative mating strength from 10% to 90% retards its spread by approximately five generations. In all cases, for a 42% transgenic release, the *Semele* allele is lost from the population. This suggests that the release threshold varies by less than 3% over the full range of assortative mating strengths.

An interesting side-note for this system is that the *A* allele also displays threshold behavior. For a 50% transgenic release, both the *T* and *A* alleles spread (Figure S2B); however for a 45% release, only the *T* allele spreads (Figure S2C). Figure S2D depicts release thresholds and equilibrium frequencies for a *Semele* allele associated with a range of fitness costs and assortative mating strengths. Assortative mating has no effect on the equilibrium frequency, increases the release threshold for the *Semele* allele by up to 2.3% and retards spread by less than six generations. These results suggest that a *Semele* allele will spread into a population even in the presence of strong assortative mating tendencies if only wild females are less attracted to transgenic males.

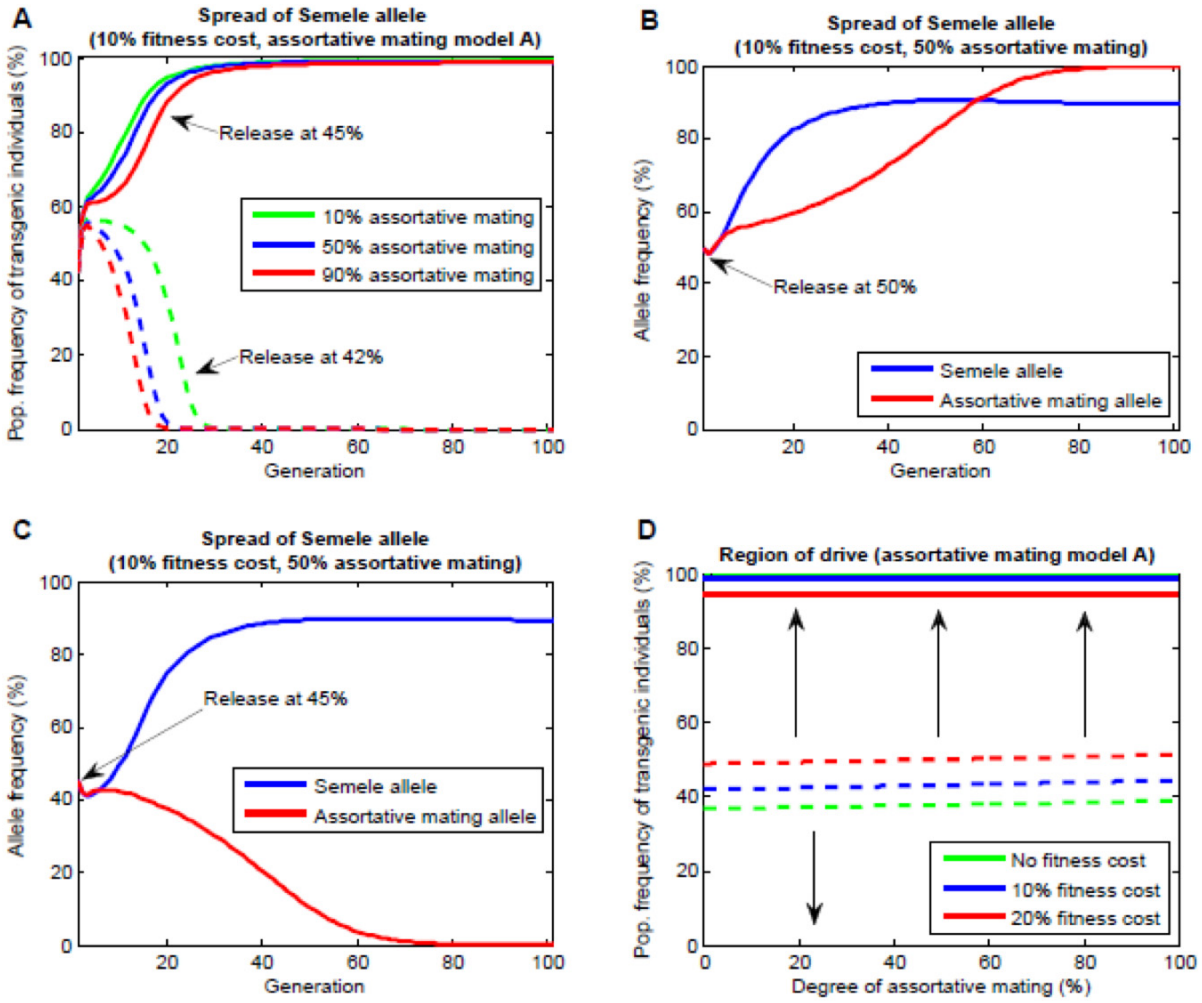

FIGURE S2.—Effects of an assortative mating allele, *A*, having the property that wild females are less attracted to *AA* males (model A). A: Time-series dynamics of an element with a 10% fitness cost and a variety of assortative mating strengths. Assortative mating increases the *T* allele release threshold by less than 3%. B-C: Time-series dynamics of both *T* and *A* alleles for an element having a 10% fitness cost and 50% assortative mating strength. The *A* allele also has a release threshold which is slightly greater than the *T* allele threshold. D: *T* allele release thresholds for several fitness costs as a function of assortative mating strength.

In model B, we assume that both transgenic and wild females are less attracted to transgenic males. In this model, all crosses involving *AA* males have a reduced weighting of $1-a$. We also assume that the *A* allele is additive in its effect, and hence all crosses involving *Aa* males have a reduced weighting of $1-0.5a$ (Figure S3). The Matlab code for both models is available from the authors upon request.

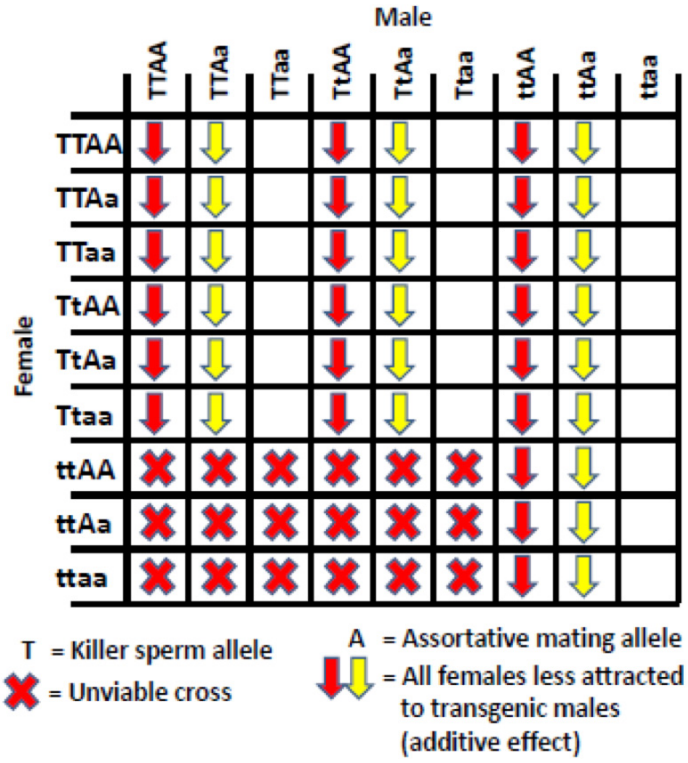

FIGURE S3.—In model B, all females are less attracted to *AA* males, and hence all crosses involving *AA* males have a reduced weighting. The *A* allele is assumed to be additive in its effect, and so crosses involving *Aa* males also have reduced weight.

Assortative mating has a significantly greater effect on *Semele* spread when transgenic females are also less attracted to transgenic males; however the *Semele* allele is still very capable of spreading into the population. Figure S4A depicts the time-series dynamics of a *Semele* allele having a 10% fitness cost for a range of assortative mating strengths. In all cases, for a 76% transgenic release, the *Semele* allele spreads to a transgenic frequency of ~99% within 50 generations. However, increasing the assortative mating strength from 10% to 90% increases the release threshold by more than 30% and retards the spread of the *Semele* allele by more than 30 generations.

Figure S2D depicts release thresholds and equilibrium frequencies for a *Semele* allele associated with a range of fitness costs and assortative mating strengths. As for model A, assortative mating has no effect on the equilibrium frequency reached by the *Semele* allele; however, unlike for model A, assortative mating increases the release threshold by up to 47% regardless of fitness costs. These results suggest that strong assortative mating tendencies may require increased introduction frequencies if all females are less attracted to transgenic males; however, provided that the new release threshold is exceeded, the *Semele* allele will spread to the same equilibrium frequency in the population.

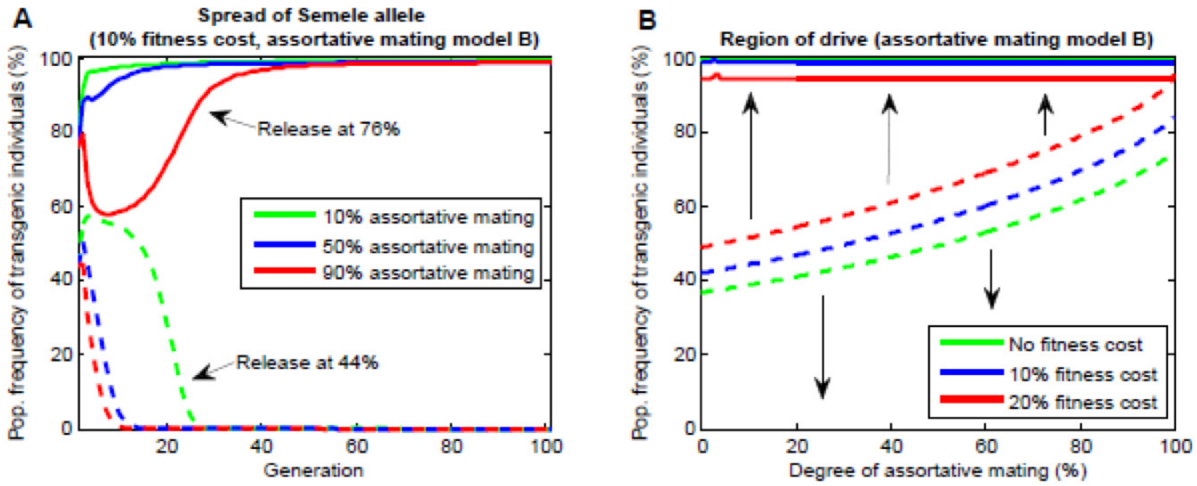

FIGURE S4.—Effects of an assortative mating allele, *A*, having the property that all females are less attracted to *AA* males (model B). A: Time-series dynamics of an element with a 10% fitness cost and a variety of assortative mating strengths. D: *T* allele release thresholds and equilibria for several fitness costs as a function of assortative mating strength. Assortative mating increases the *T* allele release threshold by up to 47%.

III. References:

Elaydi, S. N., 1995 An Introduction to Difference Equations. Springer, New York.

Jennings, H. S., 1917 The numerical results of diverse systems of breeding with respect to two pairs of characters, linked or independent, with special relation to the effect of linkage. Genetics **2**: 97-154.

Wright, S., 1920 Systems of mating. III. Assortative mating based on somatic resemblance. Genetics **6**: 144-161.