Inverse Medea as a Novel Gene Drive System for Local Population Replacement: A Theoretical Analysis

JOHN M. MARSHALL AND BRUCE A. HAY

One strategy to control mosquito-borne diseases, such as malaria and dengue fever, on a regional scale is to utilize gene drive systems to spread disease-refractory genes into wild mosquito populations. The development of a synthetic Medea element that has been shown to drive population replacement in laboratory Drosophila populations has provided encouragement for this strategy but has also been greeted with caution over the concern that transgenes may spread into countries without their consent. Here, we propose a novel gene drive system, inverse Medea, which is strong enough to bring about local population replacement but is unable to establish itself beyond an isolated release site. The system consists of 2 genetic components—a zygotic toxin and maternal antidote—which render heterozygous offspring of wild-type mothers unviable. Through population genetic analysis, we show that inverse Medea will only spread when it represents a majority of the alleles in a population. The element is best located on an autosome and will spread to fixation provided any associated fitness costs are dominant and to very high frequency otherwise. We suggest molecular tools that could be used to build the inverse Medea system and discuss its utility for a confined release of transgenic mosquitoes.

Key words: dengue fever, malaria, population replacement, transgenic mosquitoes, underdominance

Mosquito-borne diseases such as malaria and dengue fever continue to pose a major health problem through much of the world. Large-scale control of these diseases remains elusive and, consequently, there is interest in strategies that utilize gene drive systems to spread disease-refractory genes into wild mosquito populations on a regional scale (Alphey et al. 2002; Marshall and Taylor 2009). Synthetic Medea elements have been shown to drive population replacement in laboratory Drosophila populations (Chen et al. 2007) and are predicted to spread from low frequencies (Ward et al. 2010). These results provide encouragement for the population replacement strategy; but raise the possibility that Medea-linked transgenes may spread into countries without their consent (Marshall 2010). Consequently, there is also interest in gene drive systems that, while strong enough to bring about population replacement at an isolated release site, are unable to establish themselves in neighboring populations. Here, we propose a novel gene drive system, inverse Medea, which displays these properties. The system consists of 2 genetic components—a zygotic toxin and a maternal antidote—which render heterozygous offspring of wild-type mothers unviable (Figure 1A). Through population genetic analysis, we show that inverse Medea will only spread when it represents a majority of the alleles in a population and, if engineered, could provide an important vehicle for bringing about local population replacement.

To characterize the basic dynamics of the inverse Medea system, we consider the element as a single allele, which we denote by $M$, and refer to the corresponding position on the wild-type chromosome as $m$. We use a system of discrete-generation difference equations to model the spread of the element through a population, assuming random mating, infinite population size, 100% toxin efficiency, and equal fitness costs in males and females having the element. The assumption of 100% toxin efficiency is justified because there are many ways to induce cell death in the embryo—through the expression of proteins that induce apoptosis or through the expression of dsRNA that brings about the loss of essential cell death inhibitors (e.g., Wang et al. 1999; Huh et al. 2004). All mating pairs produce equal numbers of male and female offspring and so, even if the gender ratio is initially unequal, it will be identical from the second generation on. This allows us to denote the proportions of the $k$th generation that are individuals of genotypes $mm$, $Mm$, and $MM$ by $u_k$, $v_k$, and $w_k$, respectively, independent of gender. By considering all possible mating
pairs and removing unviable Mm offspring of mm females, the genotype frequencies in the next generation are given by,

\[ u_{k+1} = \left( u_k^2 + 0.25v_k^2 + u_kv_k \right) / W_{k+1}, \tag{1} \]

\[ v_{k+1} = \left( 0.5u_k^2 + v_kw_k + 0.5u_kv_k + u_kw_k \right) (1 - hs) / W_{k+1}, \tag{2} \]

\[ w_{k+1} = \left( u_k^2 + 0.25v_k^2 + v_kw_k \right) (1 - s) / W_{k+1}. \tag{3} \]

Here, \( s \) and \( hs \) represent the fitness costs associated with being homozygous or heterozygous for the inverse Medea element, and \( W_{k+1} \) is a normalizing term given by,

\[ W_{k+1} = u_k^2 + 0.25v_k^2 + u_kv_k + (u_k^2 + 0.25v_k^2 + v_kw_k)(1 - s) + (0.5u_k^2 + v_kw_k + 0.5u_kv_k + u_kw_k)(1 - hs). \tag{4} \]

We begin by calculating the equilibria that an inverse Medea allele reaches in a population by solving the equality,

\[ (u_{k+1}, w_{k+1}) = (u_k, w_k) = (u_0, w_0). \tag{5} \]

We then calculate the stabilities of these equilibrium points by calculating the eigenvalues of the Jacobian matrix.

Figure 1. Population dynamics of the inverse Medea system. (A) The element is comprised 2 genes—a zygotic toxin and maternal antidote. The expression of these genes renders heterozygous offspring of wild-type females unviable. (B) For the case of an additive fitness cost of 0.05 \((b = 0.5, s = 0.05)\), there exists a family of threshold points (separatrix), above which the element spreads to a stable equilibrium and below which the element is lost from the population. At the stable equilibrium, 99.7% of individuals are either homozygous or heterozygous for the element \((v_{u+u} = 0.997)\). (C) For the case of a dominant fitness cost of 0.05 \((b = 1, s = 0.05)\), a separatrix also exists; however, the element fixes in the population following a super-threshold release. (D) As the size of the dominant fitness cost increases \((b = 1, s \in [0, 0.05, 0.1, 0.2])\), the release threshold increases, as visualized by the separatrix moving toward the right of the de Finetti diagram, and the element still fixes in the population. (E) Spread of inverse Medea can be confined to its release site. If released at a frequency of 0.6 \((u_0 = 0.6, w_0 = 0.6)\) in a population that exchanges migrants with a neighboring population at a rate of \(\mu = 0.01\) per generation, the element is predicted to spread in the release population but to persist only at low levels \((a frequency of \approx 0.15\) in the neighboring population. (F) High migration rates lead to loss of inverse Medea from both populations. If released at a frequency of 0.8 \((u_0 = 0.2, w_0 = 0.8)\) in a population that exchanges migrants with a neighboring population at a rate of \(\mu = 0.025\) per generation, the element is predicted to be eliminated from both populations within \(\approx 100\) generations.
The equilibrium is locally stable if all eigenvalues have modulus less than one and is unstable if one or more of the eigenvalues have modulus greater than one (Elyard 1995). For simplicity, we consider 2 cases in detail—an additive fitness cost for which each allele copy is equally costly ($b = 0.5$) and a dominant fitness cost for which heterozygotes and homozygotes have the same cost ($b = 1$). For an additive fitness cost, there are 4 biologically feasible equilibria; however, only fixation, $(u_s, v_s) = (0, 1)$, and loss, $(u_s, v_s) = (1, 0)$, are simple enough to express symbolically. Loss is stable under all parameterizations ($s \in [0, 1]$), and fixation is unstable unless the fitness cost is zero ($s = 0$), in which case it is stable. The third biologically feasible equilibrium represents an unstable genotype distribution, above which the element spreads to fixation or near-fixation and below which it is lost from the population. This solution represents one of a family of thresholds, collectively referred to as a separatrix (Figure 1B). The fourth biologically feasible equilibrium represents a stable genotype distribution toward which the inverse Medea allele converges if it is initially present at super-threshold frequencies. This is the stable equilibrium depicted in Figure 1B consisting mostly of MM homozygotes, a small number of heterozygotes and minimal wild types—for $s = 0.05$, the stable equilibrium is $(u_s, v_s, w_s) = (0.003, 0.103, 0.894)$. These last 2 equilibria exist provided that the homozygous fitness cost satisfies $s < 0.138$. The split between homozygotes and heterozygotes is only relevant if refractory phenotypes are recessive; however, the refractory genes currently being investigated work through dominant mechanisms (e.g., Ito et al. 2002; Franz et al. 2006; Corby-Harris et al. 2010).

Dominant fitness costs are most likely because the zygotic toxin, maternal antidote, and refractory gene are all expected to function through a dominant mechanism, even if this involves silencing the expression of an endogenous gene. For the case of a dominant fitness cost, there are 4 equilibria, all of which can be expressed symbolically,

$$
\frac{\partial u_{s+1}}{\partial u_s} \frac{\partial u_{s+1}}{\partial v_s} \frac{\partial u_{s+1}}{\partial w_s} = \left( u_s, v_s \right)
$$

where $s = \sqrt{2 - 2r}$. The first of these equilibria represents fixation, the second represents loss, the upper sign of the third is not biologically feasible, and the lower sign of the third represents one of a family of points on a separatrix, above which the element spreads to fixation and below which the element is lost (Figure 1C). Loss is stable under all parameterizations ($s \in [0, 1]$), and fixation is stable for fitness costs that satisfy $s < 0.5$, which is also a condition for the existence of the unstable equilibrium.

We assume a release of males and females homozygous for the inverse Medea allele representing a fraction, $x$, of the total population. The initial condition for such a release is $(u_0, v_0) = (1 - x, x)$. For a population size of $N$, this corresponds to a release of $X$ mosquitoes (half male and half female) having genotype MM such that $x = X/(N + X)$. The expected outcomes for a variety of release sizes can be predicted by iterating Equations 1–4 subject to corresponding initial conditions. In the absence of a fitness cost, the inverse Medea allele is predicted to spread into a population for release frequencies greater than 0.5. For a release frequency of 0.6, transgenic individuals are expected to represent 95% of the population within 11 generations and 99% of the population within 26 generations. If the allele is then diluted through mass release of wild-type insects, it will be driven out of the population once its frequency falls below 0.5.

Dominant fitness costs are preferable to additive ones because, all else being equal, these constructs have lower release thresholds, spread more quickly, and spread to higher frequencies than constructs with additive costs. For a dominant fitness cost of $s = 0.05$, the release threshold frequency is 0.538, and following a release at a population frequency of 0.6, the population is expected to be 99% transgenic within 31 generations. For a comparable additive fitness cost ($b = 0.5, s = 0.05$), the release threshold frequency is higher (0.541) despite heterozygotes suffering a lower fitness cost ($hs = 0.025$ for additive costs cf. $hs = 0.05$ for dominant costs). Additionally, the population never becomes fully transgenic for additive fitness costs, and it is expected to only reach 99% transgenic for a release frequency of 0.6.

The trends of increasing release threshold, decreasing speed of spread, and decreasing frequency to which the inverse Medea allele equilibrate are seen over the full range of heterozygosities from $b = 1$ to $b = 0$. For a comparable recessive fitness cost ($b = 0, s = 0.05$), the release threshold frequency is higher again (0.545) and the population is expected to take 48 generations to become 98% transgenic for a release frequency of 0.6, never quite reaching a transgenic frequency of 0.99. Increasing the fitness cost produces the same trends, and population dynamics for a range of dominant fitness costs are depicted in Figure 1D. Here, as the fitness cost increases, the separatrix moves to the right of the de Finetti diagram, resulting in a higher release threshold and slower speed of spread. For a dominant fitness cost of $s = 0.1$, the release threshold frequency is 0.577, and following a release at a population frequency of 0.6, the population is expected to be 99% transgenic within 42 generations (cf. 31 generations for $b = 1, s = 0.06$).

Also of interest is the ability to confine the spread of an inverse Medea allele to a partially isolated population. To model this, we consider a 2-population model in which mosquitoes homozygous for the inverse Medea allele are released in population A, population B is initially wild type, and the mating pool of both populations is made up of individuals from both populations. For a migration rate of $\mu$ in both directions, this leads to the following substitutions in Equations 1–4 for population A,
Applying these substitutions to Equation 1, a dominant fitness cost of Medea be more isolated than this. For an inverse generation (Taylor et al. 2001), and trial sites are expected to have 10,000 individuals, and genotypes are sampled from homozygotes) and to reach 0.6 in population A, the frequency of transgenics is expected to reach 0.97 in population A (Hay BA, unpublished data). We use a multinomial distribution at each generation (Marshall JM, Mali, West Africa have been estimated at 0.008 per generation (Taylor et al. 2001), and trial sites are expected to be more isolated than this. For an inverse Medea allele with a dominant fitness cost of $s = 0.05$ and a release frequency of 0.6 in population A, the frequency of transgenics is expected to reach $0.97$ in population A ($\sim 0.8$ of which are homozygotes) and $\sim 0.015$ in population B. For a migration rate of $\mu = 0.02$ per generation, the frequency of transgenics is expected to stabilize at $\sim 0.031$ in population B.

For migration rates higher than $\mu = 0.021$ per generation, the migrants from population B dilute the transgenic individuals in population A such that the frequency of the inverse Medea allele eventually falls below the threshold required for spread, resulting in its elimination from both populations. A stochastic realization of this is depicted in Figure 1F, where the migration rate between populations is $\mu = 0.025$ per generation. This leads to the element being eliminated from both populations within $\sim 100$ generations following a release at a population frequency of 0.8. Similar dynamics occur in the absence of a fitness cost, in which case the loss threshold due to migration is $\mu = 0.027$ per generation. These results highlight the importance of accurate ecological measurements of migration rate because the inverse Medea element could be lost if the migration rate is underestimated. Preliminary analysis suggests that the relative sizes of populations A and B can also have significant effects on model predictions. For example, if population A has 15,000 individuals and population B has 5000 individuals, the loss threshold due to migration increases from $\mu = 0.021$ to $\mu = 0.027$ per generation (assuming $b = 1$, $s = 0.05$); however, if the population sizes are reversed, the loss threshold decreases to $\mu = 0.007$ per generation. An accurate ecological assessment will require understanding of fluctuating population sizes and migration rates, phenomena observed with several pest species (Triplet et al. 2005). The above modeling framework can be used to describe the population genetics of an X-linked inverse Medea allele, $X^M$. In this case, there are 5 genotypes that we need to keep track of—$X^sX^s$, $X^sY$, $X^dX^s$, $X^dY$, and $X^dX^M$—the $k$th generation proportions of which we denote by $u_{m,k}, v_{m,k}, w_{m,k}$, $v_{m,k},$ and $w_{j,k}$, respectively. By considering all possible mating pairs, the genotype frequencies in the next generation are given by,

$$u_{m,k+1} = (2u_{m,k}u_{j,k} + n_{u,k}v_{j,k})/W_{k+1},$$

$$v_{m,k+1} = (2u_{m,k}v_{j,k} + n_{u,k}v_{j,k} + 2v_{m,k}u_{j,k})/W_{k+1},$$

$$v_{m,k+1} = (2v_{m,k}u_{j,k} + n_{u,k}v_{j,k} + 2v_{m,k}u_{j,k})/W_{k+1},$$

where, $W_{k+1}$ is given by,

$$W_{k+1} = (4u_{m,k}u_{j,k} + 2v_{m,k}u_{j,k} + 2v_{m,k}v_{j,k} + 2u_{m,k}u_{j,k}) + (4u_{m,k}v_{j,k} + 2v_{m,k}v_{j,k} + 2u_{m,k}v_{j,k} + 2v_{m,k}u_{j,k})(1 - 2s).$$

Iterating Equations 12–17 for a variety of release sizes and fitness costs, we see that X-linked inverse Medea has similar properties to the autosomal element but spreads more slowly, less completely, and has no predicted benefits. In the absence of fitness costs, the X-linked element has a release threshold frequency of 0.5 and for a release frequency of 0.6 takes 45 generations to reach a transgenic frequency of 0.95 and 175 generations to reach a transgenic frequency of 0.99. Its spread is dramatically slowed in the presence of a dominant transgenic frequency of 0.967 in the presence of a dominant fitness cost of $s = 0.05$ (cf. 100% for the autosomal element). These results suggest that the inverse Medea element is best located on an autosome.

Several approaches are available to engineer the inverse Medea system. Maternal germline-specific promoters can be used to drive expression of antidotes that will be loaded into the developing oocyte/zygote, whereas transient early zygotic promoters can be used to express toxins within...
a time window in which their expression or activity can still be suppressed by a maternally deposited antidote (Chen et al. 2007). The zygotic toxin could encode a protein, in which case the antidote could be a maternally supplied microRNA that promotes degradation of the toxin transcript prior to translation. The antidote could also be a protein that sequesters, degrades or neutralizes the toxin, or limits the consequences of its activity. Alternatively, the toxin could be a microRNA that silences expression of a gene whose activity is required for early development, in which case the antidote could be a maternally expressed RNA that provides the necessary activity to the zygote and is resistant to silencing.

To illustrate the engineering process, we consider a possible example in Drosophila utilizing the caspase protease inhibitor Drosophila Inhibitor of Apoptosis 1 (DIAP1), which is required zygotically for cell and embryo viability during early gastrulation (Wang et al. 1999). DIAP1’s pro-survival activity can be inhibited by expression of proteins of the RHG family, which promote DIAP1 degradation and/or prevent DIAP1 from interacting with caspases, both of which result in caspase-dependent cell death (Wang et al. 1999; Yoo et al. 2002). The same effect can be achieved through microRNA-dependent degradation of the zygotic DIAP1 transcript (Huh et al. 2004). In either case, it should be possible to rescue cell and embryo death resulting from loss of zygotic DIAP1 through maternal expression of the baculovirus p35 protein—a broad specificity caspase inhibitor that inhibits cell death due to loss of DIAP1 (Huh et al. 2004). Finally, we note that, as with Medea, the appearance of antidote-only inverse Medea alleles through DNA breakage and rejoining can be minimized by placing the toxin and disease-refractory genes within an intron of the antidote gene, thereby requiring 2 nearby recombination events in order to produce an element that carries a functional antidote but no toxin (Chen et al. 2007).

Given the available approaches for engineering the inverse Medea system, and its predicted ability to spread at its release site without substantially contaminating neighboring populations, we propose this system as a drive system, and its predicted ability to spread at a super-threshold frequency into one cage, setting up a second cage consisting of wild types, and exchanging 1% of the individuals from each cage at each generation. If the predictions of these models are correct, then within 20 generations the transgene should be seen to spread in the release cage, with the wild-type cage remaining largely unchanged.

Funding
National Institutes of Health (DP1 OD003878 to B.A.H.).

Acknowledgments
The authors would like to thank Geoffrey Pittman for helpful discussions on how to build the inverse Medea system, and 2 anonymous reviewers whose constructive comments have improved the manuscript.

References
Alphay L, Beard CB, Billingsley P, Coetzee M, Crisanti A, Curtis C, Eggleson P, Godfray C, Hemingway J, Jacobs-Lorena M, et al. 2002. Malaria control with genetically manipulated insect vectors. Science. 298:119–121.
Chen CH, Huang H, Ward CM, Su JT, Schaeffer LV, Guo M, Hay BA. 2007. A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. Science. 316:597–600.
Corby-Harris V, Drexler A, Watkins de Jong I, Antonova Y, Palpour N, Ziegler R, Ramberg F, Lewis EE, Brown JM, Luckhart S, 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.
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.
Elyash SN. 1995. An introduction to difference equations. New York: Springer.
Franz AW, Sanchez-Vargas I, Adelman ZN, Blair CD, Beaty BJ, James AA, Olson KE. 2006. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. Proc Natl Acad Sci U S A. 103:4198–4203.
Gould F, Schliefelkam P. 2004. Population genetics of autocidal control and strain replacement. Annu Rev Entomol. 49:193–217.
Huh JR, Guo M, Hay BA. 2004. Compensatory proliferation induced by cell death in the Drosophila wind disc requires activity of the apical cell death caspase Dronc in a nonapoptotic role. Curr Biol. 14:1262–1266.
Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M. 2002. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature. 417:452–455.
Marshall JM. 2010. The Cartagena Protocol and genetically modified mosquitoes. Nat Biotechnol. 28:896–897.
Marshall JM, Taylor CE. 2009. Malaria control with transgenic mosquitoes. PLoS Med. 6:e20.
Marshall JM, Pittman GW, Buchman AB, Hay BA. 2011. Semele: a killer-male, rescue-female system for suppression and replacement of insect disease vector populations. Genetics. 187:535–551.
Robinson AS. 1976. Progress in the use of chromosomal translocations for the control of insect pests. Biol Rev. 51:1–24.
Taylor CE, Toure YT, Carnahan J, Norris DE, Dolo G, Traore SF, Edillo FE, Lanzaro GC. 2001. Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics. 157:743–750.
Triet F, Dolo G, Lanzaro G. 2005. Multilevel analyses of genetic differentiation in *Anopheles gambiae* s.s. reveal patterns of gene flow important for malaria-fighting mosquito projects. Genetics. 169:313–314.

Wang SL, Hawkins CJ, Yoo SJ, Muller HA, Hay BA. 1999. The Drosophila caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. Cell. 98:453–463.

Ward CM, Su JT, Huang Y, Lloyd AL, Gould F, Hay BA. 2010. Medea selfish genetic elements as tools for altering traits of wild populations: a theoretical analysis. Evolution. doi:10.1111/j.1558-5646.2010.01186.x.

Yoo SJ, Huh JR, Muro I, Yu H, Wang L, Wang SL, Feldman RM, Clem RJ, Muller HA, Hay BA. 2002. Hid, Rpr and Grim negatively regulate DIAP1 levels through distinct mechanisms. Nat Cell Biol. 4:416–424.

Received November 14, 2010; Revised February 10, 2011; Accepted March 1, 2011

Corresponding Editor: Dr Howard Ross