Confinement of gene drive systems to local populations: A comparative analysis

Supplemental Material:

I. *Medea*

Insect species tend to produce a large number of offspring per adult; however the environment can only sustain a limited number of individuals (Service, 1993), leading to a situation in which individuals are essentially sampled from a larger population at each generation. For this reason, we implement a stochastic version of the two-population model to capture some of the randomness inherent in this process. We illustrate this model using *Medea* (Chen *et al.*, 2007) as a case study.

Stochastic model formulation: We denote the number of individuals in population A at generation k having genotypes *mm* and *MM* by $i_{A,k}$ and $j_{A,k}$, respectively. The equivalent number of individuals in population B are denoted by $i_{B,k}$ and $j_{B,k}$, and each population has a total population size of N. Following from eqns (1-3), which describe the population dynamics of *Medea*, and eqn (18), which describe the substitutions required for the two-population model, the genotypes of embryos in the next generation in population A are described by the expected ratio $\hat{u}_{A,k+1} : \hat{v}_{A,k+1} : \hat{w}_{A,k+1}$, where,

$$\begin{aligned}\hat{u}_{A,k+1} &= [i_{A,k}(1-\mu) + i_{B,k}\mu]^2 \\ &+ 0.5[i_{A,k}(1-\mu) + i_{B,k}\mu][(N - i_{A,k} - j_{A,k})(1-\mu) + (N - i_{B,k} - j_{B,k})\mu]\end{aligned}, \qquad \text{(S1)}$$

$$\begin{aligned}\hat{v}_{A,k+1} &= 2[i_{A,k}(1-\mu)+i_{B,k}\mu][j_{A,k}(1-\mu)+j_{B,k}\mu] \\ &+0.5[(N-i_{A,k}-j_{A,k})(1-\mu)+(N-i_{B,k}-j_{B,k})\mu]^2 \\ &+[i_{A,k}(1-\mu)+i_{B,k}\mu][(N-i_{A,k}-j_{A,k})(1-\mu)+(N-i_{B,k}-j_{B,k})\mu] \\ &+[j_{A,k}(1-\mu)+j_{B,k}\mu][(N-i_{A,k}-j_{A,k})(1-\mu)+(N-i_{B,k}-j_{B,k})\mu]\end{aligned}, \quad \text{(S2)}$$

$$\begin{aligned}\hat{w}_{A,k+1} &= [j_{A,k}(1-\mu)+j_{B,k}\mu]^2 \\ &+[j_{A,k}(1-\mu)+j_{B,k}\mu][(N-i_{A,k}-j_{A,k})(1-\mu)+(N-i_{B,k}-j_{B,k})\mu] \\ &+0.25[(N-i_{A,k}-j_{A,k})(1-\mu)+(N-i_{B,k}-j_{B,k})\mu]^2\end{aligned}. \quad \text{(S3)}$$

The expected genotype frequencies in the next generation are then given by eqns (4-6) and the normalizing term, $W_{A,k+1}$, is given by eqn (7). At each generation, these expected frequencies are reduced to a population of N adults consisting of $i_{A,k+1}$ wild-types and $j_{A,k+1}$ homozygotes by sampling from the multinomial distribution,

$$\begin{aligned}&\Pr(i_{k+1}, j_{k+1} \mid i_k, j_k, N) \\ &= \frac{N!}{i_{k+1}!\, j_{k+1}!(N-i_{k+1}-j_{k+1})!}\, p_{mm,k+1}^{i_{k+1}} p_{MM,k+1}^{j_{k+1}} (1-p_{mm,k+1}-p_{MM,k+1})^{N-i_{k+1}-j_{k+1}},\end{aligned} \quad \text{(S4)}$$

with an analogous procedure applying to population B. We consider a release of $j_{A,0}$ homozygotes in a population of $i_{A,0}$ wild-types at generation 0 in population A. Population B initially consists of N wild-types. Using this initial condition and the procedure described above, we can calculate the stochastic dynamics of the *Medea* element in both populations.

II. Transposable elements

Transposable elements (TEs) are particularly interesting genomic components due to their ability to transpose replicatively and hence spread into a population (Charlesworth *et al.*, 1994). TEs differ from allelic systems in that heterozygosity and homozygosity are irrelevant – TE copy number is a more meaningful variable. We treat each TE insertion as independent and keep track of the number of mosquitoes having different TE copy numbers in both populations. In population A, the proportion of mosquitoes that have i TE copies at generation k is denoted by $q_{A,i,k}$, where $i \in \{0,1,...,T\}$ and T is the maximum number of TE insertions in the genome. The corresponding proportion for population B is $q_{B,i,k}$.

Let us begin with the case without migration, assuming random mating and an infinite

population size (Marshall, 2008). We consider a system of $T+1$ difference equations, each representing the population frequency of a different TE copy number in the next generation,

$$q_{A,i,k+1} = \frac{(1-s)^i \sum_{m,n} c_{m,n,i} q_{A,m,k} q_{A,n,k}}{W_{A,k+1}} \ . \quad \text{(S5)}$$

Here, each TE copy is associated with a fitness cost of s, and fitness costs are multiplicative (assuming independence of TE copies, and to avoid a fitness cost greater than 100% at high copy numbers). The coefficient for each mating pair, $c_{m,n,i}$, represents the probability that, when a type-m mosquito (a mosquito having m TE copies) mates with a type-n mosquito, the offspring will have i TE copies. This probability is given by,

$$c_{m,n,i} = \sum_{x=0}^{i} \gamma_{m,x} \gamma_{n,i-x} \ , \quad \text{(S6)}$$

where $\gamma_{m,x}$ represents the proportion of haploid gametes of a type-m mosquito that have x TE copies. The number of TEs in the haploid gamete is determined by the number of TEs in the diploid cell that are passed on during meiosis. For a diploid cell with m TE copies we assume, to a first approximation, that all of these TEs are far enough apart from each other that they segregate independently. The probability of having x TE copies in a gamete is then proportional to the number of ways of choosing x elements from m. If a replicative transposition event has occurred in the diploid cell, then the probability of having x TE copies in the gamete is proportional to the number of ways of choosing x elements from $m+1$, or from $m-1$ if a deletion event has occurred. Following this reasoning, a type-m host produces a proportion, $\gamma_{m,x}$, of gametes having x TE copies according to the equation,

$$\gamma_{m,x} = \frac{1}{2^{m-1}}\binom{m-1}{x}\beta_m + \frac{1}{2^m}\binom{m}{x}(1-\alpha_m-\beta_m) + \frac{1}{2^{m+1}}\binom{m+1}{x}\alpha_m \ . \quad \text{(S7)}$$

The proportion, $\gamma_{n,i-x}$, is similarly defined. Here, α_m represents the proportion of gametes that are derived from cells in which a replicative transposition event has occurred, and β_m represents the proportion of gametes that are derived from cells in which a deletion event has occurred. The replicative transposition rate for a type-m host, α_m, is equal to the replicative transposition rate per TE in a type-m host, u_m, multiplied by the number of TEs in the host genome (i.e.

$\alpha_m = mu_m$). Here, u_m is a decreasing function of m to account for suppression of transposition with increasing copy number (Townsend & Hartl, 2000). We model transposition rate as an exponentially decreasing function of copy number,

$$u_m = u_1 2^{-\lambda(m-1)} \ . \tag{S8}$$

Here, λ determines the rate at which the replicative transposition rate falls off with additional TE copies. Similarly, the deletion rate for a type-m host, β_m, is equal to the deletion rate per element, v, multiplied by the number of elements in the host genome (i.e. $\beta_m = mv$), where v is generally considered a constant. Finally, the normalizing term for eqn (S5), $W_{A,k+1}$, is obtained by summing the above terms for all TE copy numbers, and is given by,

$$W_{A,k+1} = \sum_{j,m,n} (1-s)^j c_{m,n,j} q_{A,m,k} q_{A,n,k} \ . \tag{S9}$$

Analogous equations apply for population B.

Two-population model: As expected, analytic solutions to the source model are too complex to be useful. We therefore move straight on to the two-population model. We assume that the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, we make the following substitution for mosquitoes having each copy number,

$$q_{A,i,k} \leftarrow q_{A,i,k}(1-\mu) + q_{B,i,k}\mu \ . \tag{S10}$$

Eqns (S6-S8) are unchanged, and eqn (S5) becomes,

$$q_{A,i,k+1} = \frac{(1-s)^i \sum_{m,n} c_{m,n,i} [q_{A,m,k}(1-\mu) + q_{B,m,k}\mu][q_{A,n,k}(1-\mu) + q_{B,n,k}\mu]}{W_{A,k+1}} \ . \tag{S11}$$

Finally, the normalizing term, $W_{A,k+1}$, is given by,

$$W_{A,k+1} = \sum_{j,m,n} (1-s)^j c_{m,n,j} [q_{A,m,k}(1-\mu) + q_{B,m,k}\mu][q_{A,n,k}(1-\mu) + q_{B,n,k}\mu] \ . \tag{S12}$$

Analogous equations apply for population B. Considering a release of mosquitoes having one TE copy in population A at generation 0, the initial condition for the difference equations is given by,

$$(q_{A,0,0}, q_{A,1,0}) = (1-x, x) \ , \tag{S13}$$

$$q_{B,0,0} = 1 \ . \tag{S14}$$

Here, released mosquitoes represent a proportion, x, of population A at the time of release. Using this initial condition and the difference equations described above, we can calculate the time-series dynamics of the TE in both populations through numerical iteration. We consider a default transposition rate of $u_1 = 0.1$ gen^{-1} which, although high, is realistic (Seleme *et al.*, 1999; Vasilyeva *et al.*, 1999) and several modeling approaches have recommended it as a minimum requirement for gene drive to occur within a public health time frame (Rasgon & Gould, 2005; Le Rouzic & Capy, 2006; Marshall, 2008).

III. Homing endonuclease genes

Homing endonuclease genes (HEGs) spread by expressing an endonuclease which creates a double-stranded break on versions of the homologous chromosome lacking the HEG at the position where the HEG occurs. Homologous DNA repair then copies the HEG to the cut chromosome (Fig. S1E), increasing the representation of the HEG in subsequent generations (Burt, 2003). We denote the HEG allele by "*T*" and refer to the wild-type allele as "*t*." In population A, the proportion of the k th generation that are mosquitoes of genotypes *tt*, *Tt* and *TT* are denoted by $u_{A,k}$, $v_{A,k}$ and $w_{A,k}$, respectively. The corresponding proportions for population B are $u_{B,k}$, $v_{B,k}$ and $w_{B,k}$.

As for *Medea*, we begin by considering the dynamics of HEGs without migration, assuming random mating and an infinite population size (Deredec *et al.*, 2008). We additionally assume that homing occurs at meiosis after gene expression, so that additional fitness costs are not experienced by the individual in which homing occurs. The HEG allele frequency at the k th generation is then given by,

$$q_{A,k} = w_{A,k} + 0.5v_{A,k} \ , \tag{S15}$$

and the wild-type allele frequency is,

$$p_{A,k} = 1 - q_{A,k} \ . \tag{S16}$$

By considering all possible mating pairs, the genotype frequencies in the next generation in population A are given by,

$$u_{A,k+1} = p_{A,k}^2 / W_{A,k+1} \ , \tag{S17}$$

$$v_{A,k+1} = 2p_{A,k}q_{A,k}(1-e)(1-hs)/W_{A,k+1} \ , \tag{S18}$$

$$w_{A,k+1} = (q_{A,k}^2(1-s) + 2p_{A,k}q_{A,k}e(1-hs))/W_{A,k+1} \ . \quad \text{(S19)}$$

Here, e represents the homing rate of the HEG, s and hs represent the fitness costs associated with being homozygous or heterozygous for the HEG, and $W_{A,k+1}$ is a normalizing term given by,

$$W_{A,k+1} = 1 - sq_{A,k}^2 - shp_{A,k}q_{A,k} \ . \quad \text{(S20)}$$

The HEG allele frequency at the $k+1$th generation is given by,

$$q_{A,k+1} = w_{A,k+1} + 0.5v_{A,k+1} = \frac{q_{A,k}^2(1-s) + p_{A,k}q_{A,k}(1+e)(1-hs)}{1 - sq_{A,k}^2 - shp_{A,k}q_{A,k}} \ . \quad \text{(S21)}$$

An analogous equation applies to population B.

Source model: For the source model, we begin by assuming that the HEG has already reached equilibrium in population A, ignoring migratory effects. We calculate this equilibrium by solving the equality,

$$q_{A,k+1} = q_{A,k} = q_{A,*} \ . \quad \text{(S22)}$$

For simplicity, let us assume a fitness cost with a heterozygosity of $h = 0.5$, although other cases are explored for the two-population model. This gives us two solutions,

$$q_{A,*} = \{0,1\} \ . \quad \text{(S23)}$$

The first of these represents loss and the second represents fixation. Calculating the stability of these equilibria, we see that fixation is stable for fitness costs less than $2e/(1+e)$ and is unstable for fitness costs greater than $2e/(1+e)$. Loss, on the other hand, is stable for fitness costs greater than $2e/(1+e)$ and is unstable for fitness costs less than $2e/(1+e)$. This suggests that the HEG will spread into population A irrespective of the release proportion (assuming a heterozygosity of $h = 0.5$), provided that the fitness cost is not too high. For a homing rate of $e = 0.5$, for example, the maximum tolerable fitness cost is 0.667, and for a homing rate of $e = 0.75$, the maximum tolerable fitness cost is 0.857.

We assume that the fitness cost is less than $2e/(1+e)$, that population A remains fixed for the HEG, and that this population donates mosquitoes to population B at a rate, μ, measured relative to the size of population B. The genotype frequencies in population B are then given by,

$$u_{B,k+1} = p_{B,k}^2/(W_{B,k+1} + \mu) \ , \quad \text{(S24)}$$

$$v_{B,k+1} = 2p_{B,k}q_{B,k}(1-e)(1-hs)/(W_{B,k+1} + \mu) \ , \quad \text{(S25)}$$

$$w_{B,k+1} = (q_{B,k}^2(1-s) + 2p_{B,k}q_{B,k}e(1-hs) + \mu)/(W_{B,k+1} + \mu) \ . \quad \text{(S26)}$$

Eqns (S15-S16) still apply and the normalizing term, $W_{B,k+1}$, is given by eqn (S20). Following from eqn (S21), the HEG allele frequency at the $k+1$th generation in population B is given by,

$$q_{B,k+1} = w_{B,k+1} + 0.5v_{B,k+1} = \frac{q_{B,k}^2(1-s) + p_{B,k}q_{B,k}(1+e)(1-hs) + \mu}{1 - sq_{B,k}^2 - shp_{B,k}q_{B,k} + \mu} \ . \quad \text{(S27)}$$

To characterize the dynamics of the HEG in population B, we begin by calculating its equilibria by solving the equality,

$$q_{B,k+1} = q_{B,k} = q_{B,*} \ . \quad \text{(S28)}$$

This gives us two solutions,

$$q_{B,*} = \left\{1, \frac{2\mu}{s - e(2-s)}\right\} \ . \quad \text{(S29)}$$

The first of these represents fixation, and the second represents a mixture of all three genotypes. Calculating the stability of these equilibria, we see that fixation is stable for fitness costs less than $2(e+\mu)/(1+e)$ and is unstable for fitness costs greater than $2(e+\mu)/(1+e)$. Noting that we had already assumed that the fitness cost is less than $2e/(1+e)$, and is therefore less than $2(e+\mu)/(1+e)$ for $\mu > 0$, this suggests that the HEG will fix in population B and any other population it exchanges migrants with (assuming a heterozygosity of $h = 0.5$).

Two-population model: For the two-population model, we assume that the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, this means that we need to make the following substitutions,

$$(p_{A,k}, q_{A,k}) \leftarrow (p_{A,k}, q_{A,k})(1-\mu) + (p_{B,k}, q_{B,k})\mu \ , \quad \text{(S30)}$$

Applying these substitutions to eqn (S18), for illustrative purposes, we obtain,

$$v_{A,k+1} = 2[p_{A,k}(1-\mu) + p_{B,k}\mu][q_{A,k}(1-\mu) + q_{B,k}\mu](1-e)(1-hs)/W_{A,k+1} \ . \quad \text{(S31)}$$

These substitutions apply to eqns (S17-S19) and analogous substitutions apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(u_{A,0}, w_{A,0}) = (1-x, x) \ , \quad \text{(S32)}$$

$$(u_{A,0}, w_{A,0}) = (1,0) \ . \quad \text{(S33)}$$

Here, the released mosquitoes represent a proportion, x, of population A at the time of release.

Using this initial condition and the difference equations described above, we can calculate the time-series dynamics of the HEG allele in both populations.

IV. *Wolbachia*

We consider a model in which *Wolbachia*-infected males and females are equally numerous. The *Wolbachia* bacterium can therefore be described by its overall frequency in the population. At the k th generation, we denote this frequency by $q_{A,k}$ in population A and $q_{B,k}$ in population B. We begin by considering the case without migration, assuming random mating and an infinite population size (Turelli & Hoffmann, 1991). The frequency of the *Wolbachia* bacterium at the k th generation is given by,

$$q_{A,k+1} = \frac{q_{A,k}(1-u)(1-s)}{1-q_{A,k}s-q_{A,k}(1-q_{A,k})e-q_{A,k}^2 ue(1-s)} \ . \tag{S34}$$

Here, s represents the fitness cost associated with a *Wolbachia* infection, u represents the proportion of offspring of crosses between *Wolbachia*-infected females and uninfected males that are uninfected by *Wolbachia* (the degree of imperfection in maternal transmission), and e represents the proportion of offspring of crosses between uninfected females and *Wolbachia*-infected males that are sterile (the efficiency of CI-induced sterility). An analogous equation applies to population B.

Two-population model: Analytic solutions to the source model are derived by Flor *et al.* (2007) and so we move straight on to the two-population model. We assume that the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, this means that we need to make the following substitution,

$$q_{A,k} \leftarrow q_{A,k}(1-\mu)+q_{B,k}\mu \ . \tag{S35}$$

Applying this substitution to eqn (S34), we obtain,

$$q_{A,k+1} = \frac{[q_{A,k}(1-\mu)+q_{B,k}\mu](1-u)(1-s)}{\begin{aligned}1-[q_{A,k}(1-\mu)+q_{B,k}\mu]s-[q_{A,k}(1-\mu)+q_{B,k}\mu]^2 ue(1-s)\\ -[q_{A,k}(1-\mu)+q_{B,k}\mu](1-[q_{A,k}(1-\mu)+q_{B,k}\mu])e\end{aligned}} \ . \tag{S36}$$

An analogous substitution applies to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$q_{A,0} = x \ , \tag{S37}$$

$$q_{B,0} = 0 \ . \tag{S38}$$

Here, the released mosquitoes represent a proportion, x, of population A at the time of release. Using this initial condition and the difference equations described above, we can calculate the time-series dynamics of the *Wolbachia* infection in both populations.

V. Engineered underdominance (single-locus)

A novel form of underdominance has been proposed by Davis *et al.* (2001) consisting of two transgenic constructs, each of which possesses a gene whose expression induces lethality and a gene that suppresses the expression or activity of the gene inducing lethality carried by the other construct. The constructs can either be inserted at the same locus on a pair of homologous chromosomes or at different loci on nonhomologous chromosomes. Here, we consider the single-locus system. As a three-allele system – two transgenic alleles, "*T*" and "*R*," and one wild-type allele, "*t*" – there are six possible genotypes; however since individuals possessing only one transgenic construct express a lethal gene without its suppressor, only two of these genotypes are viable – *tt* and *TR*. In population A, we denote the proportion of the k th generation that are mosquitoes having these genotypes by $u_{A,k}$ and $w_{A,k}$, respectively. The corresponding proportions for population B are $u_{B,k}$ and $w_{B,k}$. The population dynamics of single-locus engineered underdominance in a single randomly-mating population are described by eqns (47-48).

Two-population model: For the two-population model, we assume that the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, this means that we need to make the following substitutions,

$$(u_{A,k}, w_{A,k}) \leftarrow (u_{A,k}, w_{A,k})(1-\mu) + (u_{B,k}, w_{B,k})\mu \ , \tag{S39}$$

Applying these substitutions to eqn (47), for illustrative purposes, we obtain,

$$u_{A,k+1} = \frac{[u_{A,k}(1-\mu) + u_{B,k}\mu]^2}{[u_{A,k}(1-\mu) + u_{B,k}\mu]^2 + 0.5[w_{A,k}(1-\mu) + w_{B,k}\mu]^2(1-s)} \ . \tag{S40}$$

These substitutions also apply to eqn (48) and analogous substitutions apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(u_{A,0}, w_{A,0}) = (1-x, x) \ , \tag{S41}$$

$$(u_{A,0}, w_{A,0}) = (1,0) \ . \tag{S42}$$

Here, the released mosquitoes represent a proportion, x, of population A at the time of release. Using these equations, we can calculate the time-series dynamics of the engineered underdominance alleles in both populations.

VI. Underdominance (single-allele)

Molecular genetic strategies are currently being investigated to engineer underdominant alleles for which hybrids are completely unviable (Hay *et al.*, unpublished). We consider an introduced allele denoted by "*T*" and a null allele denoted by "*t*." In population A, the proportion of the k th generation that are mosquitoes of genotypes *tt* and *TT* are denoted by $u_{A,k}$ and $w_{A,k}$, respectively. The corresponding proportions for population B are $u_{B,k}$ and $w_{B,k}$. Considering the case without migration, the genotype frequencies in the next generation in population A are given by,

$$u_{A,k+1} = u_{A,k}^2 / (u_{A,k}^2 + w_{A,k}^2 (1-s)) \ , \tag{S43}$$

$$w_{A,k+1} = w_{A,k}^2 (1-s) / (u_{A,k}^2 + w_{A,k}^2 (1-s)) \ . \tag{S44}$$

Analogous equations apply to population B.

For the source model, we begin by assuming that the transgene has already reached equilibrium in population A, ignoring migratory effects, and calculate the equilibrium,

$$w_{A,k+1} = w_{A,k} = w_{A,*} \ . \tag{S45}$$

This gives us three solutions,

$$w_{A,*} = \left\{ 0, \frac{1}{2-s}, 1 \right\} \ . \tag{S46}$$

The first of these represents loss, the second represents a mixture of homozygotes and wild-types, and the third represents fixation. Calculating the stabilities of these equilibria, we see that fixation and loss are stable, while the internal equilibrium is unstable. This latter solution represents a release threshold, above which the introduced allele becomes fixed in the population. In the absence of a fitness cost, the release threshold is 50%, and with a fitness cost of $s = 0.05$, the release threshold is 51.3%.

Source model: Under the source model, we assume that population A remains fixed for

the introduced allele and that this population donates homozygotes to population B at a rate, μ, measured relative to the size of population B. The genotype frequencies in population B are then given by,

$$u_{B,k+1} = u_{B,k}^2 / (u_{B,k}^2 + w_{B,k}^2 (1-s) + \mu) \ , \tag{S47}$$

$$w_{B,k+1} = (w_{B,k}^2 (1-s) + \mu) / (u_{B,k}^2 + w_{B,k}^2 (1-s) + \mu) \ . \tag{S48}$$

To characterize the dynamics of the introduced allele in population B, we begin by calculating the equilibria,

$$w_{B,k+1} = w_{B,k} = w_{B,*} \ . \tag{S49}$$

This gives us three solutions,

$$w_{B,*} = \left\{ 1, \frac{1-\sqrt{1-4\mu(2-s)}}{4-2s}, \frac{1+\sqrt{1-4\mu(2-s)}}{4-2s} \right\} \ . \tag{S50}$$

The first of these represents fixation, and the second and third represent a mixture of homozygotes and wild-types. Calculating the stabilities of these equilibria, we see that fixation is stable, and the other two equilibria only exist when $\mu \le 1/(8-4s)$. If this condition is satisfied, then the second equilibrium is stable and the third is unstable. This implies that,

$$w_{B,*} = \begin{cases} \dfrac{1-\sqrt{1-4\mu(2-s)}}{4-2s}, \mu \le \dfrac{1}{8-4s} \\ 1, \mu > \dfrac{1}{8-4s} \end{cases} \ . \tag{S51}$$

The expression, $1/(8-4s)$, represents a migration threshold, equal to 12.5% per generation in the absence of a fitness cost and 13.2% per generation for a fitness cost of $s = 0.05$.

Two-population model: For the two-population model, we assume that the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, this means that we need to make the following substitutions,

$$(u_{A,k}, w_{A,k}) \leftarrow (u_{A,k}, w_{A,k})(1-\mu) + (u_{B,k}, w_{A,k})\mu \ . \tag{S52}$$

Applying these substitutions to eqn (S47), for illustrative purposes, we obtain,

$$u_{A,k+1} = \frac{[u_{A,k}(1-\mu) + u_{B,k}\mu]^2}{[u_{A,k}(1-\mu) + u_{B,k}\mu]^2 + [w_{A,k}(1-\mu) + w_{B,k}\mu]^2 (1-s)} \ . \tag{S53}$$

These substitutions also apply to eqn (S48) and analogous substitutions apply to population B.

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

$$(u_{A,0}, w_{A,0}) = (1-x, x) \ , \tag{S54}$$

$$(u_{A,0}, w_{A,0}) = (1,0) \ . \tag{S55}$$

Here, the released mosquitoes represent a proportion, x, of population A at the time of release. Using these equations, we can calculate the time-series dynamics of the underdominant allele in both populations.

VII. References

Burt, A., 2003. Site-specific genes as tools for the control and genetic engineering of natural populations. Proc. Biol. Sci. **270**, 921-928.

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

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

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

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

Flor, M., Hammerstein, P., Telschow, A., 2007. Dynamics and stability of *Wolbachia*-induced unidirectional cytoplasmic incompatibility in parapatric host populations. J. Evol. Biol. **20**, 696-706.

Le Rouzic, A., Capy, P., 2005. The first steps of a transposable element invasion: Parasitic strategy vs. drift. Genetics **169**, 1033-1043.

Marshall, J. M., 2008. The impact of dissociation on transposon-mediated disease control strategies. Genetics **178**, 1673-1682.

Rasgon, J. L., Gould, F., 2005. Transposable element insertion location bias and the dynamics of gene drive in mosquito populations. Insect Mol. Biol. **14**, 493-500.

Seleme, M., Busseau, I., Malinsky, S., Bucheton, A., Teninges, D., 1999. High-frequency retrotransposition of a marked *I* factor in *Drosophila melanogaster* correlates with a

dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. Genetics **151**, 761-771.

Service, M. W., 1993. Mosquito Ecology: Field Sampling Methods. Chapman and Hall, London.

Townsend, J. P., Hartl, D. L., 2000. The kinetics of transposable element autoregulation. Genetica **108**, 229-237.

Turelli, M., Hoffmann, A. A., 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature **353**, 440-442.

Vasilyeva, L. A., Bubenshchikova, E. V., Ratner, V. A., 1999. Heavy heat shock induced retrotransposon transposition in *Drosophila*. Genet. Res. **74**, 111–119.

Figure S1:

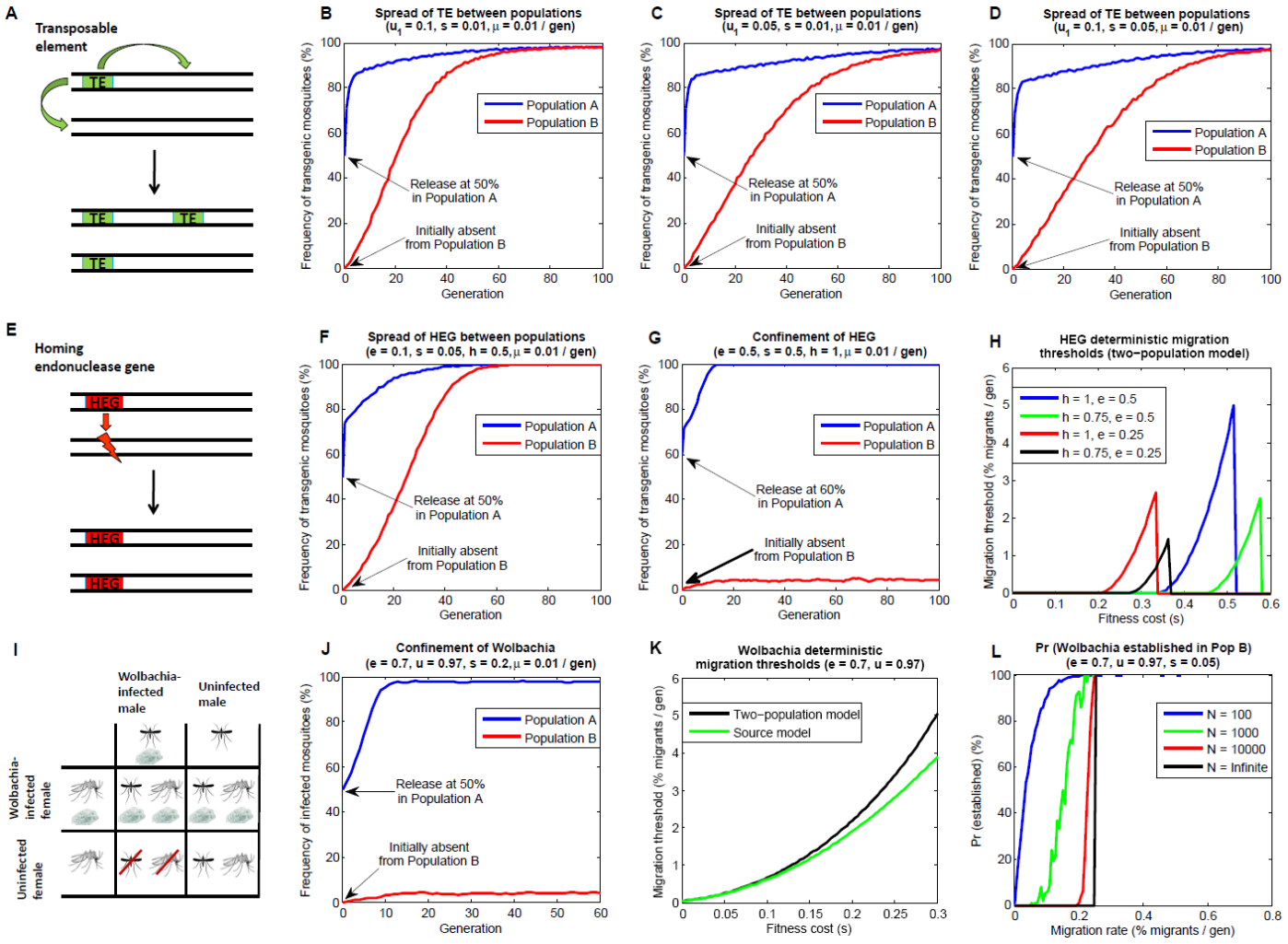

Fig. S1. Dynamics of invasive gene drive systems under the source and two-population models. (A): A transposable element (TE) expresses a transposase which facilitates its replicative transposition throughout the genome. At higher copy numbers it is represented more frequently in subsequent generations. (B): Single stochastic realization of a TE released at 50% in population A. Each of the released mosquitoes has four copies of the TE. Default parameters are used to characterize the TE ($u_1 = 0.1$, $v = 4\times10^{-6}$, $s = 0.01$). Population A exchanges migrants with population B at a rate of 0.1% per generation and the size of both populations is 10,000. The TE reaches near-fixation in both populations within 100 generations. (C): If the transposition rate is decreased (here, $u_1 = 0.05$), the TE spreads more slowly but still becomes established in

both populations. (D): If the fitness cost is increased (here, $s = 0.5$ and $u_1 = 0.1$), the TE also spreads more slowly but becomes established in both populations. However, if the fitness cost outweighs the TE replication rate, the TE is lost from both populations (not shown). (E): A homing endonucease gene (HEG) expresses an endonuclease which causes chromosome cleavage followed by gene conversion, increasing the representation of the HEG in subsequent generations. (F): Single stochastic realization of a HEG released at 50% (homozygotes) in population A. Default parameters are used to characterize the HEG ($e = 0.1,\ s = 0.05,\ h = 0.5$). Population A exchanges migrants with population B at a rate of 0.1% per generation and the size of both populations is 10,000. The HEG fixes in both populations within 90 generations. (G): Threshold dynamics are possible for HEGs with partially dominant fitness costs ($h > 0.5$). Here, the HEG is associated with a higher, dominant fitness cost ($s = 0.5,\ h = 1$) and homing efficiency ($e = 0.5$). The HEG reaches near-fixation in population A within 15 generations, but only spreads to ~5% (in the form of HEG homozygotes and heterozygotes) in population B. (H): Depending on the homing efficiency, HEGs display threshold behavior with respect to migration rate for a small range of partially or fully-dominant fitness costs. (I): *Wolbachia* is a maternally-inherited intracellular bacterium which causes offspring of matings between infected males and uninfected females to have reduced viability, thus increasing its representation in subsequent generations. (J): Single stochastic realization of a *Wolbachia* infection released at 50% in population A. Default parameters are used to characterize the bacterium ($e = 0.7,\ u = 0.97$). Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. When the infection is associated with a fitness cost of $s = 0.2$, *Wolbachia* reaches near-fixation in population A within 12 generations but only spreads to ~4% in population B. (K): *Wolbachia* displays threshold behavior with respect to migration rate. (L): For smaller population sizes, there is a chance that *Wolbachia* will become established in neighboring populations for smaller migration rates.