Medea selfish genetic elements as tools for altering traits of wild populations: A theoretical analysis

Thesis by

Catherine Marie Ward

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, CA

2011

(Defended December 20, 2010)

Acknowledgements:

Science never gets done in a vacuum and the projects described in this thesis are no exception. These folks have not only made these projects possible, but they have made enriched my Caltech experience.

Many thanks to my adviser, Dr. Bruce Hay, who allowed me to explore a variety of topics from looking at basic genetic interactions between cell death proteins and immunity, to attempting to create a gene drive system in flies, to doing microinjections with mosquitoes, and finally allowing me to get an nearly purely theoretical PhD. His willingness to let his lab go in a huge variety of directions is a testament to his intelligence (as he has the capacity to advise on an enormous number of topics) and his willingness to take risks.

My committee, Dr. Michael Elowitz, Dr. Richard Murray and Dr. Paul Sternberg, have allowed me to bring them different topics at each meeting and have opened their labs to me for any help that I have needed.

Bruce's lab has been a wonderful and collaborative place to work. I especially want to thank Kelly Dusinberre who has done the graduate school journey with me (you'll be done by June, I know you will) and has been a great source of support with extremely useful practical suggestions, great troubleshooting skills, a dry sense of humor and a willingness to get her hands into any problem. Hiaxia Huang, Jeff Copeland, Jun Huh, Israel Muro and Chun Hong Chen showed me the ropes when I joined the lab. Jessica

Su and John Marshall proved that our lab actually can model. Omar Okbari, Katie Kennedy, Ken Chan, Nikolai Kandul, Geoff Pittman and Arun Kumar have provided great conversation and fun.

I am very lucky to have a second lab family at NC State University. I worked for Dr. Fred Gould as an undergraduate trying to find QTLs associated with various phenotypes in moths. Unbelievably, when I joined Bruce's lab, the first paper Bruce gave me to read was one Fred wrote on genetic pest management. Fred has encouraged me to spend time in with is group including Mathieu Legros, Yunxin Huang, and Michael Roberts. He also introduced me to Dr. Alun Lloyd, a professor in Biomath at NC State. They have brainstormed ways to think about problems, helped me present data, allowed me to spend time in their offices, and co-authored papers with me. My graduate school experience would have been very different without them.

I also want to thank a multitude of friends in the world outside of lab including Shane Murphy, Karen Wawrousek, Christine Kara, Geoff Pittman and their kids Patu and Alela, John and Niki Galownia, folks on both the GSC and GRB (especially Eve Stenson who co-chaired the GRB with me for several years) and the people on the Caltech Animal Team.

Two teachers in high school deserve special recognition. First is Dr. Mary Eubanks who taught intro biology and then got a position at Duke and invited me to work in her lab. She demonstrated that just because your idea is crazy and/or unpopular does not mean

you are crazy or wrong. She also continues to inspire me with her breadth of knowledge from archeology to basic evolutionary genetics to her skills in creating crop improvements. Mr. Jim Ebert taught me multivariable calculus, non-Euclidean geometry and statistics. His unflinching style ("Don't get mad because you're stupid") and commitment to teaching students not just how to pass a test, but how to understand how mathematical concepts are discovered continue to help me in data analysis today.

My parents, Thomas and Denise Ward, have been enormous supporters of education and of me from the day I was born. From learning that I had excessive ATPs (too much energy) as a toddler, to feeding birds, deer and wildlife in the backyard, to keeping popular science magazines in the house, to simply having an understanding of the amount of work it takes to get an graduate degree (They have Masters Degrees in Biology and Chemistry. Mommy taught nursing students at Durham Tech and Daddy worked as a neurotoxicologist at the EPA.), they have been unwavering in their love and support. And, yes, Daddy, I might get a real job now.

And finally, thanks to my husband, John Carpenter. He has graciously listened to me discuss all manner of ideas for dealing with data and, as a chemical engineer, has provided novel ideas about how to attack the problems. We have survived (and

occasionally even enjoyed) huge amounts of homework, changes in projects, and now as I finish this thesis, writing 2 theses. He's celebrated my triumphs and given me a sounding board when problems seemed insurmountable. He keeps my life together

(dishes, laundry, car repair), tolerates me when I disappear into work for too many hours in a week, goes on crazy hikes with me and makes me laugh every day. I love you.

Abstract

One strategy for controlling transmission of insect-borne disease involves replacing the native insect population with transgenic animals unable to transmit disease. Population replacement requires a drive mechanism to ensure the rapid spread of linked transgenes conferring disease refractoriness. *Medea* selfish genetic elements have the feature that when present in a female, only offspring that inherit the element survive, a behavior that can lead to spread. We validate our model against a synthetic *Medea* element created in *Drosophila* and find that the model fits the data without parameter fitting. We use modeling to identify conditions under which *Medea* elements spread. We derive equations describing the allele frequencies required for spread of *Medea* elements with a fitness cost, and the equilibrium allele frequencies attained. We show that when *Medea* spreads, it drives the non-*Medea* genotype out of the population, and we provide estimates of the number of generations required to achieve this goal. We also characterize two contexts in which *Medea* elements with fitness costs drive the non-*Medea* allele from the population: an autosomal element in which zygotic rescue is incomplete and an X-linked element in species in which X/Y individuals are male. We explore costs and benefits associated with the introduction of multiple *Medea* elements.

Table of Contents

Acknowledgments……....i

Abstract…….……..v

Table of Contests……vi

Chapter 1: Introduction……1

Chapter 2: A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in Drosophila…….23

Chapter 3: *Medea* selfish genetic elements as tools for altering traits of wild populations: A deterministic theoretical analysis……...37

Chapter 4: Some strains of *Drosophila* show resistance to synthetic *Medea*……..144

Chapter 5: Conclusions……151

Chapter 1: Introduction

Some insects and arthropods act as vectors for important human, animal and agricultural diseases ranging from Lyme's Disease to Chagas' Disease to viral plant pathogens. Insect-borne diseases sicken and kill millions of people annually (WHO 2009, WHO 2010). The two most devastating are malaria, a *Plasmodium*, and dengue, a virus. Currently about half of the world's population is at risk for these diseases (WHO 2009, 2010). In 2008, there were 247 million cases of malaria resulting in about 1 million deaths. About 20% of childhood mortality is related to malaria in Africa (WHO 2010). Annually, dengue sickens about 50 million people, killing about 15,000 (WHO 2009). Both diseases have been expanding their ranges with dengue transmission reported in Florida in 2010 (CDC 2010).

By definition, these insect-borne diseases have a complex life cycle. They have at least two hosts, often humans and mosquitoes and the pathogen must cycle between the two hosts. That is, a person with malaria will not transmit malaria to another person without the *Plasmodium* first infecting a mosquito host. Therefore, there are currently two strategies to fight these diseases: (1) treat an infected human with pharmaceuticals and (2) prevent mosquitoes from biting humans. There are no effective drugs to target dengue (WHO 2009). While there are agents to treat malaria, they are often expensive and pathogen-resistance is an increasing problem (Hyde 2005, Greenwold *et al*. 2008). Strategies to prevent mosquito bites include spraying insecticides and repellants, wearing protective clothing or bed nets, using screen doors and air conditioning, and making alternations to the environment such that there is less standing water and fewer breeding

sites (Kitron and Spielman 1989; Trape *et al*. 2002; Gould and Schliekelman 2004). While these methods have shown limited locally efficacy, in many parts of the world mosquito population densities are extremely high and most pathogen control efforts have not seen clinical results (WHO 2010). Additionally, some methods such as air conditioning are extremely expensive while others such as draining fields or spraying large quantities of insecticides have huge ecological consequences.

Finally, as illustrated by the WHO 2010 clinical results, the incidence of disease is not linearly related with the size or the mosquito population, or number of bites a human receives (MacDonald 1957; Boete and Koella 2002, 2003). In many parts of Africa, each person receives hundreds of potentially infections bites annually. To effectively reduce the number of cases of malaria, the number of infectious bites would have to be reduced by more than 90 (and in many cases more than 99%). Even in by optimistic estimates, bed nets, screened doors and protective clothing alone cannot achieve this result. Heavy use of insecticides can dramatically reduce mosquito populations. However, once applications cease, the population will rapidly reestablish, leading to a potentially devastating disease epidemic (Roberts *et al.* 1997).

A new strategy is to modify the insect population such that is refractory to pathogens. That is, insert a gene into the insect's genome that makes that organism either immune or unable to transmit a particular pathogen. There has been success with disease inhibition (de Lara Capurro *et al.* 2000; Ito *et al.* 2002; Moreira *et al*. 2002; Franz et al. 2006; Corby-Harris, *et al.* 2010). However, creating a genetically resistant only solves one part

of the problem. In order for the strategy to work, this new genetically modified (GM) insect strain must be able to outcompete the native population such that all the mosquitoes in the population carry this particular gene of interest. Selfish genetic elements are a particular class of genes that have transmit themselves at greater than Mendelian frequencies through a genome without any apparent phenotypic benefit to that organism. If a selfish genetic element and a gene for disease refractoriness are linked in a genome, the disease refractory gene should be able to hitchhike with the selfish genetic element (Braig and Yan 2001; Gould and Schliekelman 2004; Sinkins and Gould 2006).

There are a multitude of naturally occurring selfish genetic elements including sterile male release, transposable elements, homing endonucleases, meotic drive mechanisms, underdominance, *Medea*, an intracellular bacterium *Wolbachia* and killer-rescue (Burt and Trivers 2006, Sinkins and Gould 2006, Wade and Beeman 1994, Huang *et al.* 2009). In order to choose which strategy to examine more closely, we need to describe important features of the drive system. While it is possible to release large numbers of insects, smaller releases of insects will be preferred from a financial standpoint. Additionally, because we can assume that there is a fitness cost on the insect for the inability to transmit disease (in the form of a reduced lifespan or decreased fecundity), the drive system should be able to drive even in the presence of a fitness cost (Schmid-Hempel 2005; Tripet *et al.* 2008; Vernick *et al*. 2005). The drive system should be relative stable in the face of evolution. That is, a drive systems that are vulnerable to rapid evolutionary change are probably not viable. Finally, the ability to physically create the drive system in a lab is critical for success.

Sterile male strategies have been effective in eliminating populations of screwworms (*Cochliomyia hominivorax*)and Mediterranean fruit flies (*Ceratis capitata*) (Knipling 1955; Hendrichs *et al.* 1995; Benedict and Robinson 2003). This strategy works by releasing large numbers of irradiated male insects over a target area. These insects are rendered sterile from the radiation and compete with native males to mate with any females in the area. If males are released whenever females are present, the population will crash. A strain of mosquitoes that is genetically sterile when raised in the absence of tetracycline has been developed in England (Thomas *et al.* 2000). This strategy has 2 major disadvantages. 1) Insects must be released continuously or the native insects will reemerge and 2) large numbers of insects must be released (larger if there is a fitness cost on the genetically modified insects).

Transposable elements and homing endonucleases have very high transmission rates even in the presence of large fitness costs. However, efforts to engineer them have been largely unsuccessful. Recently, an intriguing proof of principle system has been developed using homing endonucleases in mosquito. Homing endonucleases function by cutting a specific sequence of DNA and hijack the cellular machinery to repair the break by inserting itself into the break point. In this case, the I-Scel element is able to cleave mosquito chromosomes in a site-specific manner. Most of the time, it inserts itself (and a marker gene) into broken chromosome. Occasionally, the chromosome repairs itself through blunt end ligation and the site is destroyed (Windbichler *et al.* 2011). While the

strategy is highly successful initially, this rapid creation of chromosomes which do not contain the gene of interest and are not vulnerable to the endonuclease may cause the strategy to be unviable.

Meotic drive mechanisms cause population crashes by forcing the population to become entirely male or entirely female. This strategy has not been engineered in the lab. If a chromosome evolved resistance to this mechanism, it would rapidly take over the population.

Underdominance and killer-rescue are strategies that involved creating an element that kills the organism and a second element that acts as an antidote. In the case of killer-rescue strategies, the two elements are unlinked. They killer gene must be introduced a t a relatively high level. The both the killer and rescue gene will remain at intermediate to high levels for a transient period of time (Huang 2009). Underdominance takes the strategy one step further allowing 2 killer genes and 2 rescue genes to act. In this case, killer 1 is linked to rescue 2 and killer 2 is linked to rescue 1. Both sets of genes must be passed on for the organism to be viable. Underdominance requires a relatively high release threshold and is vulnerable to fitness effects, but will drive to completion (that is no native alleles will remain in the population) (Davis *et al.* 2001).

Maternal-effect lethal selfish genetic elements are members of a class of selfish genetic elements that enhance their transmission by causing the death of offspring that lack the element

(BURT and Trivers 2006). These elements were first described in the flour beetle *Tribolium castaneum* and are known by the acronym *Medea* (maternal-effect dominant embryonic arrest). *Medea* elements in *Tribolium* are chromosomally located and gametes are generated and transmitted in a strictly Mendelian manner. Tribolium *Medea* drives because when present in females, only progeny that inherit the element-containing chromosome from either the maternal and/or paternal genome survive (BEEMAN *et al.* 1992). Therefore, *Medea* enhances its transmission relative to competing non-*Medea*-bearing homologous chromosomes (hereafter referred to as the non-*Medea* allele) by causing the death of progeny that do not carry a copy of *Medea* found in the mother. One *Tribolium Medea*, *Medea*^{M1}, has been mapped and is associated with a composite Tc1 transposon insertion that includes a number of genes (LORENZEN *et al.* 2008). How this insertion confers maternal-effect lethal selfish behavior is unknown, but genetic analysis suggests a model in which *Medea* consists of two tightly linked loci: one that encodes a maternal-effect lethal activity (a toxin) inherited by all progeny of *Medea*-bearing mothers, and a second that encodes a zygotic rescue activity (an antidote) that can be inherited from either the maternal or paternal genome (BEEMAN *et al.* 1992). Loci with genetic characteristics similar to those of *Tribolium Medea* have also been described in mice, but little is known about their molecular basis (HURST 1993; PETERS and BARKER 1993; WEICHENHAN *et al.* 1996; WEICHENHAN *et al.* 1998). Synthetic *Medea* elements that drive population replacement in *Drosophila* have been generated (CHEN *et al.* 2007). The genetic and cell biological principals utilized to generate *Medea* in *Drosophila* involve maternal-specific silencing of a gene whose product is synthesized maternally and deposited into the developing oocyte, but only required during embryogenesis (the toxin), coupled with zygotic expression of a rescuing transgene (the antidote). These principles are straightforward and

conserved across the animal kingdom, and therefore should be applicable to the generation of similar elements in agricultural pest and human disease vector species.

The dynamics of *Medea* spread have been modeled by several groups. Wade and Beeman first considered this problem, focusing on several situations in which *Medea* was introduced into populations at very low frequency (WADE and BEEMAN 1994). They showed that if the presence of *Medea* does not result in a fitness (fecundity) cost to carriers, independent of maternal-effect killing, *Medea* spreads to fixation for all degrees of maternal effect lethality, though the rate of *Medea* increase is very slow when it is introduced at low frequency. They also showed that if the presence of *Medea* results in a decrease in fecundity independent of maternal-effect killing, the frequency of the *Medea* allele could still increase to a stable internal equilibrium. Whether *Medea* increased in frequency or was lost from the population was found to depend critically on the degree of dominance of the fitness costs: high levels of completely recessive fecundity costs were well tolerated, but when *Medea* was present at low frequency, even modest degrees of dominance led to loss of *Medea* from the population. Population genetic models of HASTINGS (1994), SMITH (1998) and CHEN *et al.* (2007) show that, *Medea* elements with significant dominant fitness costs can still spread, provided they are introduced above a critical introduction frequency. Previous work has focused on the fate of the *Medea* allele. However, it is the fate of *Medea*-bearing genotypes that is important for population replacement. CHEN *et al.* (2007) showed that, at least under some conditions, when *Medea* elements with fitness costs are introduced at frequencies that result in spread to an internal equilibrium allele frequency, non-*Medea* individuals are nonetheless rapidly eliminated from

the population. However, it has not been clear to what extent this conclusion can be generalized.

In summary, *Medea* elements are attractive candidates to drive population replacement for several reasons. First, they can spread (provided certain conditions detailed below are met) even if they confer a fitness cost to carriers (CHEN *et al.* 2007; HASTINGS 1994; SMITH 1998; WADE and BEEMAN 1994). Second, under at least some conditions when *Medea* spreads it eliminates the non-*Medea* genotype from the population (CHEN *et al.* 2007). Third, the synthetic form of *Medea* is the only gene drive mechanism that is both well understood at the molecular level, because it was designed with components of known behavior, and that has been demonstrated to drive population replacement. Finally, design considerations discussed by CHEN *et al.* (2007) regarding ways to prevent recombinational separation of drive and disease refractoriness functions, to prevent selfish element spread in non-target species, and to carry out multiple cycles of population replacement, provide reasons to believe that the population genetic behavior of synthetic *Medea* elements can to some extent be controlled. Therefore, *Medea* is a logical target for concerted development efforts.

My work focuses both on questions surrounding the practical aspects of the release of *Medea* and the some of the interesting mathematics behind the equilibria. We want to answer the question of how many insects must be relased given a variety of circumstances including fitness costs, location of the element, and imperfect transmission of the element as well as how long population replacement takes. In chapter 2 of this

theis, we examine how well our model behaves in comparison to our lab created Medea drive system. This chapter is a reproduction of Chen's 2007 paper with my contribution showing that our real *Medea* element behaves as predicted by our models. Chapter 3 looks properties of a *Medea* element in a large single population and is an expanded version of our *Evolution* paper (Ward *et al.*, 2011). Chapter 4 is a short discussion of work I did examining the extent to which *Medea* produced in our one lab strain of *Drosophilia* can act as a driver in various other *Drosophila* populations. Chapter 5 contains overall conclusions.

Literature Cited

Beeman, R. W., and K. S. Friesen. 1999. Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, tribolium castaneum. Heredity 82 (Pt 5):529-534.

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

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* D. K. a. B. Letourneau, B.E., ed. Genetically

Engineered Organisms: Assessing Environmental and Human Health Effects. CRC Press.

Burt, A., AND R. Trivers, 2006 *Genes in conflict*. Belknap Press, Cambridge, MA.

CDC, 2010. Morbidity and Mortality Weekly Report: Locally Acquired Dengue --- Key West, Florida, 2009—2010. 59(19);577-581.

Corby-Harris, V., Brexler, A., Watkins de Jong, L., Antonova, Y., Pakpour, N., Ziegler, R., Ramberg, F., Lewis, E.E., Brown, J.M., Luckhart, S., Riehle, M.A. 2010. Activation of Akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. PloS Pathog. 6:e1001003.

de Lara Capurro, M., J. Coleman, B. T. Beerntsen, K. M. Myles, K. E. Olson, E. Rocha, A. U. Krettli, and A. A. James. 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.

Davis S., N. Bax and P. Grewe 2001. Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations. J Theor Bio 212:83-98.

Franz, A. W., I. Sanchez-Vargas, Z. N. Adelman, C. D. Blair, B. J. Beaty, A. A. James, and K. E. Olson. 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., and P. Schliekelman. 2004. Population genetics of autocidal control and strain replacement. Annu Rev Entomol 49:193-217.

Greenwood, B. M., D. A. Fidock, D. E. Kyle, S. H. Kappe, P. L. Alonso, F. H. Collins, and P. E. Duffy. 2008. Malaria: progress, perils, and prospects for eradication. J Clin Invest 118:1266-1276.

Hendrichs, J., G. Franz, and P. Rendon. 1995. Increased effectiveness and applicability of sterile insect technique through male-only releases for control of Mediterranean fruit flies during fruiting seasons. J App Ento 119:371-377.

Hurst, L. D. 1993. scat+ is a selfish gene analogous to *Medea* of *Tribolium castaneum*. Cell 75:407-408.

Hyde, J. E. 2005. Drug-resistant malaria. Trends Parasitol 21:494-498.

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.

Kitron, U., and A. Spielman. 1989. Suppression of transmission of malaria through source reduction: antianopheline measures applied in Israel, the United States, and Italy. Rev Infect Dis 11:391-406.

Knipling, E.F. 1955. Possibilities of Insect Control or Eradication Through the Use of Sexually Sterile Males. J Econ Ento 48:459-462.

Lorenzen, M. D., A. Gnirke, J. Margolis, J. Garnes, M. Campbell, J. J. Stuart, R. Aggarwal, S. Richards, Y. Park, and R. W. Beeman. 2008. The maternal-effect, selfish genetic element *Medea* is associated with a composite Tc1 transposon. Proc Natl Acad Sci U S A 105:10085-10089.

Macdonald, 1957 *The Epidemiology and Control of Malaria*. Oxford Univ., London.

Moreira, L. A., J. Ito, A. Ghosh, M. Devenport, H. Zieler, E. G. Abraham, A. Crisanti, T. Nolan, F. Catteruccia, and M. Jacobs-Lorena. 2002. Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. J Biol Chem 277:40839-40843.

Peters, L. L., and J. E. Barker. 1993. Novel inheritance of the murine severe combined anemia and thrombocytopenia (Scat) phenotype. Cell 74:135-142.

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

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

Smith, N. G. 1998. The dynamics of maternal-effect selfish genetic elements. J Theor Biol 191:173-180.

Roberts, D.R, L. L. Laughlin, P. Hsheih, and L. J. Legters. 1997. DDT, global strategies, and a malaria control crisis in South America. Emerg Infect Dis 3: 295-302.

Thomas, D. T., C. A. Donelly, R. J. Wood and L. S. Alphey (2000). Insect population control using a dominant, repressible, lethal genetic system. Science. 287: 2474-2476.

Trape, J. F., G. Pison, A. Spiegel, C. Enel, and C. Rogier. 2002. Combating malaria in Africa. Trends Parasitol 18:224-230.

TRIPET, F., F. ABOAGYE-ANTWI and H. HURD, 2008 Ecological immunology of mosquito-malaria interactions. Trends Parasitol **24:** 219-227.

VERNICK, K. D., F. ODUOL, B. P. LAZZARO, J. GLAZEBROOK, J. XU *et al.*, 2005 Molecular genetics of mosquito resistance to malaria parasites. Curr Top Microbiol Immunol **295:** 383-415.

Wade, M. J., and R. W. Beeman. 1994. The population dynamics of maternal-effect selfish genes. Genetics 138:1309-1314.Benedict, M. Q. and A. S. Robinson, 2003 The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends in Para. 19:349-355.

Weichenhan, D., B. Kunze, W. Traut, and H. Winking. 1998. Restoration of the Mendelian transmission ratio by a deletion in the mouse chromosome 1 HSR. Genet Res 71:119-125.

Weichenhan, D., W. Traut, B. Kunze, and H. Winking. 1996. Distortion of Mendelian recovery ratio for a mouse HSR is caused by maternal and zygotic effects. Genet Res 68:125-129.

WHO, 2009. Dengue and dengue haemorrhagic fever. http://www.who.int/mediacentre/factsheets/fs117/en/index.html.

WHO, 2010. Malaria. http://www.who.int/mediacentre/factsheets/fs094/en/index.html.

Windbichler, N. M. Menichelli, P. A. Papapthanos, S. B. Thyme, H. Li, U. Y. Ulge, B. T. D. Baker, R. J. Monnat, A. Burt and A Crisanti. 2011 A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature. 12: 212-5.

Chapter 2: A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in Drosophila

First reported in 2007 in Science. Volume 316.

Chun-Hong Chen, Haixia Huang, Catherine M. Ward, Jessica T. Su, Lorian V. Schaeffer, Ming Guo, Bruce A. Hay

Abstract

One proposed strategy for controlling the transmission of insect-borne pathogens uses a drive mechanism to ensure the rapid spread of transgenes conferring disease refractoriness throughout wild populations. Here, we report the creation of maternal-effect selfish genetic elements in *Drosophila* that drive population replacement and are resistant to recombination-mediated dissociation of drive and disease refractoriness functions. These selfish elements use microRNA mediated silencing of a maternally expressed gene essential for embryogenesis, which is coupled with early zygotic expression of a rescuing transgene.

Body text:

Mosquitoes with a diminished capacity to transmit malaria or dengue have been identified in the wild and/or created in the laboratory, demonstrating that endogenous or engineered mosquito immunity can be harnessed to attack these pathogens (1–5).

However, it will be necessary to replace a large percentage of the wild mosquito population with refractory insects to achieve substantial levels of disease control (6–8). Mosquitoes carrying genes that confer disease refractoriness are not expected to have a higher fitness than native mosquitoes, implying that Mendelian transmission is unlikely to result in an increase in the frequency of transgene-bearing individuals after their initial release into the wild (4, 9). Thus, effective population replacement will require the coupling of genes conferring disease refractoriness with a genetic mechanism for driving these genes through the wild population at greater than Mendelian frequencies (10, 11).

Maternal-effect selfish genetic elements [first described in the flour beetle *Tribolium castaneum* and known by the acronym *Medea* (maternal effect dominant embryonic arrest)] select for their own survival by inducing maternal-effect lethality of all offspring not inheriting the element bearing chromosome from the maternal and/or paternal genome (12) (Fig. 1A). Current models predict that if *Medea* elements are introduced into a population above a threshold frequency, determined by any associated fitness cost, they will spread within the population (12–14) (Fig. 1, C and D). When introduced into a population at relatively high frequencies, *Medea* elements are predicted to rapidly convert the entire population into element-bearing heterozygotes and homozygotes (Fig. 1C). *Medea* in Tribolium is hypothesized to consist of a maternal lethal activity (a toxin) that kills non-*Medea*–bearing progeny and a zygotic rescue activity (an antidote) that protects *Medea*-bearing progeny from this maternal lethal effect (12, 15) (Fig. 1A).

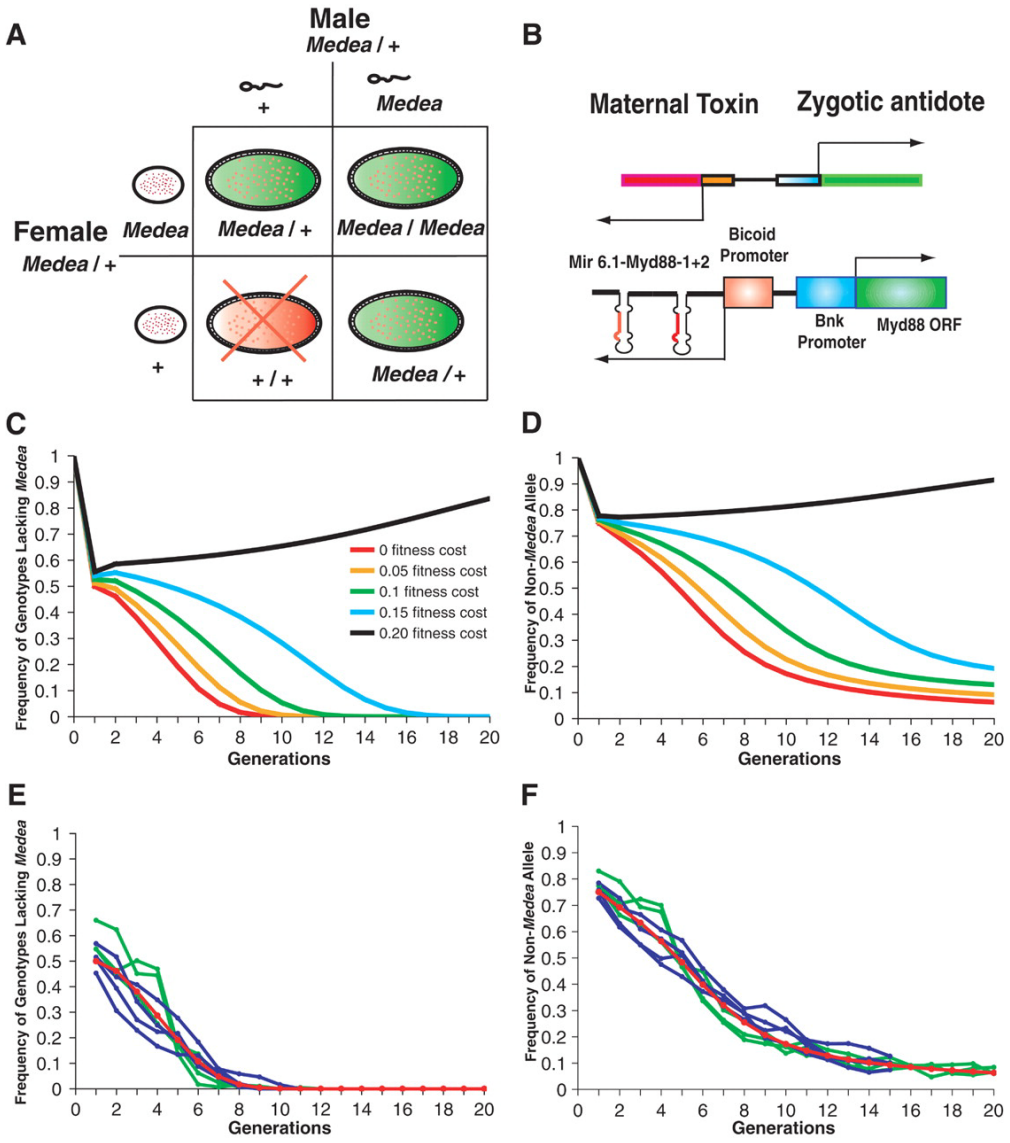

Fig. 1. Characteristics of a maternal-effect selfish genetic element (*Medea*) and a synthetic *Medea* element in *Drosophila*. (A) It is postulated that females heterozygous for *Medea* (*Medea*/+) deposit a protoxin or toxin (red dots) into all oocytes. Embryos that do not inherit a Medea-bearing chromosome die because toxin activation or activity is unimpeded (bottom left square). Embryos that inherit *Medea* from the maternal genome (top left square), the paternal genome (bottom right square), or both (top right square) survive because zygotic expression of a *Medea*-associated antidote (green background) neutralizes toxin activity. (B) (Top) Schematic of a simple molecular model that accounts for *Medea* behavior postulates the existence of two tightly linked loci. One locus consists of a maternal germline–specific promoter that drives the expression of RNA or protein that is toxic to the embryo. The second locus consists of a zygotic promoter that drives the expression of an antidote. (Bottom) Schematic of $Medea^{myd88}$ is shown. ORF, open reading frame; Mir 6.1-Myd88-1+2, transcript encoding two copies of *Drosophila* miRNA 6.1 modified to target the *myd88* 5′ untranslated region. (C) Frequency of

genotypes lacking *Medea* for an element carrying the additive fitness costs indicated, over generations, with *Medea* introduced at a 1:1:2 ratio of homozygous *Medea*-bearing males, non-*Medea*–bearing males, and females lacking *Medea*. Generation zero refers to the wild population (non-*Medea*/non-*Medea* = 1) before population seeding. Generation one refers to the progeny of crosses between these individuals and homozygous *Medea*-bearing males. (D) Frequency of the non-*Medea*–bearing chromosome for the populations described in (C). (E and F) $Medea^{myd88\text{-}1}$ drives population replacement in *Drosophila*. $Medea^{myd88\text{-}1}$ was introduced into seven population cages at an allele frequency of ~25% (16). (E) The frequency of genotypes lacking *Medea* (+/+) over generations is indicated for two separate sets of population cage experiments, involving three (green lines; 20 generations) or four (blue lines; 15 generations) population cages each. The predicted frequency of genotypes lacking *Medea*, for a *Medea* element with zero fitness cost (introduced at an allele frequency of 25%) is indicated by the red line. (F) The frequency of the non-$Medea^{myd88\text{-}1}$–bearing chromosome (+/*Medea* and +/+) over generations from the population cage experiments in (E) is indicated as above, as is the predicted frequency for an element with zero fitness cost.

To create a *Medea*-like maternal-effect selfish genetic element in *Drosophila*, we generated a P transposable element vector in which the maternal germline–specific *bicoid* (*bic*) promoter drives the expression of a polycistronic transcript encoding two microRNAs (miRNAs) designed to silence expression of *myd88* (the gene producing the toxin) [Fig. 1B and (16)]. Maternal *Myd88* is required for dorsal-ventral pattern formation in early embryo development. Females with germline loss-of-function mutations for *myd88* give rise to embryos that lack ventral structures and thus fail to hatch, even when a wild-type (WT) paternal allele is present (17). This vector (known as $Medea^{myd88}$) also carries a maternal miRNA–insensitive *myd88* transgene expressed under the control of the early embryo–specific bottleneck (bnk) promoter (the gene producing the zygotic antidote) (Fig. 1B). Our analysis focused on flies carrying a single autosomal insertion of this element, $Medea^{myd88\text{-}1}$.

Matings between heterozygous *Medea*$^{myd88\text{-}1}$/+males (where + indicates a chromosome that does not carry *Medea*$^{myd88\text{-}1}$) and homozygous +/+ females resulted in high levels of embryo viability, similar to those for the w^{1118} strain used for transformation (Table 1). In addition, 50% of the adult progeny carried *Medeamyd88*-1, as expected for Mendelian segregation without dominance. Matings among homozygous *Medea*$^{myd88\text{-}1}$ flies also resulted in high levels of egg viability. In contrast, when heterozygous *Medeamyd88*-1/+ females were mated with homozygous +/+ males, ~50% of progeny embryos had ventral patterning defects (fig. S1) and did not hatch (Table 1). All adult progeny (n >12,000 flies) carried *Medea*$^{myd88\text{-}1}$ (Table 1). On the basis of these data and the results of several other crosses (Table 1), we inferred that a single copy of *bic*-driven miRNAs targeting maternal *myd88* expression was sufficient to induce maternal-effect lethality and a single copy of zygotic *bnk*-driven *myd88* expression was sufficient for rescue. The above observations, in conjunction with the lack of any obvious fitness effects (lethality) in individuals carrying one or two copies of *Medea*$^{myd88\text{-}1}$, suggested that *Medea*$^{myd88\text{-}1}$ should be able to drive population replacement.

Table 1. *Medea*$^{myd88\text{-}1}$ shows maternal-effect selfish behavior. Progeny of crosses between parents of several different genotypes (M refers to the *Medea*$^{myd88\text{-}1}$–bearing chromosome; + refers to the non–element-bearing homolog) are shown. The maternal copy number (0 to 2) of bic-driven miRNAs targeting the endogenous *myd88* transcript (maternal toxin) and zygote copy number (0 to 2) and percentage of embryos inheriting *bnk*-driven *myd88* (zygotic antidote) are indicated, as are the adult progeny genotypes predicted for Mendelian inheritance of *Medea*$^{myd88\text{-}1}$ and the percent embryo survival. -, not measured. The asterisk denotes that embryo survival was normalized with respect to percent survival (± SD) observed in the w^{1118} stock used for transgenesis (97.1 ± 0.7%).

Parental genotype		Inherited by the		Adult M progeny (%)		
Male	Female	Oocyte (Maternal toxin)	Embryo (Zygotic antidote (n, %))	Predicted	Observed	Embroy survival (%)
M/+	+/+	0	0, 50	50	50	99.6 ± 1

			1, 50		(n>7000)	
M/M	M/M	2	2, 100	100	-	98.1 ± 0.4
+/+	M/+	1	0, 50 1, 50	50	100 (n>12000)	48.3 ± 2
M/M	M/+	1	1, 50 2, 50	100	-	98 ± 1
M/+	M/+	`	0, 25 1, 50 2, 25	75	-	74.3 ± 0.5
M/+	M/M	2	1, 50 2, 50	100	-	98.3 ± 1
+/+	M/M	2	1, 100	100	-	99.1 ± 0.4
M/M	+/+	0	1, 100	100	-	98.8 ± 0.5

To test this prediction, we mated equal numbers of WT (+/+) and *Medea*$^{myd88\text{-}1}$/*Medea*$^{myd88\text{-}1}$ males with homozygous +/+ females, giving rise to a progeny population with *Medea*$^{myd88\text{-}1}$ present at an allele frequency of ~25% (16). This level of introduction, although high, is not unreasonable, given previous insect population suppression programs (18). Replicate population cage experiments, carried out in a darkened incubator to prevent *Medea*$^{myd88\text{-}1}$–bearing flies (which are P^{w+} and thus red-eyed) from obtaining any vision-dependent advantage over their +/+ counterparts (which are *w*^{1118} and white-eyed) (19), followed three replicates for 20 generations. A second set of four replicates, which were initiated by crossing heterozygous *Medea*$^{myd88\text{-}1}$/+ males with homozygous +/+ females, was followed for 15 generations. In both experiments, non-*Medea*$^{myd88\text{-}1}$–bearing flies permanently disappeared from the population between generations 10 and 12 (Fig. 1E), without a loss of non-*Medea*–bearing + chromosomes (in *Medea*$^{myd88\text{-}1}$/+individuals) in the population (Fig. 1F). The observed changes in *Medeamyd88*-1 were not significantly different from the null hypothesis that the element had no fitness cost [(16) and fig. S2], although we cannot exclude the possibility that a

Medea$^{myd88\text{-}1}$–associated cost is counterbalanced by an unknown negative effect associated with the w^{1118} mutation in +/+ individuals. Finally, we carried out three further replicate population cage experiments in which the *Medea*$^{myd88\text{-}1}$ transgene was introduced at a frequency of 25% into the Oregon-R strain, which is WT with respect to the endogenous w locus (and thus members of which are red-eyed). Evidence for population replacement by generation 12 was observed in this context as well (16), suggesting that *Medea*$^{myd88\text{-}1}$–associated Pw+ expression is unlikely to be a major contributor to the ability of *Medeamyd*$^{88\text{-}1}$ to drive population replacement.

For any gene-drive mechanism to be successful in reducing parasite transmission, there must be tight linkage between the genes that mediate drive and effector functions (10). If the driver becomes separated from the effector gene through chromosome breakage and nonhomologous end joining (as in a reciprocal translocation) (Fig. 2A), and the effector gene carries a fitness cost, selection will favor and promote the spread of individuals carrying *Medea* elements that lack the effector (*Medea*$^{\square eff}$). Locating the effector gene between the toxin and antidote prevents a single chromosome breakage and end joining event from creating a *Medea*$^{\square eff}$-bearing chromosome(Fig. 2B). However, it does not prevent the creation of *Medea*^{ins}-bearing chromosomes that carry the antidote, and perhaps the effector, but not the toxin (Fig. 2B). *Medea*^{ins}-bearing chromosomes are insensitive to *Medea*-dependent killing. If the presence of the toxin and/or the effector results in a fitness cost, then *Medea*^{ins}-bearing chromosomes gain a fitness advantage with respect to those carrying the complete *Medea* element, thereby promoting spread of the former. This outcome can lead to the reappearance of pathogen-transmitting insects (14).

One way to prevent chromosome breakage and end joining–mediated formation of *Medea*$^{\square eff}$ and *Medea*^{ins} elements is to put the toxin and effector genes into an intron of the antidote (Fig. 2C). To test this approach, we generated flies carrying P^{w+}*Medea*$^{myd88\text{-}int}$, in which the toxin, a transcript generating maternally expressed miRNAs targeting *myd88*, was placed in an intron of the zygotically expressed antidote, bnk-driven myd88, in the opposite orientation (Fig. 2D). We characterized the behavior of one autosomal insertion of this element *Medea*$^{myd88\text{-int-}1}$, which behaved as a maternal effect selfish

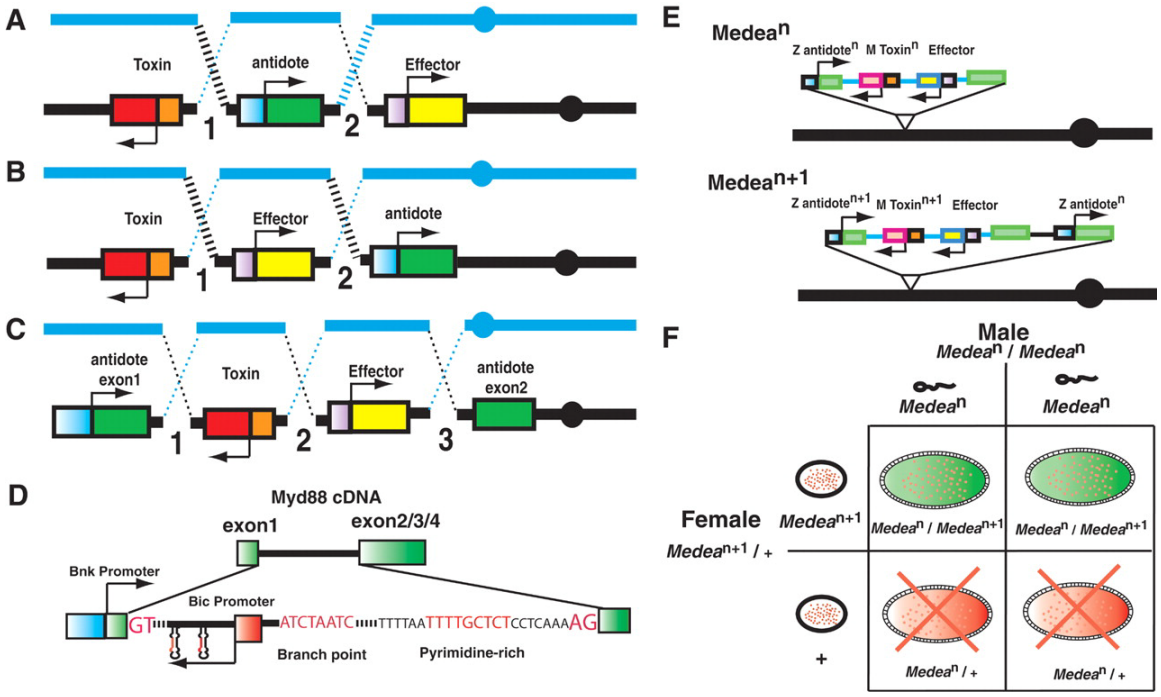

Fig. 2. A strategy for enhancing the functional lifetime of *Medea* elements in the wild and for carrying out cycles of population replacement. (A to D) Locating *Medea* toxin and effector genes in an intron of the antidote prevents chromosome breakage and end joining–mediated separation of drive and effector genes and creation of *Medea*^{ins}-bearing chromosomes. [(A) to (C)] *Medea* constructs with different gene arrangements are shown. Sites of chromosome breakage and end joining with a second nonhomologous chromosome are indicated by the crossed lines. Recombinant products referred to in the text are indicated by thick lines, the color of which indicates the centromere (solid circles) involved. (A) Recombination at site 1 generates a *Medea*^{ins}-bearing chromosome

that carries the effector. Recombination at site 2 generates a *Medea*$^{\square eff}$-bearing chromosome. (B) Recombination at site 1 or site 2 generates a *Medea*^{ins}-bearing chromosome. (C) Recombination at sites 1 to 3 generates benign chromosomes that cannot show *Medea*^{ins} or *Medea*$^{\square eff}$behavior. (D) Schematic of Medeamyd88-int. Splice donor and acceptor sites are indicated in large red letters, with the branchpoint and polypyrimidine stretch in small red letters. (E and F) A strategy for carrying out cycles of population replacement with *Medea*. (E) A first-generation *Medea* element (*Medea*^{n}), driven by Toxinn and Antidoten, is integrated into the chromosome [thick black line with centromere (solid circle) at the right] at a specific position (triangle). A second-generation *Medea* element (*Medea*$^{n+1}$), driven by *Toxin*$^{n+1}$ and *Antidote*$^{n+1}$, can be integrated at the same position using site-specific recombination (24). Locating both elements at the same position limits the possibility of homologous recombination creating chromosomes that carry both elements. (F) Because progeny carrying *Medea*^{n} are sensitive to *Toxin*$^{n+1}$, the only progeny of females heterozygous for Medean+1 that survive are those that inherit *Medea*$^{n+1}$, regardless of their status with respect to *Medea*^{n}. In contrast, the progeny of *Medea*^{n} females that fail to inherit *Medea*^{n} survive if they inherit *Antidote*^{n} as a part of *Medea*$^{n+1}$.

genetic element (Table 2). Females heterozygous for *Medea*$^{myd88\text{-}int\text{-}1}$ gave rise to a high frequency of *Medea*^{myd88}-int-1–carrying progeny (>99%), and the maternal-effect lethality associated with a single copy of P^{w+}*Medea*^{myd88}-int-1 was rescued by zygotic expression of the antidote from either the maternal or paternal genome (Table 2). However, when

Table 2. *Medea*$^{my88\text{-}int\text{-}1}$ shows maternal-effect selfish behavior. Progeny of crosses between parents of several different genotypes notations are the same as those in Table 1.

Parental genotype		Inherited by the		Adult M progeny (%)		
Male	Female	Oocyte (Maternal toxin)	Embryo (Zygotic antidote (n, %))	Predicted	Observed	Embroy survival (%)
M/+	+/+	0	0, 50 1, 50	50	51 (n=5000)	98.4 ± 0.6
M/M	M/M	2	2, 100	100	-	98.6 ± 0.8
+/+	M/+	1	0, 50 1, 50	50	99.5 (n=5000)	48.7 ± 0.6
M/M	M/+	1	1, 50 2, 50	100	-	98.4 ± 0.7
M/+	M/+	`	0, 25 1, 50 2, 25	75	-	73.6 ± 1.2

M/+	M/M	2	1, 50 2, 50	100	-	57.2 ± 1.5
+/+	M/M	2	1, 100	100	-	20.2 ± 1.1
M/M	+/+	0	1, 100	100	-	98.5 ± 0.7

homozygous element bearing females were crossed with non–element bearing males, progeny embryo survival was very poor (~20%), suggesting an inefficient zygotic rescue, perhaps resulting from inefficient splicing of the myd88 artificial intron. Population replacement for an element with these fitness characteristics is still expected to occur, though with some delay (fig. S3) as compared to that for an element in which the fitness costs are a simple function of copy number in either sex (Fig. 1, C and D). *Medea*^{ins}-bearing chromosomes can also arise if the toxin mutates to inactivity. Although toxin mutation cannot be prevented, the use of miRNAs as toxins can provide a degree of redundant protection because multiple miRNAs, each processed and functioning as an independent unit, can be linked in a polycistronic transcript [(Fig. 1B and (16)]. The use of miRNAs as toxins also provides a basis by which selfish genetic element drive can be limited to the target species. *Medea* elements only show drive when maternal-effect lethality creates an opportunity for zygotic rescue of progeny that inherit the element. Therefore, drive can be limited to a single species by the use of miRNAs that are species-specific in their ability to target the maternally expressed gene of interest.

Perhaps the most likely point of failure in any population-replacement strategy involves the effector. Effector genes can mutate to inactivity, creating *Medea*$^{\square_{eff}}$-bearing chromosomes. In addition, parasites may undergo selection for resistance to these effectors. These events, as well as the possible appearance of *Medea*^{ins}-bearing chromosomes discussed above, will lead to the reappearance of permissive conditions for

disease transmission. Therefore, it is important that strategies be available for removal of an element from the population, followed by its replacement with a new element. One potential strategy for achieving this goal, in which different Medea elements located at a common site in the genome compete with each other for germline transmission in transheterozygous females, is illustrated in Fig. 2, E and F.

Our data show de novo synthesis of a selfish genetic element able to drive itself into a population. This laboratory demonstration notwithstanding, several obstacles remain to the implementation of *Medea*-based population replacement in the wild. First, for pests such as mosquito species, there is little genetic or molecular information regarding genes and promoters used during oogenesis and early embryogenesis. This information is straightforward to generate, with the use of transcriptional profiling to identify appropriately expressed genes and transgenesis and RNA interference in adult females to identify those required for embryonic development, but it remains to be acquired. In addition, current models of the spread of *Medea* do not take into account important real-world variables, such as migration, nonrandom mating, and the fact that important disease vectors such as *Anopheles gambiae* consist of multiple partially reproductively isolated strains (20, 21). Although an understanding of the above issues is critical for the success of any population-replacement strategy, the problems are not intractable, as evidenced by past successes in controlling pests by means of sterile-male release (18) and as implied by our growing understanding of mosquito population genetics, immunity, and ecology (20–23).

Literature Cited

1. M. de Lara Capurro *et al.*, *Am. J. Trop. Med. Hyg.* 62, 427 (2000).

2. J. Ito, A. Ghosh, L. A. Moreira, E. A. Wimmer, M. Jacobs-Lorena, *Nature* 417, 452 (2002).

3. L. A. Moreira *et al., J. Biol. Chem.* 277, 40839 (2002).

4. K. D. Vernick *et al., Curr. Top. Microbiol. Immunol.* 295, 383 (2005).

5. A. W. E. Franz *et al., Proc. Natl. Acad. Sci. U.S.A.* 103, 4198 (2006).

6. G. Macdonald, *The Epidemiology and Control of Malaria* (Oxford Univ. Press, London, 1957).

7. J. M. Ribeiro, M. G. Kidwell, *J. Med. Entomol.* 31, 10 (1994).

8. C. Boëte, J. C. Koella, *Malar. J.* 1, 3 (2002).

9. P. Schmid-Hempel, *Annu. Rev. Entomol.* 50, 529 (2005).

10. A. A. James, *Trends Parasitol*. 21, 64 (2005).

11. S. P. Sinkins, F. Gould, *Nat. Rev. Genet.* 7, 427 (2006).

12. R. W. Beeman, K. S. Friesen, R. E. Denell, *Science* 256, 89 (1992).

13. M. J. Wade, R. W. Beeman, *Genetics* 138, 1309 (1994).

14. N. G. C. Smith, *J. Theor. Biol*. 191, 173 (1998).

15. R. W. Beeman, K. S. Friesen, *Heredity* 82, 529 (1999).

16. Materials and methods are available as supporting material on Science Online.

17. Z. Kambris et al., *EMBO* Rep. 4, 64 (2003).

18. F. Gould, P. Schliekelman, *Annu. Rev. Entomol.* 49, 193 (2004).

19. B. W. Geer, M. M. Green, *Am. Nat. 96,* 175 (1962).

20. A. della Torre et al., *Science* 298, 115 (2002).

21. M. Coetzee, *Am. J. Trop. Med. Hyg*. 70, 103 (2004).

22. D. Vlachou, F. C. Kafotis, *Curr. Opin. Microbiol.* 8, 415 (2005).

23. H. Pates, C. Curtis, *Annu. Rev. Entomol*. 50, 53 (2005).

24. D. D. Nimmo, L. Alphey, J. M. Meredith, P. Eggleston, *Insect Mol. Biol.* 15, 129 (2006).

25. This work did not receive specific funding. It was supported by NIH grants GM057422 and GM70956 to B.A.H. and NS042580 and NS048396 to M.G. C.-H.C. was supported by the Moore Foundation Center for Biological Circuit Design. We thank two reviewers for useful comments and improvements on the manuscript. GenBank accession numbers for *Medea*^{myd88} and *Medea*$^{myd88\text{-int}}$ are EF447106 and EF447105, respectively.

Supporting Material

Materials and methods

Construction of a modified *bicoid* promoter:

A *bicoid* promoter fragment was amplified from genomic DNA with primers Bic-5 5'- GGC CTC GAG TTA GAT CTC AAT TGT GCC ATC TCT ACA TCT CTT CGC TCA TCC CTA AAT AAA AGA ATG AAC ATC GAG GGA GG and Bic-3 5' GGC CAA TTG GGC GGC GGT TGC GCC GTT TTC C, cut with Xho1 and EcoRI [all restriction enzymes provided by New England Biolabs (NEB), Beverly MA], and ligated into pGMR (*1*) cut with the same enzymes. Two Serendipity (Sry) binding sites [CGC TCA TCC CTA AAT and GTG CCA TCT CTA CA (*2*)] were placed 5' to the bicoid promoter in order to enhance maternal expression levels. Primers 4bic-1 5'- GGC CTC GAG TTA GAT CTC AAT TGT GCC ATC TCT ACA TCT CTT CGC TCA TCC CTA AAT AAA AGA ATG AAC ATC GAG GGA GG and Bic-3 were annealed with the *bicoid* promoter fragment generated above and PCR carried out (Fig. S4A). This product was cloned into pGMR as above, generating pSry-Bic-GMR.

Construction of miRNAs targeting *myd88*.

The *Drosophila* miRNA mir6.1 is illustrated in Fig. S4B. The *Drosophila* miRNA mir6.1 stem-loop precursor and surrounding sequences is illustrated to the left, with the mature miRNA indicated in red. Processing sites for Drosha and Dicer are indicated (arrows). 22bp sequences corresponding to mature mir6.1 were replaced with sequences perfectly complementary to sequences from the *myd88* 5'UTR, generating two new miRNAs, mir6.1-myd88-1 and mir6.1-myd88-2. Two sites in the myd88 transcript were targeted to

minimize the possibility that mRNA secondary structure would prevent the miRNA-RISC complex from being able to bind and cleave the mRNA. The use of multiple miRNAs targeting a common transcript, but at distinct positions, also provides a method for limiting the possibility that a single mutational event (in either the miRNA or the target sequence) will lead to a loss of toxin efficacy. *Myd88* mRNA sequences are indicated in black. The sites in *myd88* that are targeted are Myd88-1 CGA TCG GAA AAC TCG AAA AAA T and Myd88-2 TCA CGC GCT TCA TCG TTT TAT T (Fig. S4C). To generate a mir6.1 stem-loop backbone that generates a mature miRNA complementary to one or the other of these target sites we annealed pairs of primers. For example, to make a miRNA that targets myd88-1, primers Myd881-1 and Myd881- 2 (DdMyd881- 1: 5'-GGC AGC TTA CTT AAA CTT AAT CAC AGC CTT TAA TGT CGA TCG GAA AAC TCG AAA ACA TTA AGT TAA TAT ACC ATA TC and DdMyd881- 2: 5'-AAT AAT GAT GTT AGG CAC TTT AGG TAC CGA TCG GAA AAC TCG AAA AAA TTA GAT ATG GTA TAT TAA CTT AAT GT) were annealed and filled in using PCR (Fig. S4C). This product was then amplified using primers Mir6 5’ EcoRI/BglII (5'-GGC GAA TTC CGC CAG ATC TTT TAA AGT CCA CAA CTC ATC AAG GAA AAT GAA AGT CAA AGT TGG CAG CTT ACT TAA ACT TA) and Mir6 3’ BamHI/NotI (5' GGC CGC GGC CGC ACG GAT CCA AAA CGG CAT GGT TAT TCG TGT GCC AAA AAA AAA AAA AAT TAA ATA ATG ATG TTA GGC AC). These primers add mir6.1 flanking sequences that are thought to promote miRNA processing, as well as several restriction sites (Fig. S4C). A miRNA that targets myd88-2 was generated similarly, beginning with primers Myd882-1 GGC AGC TTA CTT AAA CTT AAT CAC AGC CTT TAA TGT TCA CGC GCT TCA TCG TTT TCT TTA AGT

TAA TAT ACC ATA TC and Myd882-2 AAT AAT GAT GTT AGG CAC TTT AGG TAC TCA CGC GCT TCA TCG TTT TAT TTA GAT ATG GTA TAT TAA CTT AAA GA. PCR products were purified with Qiagen (Valencia, CA) PCR purification columns, and then digested with enzymes. For dMyd88-1 these were EcoRI and BamHI, for dMyd88-2 BglII and NotI. Digested products were then ligated into sry-bic-GMR cut with EcoRI and NotI, generating pBic-mir6.1-myd88. The structure of this construct is shown in Fig. S4D.

Zygotic promoter and *myd88* antidote rescue construct:

A 500 bp fragment of DNA containing the transient, early zygotic *bnk* promoter (*3*) was amplified from genomic DNA using primers Bnk 5' XhoI (5'-GGC CTC GAG TAT TTC ACA AAT TCA ATT TTA ATA TTT AAG) and Bnk3' EcoRI (5'-GGC GAA TTC GTT GAC GGT TGA AGT ACG AAT GTG CTG T), cut with XhoI and EcoRI and inserted into similarly cut pGMR, generating P-BNK. The myd88 coding region was amplified from a cDNA library using primers myd88-5 (5'- GGC GAA TTC ATG CGC CCT CGA TTT GTA TGC CAT C and myd88-3 (5'- GGC GCG GCC GCT CAG CCC GGC GTC TGC AGC TTC), cut with EcoRI and NotI, and ligated into similarly cut PBNK, generating P-BNK-dMYD88 (Fig. S5A). Note that because this *myd88* transcript lacks a 5' UTR present in the endogenous *myd88* transcript, it is not silenced by mir6.1-myd88-1 or mir6.1-myd88-2, which target the *myd88* 5' UTR.

Generation of *Medea*Myd88:

P-BNK-dMYD88 (Fig. S5A) was cut with XhoI and 5’ ends were dephospholated with Calf Intestinal Alkaline Phosphatase (NEB, Beverly MA). A gypsy insulatorcontaining DNA fragment was amplified from genomic DNA using a 5’ primer (gypsy 5’) that contained a number of restriction enzyme target sites (SalI, HpaI, AvrII, NheI, SpeI, KpnI, BglII) (5’- GGC GTC GAC GTT AAC CTA GGC TAG CAC TAG TGG TAA CCC CGA GAT CTT CAC GTA ATA AGT GTG CGT TGA ATT TAT TCG C) and a second primer (gypsy 3’), which contains an XhoI site, (5’- GGC CTC GAG AAT TGA TCG GCT AAA TGG TAT GGC AAG AAA AG). This PCR fragment was cut with SalI and XhoI and ligated into XhoI cut P-BNK-dMYD88. This created Pgypsy-BNK-dMYD88 (Fig. S5B), which was cut with AvrII and dephosphorylated with Calf Intestinal Alkaline Phosphatase. The modified *bicoid* promoter and myd88-targeting miRNAs in pBic-mir6.1-myd88 (Fig. S4D) were amplified using primers 4 BIC Myd 1+2 5’ AvrII- 5' GGC CCT AGG GTC GAG TTA GAT CTC AAT T and 4 BIC Myd 1+2 3’ HpaI /SpeI/ Nhe- 5' GGC GTT AAC ACT AGT GCT AGG GCC TTC TAG ACC CCG GCC GC. The PCR product was cut with AvrII and NheI and ligated into AvrII cut Pgypsy- BNK-dMYD88. One plasmid product, designated P-bicoid-myd88RNAi-gypsy-BNK-dMYD88, was selected in which the *bicoid* and *myd88* promoters were oriented so as to transcribe in opposite directions (Fig. S5C).

Generation of *Medea*myd88-int:

The plasmid P-Bnk-dMyd88 was cut was cut with EcoRI and NotI, the vector purified and then ligated with PCR fragments dMyd88 exon 1/intron 1(EcoRI, SpeI) and intron 1/

exon2 (SpeI, NotI), generating P-Bnk-dMyd88-intron. PCR fragment dMyd88 exon 1/intron contains dMyd88 exon 1and a 5' splice site. This fragment was amplified by PCR with primer Myd88 EcoRI 5' (5'-GGC GAA TTC ATG CGC CCT CGA TTT GTA TGC C) and Myd 88 exon 1 /intron SpeI /NheI 3' (5'- GGC ACT AGT GGC CGC TAG CAG CGA CTA CCA TAA GTA AAA AAT AGT TAA TGC CTA CCC AGA TTC TCC TGG ATA TCG TCG CAG). The second PCR fragment, dMyd88 intron 1/ exon2, contains intron 1, a pyrimidine-rich trait and 3' splice site as well as the 3' exons of dMyd88 derived from the cDNA. This fragment was amplified by PCR with primers myd88-EXON2–intron-SpeI-5 (5'- GGC ACT AGT TAG TAA AAC TGT TTT AAT TTT GCT CTC CTC AAA AGC CAA GGA CAC CCA GCG CTT CAT CAT G) and myd88-NotI-3 (5'- GGC GCG GCC GCT CAG CCC GGC GTC TGC AGC TTG C). PBnk-dMyd88-intron was cut with SpeI, which cuts within the intron. The product was dephosphorylated with Calf Intestinal Alkaline Phosphatase according to the manufacturers instructions (NEB, Beverley MA) and ligated with a PCR product containing the modified *bic* promoter and the Myd88- targeting miRNAs. Primers used were 4 BIC Myd 1+2 5' SpeI, which also contains a consensuses branch point site GAT TAG ATG, (5' GGC ACT AGT TAT TGA TTA GAT GTC GAG TTA GAT CTC AAT T) and 4 BIC Myd 1+2 3' NheI (5'- GGC GCT AGC GGC CTT CTA GAC CCC GGC CGC). The final product was designated as Pw+*Medea*myd88-int.

Transgenesis and Population cage experiments:

Germline transformants were generated in a w^{1118} background using standard techniques, by Rainbow Transgenic Flies, Inc (www.rainbowgene.com, Newbury Park,

CA). The exact origin of the chromosomes present in the w^{1118} strain used for transgenesis and population replacement experiments, and the relationship of these chromosomes to those present in the OR strain, is unknown. The second and third chromosomes were isogenized in a *w1118* background in the early 1990s as a prelude to a large mutagenesis screen for regulators of Ras pathway signaling (*6*). The stock has been maintained continuously in the laboratory since then. All fly experiments were carried out at 25C, ambient humidity in 250 ml bottles containing Lewis medium (*4*) supplemented with live dry yeast. Fly rearing was carried out in a light tight chamber placed in an incubator or in a darkened incubator. In a first set of experiments (three green lines in Fig. 1E,F), three populations of 50 males heterozygous for *Medea*myd88-1 (w^{1118}/Y; Pw+*Medea*$^{myd88\text{-}1}$/+) were each crossed with 50 w^{1118} females in separate bottles. In a second set of experiments (four blue lines in Fig. 1E, F) 25 wildtype (+/+) males and 25 homozygous *Medea*myd88-1/ *Medea*myd88-1 males were crossed with 50 w^{1118} females in separate bottles. For both sets of experiments flies were allowed to lay eggs for four days, after which adults were removed. Progeny were allowed to develop, eclose and mate for another 10 days. All adult progeny were collected at this single timepoint and their genotypes determined using eye color as a marker (no *Medea*, w^{1118} = white eyed; *Medea*$^{myd88\text{-}1}$/+ = yellow/orange eyed; *Medea*$^{myd88\text{-}1}$/ *Medea*$^{myd88\text{-}1}$ = darker red eyed). Note that adult progeny continued to eclose after the time of collection. These were not counted or transferred into the subsequent generation population. Numbers of adults scored per population per generation ranged between 102 and 601 (mean 333; s.d.105). Following counting, progeny were transferred to fresh bottles and allowed to lay eggs for four days, and the cycle repeated. In a second set of population cage

experiments, *Medea*$^{myd88\text{-}1}$ was first introduced into a *w+* background in order to decrease the possibility that P^{w+} expression from the vector carrying *Medea*myd88 was providing these animals with an unknown (vision independent) fitness advantage. Females homozygous for *Medea*$^{myd88\text{-}1}$ were mated with Oregon R (OR) males, which carry a wildtype copy of the *w* gene at the endogenous locus on the X chromosome. Progeny males, which are *w+*/Y; *Medea*myd88-1 were then mated as above to OR virgin females to initiate the population cage experiments. The presence of the endogenous *w+* gene completely prevents direct identification of *Medea*myd88-1-bearing individuals based on eye color, since the endogenous *w+* gene is expressed at very high levels compared with the P*w+* in our transformed strain. We determined genotypes at the end of the 12th generation in the following way. 200 males from each population were mated singly in vials with *w1118* females. All male progeny inherit the *w1118* chromosome from their mothers (making them *w1118* with respect to the endogenous locus on the X chromosome). However, those males that carry one copy of *Medea*$^{myd88\text{-}1}$ give rise to 50% red-eyed progeny, while those homozygous for *Medea*$^{myd88\text{-}1}$ give rise to all red-eyed progeny. In each of the three populations the % of non-element-bearing males was less than 1%: population #1 = 0/200 +/+, 30/200 *Medea*$^{myd88\text{-}1}$/+, 170/200 *Medea*$^{myd88\text{-}1}$/ *Medea*$^{myd88\text{-}1}$; Population #2 = 1/200 +/+, 67/200 *Medea*$^{myd88\text{-}1}$/+, 132/200 *Mede*$^{a myd88\text{-}1}$/ *Medea*$^{myd88\text{-}1}$; Population #3 = 0/200 +/+, 58/200 *Medea*$^{myd88\text{-}1}$/+, 142/200 *Medea*$^{myd88\text{-}1}$/ *Medea*myd88-1. Note that the OR genetic background into which *Medea*$^{myd88\text{-}1}$ was introduced should be assumed to be distinct from that of the *w*^{1118} strain used for transgenesis and the population replacement experiments described above. Thus, while evidence for population replacement was obtained in both sets of experiments they are

not strictly comparable in the sense that more variables have been changed than just the status of the *w* gene (w^{1118} versus w^{+}).

Embryo and adult viability determination:

Adult viability for the crosses presented in Table 1 and Table 2 was determined as follows. 50 adult males of the indicated genotype were allowed to mate with 50 virgin females in bottles supplemented with dry yeast for three days. 10 bottles were established for each cross. Adults were then removed. Adult progeny from each bottle were collected, genotyped and counted (either directly or by weighing and comparing with a standard) for 10 days following eclosion of the first progeny. For embryo viability counts, 2-4 day old adult virgin females were allowed to mate with males of the relevant genotypes for 2-3 days in egg collection chambers supplemented with wet yeast paste. On the following day, a 3 hr egg collection was carried out, after first having cleared old eggs from the females through a pre-collection period on a separate plate for three hrs. Embryos were isolated into groups of 100 and kept on an agar surface at 25oC for 48-72 hrs. The % survival was then determined by counting the number of unhatched embryos. Four groups of 100 embryos per cross were scored in each experiment, and each experiment was carried out three times. The results presented are averages from these three experiments. Embryo survival was normalized with respect to the % survival observed in parallel experiments carried out with the *w1118* strain used for transgenesis.

Modeling *Medea* population spread

In order to model the spread of *Medea* and compare this with our experimental observations, a deterministic model was created. In this model, we used as our initial conditions 25% non-element-bearing (wildtype; WT) males, 25% homozygous *Medea* bearing males, and 50% WT females for *Medea*. These initial conditions, which mimic those of the first set of population cage experiments (green lines in Fig. 1E, F) are mathematically equivalent to the second set of crosses carried out, involving heterozygous *Medea*myd88/+ males crossed to wildtype (+/+) females) (blue lines in Fig. 1E, F),. Genotypes for each generation were calculated beginning with genotype frequencies from the previous generation. We assumed random mating, nonoverlapping generations, and interfamily competition. We assume that maternal effect lethality for progeny of *Medea* that fail to inherit the element is 100%, as observed for *Medea*myd88 (Table 1).

We kept track of the frequency of homozygous, heterozygous, and WT male and female flies in each generation. For this model, aside from the first generation, the frequency of WT males and females are equal, the frequency of heterozygous males and females are equal, and the frequency of homozygous males and females are equal. From the frequency of each type of fly, we calculated the frequency of each allele type. For example, in order to calculate the frequency of WT females in generation n+1 we begin with the frequency of WT females in generation n. We calculate the percent of the next generation that will be wild type by finding the frequency of viable WT female embryos normalized (divided by) the total viable population (WT, heterozygotes and homozygotes). WT female flies can arise from (½ WT males x WT females) + (1/4 of

WT female x heterozygous males). WT embryos derived from crosses between heterozygous females x WT males will die as embryos and not contribute to the population. We calculate the total viable population by summing the WT females, WT males, heterozygous females, heterozygous males, homozygous females and homozygous males.

When we assessed a fitness cost (as in Fig. 1C, D and Fig. S3), this cost was assumed to be an additive fitness cost. The fitness cost was expressed as a fraction of the embryos that die. That is, for a 5 percent fitness cost, 95 percent of the heterozygous embryos are viable, while 90 percent of the homozygous embryos were viable. For the model presented in Fig. S3, a specific additional fitness cost of 80% was incurred by progeny of homozygous *Medea*myd88-int mothers that inherited only one copy of the element (the fathers being either *Medea*myd88-int/+ or +/+). The formulas used are shown below. WTf, Hetf, Homof, WTm, Hetm, and Homom refer to the fraction of the adult population that is WT female, heterozygous female, homozygous female, WT male, heterozygous male, homozygous male, respectively. FitCost is the fitness cost, TotalViablePopulation=the total viable population, WTembryo, Hetembryo, HomoembryoWT embryo, and Hetembryo refer to the number of WT, heterozygous, and homozygous embryos produced before the effects of the medea element are included and without normalization. All subscripts refer to the generation.

$$WTf_{n+1} = WTm_{n+} = \frac{.5WTembryo_{n+1}}{TotalViablePopulation_{n+1}}$$

$$Hetf_{n+1} = Hetm_{n+} = \frac{.5Hetembryo_{n+1}(1 - FitCost)}{TotalViablePopulation_{n+1}}$$

$$Homof_{n+1} = Homom_{n+} = \frac{.5Homoembryo_{n+1}(1 - 2FitCost)}{TotalViablePopulation_{n+1}}$$

where,

$$TotalViablePopulation_{n+1} = WTembryo_{n+1} + Hetembryo_{n+1}(1 - FitCost) + Homoembryo_{n+1}(1 - 2FitCost)$$

$$WTembryo_{n+1} = WTf_n WTm_n + 0.5WTf_n Hetm_n$$

$$Hetembryo_{n+1} = WTm_n Homof_n + 0.5WTm_n Hetf_n + 0.5Hetm_n WTf_n + 0.5Hetm_n Hetf_n + 0.5Hetm_n Homof_n + Homom_n WTf_n + 0.5Homom_n Hetf_n$$

$$Homombryo_{n+1} = 0.5Homom_n Hetf_n + 0.5Hetm_n Homof_n + 0.25Hetm_n Hetf_n + Homom_n Homof_n$$

All model calculations were carried out in Excel for Windows XP. Matrix manipulation calculations were carried out in Maple 9.01 (Maplesoft, a division of Waterloo Maple Inc, 2003).

To determine if the experimental population data presented in Fig. 1 conformed to the model, the methods of Wilson (*5*) were used. Briefly, these methods are based on a Chi-square analysis of allele frequency corrected for the non-independence of generations (the frequency of allele p at generation 3 is dependent on the frequency of allele p (the

Medea allele) in generation 2). Because of the non-independence of generations, a covariation matrix is used. The covariation matrix for a particular category (experimental trial), *c*, is denoted W_c and is calculated as

$$(W_c)_{tt} = \frac{1}{n_{c,t}} + (1 - \frac{1}{n_{c,t}}) \left[1 + (w_c - 1) \prod_{j=1}^{t-1} \left(1 - \frac{1}{N_{c,j}} \right) \right]$$

and

$$(W_c)_{t,t+r} = (W_c)_{t,t+r} = 1 + (w_c - 1) \prod_{j=1}^{t-1} \left(1 - \frac{1}{N_{c,j}} \right)$$

where

$N_{c,j}$ is the number of genes (twice the number of individuals) taken at generation *j* to be parents of the next generation. $n_{c,j}$ is the number of genes (twice the number of individuals) used for genetic analysis. To be conservative (that is, to make the assumption that our population has the largest possible variation), w*c* was calculated as $(V_c)_{O,O}/\pi_{C,O}(1-\pi_{c,o})$, where V_c is the sample variance of the initial population and π_C is the theoretical variance of the initial population.

The difference vector, γ_c^m, for any particular model, $_m$, is calculated as the difference between the normalized gene frequencies observed experimentally and the normalized gene frequencies expected from our model using a particular fitness cost. Each of these frequencies are normalized using the arcsin transformation: $2 \sin^{-1} \sqrt{p}$, modified such

that if p=0, the transform equals$2\sin^{-1}\sqrt{\frac{n_t}{2}}$ where n_t is the number of individuals in a particular generation, *t*. We calculate the Chi-square test statistic for a particular model, *m*, as $(\chi^m)^2 = \sum_c (\gamma_c^m)^T W_c^{-1} \gamma_c^m$. The degrees of freedom in each category are calculated as the number of generations observed less the number of fitted parameters in the model. To find the total degrees of freedom, the degrees of freedom in each category are summed. The values of p range from 0 (all individuals are homozygous for the construct) to 1 (all individuals are homozygous wildtype). When all 7 experimental trials were considered as a group, the Chi-square test statistic for no fitness cost was 88.3. There are 122 degrees of freedom, leading to a critical $\chi 2$=148.8, with a probability error threshold (P value) of 0.05. Our Chi-square value is less than the critical value, indicating the data from our experiments is not significantly different from the theoretical model. The minimal Chi-square value to the nearest percent occurs at 0 fitness cost. Due to the conservative nature of this test, the confidence interval is large. To the nearest 0.01, the fitness cost can range from -0.23 to 0.10 and be consistent with our observations at a P value of .05. This range is calculated by finding the first value resulting in a Chi square value falling outside the critical value, 147.7 (d.f=121).

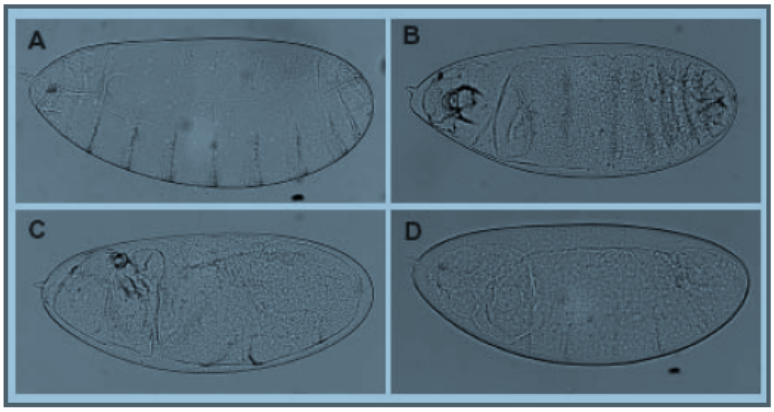

Fig. S1. Embryos of *Medeamyd88* mothers that fail to inherit *Medeamyd88* have dorsal ventral patterning defects. (A) Cuticle preparation of an embryo from wildtype parents. Anterior is to the left, and dorsal is up. Rows of denticle belts are visible on the ventral side of the embryo. (B-D) Embryos from *Medeamyd88*/+ mothers. Ventral denticle belts are decreased in size (B) or largely absent (C,D), consistent with the dorsalization expected on loss of maternal Toll signaling.

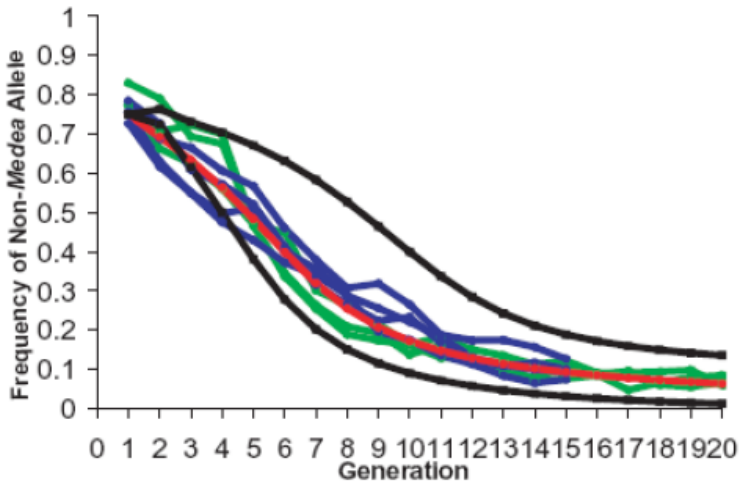

Figure S2. The population replacement behavior of *Medea*$^{myd88\text{-}1}$ is consistent with that of a *Medea* carrying little or no fitness cost. The frequency of the non-*Medea*-bearing chromosome (the + allele) is plotted over generations, from Fig. 1F. The black lines identify boundary conditions for a *Medea* with a fitness gain of 23% or a fitness cost of 10%, values at which we would reject the hypothesis that *Medea*myd88-1 conforms to the model. See SOM methods for details.

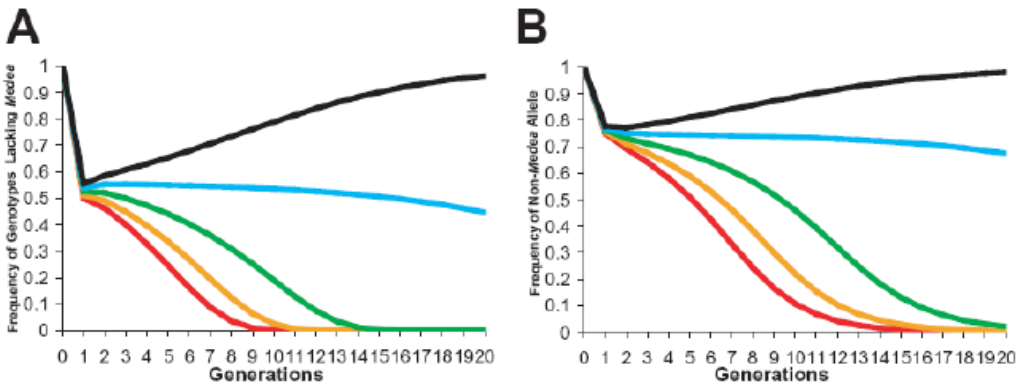

Figure S3. Drive characteristics of a *Medea* elements with the fitness characteristics of *Medeamyd88-int*. *Medea* is modeled as being introduced into the population at an allele frequency of 25%, representing a scenario in which equal numbers of wildtype and homozygous *Medea*-bearing males are mated with wild, non-element- bearing females, as in Fig. 1C, D. (A) Frequency of individuals lacking *Medea* for an element in which progeny of homozygous *Medea* females that inherit only one copy of the *Medea* suffer an 80% fitness cost (embryo mortality), in addition to either no fitness cost (red line), a five percent additive fitness cost (yellow line), a 10 percent additive fitness cost (green line), a 15% additive fitness cost (blue line), or a 20% additive fitness cost (black line) over generations. (B) Frequency of the + allele (non-element bearing chromosome) for the populations described in (A). For a *Medea* that has a fixed, additive fitness cost (such as described in Fig. 1C, D), the frequency of non-*Medea* bearing chromosomes initially decreases rapidly, but slows and eventually reaches a stable equilibrium value. This is due to *Medea*-dependent selection against the non element-bearing chromosome, which is balanced by its increased fitness relative to that of *Medea*-bearing counterparts, as detailed by Wade and Beeman (*13*). In contrast, when the progeny of homozygous *Medea*-bearing females cannot be rescued by a single *Medea*, then the non-element bearing homolog comes under selective pressure that increases with the frequency of homozygous *Medea*-bearing females in the population. For situations in which the fixed additive fitness costs are modest (0-10%) this results in a more dramatic reduction of the non-Medea-bearing chromosome from the population. *Medea* elements with this characteristic may be useful for population replacement since the presence of two effector copies in each individual should also delay the re-appearance of disease carriers when the effector mutates to inactivity.

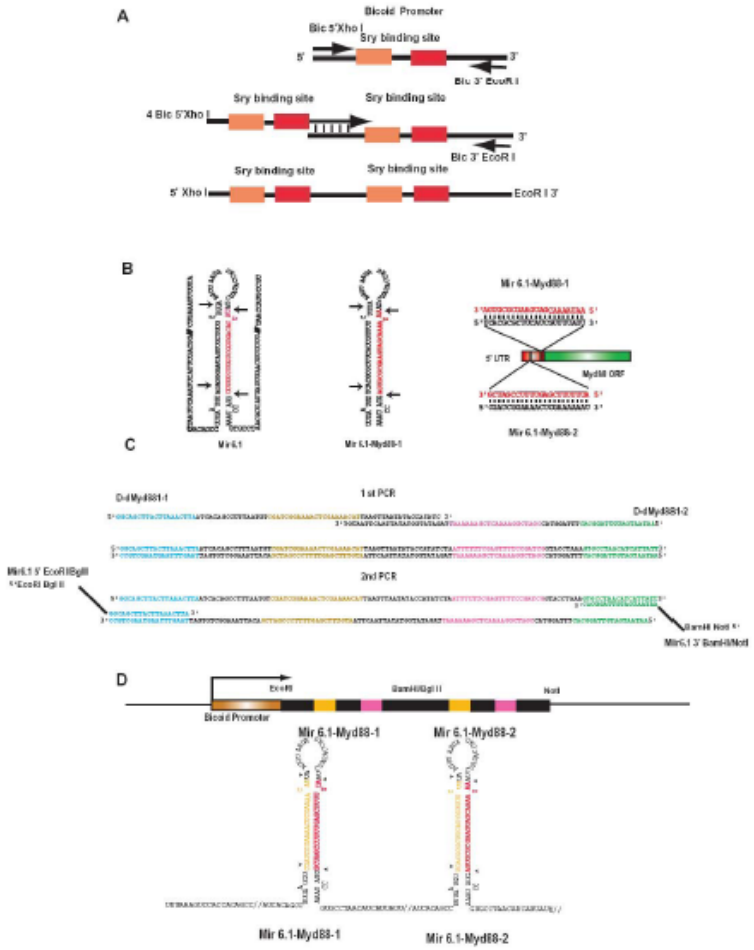

Figure S4. Scheme for the generation of a stronger version of the *bic* promoter, and for the generation of a transcript carrying two miRNAs designed to silence maternal *myd88*. See SOM methods for details. (A) Generation of a stronger bic promoter. (A, top) PrimersBic 5'Xho1 and Bic 3' EcoR1 were used to amplify the bicoid promoter. (A, middle) This fragment was re-amplified with Bic 3' EcoR1 and a second 5' primer (4 Bic 5' Xho1) which included a dimer of Sry binding sites. (A, bottom) The final product, which carries an Xho1 site at the 5' end, and an EcoR1 site at the 3' end, constitutes the modified bicoid promoter. (B) The *Drosophila* miRNA mir6.1 stem-loop precursor and surrounding sequences is illustrated to the left, with the mature miRNA indicated in red. Processing sites for Drosha and Dicer are indicated (arrows). 22bp sequences corresponding to mature mir6.1 were replaced with sequences perfectly complementary to sequences from the *myd88* 5'UTR, generating two new miRNAs, mir6.1-myd88-1 (shown) and mir6.1-myd88-2. Mature mir6.1-myd88-1 and mir6.1-myd88-2 are indicate in red, and complementary *Myd88* mRNA sequences targeted by these miRNAs are indicated in black. (C) Strategy for the synthesis of mir6.1-myd88-1 using two rounds of PCR. The first round of PCR amplifies the miRNA stem loop (miRNA and miRNA*strand indicated in pink and yellow, respectively). This product was amplified in a second round of PCR using oligonucleotides that provide mir6.1 flanking sequences (not shown) and restriction sites for cloning. (D, upper) Schematic of the sequences that

make up the bic-driven miRNAs that target myd88. Cloning sites are indicated. (D, lower) Stem loop regions and surrounding sequences of bic-mir6.1-myd88-1 + myd88-2. The *bic* promoter is located to the left.

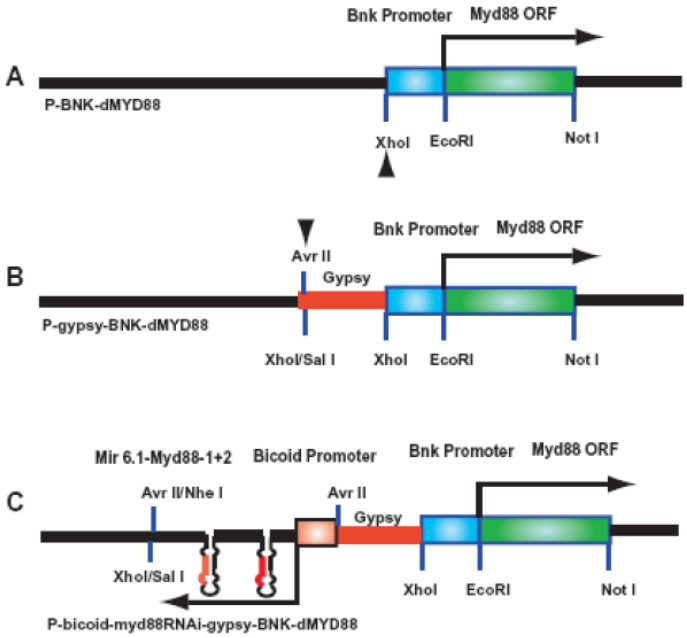

Figure S5. Schematic depicting key intermediates in the generation of a P element expressing *myd88*-silencing miRNAs under the control of the maternal *bic* promoter, and a miRNA-insensitive version of *myd88* under the control of the transient, early *bnk* promoter. See SOM methods for details.

Supplementary References

S1. B. A. Hay, T. Wolff, G. M. Rubin, *Development* **120**, 2121 (1994).

S2. C. Ruez, F. Payre, A. Vincent, *Mech. Dev.* **78**, 125 (1998).

S3. E. D. Schejter, E. Wieschaus, *Cell* **75**, 373 (1993).

S4. E.B. Lewis, *Dros. Inf. Serv*. **34**, 117 (1960).

S5. S. R. Wilson, *Genetics* **95**, 489 (1980).

S6. F. D. Karim, H. C. Chang, M. Therrien, D. A. Wassarman, T. Laverty, and G. M. Rubin, *Genetics* **143**, 315 (1996).

Chapter 3: *Medea* selfish genetic elements as tools for altering traits of wild populations: A theoretical analysis

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

Catherine M. Ward[1], Jessica T. Su[1], Yunxin Huang[2*], Alun L. Lloyd[3], Fred Gould[2], and Bruce A. Hay[1]

[1]Division of Biology, MC156-29, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125.

[2]Department of Entomology, North Carolina State University, Raleigh, North Carolina, 27695

[3]Department of Mathematics, North Carolina State University, Raleigh, North Carolina, 27695

*Current address: Department of Entomology, University of Kentucky, Lexington, KY 40546

Key Words: population replacement, mosquito, maternal effect, introgression, dengue, malaria

Running head: *Medea* selfish elements and population replacement

Abstract

One strategy for controlling transmission of insect-borne disease involves replacing the native insect population with transgenic animals unable to transmit disease. Population replacement requires a drive mechanism to ensure the rapid spread of linked transgenes conferring disease refractoriness. *Medea* selfish genetic elements have the feature that when present in a female, only offspring that inherit the element survive, a behavior that can lead to spread. Here we use modeling to identify conditions under which *Medea* elements spread. We derive equations describing the allele frequencies required for spread of *Medea* elements with a fitness cost, and the equilibrium allele frequencies attained. We show that when *Medea* spreads, it drives the non-*Medea* genotype out of the population, and we provide estimates of the number of generations required to achieve this goal. We also characterize two contexts in which *Medea* elements with fitness costs drive the non-*Medea* allele from the population: an autosomal element in which zygotic rescue is incomplete and an X-linked element in species in which X/Y individuals are male. Finally, we explore costs and benefits associated with the introduction of multiple *Medea* elements. Our results suggest that *Medea* elements can drive population replacement under a wide range of conditions.

Introduction

Insects act as vectors for many human, animal and agricultural diseases. Mosquitoes are particularly important vectors for a number of important human diseases, including malaria and dengue fever. Effective vaccines against these pathogens do not exist, and in the case of malaria the appearance and spread of drug-resistant *Plasmodium* is a constant source of concern (Hyde 2005; Greenwood et al. 2008). Vector suppression through the release of sterile males, the use of insecticides, or modification of the environment, provides an important approach for limiting mosquito-borne disease (Kitron and Spielman 1989; Trape et al. 2002; Gould and Schliekelman 2004). However, each method has limitations. Release of sterile males and the use of insecticides provide only transient population suppression. Insecticides affect many non-target species and mosquitoes often evolve resistance to these compounds. Wholesale modification of the environment may not be feasible or desirable in some situations based on ecological concerns.

Replacement of insect disease vectors with modified counterparts refractory to pathogen transmission is a long-established concept for disease prevention (reviewed in Braig and Yan 2001; Gould and Schliekelman 2004; Sinkins and Gould 2006). Mosquitoes with a diminished capacity to transmit *Plasmodium* or dengue have been identified in the wild and/or created in the laboratory (de Lara Capurro et al. 2000; Ito et al. 2002; Moreira et al. 2002; Franz et al. 2006), demonstrating that the insect immune system can be harnessed to attack these pathogens. However, insect disease vectors are spread over wide areas, and in order to achieve significant levels of disease control it will be necessary to replace a large percentage of the

wild mosquito population with refractory insects (Macdonald 1957; Boete and Koella 2002; Boete and Koella 2003). In addition, enhancement of immune function in insects is generally thought to be costly, requiring tradeoffs with other life history traits such as longevity and fecundity. This fitness cost is likely to result in a decreased ability to pass on alleles conferring pathogen resistance to future generations relative to alternative, disease transmission-permissive alleles (Schmid-Hempel 2005; Vernick et al. 2005; Tripet et al. 2008). Therefore, simple mass release of transgene-bearing insects is unlikely to result in a high enough frequency of disease refractory individuals to prevent disease transmission. Together, these observations imply effective population replacement will require that genes conferring disease refractoriness be linked with a genetic mechanism for driving them through the wild population at greater than Mendelian frequencies.

Selfish genetic elements have increased rates of transmission relative to the rest of the genome of the individual in which they appear. This can result in their spread within a population even if they are neutral or lead to fitness costs in the organisms in which they reside (Burt and Trivers 2006). Such elements are said to "drive" themselves into a population. Many people have proposed linking genes for disease refractoriness with a selfish genetic element. The hope is that these beneficial genes will hitchhike with the selfish genetic element as it drives itself through the native insect population, thereby replacing the wild, disease-permissive population with one that is refractory to disease transmission. A number of selfish genetic elements have been considered as vehicles for driving genes into populations. These include transposons, male post meiotic segregation distorters, homing endonucleases, B-chromosomes, *Medea* elements, and the intracellular bacterial symbiont *Wolbachia*. Translocations, compound

chromosomes, 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, have also been proposed as ways of driving linked genes into populations (reviewed in Braig and Yan 2001; Gould and Schliekelman 2004; Sinkins and Gould 2006).

Maternal-effect lethal selfish genetic elements are members of a class of selfish genetic elements that enhance their transmission by causing the death of offspring that lack the element (Burt and Trivers 2006). These elements were first described in the flour beetle *Tribolium castaneum* and are known by the acronym *Medea* (maternal-effect dominant embryonic arrest). *Medea* elements in *Tribolium* are chromosomally located and gametes are generated and transmitted in a strictly Mendelian manner. *Tribolium Medea* has the feature that when present in females, only progeny that inherit the element-containing chromosome from either the maternal and/or paternal genome survive (Fig. 1A) (Beeman et al. 1992). In contrast, *Medea*-bearing males give rise to wildtype and *Medea*-bearing progeny with equal frequency when mated to wildtype females. Therefore, *Medea* enhances its transmission relative to competing non-*Medea*-bearing homologous chromosomes (hereafter referred to as the non-*Medea* allele) by causing the death of progeny that do not carry a copy of *Medea* found in the mother (Fig. 1A). One *Tribolium Medea*, $Medea^{M1}$, has been mapped and is associated with a composite Tc1 transposon insertion that includes a number of genes (Lorenzen et al. 2008). How this insertion confers maternal-effect lethal selfish behavior is unknown, but genetic analysis suggests a model in which *Medea* consists of two tightly linked loci: one that encodes a maternal-effect lethal activity (a toxin) inherited by all progeny of *Medea*-bearing mothers, and a second that encodes a zygotic rescue activity (an antidote) that can be inherited from either

the maternal or paternal genome (Beeman et al. 1992; Beeman and Friesen 1999) (Fig. 1B). Loci with genetic characteristics similar to those of *Tribolium Medea* have also been described in mice, but little is known about their molecular basis (Hurst 1993; Peters and Barker 1993; Weichenhan et al. 1996). Synthetic *Medea* elements that drive population replacement in *Drosophila* have been generated (Chen *et al.* 2007). The genetic and cell biological principles utilized to generate these elements involve maternal-specific silencing of a gene whose product is synthesized maternally and deposited into the developing oocyte, but only required during embryogenesis (the toxin), coupled with zygotic expression of a rescuing transgene (the antidote). These principles are straightforward and conserved across the animal kingdom, and therefore should be applicable to the generation of similar elements in agricultural pest and human disease vector species.

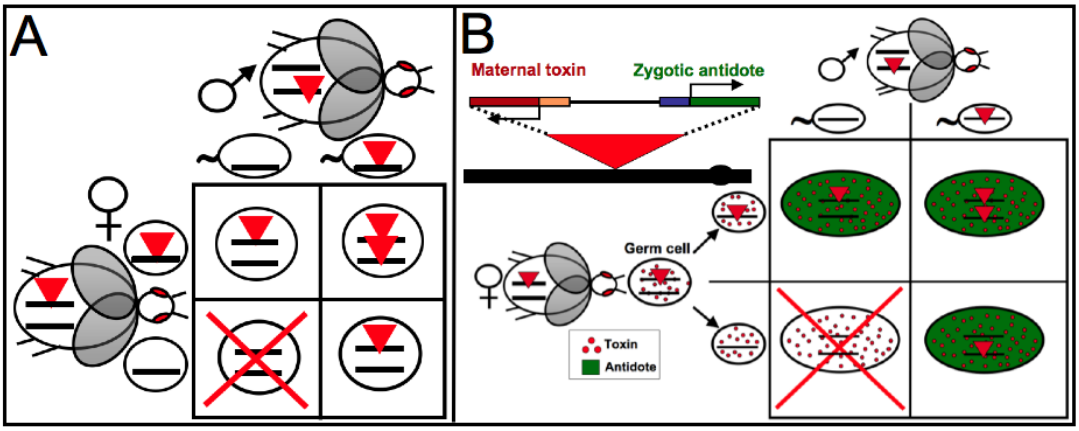

Figure 1.

Genetic behavior of *Medea*, and the molecular logic underlying synthetic *Medea* elements in *Drosophila*. See text for details.

The dynamics of *Medea* spread have been modeled by several groups. Wade and Beeman first considered this problem, focusing on several situations in which *Medea* was introduced into populations at very low frequency (Wade and Beeman 1994). They showed that if the presence of *Medea* does not result in a fitness (fecundity) cost to carriers, independent of maternal-effect killing, *Medea* spreads to fixation for all degrees of maternal effect lethality, though the rate of *Medea* increase is very slow when it is introduced at low frequency. They also showed that if the presence of *Medea* results in a decrease in fecundity independent of maternal-effect killing, the frequency of the *Medea* allele could still increase to a stable internal equilibrium. Whether *Medea* increased in frequency or was lost from the population was found to depend critically on the degree of dominance of the fitness costs: high levels of completely recessive fecundity costs were well tolerated, but when *Medea* was present at low frequency, even modest degrees of dominance led to loss of *Medea* from the population. These authors, and Smith (Smith 1998), showed that *Medea*'s ability to spread in the face of fitness costs could be enhanced if progeny of a *Medea*-bearing mother compete with each other for resources. In this context, known as family-level, or soft selection (Wade, 1985; Kelly 1992), the death of non-*Medea* offspring within the family of a *Medea*-bearing mother frees limited resources for sibling *Medea*-bearing progeny, providing them with a real fitness benefit that increases their likelihood of survival. Any level of family-level selection promotes the spread of *Medea*. In the work below we assume no family-level selection because this assumption provides a more conservative estimate of *Medea*'s potential as a population replacement drive mechanism and because there is no evidence that mosquitoes show family-level selection. That said, some mosquitoes, such as *Aedes aegypti*, an important vector of dengue, breed in small containers that may often be resource-limited for larval growth (Clements 1999), suggesting that family-

level selection could be important in some contexts, a topic that should be further explored. Population genetic models of Hastings (1994), Smith (1998) and Chen *et al*. (2007) show that, in the absence of family-level selection, *Medea* elements with significant dominant fitness costs can still spread, provided they are introduced above a critical introduction frequency. Previous work has focused on the fate of the *Medea* allele. However, it is the fate of *Medea*-bearing genotypes that is important for population replacement. Chen *et al.* (2007) showed that, at least under some conditions, when *Medea* elements with fitness costs are introduced at frequencies that result in spread to an internal equilibrium allele frequency, non-*Medea* individuals are nonetheless rapidly eliminated from the population. However, it has not been clear to what extent this conclusion can be generalized.

In summary, *Medea* elements are attractive candidates to drive population replacement for several reasons. First, they can spread (provided certain conditions detailed below are met) even if they confer a fitness cost to carriers (Hastings 1994; Wade and Beeman 1994; Smith 1998; Chen et al. 2007). Second, under at least some conditions when *Medea* spreads it eliminates the non-*Medea* genotype from the population (Chen et al. 2007). Third, the synthetic form of *Medea* is the only gene drive mechanism that is both well understood at the molecular level, because it was designed with components of known behavior, and that has been demonstrated to drive population replacement. Finally, design considerations discussed by Chen *et al.* (Chen *et al.* 2007) regarding ways to prevent recombinational separation of drive and disease refractoriness functions, to prevent selfish element spread in non-target species, and to carry out multiple cycles of population replacement, provide reasons to believe that the population genetic behavior of synthetic *Medea* elements can to some extent be

controlled. Therefore, *Medea* is a logical target for concerted development efforts (reviewed in Hay et al. 2010). The aim of this paper is to describe the conditions under which *Medea* elements are predicted to succeed as drivers of population replacement in large, unstructured populations.

The model.

We use a deterministic model to examine the invasion of synthetic *Medea* elements into populations. Terms are defined in Table 1. Table 2 presents the frequencies of parental genotypes in the population, and offspring frequencies and genotypes produced by each mating. The deterministic model assumes an infinite population with random mating and discrete, non-overlapping generations. Expression of the toxin/antidote genes that make up *Medea*, and/or the cargo genes linked to *Medea* (genes conferring resistance to pathogen transmission), may result in a fitness cost to carriers. Fitness costs may also arise through tight linkage between *Medea* and a nearby deleterious allele or insertion-dependent effects on the expression of nearby genes. These are fitness costs not associated with the *Medea*-killing itself. We consider three types of fitness cost: an embryonic fitness cost ($c_{E,Het}$ and $c_{E,Homo}$), a maternal fecundity loss ($c_{D,Het}$ and $c_{D,Homo}$), and a paternal fertility loss ($c_{S,Het}$ and $c_{S,Homo}$). $c_{D,Het}$, $c_{S,Het}$, $c_{D,Homo}$, and $c_{S,Homo}$ act on the genotypes of the parents, causing a fecundity loss in females and a fertility loss in males. For example, if wildtypes have a fertility/fecundity of 1, then in heterozygous females this cost can be interpreted as meaning heterozygous females lay only (1-$c_{D,Het}$) fertile eggs, while heterozygous males only successfully fertilize (1-$c_{S,Het}$) eggs. An embryonic cost refers to the fraction of *Medea*-bearing embryos dying as juveniles. Since

costs are likely to be borne by both parents (e.g. insertion site-dependent effects; consequences of toxin/antidote expression) or by the female alone (e.g. costs associated with expression of maternal toxin or disease resistance effector), we do not consider paternal fitness costs in isolation. In some of what follows, it is more convenient to frame the discussion in terms of fitness (V_x) rather than the fitness costs induced by the construct. For example, the fitness of a homozygous female is $V_{D,Homo}=1-c_{D,Homo}$. In our simulations, all fitness costs are assumed to be multiplicative both at a single locus and between loci, and are restricted to values of no more than 20% per copy. At these levels of fitness costs, additive and multiplicative fitness costs behave qualitatively similarly. In addition, the use of multiplicative fitness costs allows us to analyze the behavior of multiple elements without a hard limit on the number of copies per organism. We briefly consider recessive fitness costs but find, following Wade and Beeman, that when costs are completely recessive, *Medea* spreads following any finite introduction, regardless of the cost to homozygotes (Wade and Beeman 1994). Instead, we focus on the more conservative scenario, in which heterozygotes experience a fitness cost. This is likely to be more relevant to population replacement since many of the costs associated with synthetic *Medea* elements, including costs associated with the expression of toxin and antidote, as well as costs associated with expression of transgene effectors, are likely to be dominant.

Table 1: A list of abbreviations.

All abbreviations in the text are listed here.

S_{MM}	Fraction of the male population homozygous for *Medea*
S_{M+}	Fraction of the male population heterozygous for *Medea*
S_{++}	Fraction of the male population homozygous for the non-*Medea* allele
D_{MM}	Fraction of the female population homozygous for *Medea*
D_{M+}	Fraction of the female population heterozygous for *Medea*
D_{++}	Fraction of the female population homozygous for the non-*Medea* allele
G_{MM}	Fraction of the population homozygous for *Medea*
G_{M+}	Fraction of the population heterozygous for *Medea*
G_{++}	Fraction of the population homozygous for the non-*Medea* allele
‘	A prime in all above refers to the next generation
$c_{S,genotype}$	Male fertility loss at a given genotype
$c_{D,genotype}$	Female fecundity loss at a given genotype
$c_{E,genotype}$	Rate of embryonic death at a given genotype, independent of the maternal effect *Medea* killing
p	Allele frequency of the *Medea* element
q	Allele frequency of the WT element
W	Mean fitness
$V_{S,genotype}$	Male fertility retained at a given genotype($1-c_{S,genotype}$). When homozygous fitness is equal to the square of heterozygous fitness, the fitness cost is said to be multiplicative.
$V_{D,genotype}$	Female fecundity retained at a given genotype($1-c_{D,genotype}$). When homozygous fitness is equal to the square of heterozygous fitness, the fitness cost is said to be multiplicative.
$V_{E,genotype}$	Embryonic viability retained at a given genotype($1-c_{E,genotype}$). When homozygous fitness is equal to the square of heterozygous fitness, the fitness cost is said to be multiplicative.
t_0	Fraction of WT offspring of a *Medea*-bearing mother that die.
t_1	Fraction of heterozygous offspring of a homozygous *Medea* mother that live.

Table 2: Parental genotype frequency, fitness effects, and offspring frequency.

Mating frequencies are the product of the genotype frequencies of the male and female parents. A reduction of fertility, fecundity or embryonic viability leads to a reduction in the frequency of viable offspring of some genotypes. When m_0=1, all non-*Medea* offspring of *Medea* heterozygous mothers will live. When m_1=1, all heterozygous *Medea* offspring of *Medea* homozygous mothers will live. By summing all the families and dividing by the mean fitness of the population, we find the genotype frequencies of the offspring. This is explicitly done in the text.

	Parental Genotype Frequency			Fitness/Fecundity/Fertility		Offspring Frequency		
Family	Male	Female	Mating	Male	Female	Homo	Het	WT
1	S_{MM}	D_{MM}	$S_{MM}D_{MM}$	$V_{S, Homo}$	$V_{D, Homo}$	$V_{E, Homo}$		
2	S_{M+}	D_{MM}	$S_{M+}D_{MM}$	$V_{S, Het}$	$V_{D, Homo}$	½ $V_{E, Homo}$	½$V_{E,Het}m^2_1$	
3	S_{++}	D_{MM}	S_{++} D_{MM}	1	$V_{D, Homo}$		$V_{E, Het}$ m_1	
4	S_{MM}	D_{M+}	$S_{MM}D_{M+}$	$V_{S, Homo}$	$V_{D, Het}$	½ $V_{E, Homo}$	½ $V_{E, Het}$	
5	S_{M+}	D_{M+}	$S_{M+}D_{M+}$	$V_{S, Het}$	$V_{D. Het}$	¼ $V_{E,Homo}$	½ $V_{E. Het}$	¼m_0
6	S_{++}	D_{M+}	$S_{++}D_{M+}$	1	$V_{D, Het}$		½ $V_{E, Het}$	½ m_0
7	S_{MM}	D_{++}	$S_{MM}D_{++}$	$V_{S, Homo}$	1		$V_{E. Het}$	
8	S_{M+}	D_{++}	$S_{M+}D_{++}$	$V_{S. Het}$	1		½ $V_{E, Het}$	½
9	S_{++}	D_{++}	S_{++} D_{++}	1	1			1

We consider two types of *Medea*-dependent lethality. The term t_0 refers to the fraction of non-*Medea* progeny from heterozygous *Medea* mothers that die. Typically, we will

consider t_0=1, meaning that embryos from *Medea*-bearing mothers that fail to inherit *Medea* always die. We also consider situations in which heterozygous offspring of homozygous *Medea* mothers have a probability of dying. This death, also considered by Smith (Smith 1998), and observed by Beeman et al. (Beeman et al. 1992) in *Tribolium*, and by Chen et al. (Chen et al. 2007) in *Drosophila*, may represent incomplete zygotic rescue of maternal-effect lethality associated with two maternal copies of the toxin gene and one copy of the zygotic antidote. The fraction that die by this mechanism is modeled as t_1, with t_1=1 meaning that all heterozygous progeny die. Except where noted specifically, t_1=0. It is sometimes useful to consider the fraction of progeny that live (μ): μ_0=1-t_0 and μ_1=1-t_1.

Given the assumptions above, the equations for the genotype frequencies in generation $n + 1$ from those in generation n are

$$G'_{MM} = S'_{MM} = D'_{MM} = \frac{V_{E,Homo}}{W}(S_{MM}D_{MM}V_{S,Homo}V_{D,Homo} + \tfrac{1}{2}S_{M+}D_{MM}V_{S,Het}V_{D,Homo} + \tfrac{1}{2}S_{MM}D_{M+}V_{S,Homo}V_{D,Het} + \tfrac{1}{4}S_{M+}D_{M+}V_{S,Het}V_{D,Het})$$

$$G'_{M+} = S'_{M+} = D'_{M+} = \frac{V_{E,Het}}{W}(\tfrac{1}{2}S_{M+}D_{MM}V_{S,Het}V_{D,Homo}\mu_1 + S_{++}D_{MM}V_{D,Homo}\mu_1 + \tfrac{1}{2}S_{MM}D_{M+}V_{S,Homo}V_{D,Het} + \tfrac{1}{2}S_{M+}D_{M+}V_{S,Het}V_{D,Het} + \tfrac{1}{2}S_{++}D_{M+}V_{D,Het} + S_{MM}D_{++}V_{S,Homo} + \tfrac{1}{2}S_{M+}D_{++}V_{S,Het})$$

$$G'_{++} = S'_{++} = D'_{++} = \frac{1}{W}\left(\frac{1}{4} S_{M+} D_{M+} V_{S,Het} V_{D,Het} \mu_0 + \frac{1}{2} S_{++} D_{M+} V_{D,Het} \mu_0 + \frac{1}{2} S_{M+} D_{++} V_{S,Het} + S_{++} D_{++}\right)$$

where the mean fitness, *W*, equals the sum of the right sides of these three equations multiplied by *W*,

$$\begin{aligned} W = {} & V_{E,Homo}(S_{MM} D_{MM} V_{S,Homo} V_{D,Homo} + \tfrac{1}{2} S_{M+} D_{MM} V_{S,Het} V_{D,Homo} + \tfrac{1}{2} S_{MM} D_{M+} V_{S,Homo} V_{D,Het} + \\ & \tfrac{1}{4} S_{M+} D_{M+} V_{S,Het} V_{D,Het}) + \\ & V_{E,Het}(\tfrac{1}{2} S_{M+} D_{MM} V_{S,Het} V_{D,Homo} \mu_1 + S_{++} D_{MM} V_{D,Homo} \mu_1 + \tfrac{1}{2} S_{MM} D_{M+} V_{S,Homo} V_{D,Het} + \\ & \tfrac{1}{2} S_{M+} D_{M+} V_{S,Het} V_{D,Het} + \tfrac{1}{2} S_{++} D_{M+} V_{D,Het} + S_{MM} D_{++} V_{S,Homo} + \tfrac{1}{2} S_{M+} D_{++} V_{S,Het}) + \\ & \tfrac{1}{4} S_{M+} D_{M+} V_{S,Het} V_{D,Het} \mu_0 + \tfrac{1}{2} S_{++} D_{M+} V_{D,Het} \mu_0 + \tfrac{1}{2} S_{M+} D_{++} V_{S,Het} + S_{++} D_{++} \end{aligned} .$$

In all generations after the introduction, male and female genotype distributions are the same.

Understanding the relationship between genotype frequency, fitness, and *Medea* allele frequency.

Throughout the text we follow allele fitness as a way understanding the fate of *Medea* in populations. By allele fitness we mean the probability that a given allele in a zygote that has survived possible *Medea*-dependent killing will be passed in the next generation to a zygote that also survives *Medea*-dependent killing, given a specific set of population genotype frequencies. This term incorporates fitness losses associated with *Medea*-dependent maternal-

effect killing, as well as killing-independent fitness costs associated with *Medea*. In order to understand the dynamics of *Medea* spread, and how this depends on allele frequency and fitness, we need to be able to describe *Medea* allele and genotype frequencies, and allele and genotype fitnesses, over generations. In a *Medea*-bearing population the fate of an individual depends on the genotype of its mother as well as its own genotype. Thus, knowledge of one genotype frequency after a single round of random mating is not sufficient to characterize the population, as it would be if the population was in Hardy-Weinberg equilibrium.

We approach this problem first by presenting an example, a Medea with a 20% multiplicative embryonic fitness cost. We plot, on a DeFinetti diagram, the trajectories of genotype frequencies over 1000 generations when present in a population initially composed of different proportions of *Medea* homozygotes and non-*Medea* individuals (points along the horizontal axis), non-*Medea* individuals and *Medea* heterozygotes (points along the left axis), or *Medea* homozygotes and heterozygotes (points along the right axis) (Fig. 2A). For this set of parameters, all populations converge to one of two stable equilibrium points, composed of either non-*Medea* individuals, or of two thirds *Medea* homozygotes and one third *Medea* heterozygotes, the stable internal equilibrium allele frequency (SIEAF) (Fig. 2A). The regions of initial conditions that converge to each stable equilibrium are separated by a set of gamete frequencies, known as a separatrix, that define a threshold between *Medea* allele loss and fixation. The separatrix is the stable manifold of the unstable equilibrium (a saddle). This family of points includes one, the unstable internal equilibrium allele frequency (the UIEAF), discussed further below. Importantly, all populations initiating on either side of the separatrix approach and ultimately follow a common trajectory in moving towards one or the other stable

equilibrium (the common trajectory is the unstable manifold of the unstable equilibrium). This observation implies that one can calculate genotype frequencies, and thus allele fitness, as a function of *Medea* allele frequency, by calculating the approximate positions of points on this common trajectory. To do this we take a number of starting parental genotypes distributed throughout the parameter space of all possible parental genotypes, indicated by the black dots in the DeFinetti diagrams in Fig. 2B. Each genotype in the distribution is advanced one generation and all possible genotype distributions for that generation, indicated by the green region, are plotted. The procedure is repeated for a second generation, resulting in the region of possible genotypes indicated in red; for a third generation, resulting in the region of possible genotypes indicated in yellow; and for a fourth generation, resulting in the region of possible genotypes indicated in blue. After four generations the genotype space distribution is very tight (the blue region that resembles a line in Fig. 2B). Throughout the remainder of the text we use the constrained values of genotype space during the fourth generation to calculate genotype frequencies and fitness values with respect to *Medea* allele frequency. Plots of genotype or fitness as a function of *Medea* allele frequency (as in Fig. 3A,C) which appear line-like, are not one-dimensional lines, but narrow two-dimensional bands around a line. Places where the bands cross are not points but small areas.

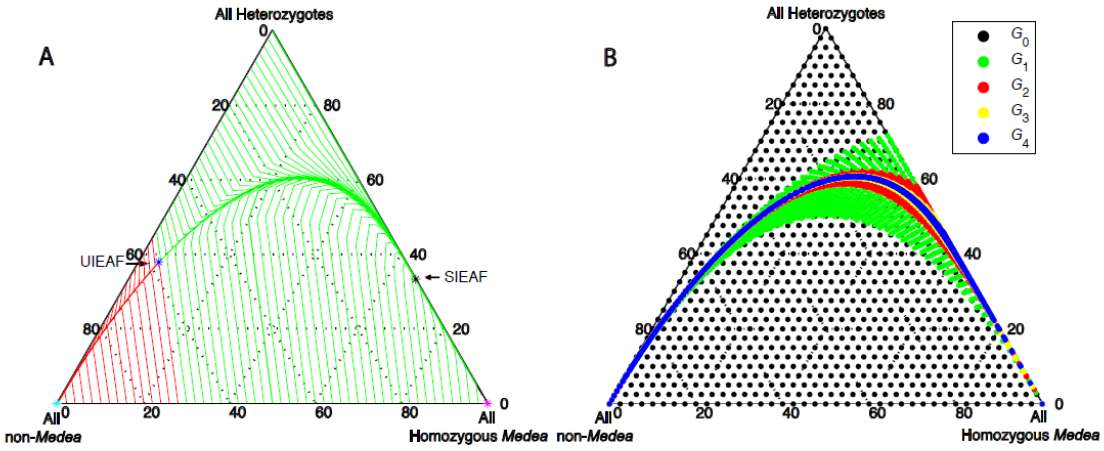

Figure 2.

DeFinetti diagrams showing genotype trajectories for a Medea with a fitness cost. (A) The DeFinetti diagram plots the change in genotype frequencies over generations for a *Medea* with a 20% embryonic, multiplicative fitness cost, and values of t_1=0 and t_1=1. Population trajectories start with different ratios of two of the three genotypes (genotypes corresponding to points along each of the sides of the triangle). Green lines show trajectories that end at 2/3 *Medea* homozygotes, 1/3 *Medea* heterozygotes and no non-*Medea* individuals, the SIEAF (the stable internal equilibrium allele frequency). Red lines indicate population trajectories that end with loss of *Medea* individuals from the population. The unstable internal equilibrium frequency (UIEAF) is a point on the common trajectory taken by *Medea*-bearing populations that separates populations in which *Medea* spreads from those in which *Medea* is lost. (B) Plot of genotype frequencies over four generations for the *Medea* allele in (A), introduced into a population at a number of different starting genotype frequencies (black circles). When adults from within the G_0 genotype distributions (each of the black circles) mate randomly with each other, a range of possible G_1 genotype distributions, indicated by the green region, is obtained. When adults from G_1 genotype distributions mate randomly, a set of possible G_2 offspring genotype distributions defined by the red region is obtained; matings within each G_2 genotype distribution result in the set of possible G_3 offspring distributions defined by the yellow region; and G3 matings result in the G_4 (blue) distribution. The G_4 distribution, which is highly constrained, can be used to approximate genotype frequencies and allele fitness for specific *Medea* allele frequencies.

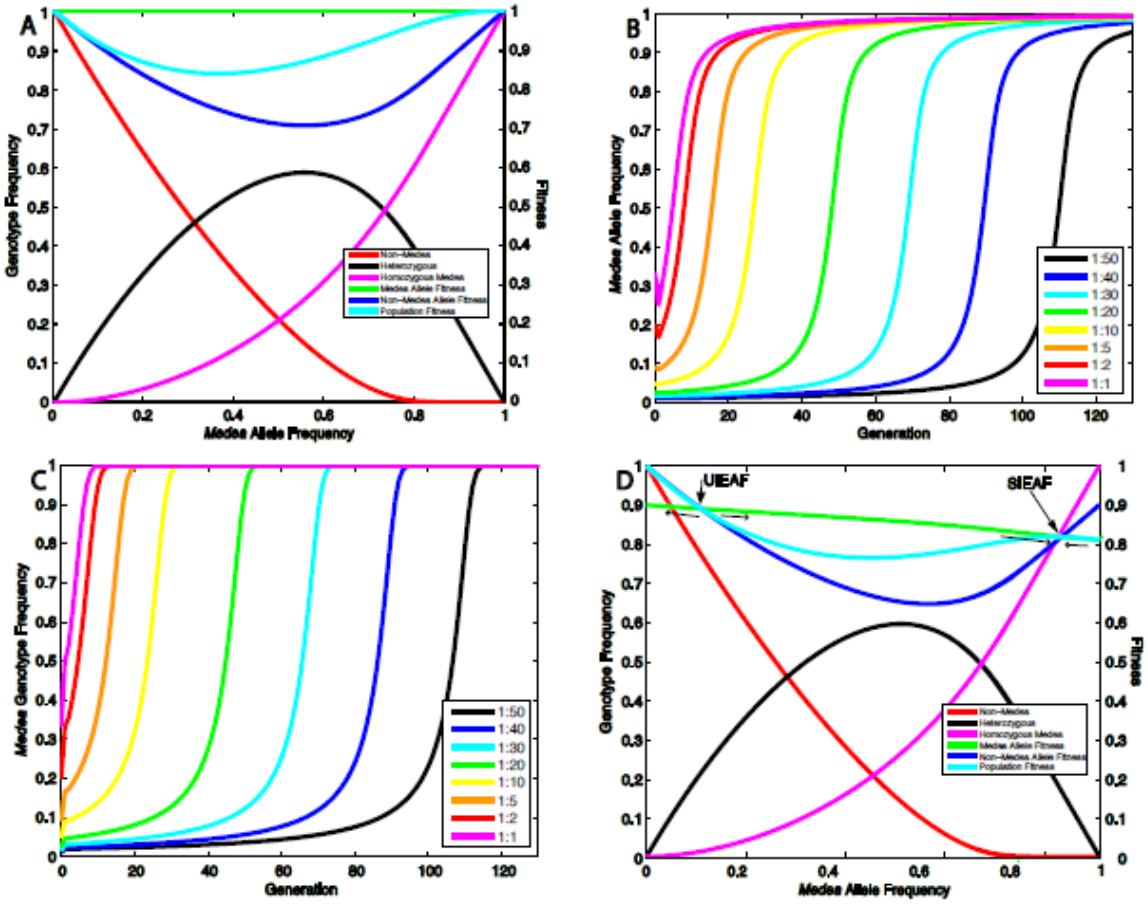

Figure 3.

Characteristics of *Medea* allele and genotype spread as a function of introduction frequency, and of allele fitness as a function of *Medea* allele frequency. (A) The frequency of individuals lacking *Medea* (Non-*Medea*), heterozygotes (Heterozygous), and homozygotes for *Medea* (Homozygous *Medea*), are plotted with respect to *Medea* allele frequency. The fitness of the *Medea* allele, the non-*Medea* allele, and the population is also shown. The presence of *Medea* in the female is assumed to result in 100% death of non-*Medea*-bearing progeny, with all heterozygous *Medea* progeny of homozygous *Medea* mothers being viable. (B) *Medea* allele frequency is plotted as a function of the number of generations for different introduction ratios of homozygous *Medea*/non-*Medea* males into a population of non-*Medea* females. (C) The population frequency of individuals with *Medea* (*Medea* Genotype Frequency) is plotted as a function of generations for different introduction ratios of homozygous *Medea*/non-*Medea* males into a population of non-*Medea* females. (D) Plot of allele and genotype fitness, and genotype frequency, as a function of *Medea* allele frequency, for a *Medea* that carries a 10% embryonic fitness cost. The unstable internal equilibrium allele frequency (UIEAF) and stable internal equilibrium allele frequency (SIEAF) are indicated. Thin arrows indicate the directions in which the *Medea* allele frequency moves on either side of the UIEAF and SIEAF.

An autosomal *Medea* increases to an equilibrium allele frequency when introduced above a threshold frequency determined by fitness costs, and this increase is accompanied by the elimination of non-*Medea* individuals from the population.

Medea increases in frequency by killing alternative non-*Medea* alleles, thereby causing a relative increase in the population frequency of the *Medea* allele. *Medea*-bearing individuals and alleles experience no direct benefit from this killing, but non-*Medea* alleles experience *Medea*-dependent death (a fitness loss) in each generation that is dependent on the *Medea* allele frequency. The relationship between genotype and *Medea* allele frequency, and between allele and population fitness and *Medea* allele frequency, for a *Medea* with no fitness cost, is illustrated in Figure 3A. The *Medea* allele spreads to fixation because its fitness is always greater than that of the non-*Medea* allele. This result agrees with that of earlier works (Hastings 1994; Wade and Beeman 1994; Smith 1998). The rate of *Medea* spread depends dramatically on the introduction frequency. If *Medea* is released into a population at low frequencies there is a long lag phase during which the frequency of *Medea* alleles and individuals increases only slowly because the frequency of *Medea*-dependent killing is low (Fig. 3B, C) (Wade and Beeman 1994). This lag phase is followed by roughly 15 generations that accounts for a dramatic loss of non-*Medea* alleles and individuals (Fig. 3B, C). If *Medea* is released at higher frequencies the lag phase is shortened, but in other respects the population trajectories are very similar (Fig. 3B, C).

We now consider the fate of autosomal *Medea* alleles that have a fitness cost not associated with *Medea*-dependent killing. As an example, we begin by considering the fate of a *Medea* carrying a 10% embryonic fitness cost (Fig. 3D). Recall that fitness costs are multiplicative, so homozygotes carry a 19% fitness cost. Note that the fitness curves for the *Medea* and non-*Medea* allele cross at two positions. These define small regions that include the unstable internal equilibrium allele frequency (UIEAF), a point on the separatrix of Fig. 2A, and the stable internal equilibrium allele frequency (SIEAF) (the stable internal equilibrium of Wade and Beeman (1994), and Smith (1998)), the point to which all gamete frequencies converge in Fig. 2A when *Medea* spreads. At these two points *Medea*-dependent killing of non-*Medea* alleles is balanced by natural selection-dependent loss of *Medea* alleles. If *Medea* is present at a frequency below the UIEAF, the *Medea*-dependent cost to the non-*Medea* allele is less than that associated with carrying the *Medea* allele (the non-*Medea* allele has a higher average fitness than the *Medea* allele), and *Medea* is driven out of the population. In contrast, if the *Medea* allele frequency is just above the UIAEF but below the SIAEF, the non-*Medea* allele has a lower fitness (due to increased *Medea*-dependent killing) than the *Medea* allele, and the frequency of *Medea* increases over the generations towards the SIEAF, even though this is associated with a decrease in overall population fitness. The SIEAF represents a stable upper limit on the *Medea* allele frequency since when the *Medea* allele frequency is higher than the SIEAF, non-*Medea* alleles have, on average, a higher relative fitness; they lack the fitness costs associated with being in *Medea* homozygotes (they are in heterozygotes), and because *Medea* is now common they only rarely suffer the cost of death due to maternal-effect lethality in non-*Medea* progeny. As a result, fitness costs associated with carrying *Medea* cause the *Medea* allele frequency to decrease through natural selection, back towards the SIEAF (Fig. 3D). In

consequence, the SIEAF also represents a local population fitness maximum. If the fitness cost associated with *Medea* is increased, the UIEAF will increase, and the SIEAF (discussed further below) will decrease.

In order to solve for the equilibrium values of this system, we set $G'_{MM} = G_{MM}$ and $G'_{++} = G_{++}$. Recalling that genotype frequencies must sum to one, $G_{MM} + G_{M+} + G_{++} = 1$, and assuming that all fitness values are greater than 0, we find 4 biologically possible equilibria. Two of these equilibria, a population that contains only non-*Medea* individuals and a population that contains only homozygous *Medea* individuals, are trivial. The third contains all three genotypes. This is an unstable equilibrium and is the UIEAF. The final equilibrium includes a mix of heterozygous and homozygous *Medea* individuals and is the SIEAF. The general forms of the equations corresponding to these equilibria are too complicated to be useful. The general forms of the expressions that describe the equilibria and the corresponding Maple code are left to the supplemental materials.

In order to analyze the stability of these equilibria, we use the standard linear stability analysis for difference equations. We calculate the eigenvalues of the Jacobian matrix,

$$\begin{pmatrix} \frac{\partial G'_{MM}}{\partial G_{MM}} & \frac{\partial G'_{MM}}{\partial G_{++}} \\ \frac{\partial G'_{++}}{\partial G_{MM}} & \frac{\partial G'_{++}}{\partial G_{++}} \end{pmatrix}$$

which are evaluated at each equilibrium in turn. The equilibrium is stable if both eigenvalues have modulus less than one. If the eigenvalue with the largest modulus lies on the unit circle (e.g. equals 1 or −1) then linear stability analysis is inconclusive. If either eigenvalue has modulus larger than one then the equilibrium is unstable.

For the trivial equilibrium where there are no *Medea* individuals in the population, we can find the general stability criteria. The eigenvalues of the Jacobian are 0 and $V_{E,Het}(V_{S,Het}+V_{D,Het})$. Notice that the stability does not depend on the fitness of homozygotes. This equilibrium is stable except when heterozygotes have no fitness cost, in which case the analysis is inconclusive. In the case where the population consists of only homozygous *Medea* individuals, the eigenvalues are 0 and

$$\frac{V_{E,Het}(V_{D,Het}V_{S,Homo}+V_{S,Het}V_{D,Homo}-V_{S,Het}V_{D,Homo}t_1)}{V_{E,Homo}V_{D,Homo}V_{S,Homo}}.$$

If t_1=0, $V_{E,Het}\geq V_{E,Homo}$, $V_{D,Het}\geq V_{D,Homo}$, $V_{S,Het}\geq V_{S,Homo}$, and at least one homozygous fitness is less than 1, this equilibrium is unstable. These results imply that for a *Medea* with a dominant fitness cost, for which heterozygotes are more fit than homozygotes, and heterozygous offspring of homozygous *Medea* females always survive, *Medea* will not spread from very low

frequency; if it is present at a frequency from which it does spread, the population will always contain some non-*Medea* alleles.

The general expressions for the equilibrium values for the three genotype (G_{++}, G_{M+}, G_{MM}) or two genotype (G_{++}=0, G_{M+}, G_{MM}) scenarios in which all parameters vary are too cumbersome to be useful. Here we consider a simpler case, in which t_0=1, t_1=0, and there is an embryonic fitness cost associated with *Medea*. We have also considered maternal and parental fitness costs with t_0=1, t_1=0. Explicit expressions for equilibria and stability conditions incorporating these costs are shown in the Supplementary Materials. For each of the three fitness scenarios a maximum of 4 biologically relevant equilibria exist. These are (1) the equilibrium with no *Medea*, which is stable if heterozygotes have a fitness cost; (2) an unstable equilibrium with all genotypes; (3) an equilibrium with heterozygous and homozygous *Medea* individuals (no non-*Medea* individuals) that is stable if the fitness cost to homozygotes is higher than that to heterozygotes, and homozygotes have a fitness of at least 0.3 (situations in which fitness values are below 0.3 and/or homozygotes are more fit than heterozygotes are likely not biologically relevant); and (4) an equilibrium with only homozygous *Medea* that is unstable if the fitness cost to homozygotes is greater than that to heterozygotes.

To simplify the analysis of the model we limit our analysis to cases where there are 2 independent variables (homozygous and heterozygous fitness, or t_1 and fitness). We plot these variables against each other to create a parameter space diagram in which regions of feasibility

and stability or instability are indicated. In the current case, which incorporates an embryonic fitness cost, our equations simplify to:

$$G'_{MM} = V_{E,Homo}(G^2_{MM} + G_{M+}G_{MM} + \tfrac{1}{4}G^2_{M+})/W$$

$$G'_{M+} = V_{E,Het}(2G_{MM}G_{++} + G_{MM}G_{M+} + \tfrac{1}{2}G^2_{M+} + G_{M+}G_{++})/W$$

$$G'_{++} = (\tfrac{1}{2}G_{M+}G_{++} + G^2_{++})/W$$

where the mean fitness, *W*, equals

$$W = V_{E,Homo}(G^2_{MM} + G_{M+}G_{MM} + \tfrac{1}{4}G^2_{M+}) + V_{E,Het}(2G_{MM}G_{++} + G_{MM}G_{M+} + \tfrac{1}{2}G^2_{M+} + G_{M+}G_{++}) + \tfrac{1}{2}G_{M+}G_{++} + G^2_{++}$$

We find 4 equilibria. Regions of stability and feasibility are indicated in Fig. 4A, which plots homozygous fitness versus heterozygous fitness.

(1) All non-*Medea* individuals:

G_{++} = 1, G_{M+} = G_{MM}= 0;

As noted above for the general case, this equilibrium is always stable (Figure 4A, regions A, B, and C) unless $V_{E,Het}$=1 (Fig. 4A, line a), at which value the linear stability analysis is inconclusive. In other words, if the presence of *Medea* results in a fitness cost to heterozygotes, very low-level introductions of *Medea* will result in loss of the *Medea* allele. Numerical results indicate that when fitness costs are purely recessive ($V_{E,Het} = 1$; $V_{E,Homo} \leq 1$), this equilibrium is unstable, implying that low-frequency introductions of such a *Medea* can result in spread, even if the fitness of homozygotes is close to zero. Throughout the text we consider the more realistic case in which heterozygotes experience a fitness cost ($V_{E,Het}$<1).

(2) All three genotypes present in population:

$$G_{++} = -\frac{V_{E,Het}^2 - V_{E,Het} + V_{E,Homo}}{-V_{E,Homo} + V_{E,Het} - 1};\ G_{MM} = -\frac{1 + V_{E,Het}^2 - 2V_{E,Het}}{-V_{E,Homo} + V_{E,Het} - 1};\ G_{M+} = 1 - G_{++} - G_{MM}$$

Equilibrium 2 is only biologically feasible if $V_{E,Homo} \geq V_{E,Het}(1\text{-}V_{E,Het})$ (Fig. 4A, regions A and B). By linear stability analysis this equilibrium is unstable when $V_{E,Het}$<1 and $V_{E,Homo} > V_{E,Het}(1\text{-}V_{E,Het})$ (Fig. 4A, regions A and B). At $V_{E,Het}$=1 (Fig. 4A, line a), this equilibrium is coincident

with equilibrium 1 (G_{++}=1, G_{M+}=0 and G_{MM}=0) and unstable as determined through numerical simulations.

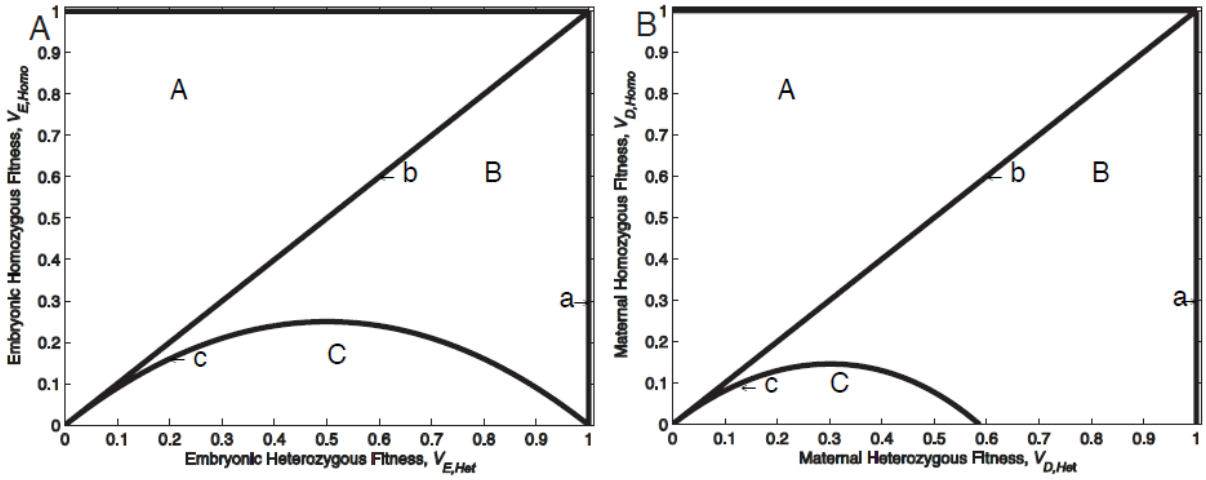

Figure 4.

Diagrams partitioning (V_{Het}, V_{Homo}) fitness parameter space into regions in which linear stability analysis indicates qualitatively similar behaviors are observed. (A) Parameter space diagram of ($V_{E,Het}$, $V_{E,Homo}$) space (this diagram is identical for a *Medea* with parental fitness cost) . Qualitative behavior changes as each curve is crossed, with the occurrence of a bifurcation. Equilibrium 1, which consists of only the non-*Medea* genotype, is stable in all regions except at line a where the analysis is inconclusive. Equilibrium 2, which consists of all genotypes, is unstable in regions A and B and infeasible in C. Equilibrium 3, which consists of heterozygous and homozygous *Medea*, is infeasible in A, stable in B and unstable in C. Equilibrium 4, which consists of only the homozygous *Medea* genotype, is stable in A and unstable in B and C. Line a corresponds to a region in which Equilibrium 1 and 2 are coincident. Line b separates regions A and B. On this line, Equilibrium 3 and 4 are coincident. Transcritical bifurcation occurs as Equilibrium 3 moves through Equilibrium 4 (i.e. the two collide), with the two equilibria exchanging stability. Curve c separates regions B and C. On this curve, Equilibrium 2 and 3 are coincident. Transcritical bifurcation occurs as the two equilibria collide, with the two equilibria exchanging stability. (B) Parameter space diagram of ($V_{D,Het}$, $V_{D,Homo}$) space. Explanations are as in (A).

(3) No non-*Medea* individuals in the population:

$$G_{++} = 0;\ G_{MM} = \frac{V_{E,Homo}}{2V_{E,Het} - V_{E,Homo}};\ G_{M+} = 1 - G_{MM} - G_{++}$$

This equilibrium is biologically feasible if $V_{E,Homo} \leq V_{E,Het}$ (Fig. 4A, regions B and C). By linear stability analysis, this equilibrium is unstable for $V_{E,Homo} < V_{E,Het}(1-V_{E,Het})$ (Fig. 4A, region C), and stable for $V_{E,Homo} > V_{E,Het}(1-V_{E,Het})$ (Fig. 4A, region B). For parameter sets with values in region B, there is an equilibrium which we define as the SIEAF, at which non-*Medea* individuals, but not non-*Medea* alleles, have been eliminated from the population. When $V_{E,Homo} = V_{E,Het}$ (Fig. 4A, line b) this equilibrium is coincident with equilibrium 4. When $V_{E,Homo} = V_{E,Het}(1-V_{E,Het})$ (Fig. 4A, line c), this equilibrium is coincident with equilibrium 2.

(4) All *Medea* homozygous individuals in the population:

$G_{++} = 0$, $G_{M+} = 0$, $G_{MM} = 1$;

By linear analysis, this equilibrium is stable if $V_{E,Homo} > V_{E,Het}$ (Fig. 4A, region A), and unstable if $V_{E,Homo} < V_{E,Het}$ (Fig. 4A, regions B and C). When $V_{E,Homo} = V_{E,Het}$, this equilibrium is coincident with equilibrium 3 (Fig. 4A, line b) and stable. Therefore, when a *Medea* carries a fitness cost, and homozygotes are less fit than heterozygotes, unless a population begins with no wildtype alleles, the population will always contain wildtype alleles.

A *Medea* with parental fitness costs has identical regions of stability to a *Medea* with an embryonic cost. Equilibrium 2 is $G_{++} = 0$, $G_{MM} = \frac{V_{D,Het}}{3V_{D,Het} - 2V_{D,Homo}}$, $G_{M+} = 1 - G_{MM}$.

Equilibrium 3 is $G_{++} = \frac{(V_{P,Het}^2 - V_{P,Het} + V_{D,Homo})V_{P,Homo}}{V_{P,Homo}^2 + V_{P,Homo}V_{V,Het}^2 - 3V_{P,Homo}V_{P,Het} + 2V_{P,Homo} + V_{P,Het}^2 - 2V_{P,Het} + 1}$,

$G_{MM} = \frac{V_{P,Het}^2 - 2V_{P,Het} + 1}{V_{P,Homo}^2 + V_{P,Homo}V_{V,Het}^2 - 3V_{D,Homo}V_{D,Het} + 2V_{D,Homo} + V_{D,Het}^2 - 2V_{D,Het} + 1}$,

$G_{M+} = 1 - G_{MM} - G_{++}$. In the case of a maternal only fitness cost, there are again four equilibria. Equilibrium 1 has only non-*Medea* individuals and is stable if $V_{E,Het} < 1$ (Fig. 4B, regions A,B, and C, but not line a). Equilibrium 2 has all three genotypes in an analytically tractable ratio, but the formula is complicated and therefore not reproduced here (see Supplemental Materials for details). Equilibrium 2 is feasible, except at extremely low fitness values (Fig. 4B, regions A and B, but not C). When it is feasible, it is unstable. Equilibrium 3 has no non-*Medea* individuals,

$$G_{++} = 0,\ G_{MM} = \frac{V_{D,Homo} - V_{D,Het} + \sqrt{V_{D,Homo}^2 - 2V_{D,Het}V_{D,Homo} + 2V_{D,Het}^2}}{V_{D,Het}},\ G_{M+} = 1 - G_{MM}$$

and is only biologically feasible when $V_{E,Homo} < V_{E,Het}$ (Figure 4B, regions B and C). The equilibrium is stable in region B. Equilibrium 4 contains only homozygous *Medea* individuals and is only stable when heterozygotes are more fit than homozygotes (Figure 4B, region A).

The above observations suggest that if an autosomal *Medea* with a fitness cost is present above a critical frequency, it will spread, ultimately reaching the SIEAF, at which point non-*Medea* individuals will no longer be present in the population. This feature of *Medea*, that if spread occurs non-*Medea* individuals are driven from the population, is important because the epidemiology of insect-borne diseases such as malaria indicates that disease prevention through population replacement will require that most insects be refractory to disease transmission (Macdonald 1957; Boete and Koella 2002; Boete and Koella 2003). In this context the quantity of most interest is not the allele frequency of the transgene, but the frequency of transgene-bearing individuals.

The rate at which *Medea* spreads and eliminates non-*Medea* individuals is a function of its introduction frequency and the nature and size of the fitness cost. This point is illustrated for the relationship between a parental fertility/fecundity cost, an embryonic fitness cost, or maternal fecundity cost, on the X axis, the ratio of homozygous *Medea*/*non-Medea* males introduced into a population of non-*Medea* females, indicated on the Y axis, and number of generations required to bring non-*Medea* individuals below a specific frequency (1%) in Fig. 5A, 5B and 5C, respectively. The lower bound on these graphs (the black border) defines for any given fitness cost, the critical homozygous *Medea*/non-*Medea* male introduction ratio (CMIR), below which *Medea* will not spread. This number is of practical interest because the release of *Medea*-bearing males is most appropriate for population replacement in mosquito populations since it is technically feasible to release only males (Knipling et al. 1968; Asman

et al. 1981; Catteruccia et al. 2005; Smith et al. 2007), and females bite and transmit disease while males do not. In addition, the release of homozygous males only is likely to maximize spread since it forces all *Medea*-bearing individuals to mate with non-*Medea*-bearing individuals, maximizing drive in the next, heterozygous, generation. A sex-independent parental fitness cost (Fig. 5A) requires the highest introduction frequency. In the first generation two copies of the fitness cost are born by homozygous *Medea* fathers used to initiate population replacement, leading to a decrease in their reproductive output and the effective introduction frequency. In the second generation each heterozygous parent bears one copy of the cost. In future generations the costs are born in a dose-dependent manner by *Medea*-bearing individuals of both sexes. *Medeas* carrying an embryonic fitness cost (Fig. 5B) require a somewhat lower introduction frequency because there is no cost in the first, parental generation. In subsequent generations the costs are born in a dose-dependent manner by all *Medea* carriers, as with a parental cost. A female-specific fitness cost (Fig. 5C) requires the lowest introduction frequency because no fitness cost is incurred in the first, parental generation, while in subsequent generations the costs are limited to females.

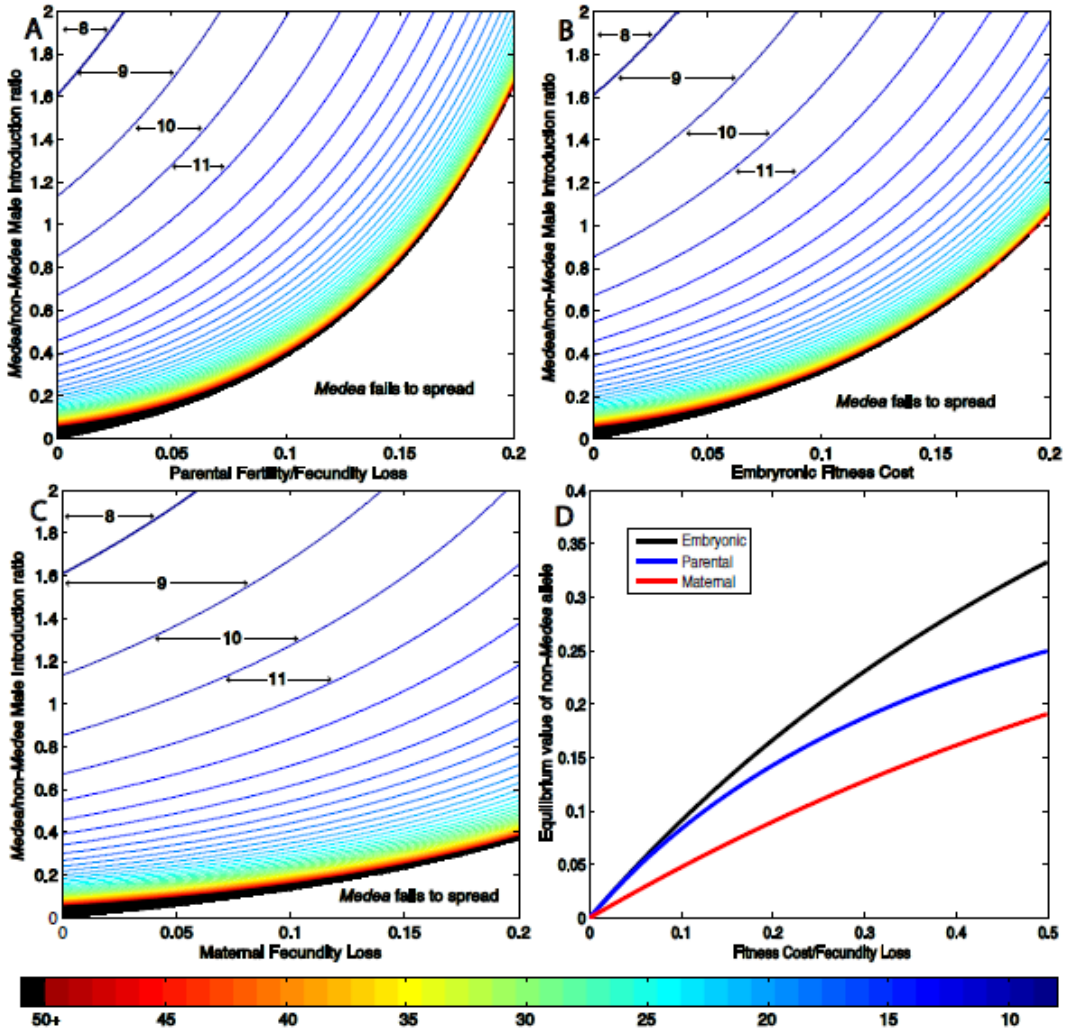

Figure 5.

When an autosomal *Medea* with a fitness cost and t_1=0 spreads, it drives the elimination of non-*Medea* individuals, but not non-*Medea* alleles, from the population. (A) Plot describing the number of generations required for *Medea* to be present in 99% of individuals, for a *Medea* element with a parental fertility/fecundity cost. Homozygous *Medea* male:non-*Medea* male introduction ratios are indicated on the Y axis, and parental fertility/fecundity cost on the X axis. Area between lines indicates regions of parameter space within which a specific number of generations (indicated by numbers and arrows) are required for the frequency of *Medea* individuals to reach 99% or greater. Line color, shown in the heat map at the bottom, provides a measure of how many generations are required. Black lines (50+) indicate that fifty or more generations are required. The border between the black-lined region and the lower unlined region defines the critical male introduction ratio (CMIR). (B) Plot describing the number of generations required for *Medea* to be present in 99% of individuals, for a *Medea* element with an embryonic fitness cost. (C) Plot describing the number of generations required for *Medea* to be present in 99% of individuals, for a *Medea* element with a maternal fecundity cost. (D)

Stable internal equilibrium values of the non-*Medea* allele are plotted as a function of fitness cost/fecundity loss for embryonic, sex-independent parental, or maternal costs.

The fate of the autosomal non-*Medea* allele

As discussed above, when an autosomal *Medea* allele carrying a fitness cost is introduced above a critical value and complete zygotic rescue requires only one copy of the antidote, the frequency of the *Medea* allele increases and approaches an equilibrium, the SIEAF, with the non-*Medea* allele. Equilibrium allele frequency values can be algebraically derived. Since we are focused on the equilibrium conditions when there are no non-*Medea* individuals in the population, G_{++}=0, and the proportion of heterozygotes is directly related to q, the fraction of non-*Medea* alleles in the adult population, G_{M+}=2q, and G_{MM}=1-2q.

For the parental (two parent) fitness cost the stable equilibrium occurs at

$$q^* = \frac{1 - V_{P,Het}}{3 - 2V_{P,Het}} = \frac{c_{P,Het}}{2c_{P,Het} + 1}$$

For the embryonic fitness cost the stable equilibrium occurs at

$$q^* = \frac{1 - V_{E,Het}}{2 - V_{E,Het}} = \frac{c_{E,Het}}{c_{E,Het} + 1}$$

For the maternal fitness cost the stable equilibrium occurs at

$$q^* = \tfrac{1}{2}\left(2 - V_{D,Het} - \sqrt{V_{D,Het}^2 - 2V_{D,Het} + 2}\right)$$

These equilibrium values, for *Medea* elements carrying parental, embryonic, or maternal fitness costs, are plotted as a function of fitness cost in Fig. 5D.

We now relax the assumption that heterozygous *Medea* offspring of homozygous *Medea* mothers always survive. Such a situation could easily arise if progeny of homozygous *Medea*-bearing mothers (which inherit two dosages of the toxin) cannot, or can only imperfectly be rescued from death by a single copy of *Medea* (a single copy of the antidote) in the zygote ($0<t_1<1$) (see also Fig. S3 in (Chen *et al.* 2007)). Smith makes a related point, though details of his model differ from ours with respect to the fate of homozygous *Medea* progeny of homozygous *Medea* mothers (Smith 1998): we assume that homozygous progeny of homozygous *Medea* mothers show 100% survival, as observed in (Chen *et al.* 2007), while he does not. We calculate the equilibrium values assuming that $V_{S,Het}=V_{S,Homo}=V_{D,Het}=V_{D,Homo}=1$, $V_{E,Homo}=V_{E,Het}^2$ and $t_0=1$. Again there are four possible equilibria. We plot the feasibility and stability of equilibria within t_1, embryonic fitness parameter space in Fig. 6A.

(1) No *Medea*-bearing individuals in the population:

$G_{++}=1$, $G_{MM}=0$, $G_{M+}=0$.

This equilibrium is stable (Fig. 6A, regions A, B, and C) except when $V_{E,Het}$=1 (Fig. 6A, line a), where the stability analysis is inconclusive. Simulations indicate that at this latter point the equilibrium is unstable. This equilibrium implies that if there is a fitness cost, regardless of t_1, with very low introduction frequencies *Medea* will be lost.

(2) All three genotypes in the population:

$$G_{++} = \frac{\left(V_{E,Het}^3 t_1^2 - 4V_{E,Het}^2 t_1 - 2V_{E,Het} t_1 + 8V_{E,Het} + 2t_1 - 4\right)V_{E,Het}}{4V_{E,Het}^2 - 4V_{E,Het} + V_{E,Het}^4 t_1^2 + 2V_{E,Het}^2 t_1 - 4V_{E,Het}^3 t_1 - 2V_{E,Het} t_1 + 4}$$

$$G_{MM} = \frac{4\left(V_{E,Het}^2 - 2V_{E,Het} + 1\right)}{4V_{E,Het}^2 - 4V_{E,Het} + V_{E,Het}^4 t_1^2 + 2V_{E,Het}^2 t_1 - 4V_{E,Het}^3 t_1 - 2V_{E,Het} t_1 + 4}$$

$$G_{M+} = 1 - G_{MM} - G_{++}$$

This equilibrium is not feasible for low fitness values (Figure 6A, region C), and is biologically feasible, but unstable when

$$t_1 \geq \frac{2V_{E,Het}^2 + V_{E,Het} + 1 - \sqrt{-4V_{E,Het}^4 + 8V_{E,Het}^3 - 3V_{E,Het}^2 - 2V_{E,Het} + 1}}{V_{E,Het}^3}$$

(Fig. 6A, regions A and B).

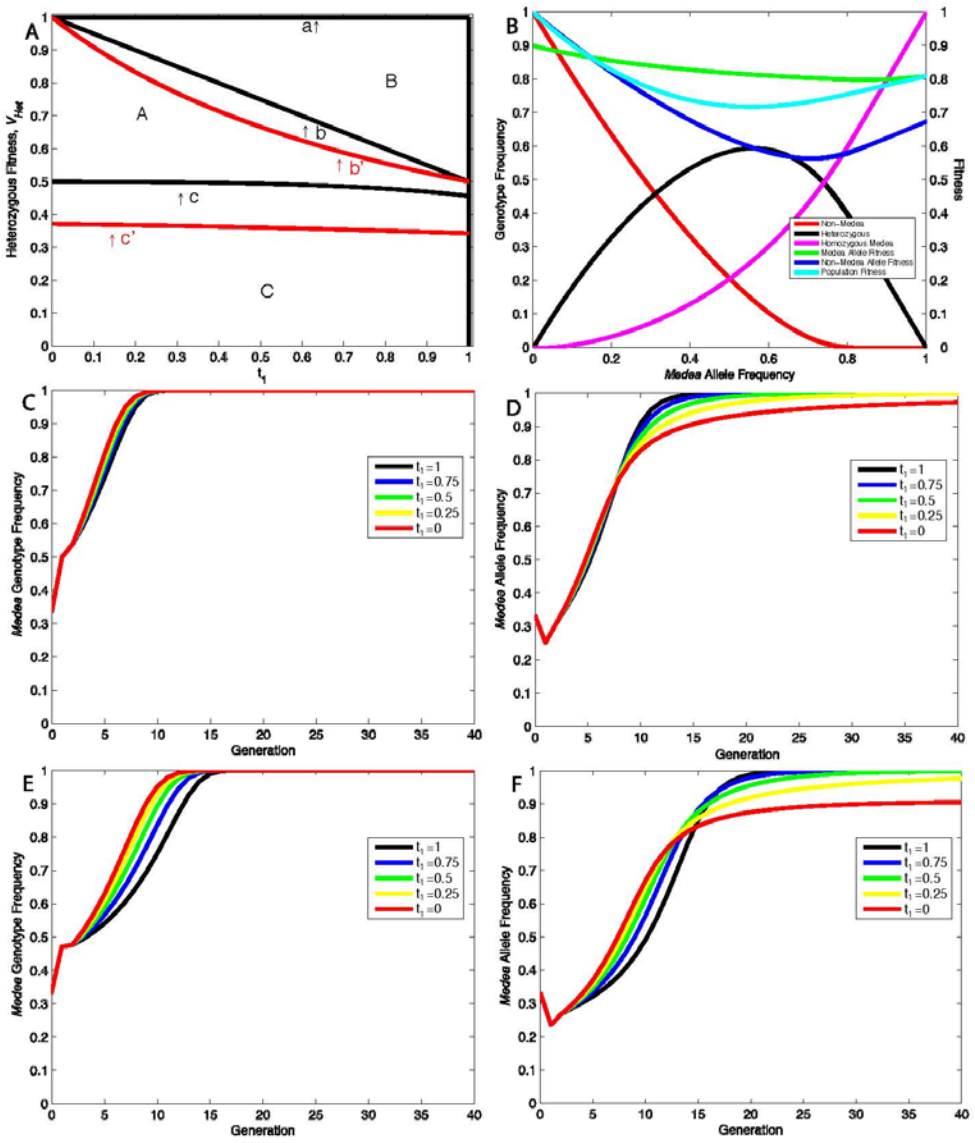

Figure 6. When *Medea* is located on an autosome, and heterozygous *Medea* offspring of homozygous *Medea* mothers do not always survive, the *Medea* allele achieves a lower equilibrium frequency or is eliminated from the population. (A) Diagram partitioning (t_1, V_{Het}) parameter space into regions in which linear stability analysis indicates qualitatively similar behaviors are observed. Qualitative behavior changes as we cross each of these curves, with the occurrence of a bifurcation. Black lines partition parameter space for *Medea* elements with an embryonic or parental fitness costs, while red lines partition parameter space for *Medea* elements with a maternal fitness cost. Equilibrium 1, which consists of only the non-*Medea* genotype, is stable in all regions. Equilibrium 2, which consists of all genotypes, is unstable in regions A and B and infeasible in C. Equilibrium 3, which consists of heterozygous and homozygous *Medea* genotypes, is infeasible in C, stable in A and unstable in B. Equilibrium 4, which consists of only the homozygous *Medea* genotype, is stable in B and unstable in A

and C. Line a corresponds to a *Medea* with no fitness cost. At line a, the stability of equilibrium 1, the all non-*Medea* equilibrium, is inconclusive. Line b separates regions A and B. On this line, Equilibrium 3 and 4 are coincident. Transcritical bifurcation occurs as Equilibrium 3 moves through Equilibrium 4 (i.e. the two collide), with the two equilibria exchanging stability. Curve c separates regions A and C. On this curve, the Equilibrium 2 and 3 are coincident. (B) Plot of allele fitness, and genotype frequency, as a function of *Medea* allele frequency, for a *Medea* that carries a 10% embryonic fitness cost and has t_1=0.5. Compare with the *Medea* shown in Fig. 3D, in which t_1=0. (C) *Medea*-bearing genotype frequency is plotted as a function of the number of generations for *Medea* elements with zero fitness cost and different levels of heterozygous offspring lethality (t_1), introduced into a population of non-*Medea* females using a fixed 1:1 ratio of *Medea*:non-*Medea* males. (D) Plot of *Medea* allele frequency as a function of the number of generations for the zero fitness cost *Medea* elements in (C). (E) Plot as in (C) for *Medea* elements with a 10% embryonic fitness cost. (F) Plot of *Medea* allele frequency as a function of the number of generations for *Medea* elements with a 10% embryonic fitness cost. Note that the *Medea* allele frequency increases to higher equilibrium values as t_1 increases.

(3) No non-*Medea* individuals in the population:

$$G_{++} = 0;\ G_{MM} = \frac{1 - V_{E,Het} - \sqrt{1 - 2V_{E,Het}t_1}}{V_{E,Het} + 2t_1 - 2};\ G_{M+} = 1 - G_{MM} - G_{++}$$

The equilibrium population consists only of *Medea*-bearing genotypes and is only feasible when $t_1 \leq 2\ (1\text{-}V_{E,Het})$, and is stable in Fig. 6A region A but unstable in Fig. 6A region C. The algebraic expression describing curve c is the equality shown in equilibrium 2. This equilibrium indicates that there are situations in which the death of heterozygous progeny of homozygous mothers causes a decrease in the equilibrium non-*Medea* allele frequency as compared to the case in which such heterozygous progeny never die. There is an additional curve where the linear stability analysis is inconclusive,

$t_1 = \frac{2(1+V_{E,Het})}{4V_{E,Het}^2 + 4V_{E,Het} + 1}, 0.481 \le V_{E,Het} \le \frac{\sqrt{2}}{2}$. The stability does not change when this curve is crossed, and therefore it is not plotted.

(4) All *Medea* homozygous individuals in the population:

G_{++}=0, G_{M+}=0, and G_{MM}=1.

The population is entirely homozygous *Medea*. This equilibrium is stable when t_1>2(1-$V_{E,Het}$) (Figure 6A, region B) and unstable when t_1<2(1-$V_{E,Het}$) (Figure 6A, regions A and C). At the equality, the analysis is inconclusive but the equilibrium is coincident with equilibrium 3. This equilibrium indicates that there are biologically relevant situations (high *Medea* fitness and high t_1) in which the non-*Medea* allele is driven from the population.

Figs. 6B-F illustrate the behavior of Medea elements with t_1>0. In Figs. 6C-F homozygous *Medea* and non-*Medea* males are introduced at a 1:1 ratio into a population of non-*Medea* females. Figs. 6C, D show the behavior of *Medea* elements with no fitness cost, in which heterozygous progeny of homozygous mothers do not die (t_1=0), sometimes die, (t_1=0.25; t_1=0.5; t_1=0.75), or always die (t_1=1.0). Figs. 6E, F show the behavior of *Medea* elements that carry a 10% embryonic fitness cost, with t_1 values as above. In both cases non-*Medea* individuals are eliminated from the population, though values of t_1>0 result in modest delays. The fate of the non-*Medea* alleles is more complicated. As illustrated in Fig. 3, and noted by

previously by others (HASTINGS 1994; WADE and BEEMAN 1994), a *Medea* with no fitness cost, and with t_1=0, will ultimately spread to fixation. Values of t_1>0 speed this process, with greater t_1 values leading to more rapid loss of the non-*Medea* allele (Fig. 6D) (with t_1 values of 0, 0.25, and 1 resulting in non-*Medea* allele frequencies of 0.0279, 0.00161, and 1.03×10^{-10}, respectively after 40 generations). When a *Medea* carries a fitness cost, values of t_1>0 result in a decrease in the non-*Medea* equilibrium allele frequency, which can go to 0 for high values of t_1, resulting in loss of the non-*Medea* allele (Fig. 6F). For example, for the *Medea* shown in Fig. 6F, which has a 10% embryonic fitness cost and t_1=0, the non-*Medea* equilibrium allele frequency is 9.09%. In contrast, t_1 values of 0.25, 0.5, 0.75, or 1 result in non-*Medea* allele frequencies of 0.0220, 8.74×10^{-4}, 9.24×10^{-6}, or 3.65×10^{-8}, respectively, after only 40 generations. The curves shown in Fig. 6A demarcate regions of fitness, t_1 space within which non-*Medea* alleles are lost (region B) or retained (region A) in the population, for different kinds of fitness costs.

The mechanism by which values of t_1>0 can lead to loss of the non-*Medea* allele from the population can be understood by considering the changing fitness of the non-*Medea* allele for t_1=0 and t_1>0, as the frequency of *Medea* increases. Fig. 6B provides an example, for a *Medea* with a 10% embryonic fitness cost and t_1= 0.5. When such a *Medea* allele is present at frequencies just above the UIAEF, most *Medea* alleles are in heterozygotes, and non-*Medea* alleles experience a *Medea*-dependent fitness cost similar to that for a *Medea* with a 10% embryonic fitness cost and t_1=0 (Fig. 6B, compare with Fig. 3D). For the element with t_1=0, the fitness of the non-*Medea* allele decreases initially as *Medea* spreads, but then recovers significantly as the frequency of males homozygous for *Medea* (which protect non-*Medea*

alleles in heterozygous female parents from death in non-*Medea* progeny) increases (Fig. 3D). Homozygous *Medea* females always allow survival of non-*Medea* alleles in progeny. In contrast, for a *Medea* with a 10% fitness cost and t_1=0.5, the fitness of the non-*Medea* allele continues to remain low, never becoming greater than that of the *Medea* allele (Fig. 6B). Homozygous male parents continue to protect non-*Medea* alleles in heterozygous mothers from death in progeny, as above. However, for *Medea* elements with t_1=0.5, non-*Medea* alleles from heterozygous males face a new, 50% risk of death in heterozygous progeny when female parents are homozygous for *Medea*. As the frequency of homozygotes increases, so does the frequency of killing through this mechanism: 0% killing when non-*Medea* alleles are in females; 50% killing when they are in males, resulting in up to 25% killing as the *Medea* allele frequency approaches fixation. *Medea* alleles in heterozygous progeny are also killed through this mechanism. But as *Medea* spreads, the frequency of heterozygotes becomes rare with respect to the frequency of *Medea*-bearing individuals (which are mostly homozygotes), while heterozygotes make up the majority of individuals carrying the non-*Medea* allele. Therefore, the fitness costs from this form of death are born primarily by the non-*Medea* allele. In summary, when t_1>0, non-*Medea* alleles experience a novel *Medea*-dependent fitness cost that increases as *Medea* spreads and homozygous females become common. If t_1 is large this can have a dramatic effect on non-*Medea* allele fitness. In particular, high t_1 values can result in the fitness of the *Medea* allele (when present at frequencies above the UIAEF) being always greater than that of the non-*Medea* allele (Fig. 6B). This results in the non-*Medea* allele being eliminated from the population, an outcome also reflected in population fitness, which - following an initial decrease resulting from high levels of *Medea*-dependent killing - increases continuously as *Medea* spreads to fixation.

When a *Medea* located on the X chromosome in a X/Y male heterogametic species spreads, it always drives the non-*Medea* chromosome to extinction.

The spread of a *Medea* on the X, in a X/Y male heterogametic species, also results in non-*Medea* alleles experiencing a novel cost that leads to their loss from the population. S_x represents the fraction of the total population that is male of genotype x and D_x the fraction of the total population that is female of genotype x. For a *Medea* on the X, the standard iterative *Medea* equations are modified as follows:

$$S'_{MY} = \frac{V_{E,Het}}{W}\left(\tfrac{1}{2}S_{MY}D_{MM} + \tfrac{1}{4}S_{MY}D_{M+} + \tfrac{1}{2}S_{+Y}D_{MM} + \tfrac{1}{4}S_{+Y}D_{M+}\right)$$

$$S'_{+Y} = \frac{1}{W}\left(\tfrac{1}{2}S_{MY}D_{++} + \tfrac{1}{2}S_{+Y}D_{++}\right)$$

$$D'_{MM} = \frac{V_{E,Het}^2}{W}\left(\tfrac{1}{2}S_{MY}D_{MM} + \tfrac{1}{4}S_{MY}D_{M+}\right)$$

$$D'_{M+} = \frac{V_{E,Het}}{W}\left(\tfrac{1}{4}S_{MY}D_{M+} + \tfrac{1}{2}S_{MY}D_{++} + \tfrac{1}{4}S_{+Y}D_{M+} + \tfrac{1}{2}S_{+Y}D_{MM}\right)$$

$$D'_{++} = \frac{1}{W}\left(\tfrac{1}{2} S_{+Y} D_{++}\right)$$

where the mean fitness, *W*, equals

$$W = V_{E,Het}^2\left(\tfrac{1}{2} S_{MY} D_{MM} + \tfrac{1}{4} S_{MY} D_{M+}\right) + V_{E,Het}\left(\tfrac{1}{2} S_{MY} D_{MM} + \tfrac{1}{2} S_{MY} D_{M+} + \tfrac{1}{2} S_{MY} D_{++} + S_{+Y} D_{MM} + \tfrac{1}{2} S_{+Y} D_{M+}\right) + \tfrac{1}{2} S_{MY} D_{++} + S_{+Y} D_{++}$$

Recalling $S_{+Y}+S_{MY}+D_{++}+D_{M+}+D_{MM}=1$, and setting $S'_{+Y}=S_{+Y}$, $D'_{MM}=D_{MM}$, $D'_{M+}=D_{M+}$, and $D'_{++}=D_{++}$, we find 4 equilibria.

(1) $D_{MM}=0$; $D_{M+}=0$; $D_{++}=1/2$; $S_{MY}=0$; $S_{+Y}=1/2$

Equilibrium (1) is stable for all values of $V_{E,Het}$ except $V_{E,Het}=1$. At $V_{E,Het}=1$ the linear stability analysis is inconclusive, but simulations indicate that the equilibrium is unstable. In other words, as with the case of an autosomal element carrying a fitness cost, if the presence of *Medea* results in a fitness cost to carriers, very low-level introductions of *Medea* will result in loss of the *Medea* allele.

Equilibrium (2) is a mixture of all genotypes. The equilibrium expressions are cumbersome and can be found in Supplementary Materials. There is a cumbersome analytical solution for when the equilibrium is biologically feasible. Numerically the equilibrium is feasible, but unstable, when $V_{E,Het}$ is less than about 0.54. In other words, when the fitness of *Medea* is low there is an equilibrium at which non-*Medea* individuals exist with *Medea*-bearing individuals, but it is unstable. This equilibrium is unlikely to be biologically relevant since we are primarily interested in elements with fitness greater than 0.54.

(3) $D_{MM} = \frac{V_{E,Het}}{2 - V_{E,Het}};\ D_{M+} = \frac{1 - 2V_{E,Het}}{2 - V_{E,Het}};\ D_{++} = 0;\ S_{MY} = \frac{1}{2 - V_{E,Het}}; S_{+Y} = 0$

Equilibrium (3) is stable for $V_{E,Het}<0.5$, unstable for $V_{E,Het}>0.5$, and the analysis is inconclusive at $V_{E,Het}=0.5$. In other words, when the fitness of *Medea* is below 0.5 (probably of little biological relevance), there are conditions under which *Medea* can still spread, such that at equilibrium all individuals carry *Medea*, and some females carry non-*Medea* alleles. Wildtype alleles cannot exist in males because, as discussed further below, X/Y males cannot be rescued from death when they inherit a non-*Medea* X chromosome from a *Medea*-bearing mother.

(4) $D_{MM} = \frac{V_{E,Het}}{V_{E,Het} + 1}\ ;\ D_{M+} = 0; D_{++} = 0;\ D_{M+} = 0;\ S_{MY} = \frac{1}{V_{E,Het} + 1};\ S_{+Y} = 0$

This equilibrium is stable for $V_{E,Het}>1/2$, unstable for $V_{E,Het}<1/2$, and the analysis is inconclusive at $V_{E,Het}=1/2$. This equilibrium is particularly interesting because it suggests that if *Medea* becomes established, there is a broad range of physiologically relevant conditions ($V_{E,Het}>1/2$) under which the non-*Medea* allele is eliminated from the population.

Loss of the non-*Medea* allele occurs under the conditions of equilibrium (4) because this allele experiences a unique cost. In each generation X-linked non-*Medea* alleles present in heterozygous *Medea* female parents have a 50% probability of ending up in a male progeny, which, as noted above, are doomed to death because they cannot be rescued by a paternally derived *Medea* allele. Thus, the non-*Medea* X allele experiences a minimum 50% probability of death in the subsequent generation each time it finds itself in a heterozygous *Medea* female, and the probability of finding itself in a heterozygous *Medea* female (as opposed to a non-*Medea* female) increases as the frequency of *Medea* increases. This stands in contrast to the case of an autosomal *Medea* with $t_1=0$, in which the spread of *Medea* serves to protect the non-*Medea* allele from death (see discussion of Fig. 3D). As a result of this X-linkage effect, the fitness costs associated with the non-*Medea* allele remain higher than those associated with the *Medea* allele for all *Medea* allele frequencies above the UIEAF. This forces the non-*Medea* allele out of the population, and is also reflected in changes in population fitness as *Medea* spreads. Population fitness initially decreases as a result of *Medea*-dependent killing of non-*Medea* alleles; it then increases continuously as *Medea* spreads and the killing of non-*Medea* male progeny declines. These features of an X-linked *Medea* are illustrated in Fig. 7A, which

plots allele fitness versus *Medea* allele frequency for a X-linked *Medea* with a 10% embryonic fitness cost. The dynamics of *Medea* spread for representative elements located on the X carrying an embryonic fitness cost are illustrated in Figs. 7B-D. Non-*Medea* individuals are rapidly eliminated from the population, though the times required are somewhat longer, and CMIRs somewhat higher than those required for an autosomal element with a similar fitness cost, since introduced males carry only one copy of *Medea*.

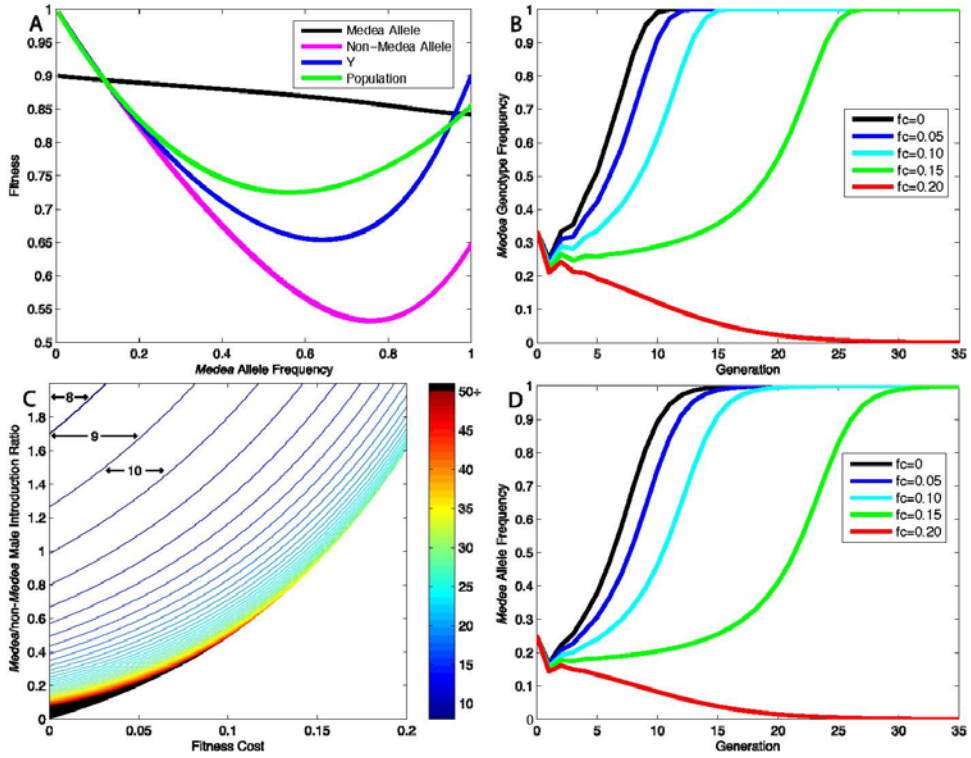

Figure 7.

When *Medea* is located on the X chromosome in a male heterogametic species, and *Medea* spreads, the non-*Medea* allele is eliminated from the population. (A) Plot of allele and population fitness as a function of *Medea* allele frequency, for a *Medea* that carries a 10% embryonic fitness cost, located on the X chromosome. Lines and labels and other conditions are as in Fig. 3D. (B) *Medea* genotype frequency is plotted as a function of the number of generations for *Medea* elements on the X with different levels of an embryonic fitness cost, introduced into a population of non-*Medea* females using a fixed 1:1 ratio of *Medea*:non-

Medea males. (C) Plot describing the number of generations required for *Medea* to be present in 99% of individuals, for a *Medea* element on the X with an embryonic fitness cost. Compare with Fig. 4B. (D) *Medea* allele frequency is plotted as a function of the number of generations for the elements shown in (B).

Population replacement with multiple *Medea* elements

Insects used for population replacement may carry multiple, unlinked, versions of a single *Medea* element as a way of increasing the introduction frequency of the element given a particular organism introduction frequency, assuming that a single copy of the element in progeny guarantees their survival regardless of how many copies of *Medea* (and thus the toxin) the mother carries. Because the number of possible genotypes scales as 3^n, where n is the number of copies of the construct, it is not convenient to examine high copy number deterministically. However, up to 3 copies is practical deterministically, and probably realistic in practice. Details of this model are provided in Supplementary Materials. The invasion dynamics for three unlinked autosomal elements, with each element bearing the same embryonic fitness cost, are shown in Fig. 8A (compare with Fig. 5B). For elements with low fitness costs a major effect of introducing multiple elements is to increase the rate of *Medea* spread, while at high fitness costs a primary effect is to lower the critical introduction frequency. Both of these effects occur because with multiple unlinked elements more *Medea* heterozygotes are generated during early generations, leading to increased killing of non-*Medea* alleles.

In reality it is likely that individual elements will have distinct fitness costs as a result of insertion site-dependent effects. Transgene insertions may have effects on neighboring genes,

and neighboring genes or local chromatin structure may have effects on the expression of toxins, antidotes, or genes conferring disease refractoriness. We consider the simple case of a *Medea* element located at two different positions in the genome, in which the two elements have different fitness costs. For details see Supplementary Materials. As illustrated in Fig. 8B, while both *Medea* elements initially increase in frequency, ultimately the *Medea* with the lowest fitness cost spreads through the population, while the other element is eventually lost. Loss of the element(s) with the higher fitness cost occurs because once population replacement has gone to completion all viable genotypes (individuals) must inherit *Medea*. In this situation *Medea*'s selfish behavior is not relevant, since all *Medea* elements in an individual are proposed to function equally well in terms of *Medea*-dependent killing (t_0=1) and rescue (t_1=0). Instead, those genotypes carrying the *Medea* element with the lowest fitness cost will have a selective advantage over those carrying *Medea* elements with higher fitness costs, resulting in a generation-by-generation increase in the frequency of the former. An important implication of this is that while the introduction of multiple copies of a single *Medea* element can facilitate population replacement, one *Medea* is ultimately likely to predominate. Since a low fitness cost associated with a specific *Medea* insertion could reflect decreased expression of genes conferring disease refractoriness, it will be important to be confident that the presence of each *Medea* in isolation is sufficient to confer disease refractoriness.

Population replacement may also utilize individuals that carry several different *Medea* elements that fail to rescue each other (because they utilize different toxins and antidotes). In one possible arrangement, two such *Medea* elements may be inserted next to each other (two *Medea* - one insertion), creating a composite element. Such an arrangement could be used as a

way of generating redundancy that prevents the appearance of antidote-only alleles. Antidote-only alleles are resistant to *Medea*-dependent killing, and can lead to the reappearance of non-*Medea* individuals since a female carrying such an allele and a non-*Medea* allele can give rise to non-*Medea* progeny (Smith 1998). Antidote-only alleles can arise through mutational inactivation of the maternal toxin. They can also be created if unlinked copies of the antidote gene, perhaps generated through retrotransposition, come under the control of an early zygotic promoter. However, with the two *Medea* - one insertion arrangement, both toxins, or both antidotes would need to undergo such alterations in order for antidote-only alleles to appear. To the extent that such alterations result from independent mutational events, the probability of two such events occurring in a single genome becomes quite low. The dynamics of a two *Medea* - one insertion composite element are the same as those for one *Medea* at a single locus. All individuals gain the benefits of *Medea*, but each *Medea*-bearing individual also incurs all fitness costs.

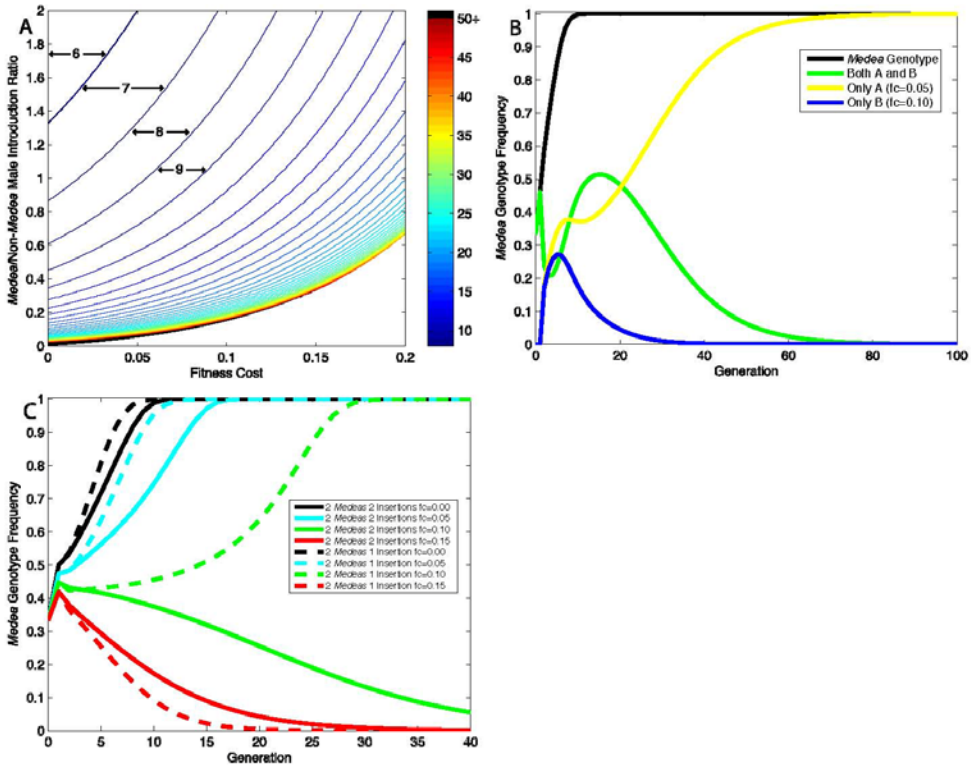

Figure 8. Population replacement with multiple *Medea* elements can serve to enhance or retard the spread of *Medea*, depending on how the elements are organized. (A) Plot describing the number of generations required for *Medea* to be present in 99% of individuals, when three, unlinked *Medea* elements carrying identical embryonic fitness costs are introduced into the population. Compare with Fig. 4B. (B) Plot of the population frequency of individuals with *Medea* as a function of generations, when two identical *Medea* elements with different levels of an embryonic fitness cost are introduced into a population. (C) Plot of the population frequency of individuals with *Medea* as a function of generations. Examples in which two different *Medea* elements are located at a common position, creating a composite element (2 *Medea*s 1 Insertion), are indicated by dotted lines. Examples in which two different *Medea* elements are unlinked (2 *Medea*s 2 Insertion), are indicated by solid lines. When both *Medea* elements are located at the same position spread occurs more rapidly than with two unlinked elements, and under fitness cost conditions (c_E=0.10) in which two unlinked *Medea* elements fail to spread.

In a second possible arrangement, known as two *Medea* - two insertions, *Medea* elements that fail to rescue each other are unlinked. Both *Medea* elements, if introduced above a critical value, will spread. Redundancy is still preserved since inactivation of both toxins, creation of

two new antidotes, or inactivation of both copies of the disease resistance gene would be necessary for the appearance of individuals potentially able to transmit disease. We compared the dynamics of *Medea* spread for the two-*Medea* one insertion and two-*Medea* two insertion strategies. We consider the case in which fitness costs are embryonic. Interestingly, the two-*Medea*-two insertion strategy consistently results in *Medea* entering the population more slowly and requiring a higher introduction frequency, than with the two-*Medea* one insertion strategy. This occurs even though the two-*Medea* one insertion strategy requires that each *Medea*-bearing individual bear the full costs associated with two or four copies of the disease resistance transgene, while the two-*Medea* two insertion strategy does not (Fig. 8C). The reason the two-*Medea* two insertion strategy performs poorly reflects the fact that females heterozygous for both elements experience a unique cost - a significant probability that offspring will inherit one, but not the other element, resulting in progeny death. In contrast, in a two *Medea* - one insertion strategy, inheritance of *Medea* guarantees progeny survival. In short, a requirement that some individuals inherit several different unlinked *Medea*s in order to survive results in an increased frequency of progeny genotypes in which rescue fails, while decreasing the number in which rescue succeeds, thereby slowing *Medea* spread.

Discussion

Here we use a deterministic model to show that *Medea* selfish genetic elements can drive rapid population replacement under a wide range of conditions, provided that they are introduced above a critical introduction frequency (or critical male introduction ratio), determined by the fitness costs associated with *Medea*. Stochastic effects (drift, founder effects) will soften this transition such that *Medea* will sometimes spread when introduced below the CMIR, and sometimes fail to spread when introduced above it, as recently modeled for the case of the

Wolbachia drive system (Jansen et al. 2008). A detailed analysis of *Medea* behavior in finite populations remains to be carried out. A critical feature of *Medea*'s potential as a drive mechanism, highlighted throughout this work, is that under all conditions in which spread occurs, even when *Medea* carries a fitness cost and non-*Medea* alleles remain in the population, non-*Medea* individuals are permanently eliminated from the population. In some cases, when autosomal *Medea* elements have a $t_1>0$, or when *Medea* is located on the X, the non-*Medea* allele can also be eliminated from the population. The rate of *Medea* spread is a function of introduction ratio, fitness costs, and number of elements. Low fitness costs allow rapid spread at relatively low *Medea*/non-*Medea* male introduction ratios, while high fitness costs require higher introduction ratios in order for spread to occur quickly, or at all. The use of introduction strains that carry multiple copies of a particular element can further increase the rate of spread and allow the use of lower introduction ratios, though one element is likely to ultimately predominate within the population. In summary, the key to rapid population replacement with *Medea* is to have a high enough introduction ratio so that *Medea*-dependent killing of non-*Medea* alleles on a generation-by-generation basis, particularly for the early generations, is substantially greater than the coincident loss of fitness-compromised *Medea* alleles through natural selection.

Practical population replacement requires that transgenic individuals be refractory to disease transmission for many generations. *Medea* elements that are autosomally linked with $t_1>0$, or that are X-linked in a male heterogametic species (such as the malaria vector, *Anopheles gambiae*) could be useful in this regard since having no non-*Medea* alleles in the population serves to maximize the number of genes for disease refractoriness in individual females in the

population. These genes, because they are proposed not to confer a fitness benefit to carriers, will eventually undergo mutational decay. If such mutations result in a fitness increase to carriers (a loss of a fitness cost associated with their anti-disease function), and the non-*Medea* allele has a significant equilibrium frequency in the population, then insects permissive for disease transmission will appear. This appearance is delayed if each individual necessarily carries two *Medea* elements.

The ability to eliminate a specific allele from the population also provides a basis from which to carry out modifications of a transgenic population. For example, the toxin component of the drive mechanism also does not provide a fitness benefit to the organisms in whose genomes it resides. Therefore, it too will eventually undergo mutational decay to inactivity, resulting in the appearance of antidote-only alleles. Pre-existing diversity and mutation within the pathogen population may also contribute to the emergence of pathogen populations resistant to first generation effectors. It is also possible, though probably unlikely, that the presence of specific effectors will facilitate the emergence of new pathogens and forms of disease. For all of these reasons, it is essential that strategies be available to remove first-generation elements from the population, replacing them with second-generation elements, if desired (Braig and Yan 2001; Christophides 2005; James 2005; Hay et al. 2010). As discussed in Chen et al. (2007), second generation *Medea* elements can be generated that will spread at the expense of first generation elements, when both elements are located at the same chromosomal position. If second generation autosomal elements having $t_1>0$ are used, or the elements are X-linked, first generation elements can be eliminated from the population during this process. The use of such second generation elements carries a price in that somewhat higher introduction frequencies

are needed than with an autosomal *Medea* element having t_1=0 (Chen et al. 2007, Fig. S3; and compare Fig. 5B with Fig. 7C). But given the importance that control over the fate of released transgenes is likely to have for the acceptance of population replacement as a viable strategy for disease prevention, this may be a small price to pay. This strategy does not restore the population to its pre-transgenic state, but it does provide a method for removing specific transgenes from the population.

What are the contexts in which area-wide population replacement with *Medea* can realistically be carried out? Our results suggest that in order for *Medea* to drive rapid population replacement within 10-20 generations (roughly 1-2 years), *Medea*/non-*Medea* male introduction ratios of between 1:10 and 1:1 are needed, depending on the nature and size of fitness costs, and the number of elements carried in the introduction strain. These numbers represent optimistic estimates because they assume that *Medea* males are competitive with wild males. However, *Medea* male mating competitiveness may be decreased (thereby lowering the effective *Medea*/non-*Medea* male introduction ratio) through several mechanisms. For example, factory rearing can select for distinct mating and other life history traits that are maladaptive in the wild. The factory strain may also be partially reproductively isolated from the wild population of interest. It should be possible to remove some of these genetic barriers to male mating competitiveness and offspring fitness by introgressing *Medea* into the wild genetic background prior to, or during factory rearing. However, non-genetic fitness costs associated with factory rearing (food, stress associated with handling and distribution to sites of release) may still result in loss of male mating competitiveness (reviewed in Dyck et al. 2005). The importance of each of these variables will need to be evaluated, ideally in field

tests, before accurate estimates of *Medea*'s ability to penetrate a specific population can be obtained.

Wild populations of *Aedes aegypti* and some *Anopheles* species have been estimated to range from 10,000-20,000 adults per village (Scott et al. 2000; Taylor et al. 2001). These sizes are small compared with those associated with classical sterile male release in other insects; 68,000 per week in the case of the screw worm fly, and ~10^9 in the case of ongoing Mediterranean fruit fly suppression programs (Dyck et al. 2005). With respect to mosquitoes, weekly factory production of 1,000,000 *Aedes aegypti* could be achieved routinely in the 1960s. Large numbers of *Anopheles* males have also been produced in factory environments using mid-twentieth century technologies (Knipling et al. 1968). In some contexts it may also be possible to take advantage of naturally-occurring changes in mosquito population size to provide an environment in which *Medea* can more easily gain a foothold within a population. For example, while wet season populations of *Anopheles* adults per village in Mali can reach ~15,000, in the dry season these populations consist of only 1,000-3,000 adults (Taylor et al. 2001). These encouraging points notwithstanding, it is important to emphasize that area-wide population replacement remains a daunting task. Disease-endemic regions can be very large (thousands of square miles), and consist of many villages, requiring that the number of *Medea* males to be released be scaled accordingly. Modest migration rates can result in rapid spread of *Medea* in space (J. M. Marshall, C. M. Ward and J. T. Su, unpublished observations). However, if inter-village distances and other environmental obstacles are such that mosquito migration between villages is minimal, *Medea* will need to be introduced at many sites in order to create area-wide protection within a reasonable timeframe. These facts, coupled with the

issues associated with factory rearing and distribution noted above, argue that a detailed analysis of the local mosquito population, the environment, and the amount of resources available to raise and distribute *Medea*-bearing insects will need to be carried out before population replacement is attempted in any given area. Finally, we note that the models examined here make a number of assumptions: infinite population size, non-overlapping generations, no age structure within the population, random mating, and no migration. This kind of model is often used to gain basic insights into population genetic processes. However, it provides only a qualitative snapshot of the conditions under which *Medea* can succeed in driving population replacement. It will be important to carry out more detailed modeling that takes account of the biology of specific pest species, as well as other variables that can influence rate of spread and functional lifetime in the wild.

Acknowledgements

Y.H., A.L.L. and F.G. were supported by NIH grant R01-AI54954-0IA2 and a grant to the Regents of the University of California from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health initiative. B.A.H, C.M.W. and J.T.S were supported by grants to B.A.H from NIH (R01 GM072879; R01 GM070956; DP1 OD003878), and the Weston Havens Foundation. Support was also provided by a grant to the Regents of the University of California from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health initiative. We thank M. Legros for making helpful suggestions on model analysis.

Literature Cited

Asman, S. M., P. T. McDonald, and T. Prout. 1981. Field studies of genetic control systems for mosquitoes. Annu Rev Entomol 26:289-318.

Beeman, R. W., and K. S. Friesen. 1999. Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, tribolium castaneum. Heredity 82 (Pt 5):529-534.

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

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* D. K. a. B. Letourneau, B.E., ed. Genetically Engineered Organisms: Assessing Environmental and Human Health Effects. CRC Press.

Burt, A., and R. Trivers. 2006. Genes in conflict. Belknap Press, Cambridge, MA.

Catteruccia, F., J. P. Benton, and A. Crisanti. 2005. An *Anopheles* transgenic sexing strain for vector control. Nat Biotechnol 23:1414-1417.

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

Christophides, G. K. 2005. Transgenic mosquitoes and malaria transmission. Cell Microbiol 7:325-333.

Clements, A. N. 1999. The Biology of Mosquitoes: Sensory Reception and Behavior. CABI, Oxfordshire, UK.

de Lara Capurro, M., J. Coleman, B. T. Beerntsen, K. M. Myles, K. E. Olson, E. Rocha, A. U. Krettli, and A. A. James. 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.

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

Franz, A. W., I. Sanchez-Vargas, Z. N. Adelman, C. D. Blair, B. J. Beaty, A. A. James, and K. E. Olson. 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., and P. Schliekelman. 2004. Population genetics of autocidal control and strain replacement. Annu Rev Entomol 49:193-217.

Greenwood, B. M., D. A. Fidock, D. E. Kyle, S. H. Kappe, P. L. Alonso, F. H. Collins, and P. E. Duffy. 2008. Malaria: progress, perils, and prospects for eradication. J Clin Invest 118:1266-1276.

Hastings, I. M. 1994. Selfish DNA as a method of pest control. Philos Trans R Soc Lond B Biol Sci 344:313-324.

Hay, B.A., C-H. Chen, C.M. Ward, H. Huang, Su, J.T., and M. Guo. Engineering the genomes of wild insect populations: Challenges, and opportunities provided by synthetic *Medea* selfish genetic elements. J. Insect Physiol. (in press)

Hurst, L. D. 1993. scat+ is a selfish gene analogous to *Medea* of *Tribolium castaneum*. Cell 75:407-408.

Hyde, J. E. 2005. Drug-resistant malaria. Trends Parasitol 21:494-498.

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.

Jansen, V. A., M. Turelli, and H. C. Godfray. 2008. Stochastic spread of Wolbachia. Proc Biol Sci 275:2769-2776.

Kelly, J. K. 1992. Kin selection in density regulated populations. J Theor Biol 157:447-461.

Kitron, U., and A. Spielman. 1989. Suppression of transmission of malaria through source reduction: antianopheline measures applied in Israel, the United States, and Italy. Rev Infect Dis 11:391-406.

Knipling, E. F., H. Laven, G. B. Craig, R. Pal, J. B. Kitzmiller, C. N. Smith, and A. W. Brown. 1968. Genetic control of insects of public health importance. Bull World Health Organ 38:421-438.

Lorenzen, M. D., A. Gnirke, J. Margolis, J. Garnes, M. Campbell, J. J. Stuart, R. Aggarwal, S. Richards, Y. Park, and R. W. Beeman. 2008. The maternal-effect, selfish genetic element *Medea* is associated with a composite Tc1 transposon. Proc Natl Acad Sci U S A 105:10085-10089.

Macdonald, G. 1957. The Epidemiology and Control of Malaria. Oxford Univ., London.

Moreira, L. A., J. Ito, A. Ghosh, M. Devenport, H. Zieler, E. G. Abraham, A. Crisanti, T. Nolan, F. Catteruccia, and M. Jacobs-Lorena. 2002. Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. J Biol Chem 277:40839-40843.

Peters, L. L., and J. E. Barker. 1993. Novel inheritance of the murine severe combined anemia and thrombocytopenia (Scat) phenotype. Cell 74:135-142.

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

Scott, T. W., A. C. Morrison, L. H. Lorenz, G. G. Clark, D. Strickman, P. Kittayapong, H. Zhou, and J. D. Edman. 2000. Longitudinal studies of *Aedes aegypti* (Diptera:

Culicidae) in Thailand and Puerto Rico: population dynamics. J Med Entomol 37:77-88.

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

Smith, N. G. 1998. The dynamics of maternal-effect selfish genetic elements. J Theor Biol 191:173-180.

Smith, R. C., M. F. Walter, R. H. Hice, D. A. O'Brochta, and P. W. Atkinson. 2007. Testis-specific expression of the beta2 tubulin promoter of *Aedes aegypti* and its application as a genetic sex-separation marker. Insect Mol Biol 16:61-71.

Taylor, C., Y. T. Toure, J. Carnahan, D. E. Norris, G. Dolo, S. F. Traore, F. E. Edillo, and G. C. Lanzaro. 2001. Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics 157:743-750.

Trape, J. F., G. Pison, A. Spiegel, C. Enel, and C. Rogier. 2002. Combating malaria in Africa. Trends Parasitol 18:224-230.

Tripet, F., F. Aboagye-Antwi, and H. Hurd. 2008. Ecological immunology of mosquito-malaria interactions. Trends Parasitol 24:219-227.

Vernick, K. D., F. Oduol, B. P. Lazzaro, J. Glazebrook, J. Xu, M. Riehle, and J. Li. 2005. Molecular genetics of mosquito resistance to malaria parasites. Curr Top Microbiol Immunol 295:383-415.

Wade, M. J. 1985. Hard selection, soft selection, kin selection, and group selection. Am Nat. 125:61-73.

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

Weichenhan, D., B. Kunze, W. Traut, and H. Winking. 1998. Restoration of the Mendelian transmission ratio by a deletion in the mouse chromosome 1 HSR. Genet Res 71:119-125.

Weichenhan, D., W. Traut, B. Kunze, and H. Winking. 1996. Distortion of Mendelian recovery ratio for a mouse HSR is caused by maternal and zygotic effects. Genet Res 68:125-129.

Supplementary Material

The supplementary material is intended to provide mathematical details the text leaves out for clarity. We begin with a section describing how we calculate fitness for the autosomal case. We then repeat the calculations for an X-linked allele. We give the equations for 2 kinds of *Medea* and 3 copies of a single *Medea*. Finally, we attach the Maple code for the equilibria calculations.

Fitness Calculations

By fitness of a particular genotype we mean the average number of progeny a zygote of that genotype will have, given a particular zygote genotype distribution. A zygote with a fitness of 1 exactly replaces itself (has one progeny). Fitness of a particular allele refers to the average number of progeny an individual with that allele will have, given a particular genotype distribution. Fitness has three components. 1) The ability of an organism to survive to reproductive maturity, $l_{genotype}$. This is the embryonic fitness. 2) The ability of an organism to make gametes (a parental fertility or fecundity loss), $m_{genotype}$. 3) A component specific to *Medea*, the ability of the gametes to survive fusion to form a viable zygote, $n_{gametetype}$. In order to calculate fitness we must track the fate of the 8 types of gametes. Gametes have 3 essential attributes, 1) whether they are sperm or egg, 2) whether they carry the *Medea* or non-*Medea* allele and 3) the genotype of the gamete's parent.

To find fitnesses, we begin by finding the distribution of gametes given a distribution of zygotes. We start by introducing the following terminology. A zygote has already undergone death by the *Medea* mechanism but has not experienced any fitness costs. Zygotes can be $zygote_{++}$, $zygote_{M+}$, or $zygote_{MM}$ for the fraction of zygotes that are homozygous non-*Medea*, heterozygous for *Medea*, or homozygous *Medea*, respectively. Egg/sperm sub gamete genotype, gamete's parent's genotype. Gamete genotype can be *p* or *q* for *Medea* and non *Medea* respectively. Gamete's parent's genotype can be *MM*, *M+*, or ++ for homozygous *Medea*, heterozygous *Medea* and homozygous non-*Medea*, respectively. For example, we define spm_{q++} as the the fraction of male gametes that are non-*Medea* from a non-*Medea* parent. V_P is the parental fitness cost. In the case of an

egg, it is V_D and in the case of a sperm it is V_S. We do not consider the case where V_D is not equal to V_S. Mathematically,

$$egg_{q++} = spm_{q++} = \frac{zygote_{++}}{zygote_{++} + zygote_{M+}V_E V_P + zygote_{MM}V_E^2 V_P^2}$$

$$egg_{qM+} = egg_{pM+} = spm_{qM+} = spm_{pM+} = \frac{\frac{1}{2} zygote_{M+}V_E V_P}{zygote_{++} + zygote_{M+}V_E V_P + zygote_{MM}V_E^2 V_P^2}$$

$$egg_{pMM} = spm_{pMM} = \frac{zygote_{MM}V_E^2 V_P^2}{zygote_{++} + zygote_{M+}V_E V_P + zygote_{MM}V_E^2 V_P^2}$$

Now we examine the fitness of each type of gamete (part 3). To find fitness, we examine the fate of the gamete when it joins with all other possible gametes. For example, a non-*Medea* sperm from a non-*Medea* parent will always survive when it joins a non-*Medea* egg from a non-*Medea* parent, will die a fraction (1-t_0) of the time when it joins a non-*Medea* egg from a heterozygous parent, will always survive when it joins a *Medea* egg from a heterozygous parent, and will die a fraction (1-t_1) of the time when it joins a *Medea* egg from a homozygous *Medea* female. To find the fitness of the genotype, we find the mean of the fitness of sperm and egg of the same genotype.

$$n_{spermq++} = egg_{q++} + egg_{qM+}(1-t_0) + egg_{pM+} + egg_{pMM}(1-t_1)$$

$$n_{eggq++} = spm_{q++} + spm_{qM+} + spm_{pM+} + spm_{pMM}$$

$$n_{q++} = \tfrac{1}{2}\left(n_{spermq++} + n_{eggq++}\right)$$

$$n_{spermqM+} = egg_{q++} + egg_{qM+}(1-t_0) + egg_{pM+} + egg_{pMM}(1-t_1)$$

$$n_{eggqM+} = spm_{q++}(1-t_0) + spm_{qM+}(1-t_0) + spm_{pM+} + spm_{pMM}$$

$$n_{qM+} = \tfrac{1}{2}\left(n_{spermqM+} + n_{eggqM+}\right)$$

$$n_{spermpM+} = egg_{q++} + egg_{qM+} + egg_{pM+} + egg_{pMM}$$

$$n_{eggpM+} = spm_{q++} + spm_{qM+} + spm_{pM+} + spm_{pMM}$$

$$n_{qM+} = \tfrac{1}{2}\left(n_{spermqM+} + n_{eggqM+}\right)$$

$$n_{spermpMM} = egg_{q++} + egg_{qM+} + egg_{pM+} + egg_{pMM}$$

$$n_{eggpMM} = spm_{q++}(1-t_1) + spm_{qM+}(1-t_1) + spm_{pM+} + spm_{pMM}$$

$$n_{qMM} = \tfrac{1}{2}\left(n_{spermqMM} + n_{eggqMM}\right)$$

Genotype fitness

The genotype fitness is calculated by multiplying each component of fitness.

$$fitness_{\text{hom}\,ozygousMedea} = l_{MM} m_{MM} n_{pMM}$$
$$fitness_{heterozygousMedea} = \tfrac{1}{2} l_{M+} m_{M+} \left(n_{pM+} + n_{qM+} \right)$$
$$fitness_{\text{hom}\,ozygousnonMedea} = l_{++} m_{++} n_{p++}$$

Allele fitness

The *Medea* allele fitness is calculated by finding the fitness of the heterozygote multiplied by the fraction of *Medea* alleles in heterozygotes and adding the fitness of homozygous *Medea* multiplied by the fraction of *Medea* alleles in homozygotes. Fitness of the non-*Medea* allele is calculated similarly.

$$fitness_{MedeaAllele} = l_{M+} m_{M+} n_{pM+} \frac{zygote_{M+}}{zygote_{M+} + 2zygote_{MM}} + l_{MM} m_{MM} n_{pMM} \frac{2zygote_{MM}}{zygote_{M+} + 2zygote_{MM}}$$

$$fitness_{nonMedeaAllele} = l_{M+} m_{M+} n_{pM+} \frac{2zygote_{M+}}{zygote_{M+} + 2zygote_{++}} + l_{++} m_{++} n_{p++} \frac{2zygote_{++}}{zygote_{M+} + 2zygote_{++}}$$

Population fitness

The population fitness is the sum of the products of each genotype and the fraction of zygotes with that genotype.

$$fitness_{population} = zygote_{MM}\, fitness_{\mathrm{hom}\, ozygousMedea} + zygote_{M+}\, fitness_{heterozygousMedea} + zygote_{++}\, fitness_{\mathrm{hom}\, ozgousnonMedea}$$

X chromosome

An X-linked *Medea* is different from autosomal *Medea* in that the ratio of males to females is not 1 to 1. There are only 2 male genotypes *Medea* Y and non-*Medea* Y.

	Parental Genotype Frequency			Male Offspring Frequency		Female Offspring Frequency		
Family	Male	Female	Mating	*Medea*	non-*Medea*	Homo	Het	WT
1	S_{MY}	D_{MM}	$S_{MY}*D_{MM}$	V_E		V_E^2		
2	S_{+Y}	D_{MM}	$S_{+Y}*D_{MM}$	V_E			V_E	
3	S_{MY}	D_{M+}	$S_{MY}*D_{M+}$	½ V_E	½	½ V_E^2	½ V_E	
4	S_{+Y}	D_{M+}	$S_{+Y}*D_{M+}$	½ V_E	½		½ V_E	½
5	S_{MY}	D_{++}	$S_{MY}*D_{++}$		1		V_E	
6	S_{+Y}	D_{++}	$S_{+Y}*D_{++}$		1			1

Equations are shown in the text.

Fitness:

We use the same definitions of fitness and symbols as defined in the autosomal fitness cost case.

$$egg_{q++} = \frac{femalezygote_{++}}{femalezygote_{++} + femalezygote_{M+}V_E + femalezygote_{MM}V_E^2}$$

$$egg_{qM+} = egg_{pM+} = \frac{\frac{1}{2} femalezygote_{M+}V_E}{femalezygote_{++} + femalezygote_{M+}V_E + femalezygote_{MM}V_E^2}$$

$$egg_{pMM} = spm_{pMM} = \frac{femalezygote_{MM}V_E^2V_P^2}{femalezygote_{++} + femalezygote_{M+}V_E + femalezygote_{MM}V_E^2}$$

$$spm_{q+Y} = \frac{\frac{1}{2} malezygote_{+Y}}{malezygote_{+Y} + malezygote_{MY}V_E}$$

$$spm_{Y+Y} = \frac{\frac{1}{2} malezygote_{+Y}}{malezygote_{+Y} + malezygote_{MY}V_E}$$

$$spm_{pMY} = \frac{\frac{1}{2} malezygote_{MY}V_E}{malezygote_{+Y} + malezygote_{MY}V_E}$$

$$spm_{YMY} = \frac{\frac{1}{2} malezygote_{MY}V_E}{malezygote_{+Y} + malezygote_{MY}V_E}$$

Now we examine the fitness of each type of gamete (part 3). To find fitness, we examine the fate of the gamete when it joins with all other possible gametes.

$$n_{spermq+Y} = egg_{q++} + egg_{pM+} + egg_{pMM}$$

$$n_{eggq+Y} = spm_{q+Y} + spm_{Y+Y} + spm_{pMY} + spm_{YMY}$$

$$n_{spermYM+} = egg_{q++} + egg_{pM+} + egg_{pMM}$$

$$n_{eggqM+} = spm_{pMY}$$

$$n_{spermpM+} = egg_{q++} + egg_{qM+} + egg_{pM+} + egg_{pMM}$$

$$n_{eggpM+} = spm_{q+Y} + spm_{YMY} + spm_{pMY} + spm_{Y+Y}$$

$$n_{spermpMY} = egg_{q++} + egg_{qM+} + egg_{pM+} + egg_{pMM}$$

$$n_{eggpMM} = spm_{q+Y} + spm_{qMY} + spm_{pMY} + spm_{pMY}$$

Allele fitness

The *Medea* allele fitness is calculated by finding the fitness of the heterozygous females multiplied by the fraction of *Medea* alleles in heterozygous, adding the fitness of homozygous *Medea* females multiplied by the fraction of *Medea* alleles in a homozygous female *Medea* background and adding the fitness of male *Medea* individuals and multiplying by the fraction of *Medea* alleles in a male *Medea* background. Fitness of the non-*Medea* allele and Y are calculated similarly.

Population fitness

The population fitness is the sum of the fitness of each genotype multiplied by the fraction of zygotes with that genotype.

Two *Medea*, Two Insertions

We begin by defining variables. We do not define all possible genotypes, but explain how the naming scheme works.

S_{AABB} Fraction of the male population homozygous for *Medea*^{A} and *Medea*^{B}, a + in the subscript refers to being non-*Medea* at that locus.

S_{A+BB} is the fraction of the male population heterozygous for *Medea*^{A} and homozygous for *Medea*^{B}.

D_{AABB} Fraction of the female population homozygous for *Medea*^{A} and *Medea*^{B}, a + in the subscript refers to being non-*Medea* at that locus.

D_{A+BB} is the fraction of the female population heterozygous for *Medea*^{A} and homozygous for *Medea*^{B}.

s_A Fraction of embryos that die per copy of the *Medea*^{A} construct

s_B Fraction of embryos that die per copy of the *Medea*^{B} construct

W Mean fitness

V_A Fraction of embryos that live per copy of the A construct; $(1\text{-}s_A)$

V_B Fraction of embryos that live per copy of the B construct; $(1\text{-}s_B)$

M_{A+} Male gamete with *Medea*^{A} and non-*Medea*^{B}

F_{A+} Female gamete with *Medea*^{A} and non-*Medea*^{B}

Tables of multiple families are extremely large. Instead, we find the probability of each type of gamete. For male gametes, the genotype of the father is unimportant. However, for female gametes, if the gamete is non-*Medea*, the genotype of the female parent is critical.

We find the gamete population:

$$M_{AB}=S_{AABB}+\tfrac{1}{2}S_{AAB+}+\tfrac{1}{2}S_{A+BB}+\tfrac{1}{4}S_{A+B+}$$

$$M_{A+}=\tfrac{1}{2}S_{AAB+}+S_{AA++}+\tfrac{1}{4}S_{A+B+}+\tfrac{1}{2}S_{A+++}$$

$$M_{+B}=\tfrac{1}{2}S_{A+BB}+\tfrac{1}{4}S_{A+B+}+S_{++BB}+\tfrac{1}{2}S_{++B+}$$

$$M_{++}=\tfrac{1}{4}S_{A+B+}+\tfrac{1}{2}S_{A+++}+\tfrac{1}{2}S_{++B+}+S_{++++}$$

$$F_{AB}=D_{AABB}+\tfrac{1}{2}D_{AAB+}+\tfrac{1}{2}D_{A+BB}+\tfrac{1}{4}D_{A+B+}$$

$$F_{A+}=\tfrac{1}{2}D_{AAB+}+D_{AA++}+\tfrac{1}{4}D_{A+B+}+\tfrac{1}{2}D_{A+++}$$

$$F_{+B}=\tfrac{1}{2}D_{A+BB}+\tfrac{1}{4}D_{A+B+}+D_{++BB}+\tfrac{1}{2}D_{++B+}$$

$$F_{++}=\tfrac{1}{4}D_{A+B+}+\tfrac{1}{2}D_{A+++}+\tfrac{1}{2}D_{++B+}+D_{++++}$$

We allow the gametes to combine and *Medea* to kill. Recall that *Medea* mothers kill any wildtype offspring. *Medea* A and *Medea* B act independently. An embryo with a mother that has *Medea* A will not be rescued by a *Medea* B allele.

$G_{AABB}' = V_A^2 V_B^2 (F_{AB} M_{AB})/W$

$G_{AAB+}' = V_A^2 V_B \, (M_{AB} F_{A+} + M_{A+} F_{AB}) \, /W$

$G_{AA++}' = V_A^2 \, (M_{A+} (D_{AA++} + \tfrac{1}{2} D_{A+++}))/W$

$G_{A+BB}' = V_A V_B^2 \, (M_{AB} F_{+B} + M_{+B} F_{AB}) \, /W$

$G_{A+B+}' = V_A V_B (M_{++} F_{AB} + M_{A+} F_{+B} + M_{+B} F_{A+} + M_{AB} F_{++})/W$

$G_{A+++}' = V_A (M_{A+} (\tfrac{1}{2} D_{A+++} + D_{++++}) + M_{++} (\tfrac{1}{2} D_{A+++} + D_{++++}))/W$

$G_{++BB}' = V_B^2 (M_{+B} (D_{++BB} + \tfrac{1}{2} D_{++B+}))/W$

$G_{++B+}' = V_B (M_{+B} (\tfrac{1}{2} D_{++B+} + D_{++++}) + M_{++} (D_{++BB} + \tfrac{1}{2} D_{++B+}))/W$

$G_{++++}' = M_{++} D_{++++}/W$

Where,

$W = V_A^2 V_B^2 (F_{AB} M_{AB}) + V_A^2 V_B (M_{AB} F_{A+} + M_{A+} F_{AB}) + V_A^2 (M_{A+} (D_{AA++} + \tfrac{1}{2} D_{A+++})) + V_A V_B^2 (M_{AB} F_{+B} + M_{+B} F_{AB}) +$

$$V_A V_B(M_{++}F_{AB}+M_{A+}F_{+B}+M_{+B}F_{A+}+M_{AB}F_{++})+V_A(M_{A+}(\tfrac{1}{2}D_{A+++}+D_{++++})+M_{++}(\tfrac{1}{2}D_{A+++}+D_{++}$$

$$_{++}))+$$

$$V_B^{\,2}(M_{+B}(D_{++BB}+\tfrac{1}{2}D_{++B+}))+V_B(M_{+B}(\tfrac{1}{2}D_{++B+}+D_{++++})+M_{++}(\,D_{++BB}+\tfrac{1}{2}D_{++B+}))+M_{++}D_{++++}$$

Three copies of *Medea*

We begin by defining variables. We do not define all possible genotypes, but explain how the naming scheme works.

S_{AABBCC} Fraction of the male population homozygous for *Medea*^{A} and *Medea*^{B}, a + in the subscript refers to being non-*Medea* at that locus.

S_{A+BBC+} Fraction of the male population heterozygous at for *Medea*^{A}, homozygous for *Medea*^{B}, and heterozygous for *Medea*^{C}.

D_{AABBCC} Fraction of the female population homozygous for *Medea*^{A} and *Medea*^{B}, a + in the subscript refers to being non-*Medea* at that locus.

S_{A+BB++} Fraction of the female population heterozygous at for *Medea*^{A}, homozygous for *Medea*^{B} and homozygous non-*Medea* for *Medea*^{C}.

s_A Fraction of embryos that die per copy of the *Medea*^{A} construct

s_B Fraction of embryos that die copy of the *Medea*^{B} construct

W Mean fitness

V_A Fraction of embryos that live per copy of the A construct; $(1-s_A)$

V_B Fraction of embryos that live per copy of the B construct; $(1-s_B)$

M_{A++} Male gamete with *Medea*^{A}, non-*Medea*^{B} and non-*Medea*^{C}

F_{+BC} Female gamete with non-*Medea*^{A}, *Medea*^{B} and *Medea*^{C}

Tables of multiple families are extremely large. Instead, we make find the probability of each type of gamete. For male gametes, the genotype of the father is unimportant. However, for female gametes, if the gamete is non-*Medea*, the genotype of the female parent is critical.

We find the gamete population:

$$M_{ABC}=S_{AABBCC}+\tfrac{1}{2}S_{AABBC+}+\tfrac{1}{2}S_{AAB+CC}+\tfrac{1}{4}S_{AAB+C+}+\tfrac{1}{2}S_{A+BBCC}+\tfrac{1}{4}S_{A+BBC+}+\tfrac{1}{4}S_{A+B+CC}+\tfrac{1}{8}S_{A+B+C+}$$

$$M_{AB+}=\tfrac{1}{2}S_{AABBC+}+S_{AABB++}+\tfrac{1}{4}S_{AAB+C+}+\tfrac{1}{2}S_{AAB+++}+\tfrac{1}{4}S_{A+BBC+}+\tfrac{1}{2}S_{A+BB++}+\tfrac{1}{8}S_{A+B+C+}+\tfrac{1}{4}S_{A+B++}+$$

$$M_{A+C}=S_{AABBCC}+\tfrac{1}{2}S_{AABBC+}+\tfrac{1}{2}S_{AAB+CC}+\tfrac{1}{4}S_{AAB+C+}+\tfrac{1}{2}S_{A+BBCC}+\tfrac{1}{4}S_{A+BBC+}+\tfrac{1}{4}S_{A+B+CC}+\tfrac{1}{8}S_{A+B+C+}$$

$$M_{A++}=\tfrac{1}{4}S_{AAB+C+}+\tfrac{1}{2}S_{AAB+++}$$
$$+\tfrac{1}{2}S_{AA++C+}+S_{AA++++}+\tfrac{1}{8}S_{A+B+C+}+\tfrac{1}{4}S_{A+B+++}+\tfrac{1}{4}S_{A+++C+}+\tfrac{1}{2}S_{A+++++}$$

$$M_{+BC}=\tfrac{1}{2}S_{A+BBCC}+\tfrac{1}{4}S_{A+BBC+}+\tfrac{1}{4}S_{A+B+CC}+\tfrac{1}{8}S_{A+B+C+}+S_{++BBCC}+\tfrac{1}{2}S_{++BBC+}+\tfrac{1}{2}S_{++B+CC}+\tfrac{1}{4}S_{++B+C+}$$

$$M_{+B+}=$$

$$¼S_{A+BBC+}+½S_{A+BB++}+⅛S_{A+B+C+}+¼S_{A+B+++}+½S_{++BBC+}+S_{++BB++}+¼S_{++B+C+}+½S_{++B+++}$$

$$M_{++C}=¼S_{A+B+CC}+⅛S_{A+B+C+}+½S_{A+++CC}+¼S_{A+++C+}+½S_{++B+CC}+¼S_{++B+C+}+S_{++++CC}+½S_{+++}$$

$$+C+$$

$$M_{+++}=$$

$$⅛S_{A+B+C+}+¼S_{A+B+++}+¼S_{A+++C+}+½S_{A+++++}+¼S_{++B+C+}+½S_{++B+++}+½S_{++++C+}+S_{++++++}$$

$$F_{ABC}=D_{AABBCC}+½D_{AABBC+}+½D_{AAB+CC}+¼D_{AAB+C+}+½D_{A+BBCC}+¼D_{A+BBC+}+¼D_{A+B+CC}+⅛D_{A}$$

$$+B+C+$$

$$F_{AB+}=½D_{AABBC+}+D_{AABB++}+¼D_{AAB+C+}+½D_{AAB+++}+¼D_{A+BBC+}+½D_{A+BB++}+⅛D_{A+B+C+}+¼D_{A}$$

$$+B+++$$

$$F_{A+C}=D_{AABBCC}+½D_{AABBC+}+½D_{AAB+CC}+¼D_{AAB+C+}+½D_{A+BBCC}+¼D_{A+BBC+}+¼D_{A+B+CC}+⅛D_{A}$$

$$+B+C+$$

$$F_{A++}=¼D_{AAB+C+}+½D_{AAB+++}$$

$$+½D_{AA++C+}+D_{AA++++}+⅛D_{A+B+C+}+¼D_{A+B+++}+¼D_{A+++C+}+½D_{A+++++}$$

$$F_{+BC}=½D_{A+BBCC}+¼D_{A+BBC+}+¼D_{A+B+CC}+⅛D_{A+B+C+}+D_{++BBCC}+½D_{++BBC+}+½D_{++B+CC}+¼D$$

$$++B+C+$$

$$F_{+B+}=$$

$$¼D_{A+BBC+}+½D_{A+BB++}+⅛D_{A+B+C+}+¼D_{A+B+++}+½D_{++BBC+}+D_{++BB++}+¼D_{++B+C+}+½D_{++B+++}$$

F_{++C}=¼D_{A+B+CC}+⅛D_{A+B+C+}+½D_{A+++CC}+¼D_{A+++C+}+½D_{++B+CC}+¼D_{++B+C+}+D_{++++CC}+½D_{++++C+}

F_{+++}=

⅛D_{A+B+C+}+¼D_{A+B+++}+¼D_{A+++C+}+½D_{A+++++}+¼D_{++B+C+}+½D_{++B+++}+½D_{++++C+}+D_{+++++}

We allow the gametes to combine and *Medea* to kill. Recall that *Medea* mothers kill any wildtype offspring. *Medea* A and *Medea* B are the same type of *Medea*. An embryo with a mother that has *Medea* A will be rescued by a *Medea* B allele.

G_{AABBCC}'=$V_A^2V_B^2V_C^2(M_{ABC}F_{ABC})/W$

G_{AABBC+}'= $V_A^2V_B^2V_C(M_{ABC}F_{AB+}+M_{AB+}F_{ABC})/W$

G_{AABB++}'= $V_A^2V_B^2$ $(M_{AB+}F_{AB+})/W$

G_{AAB+CC}'= $V_A^2V_BV_C^2(M_{ABC}F_{A+C}+M_{A+C}F_{ABC})/W$

G_{AAB+C+}'= $V_A^2V_BV_C(M_{ABC}F_{A++}+M_{AB+}F_{A+C}+M_{A+C}F_{AB+}+M_{A++}F_{ABC})/W$

G_{AAB+++}'= $V_A^2V_B$ $(M_{AB+}F_{A++}+M_{A++}F_{AB+})/W$

G_{AA++CC}'= $V_A^2V_C^2(M_{A+C}F_{A+C})/W$

G_{AA++C+}'= $V_A^2V_C(M_{A+C}F_{A++}+M_{A++}F_{A+C})/W$

G_{AA++++}'= $V_A^2(M_{A++}F_{A++})/W$

$G_{A+BBCC}' = V_A V_B^2 V_C^2 (M_{ABC} F_{+BC} + M_{+BC} F_{ABC})/W$

$G_{A+BBC+}' = V_A V_B^2 V_C (M_{ABC} F_{+B+} + M_{AB+} F_{++C} + M_{A++} F_{+BC} + M_{+B+} F_{ABC})/W$

$G_{A+BB++}' = V_A V_B^2 (M_{AB+} F_{+B+} + M_{+B+} F_{AB+})/W$

$G_{A+B+CC}' = V_A V_B V_C^2 (M_{ABC} F_{++C} + M_{A+C} F_{+BC} + M_{+B+} F_{A+C} + M_{++C} F_{ABC})/W$

$G_{A+B+C+}' = V_A V_B V_C (M_{ABC} F_{+++} + M_{AB+} F_{++C} + M_{A+C} F_{+B+} + M_{A++} F_{+BC} + M_{+BC} F_{A++} + M_{+B+} F_{A+C} + M_{++C} F_{AB+} + M_{+++} F_{ABC})/W$

$G_{A+B+++}' = V_A V_B (M_{AB+} F_{+++} + M_{A++} F_{+B+} + M_{+B+} F_{A++} + M_{+++} F_{AB+})/W$

$G_{A+++CC}' = V_A V_C^2 (M_{A+C} F_{++C} + M_{++C} F_{A+C})/W$

$G_{A+++C+}' = V_A V_C (M_{A+C} F_{+++} + M_{A++} F_{++C} + M_{++C} F_{A++} + M_{+++} F_{A+C})/W$

$G_{A+++++}' = V_A (M_{A++} F_{+++} + M_{+++} F_{A++})/W$

$G_{++BBCC}' = V_B^2 V_C^2 (M_{+BC} F_{+BC})/W$

$G_{++BBC+}' = V_B^2 V_C (M_{+BC} F_{+B+} + M_{+B+} F_{+BC})/W$

$G_{++BB++}' = V_B^2 (M_{+B+} F_{+B+})/W$

$G_{++B+CC}' = V_B V_C^2 (M_{+BC} F_{++C} + M_{++C} F_{+BC})/W$

$G_{++B+C+}' = V_B V_C (M_{+BC} F_{+++} + M_{+B+} F_{++C} + M_{++C} F_{+B+} + M_{+++} F_{+BC})/W$

$G_{++B+++}' = V_B (M_{+B+} F_{+++} + M_{+++} F_{+B+})/W$

$G_{++++CC}' = V_C^2(M_{++C}F_{++C})/W$

$G_{++++C+}' = V_C(M_{++C}F_{+++}+M_{+++}F_{++C})/W$

$G_{++++++}' = (M_{+++}D_{++++++})/W$

$W=V_A^2V_B^2V_C^2(M_{ABC}F_{ABC})+V_A^2V_B^2V_C(M_{ABC}F_{AB+}+M_{AB+}F_{ABC})+V_A^2V_B^2(M_{AB+}F_{AB+})+V_A^2V_BV_C^2$
$(M_{ABC}F_{A+C}+M_{A+C}F_{ABC})+V_A^2V_BV_C(M_{ABC}F_{A++}+M_{AB+}F_{A+C}+M_{A+C}F_{AB+}+M_{A++}F_{ABC})+V$
$_A^2V_B(M_{AB+}F_{A++}+M_{A++}F_{AB+})+V_A^2V_C^2(M_{A+C}F_{A+C})+V_A^2V_C(M_{A+C}F_{A++}+M_{A++}F_{A+C})+V_A$
$^2(M_{A++}F_{A++})+V_AV_B^2V_C^2(M_{ABC}F_{+BC}+M_{+BC}F_{ABC})+V_AV_B^2V_C(M_{ABC}F_{+B+}+M_{AB+}F_{++C}+M$
$_{A++}F_{+BC}+M_{+B+}F_{ABC})+V_AV_B^2(M_{AB+}F_{+B+}+M_{+B+}F_{AB+})+V_AV_BV_C^2(M_{ABC}F_{++C}+M_{A+C}F_{+B}$
$_C+M_{+B+}F_{A+C}+M_{++C}F_{ABC})+V_AV_BV_C(M_{ABC}F_{+++}+M_{AB+}F_{++C}+M_{A+C}F_{+B+}+M_{A++}F_{+BC}+$
$M_{+BC}F_{A++}+M_{+B+}F_{A+C}+M_{++C}F_{AB+}+M_{+++}F_{ABC})+V_AV_B(M_{AB+}F_{+++}+M_{A++}F_{+B+}+M_{+B+}$
$F_{A++}+M_{+++}F_{AB+})+V_AV_C^2(M_{A+C}F_{++C}+M_{++C}F_{A+C})+V_AV_C(M_{A+C}F_{+++}+M_{A++}F_{++C}+M_{+}$
$_{+C}F_{A++}+M_{+++}F_{A+C})+V_A(M_{A++}F_{+++}+M_{+++}F_{A++})+V_B^2V_C^2(M_{+BC}F_{+BC})+V_B^2V_C(M_{+BC}$
$F_{+B+}+M_{+B+}F_{+BC})+V_B^2(M_{+B+}F_{+B+})+V_BV_C^2(M_{+BC}F_{++C}+M_{++C}F_{+BC})+V_BV_C(M_{+BC}F_{+++}$
$+M_{+B+}F_{++C}+M_{++C}F_{+B+}+M_{+++}F_{+BC})+V_B(M_{+B+}F_{+++}+M_{+++}F_{+B+})+V_C^2(M_{++C}F_{++C})$

$+V_C(M_{++C}F_{+++}+M_{+++}F_{++C})+M_{+++}D_{++++++}$

Equilibria Calculations (Maple Code)

The attached code calculates equilibrium values and stability for both autosomal and X-linked *Medea*. The code contains much of the output. Some of the equilibria take many pages to output; therefore that output has been suppressed. Some calculations take minutes to days to run on a PC with 2 gigabytes of RAM with and an Intel® Core2™ CPU . We provide appropriate warnings.

Here we provide a summary of the calculations with more details than the text. Some cumbersome equations are not reproduced. Equilibria are calculated by simultaneously solving $G'_{++} = G_{++}$ and $G'_{MM} = G_{MM}$. To find stability, the modulus of the eigenvalues of the Jacobian must be less than 1.

Recall the Jacobian matrix is defined as

$$\begin{pmatrix} \frac{\partial G'_{MM}}{\partial G_{MM}} & \frac{\partial G'_{MM}}{\partial G_{++}} \\ \frac{\partial G'_{++}}{\partial G_{MM}} & \frac{\partial G'_{++}}{\partial G_{++}} \end{pmatrix}.$$

Embrynoic Fitness Costs

$V_{D,Het}=V_{D,Homo}=V_{S,Het}=V_{S,Homo}=1$, $t_1=0$, $t_0=1$

There are 4 equilibria.

1. $G_{++} = 1$, $G_{M+} = G_{MM}= 0$

The eigenvalues are

$$\begin{pmatrix} 0 \\ V_{E,Het} \end{pmatrix}$$

2. $$G_{++} = -\frac{V_{E,Het}^2 - V_{E,Het} + V_{E,Homo}}{-V_{E,Homo} + V_{E,Het} - 1}$$

$$G_{MM} = -\frac{1 + V_{E,Het}^2 - 2V_{E,Het}}{-V_{E,Homo} + V_{E,Het} - 1}$$

Feasibility:

Using $G_{++} = 0$

$$V_{E,Homo} \geq V_{E,Het} - V_{E,Het}^2$$

Using other genotype boundaries, no additional feasibility conditions are found.

Stability: the eigenvalues are cumbersome expressions (see expression 22 in the maple code). In the biologically feasible realm, the modulus of each eigenvalue is equal to 1 when $V_{E,Het} = 1$ and $V_{E,Homo} = V_{E,Het} - V_{E,Het}^2$. These boundaries are coincident with feasibility. Except at boundaries, all feasible solutions are unstable.

3. $G_{++} = 0$

$$G_{MM} = \frac{V_{E,Homo}}{2V_{E,Het} - V_{E,Homo}}$$

Biological feasibility:

$$V_{E,Het} \geq V_{E,Homo}$$

The eigenvalues are

$$\begin{pmatrix} \frac{-V_{E,Homo} + V_{E,Het}}{V_{E,Het}^2} \\ \frac{V_{E,Homo}}{V_{E,Het}} \end{pmatrix}$$

The second eigenvalue shows a change in stability that is coincident with feasibility. Therefore, no examination $V_{E,Het} \geq V_{E,Homo}$ is necessary.

The modulus of the first eigenvalue equals 1 when

$V_{E,Homo} = V_{E,Het} + V_{E,Het}^2$ and $V_{E,Homo} = V_{E,Het} - V_{E,Het}^2$

The first solution is never biologically feasible. The second solution is stable when $V_{E,Homo} > V_{E,Het} - V_{E,Het}^2$.

4. $G_{++} = 0$, $G_{M+} = 0$, $G_{MM} = 1$

The eigenvalues are

$$\begin{pmatrix} 0 \\ \dfrac{V_{E,Het}}{V_{E,Homo}} \end{pmatrix}$$

The stability boundary is

$V_{E,Het} = V_{E,Homo}$

Stability occurs when

$V_{E,Het} < V_{E,Homo}$

Parental Fitness Costs

$V_{D,Het}=V_{S,Het}$, $V_{D,Homo}=V_{S,Homo}$, $V_{E,Homo}=V_{E,Het}=1$, $t_1=0$, $t_0=1$

Stability and feasibility analysis yields the same boundaries as embryonic. Detailed analysis is shown in Maple Code. As noted in the text, the equilibrium values are different.

Maternal Fitness Costs

$V_{E,Het}=V_{S,Het}=V_{E,Homo}=V_{S,Homo}=1$, $t_1=0$, $t_0=1$

1. $G_{++} = 1$, $G_{M+} = G_{MM} = 0$

The eigenvalues are

2. $G_{++} = 0$

$$G_{MM} = \frac{V_{D,\,Homo} - V_{D,\,Het} \pm \sqrt{V_{D,\,Homo}^2 - 2V_{D,\,Het}V_{D,\,Homo} + 2V_{D,\,Het}^2}}{V_{D,\,Het}}$$

Only the (+) solution is relevant, when $V_{D,Het} \geq V_{D,Homo}$

Stability:

The only boundary condition other than feasibility is

$V_{D,Homo} = \frac{V_{D,Het}^2 - V_{D,Het} + 1 - \sqrt{4V_{D,Het}^3 - 7V_{D,Het}^2 + 2V_{D,Het} + 1}}{V_{D,Het} - 2}$. This equilibrium is stable when homozygous fitness is greater than the expression.

3. The all genotypes equilibrium is a very cumbersome expression. However, by solving for no non-*Medea* individuals in the population, we find that the biological feasibility boundary is

the same as the stability boundary for equilibrium 2. There are no other stability boundaries. The equilibrium is always unstable when feasible.

4. $G_{++} = 0$, $G_{MM} = 1$

The eigenvalues are

$$\begin{pmatrix} 0 \\ \frac{V_{D,Homo} + V_{D,Het}}{2V_{D,Homo}} \end{pmatrix}$$

This equilibrium is stable when $V_{D,Het} > V_{D,Homo}$

Embryonic Fitness Costs and t_1

$V_{D,Het}=V_{D,Homo}=V_{S,Het}=V_{S,Homo}=1$, $t_1=0$, $t_0=1$

There are 4 equilibria.

1. $G_{++} = 1$, $G_{M+} = G_{MM} = 0$

The eigenvalues are

$$\begin{pmatrix} 0 \\ V_{E,Het} \end{pmatrix}$$

2. $G_{++} = -\dfrac{V_{E,Het}^3 t_1^2 - 4V_{E,Het}^2 t_1 - 2V_{E,Het} t_1 + 8V_{E,Het} - 4 + 2t_1}{t_1^2 V_{E,Het}^4 - 4V_{E,Het}^3 t_1 + 4V_{E,Het}^2 + 2V_{E,Het}^2 t_1 - 4V_{E,Het} - 2V_{E,Het} t_1 + 4}$

$$G_{MM} = -\frac{4(-2V_{E,Het}) + 1 + V_{E,Het}^2}{t_1^2 V_{E,Het}^4 - 4V_{E,Het}^3 t_1 + 4V_{E,Het}^2 + 2V_{E,Het}^2 t_1 - 4V_{E,Het} - 2V_{E,Het} t_1 + 4}$$

Feasibility:

Using $G_{++} = 0$

$$t_1 = \frac{2V_{E,Het}^2 + V_{E,Het} + 1 \pm \sqrt{-4V_{E,Het}^4 + 8V_{E,Het}^3 - 3V_{E,Het}^2 - 2V_{E,Het} + 1}}{V_{E,Het}^3}$$

Only the (-) solution is relevant.

Stability: No eigenvalues are less than or equal to 1 within the biologically feasible region. Therefore the equilibrium is unstable.

3. $G_{++} = 0$

$$G_{MM} = -\frac{V_{E,Het} - 1 \pm \sqrt{1 - 2V_{E,Het} t_1}}{V_{E,Het} + 2t_1 - 2}$$

Only the (+) solution is biologically relevant.

Biological feasibility:

$V_{E,Het} \leq 1 - \frac{1}{2} t_1$

The eigenvalues are cumbersome functions that are not reproduced here – see Maple code.

The modulus of the first eigenvalue equals 1 when

(a) $V_{E,Het} = 1 - \frac{1}{2} t_1$,

(b) $V_{E,Het} = \frac{1}{2t_1}$,

(c) $V_{E,Het} = -\frac{-\frac{1}{2} - \frac{1}{2}\sqrt{1+4t_1} + t_1}{t_1}$,

(d) $t_1 = \frac{2V_{E,Het}^2 + V_{E,Het} - 1 + \sqrt{-4V_{E,Het}^4 + 8V_{E,Het}^3 - 3V_{E,Het}^2 - 2V_{E,Het} + 1}}{V_{E,Het}^3}$

In case (a), this is the feasibility boundary.

In case (b), this solution is entirely outside the range of biological feasibility.

In case (c), no change of stability is found after passing this curve.

In case (d), solutions are stable above the curve and unstable below it.

No additional boundaries are found with solutions of the second eigenvalue.

4. $G_{++} = 0$, $G_{M+} = 0$, $G_{MM} = 1$

The eigenvalues are

$$\begin{pmatrix} 0 \\ -\dfrac{t_1 - 2}{2V_{E,Het}} \end{pmatrix}$$

The stability boundary is

$$V_{E,Het} = 1 - \tfrac{1}{2} t_1$$

Stability occurs when

$$V_{E,Het} < 1 - \tfrac{1}{2} t_1$$

X-linked Element

1. D_{MM}=0; D_{M+}=0; D_{++}=1/2; S_{MY}=0; S_{+Y}=1/2

The eigenvalues are 0, -.5V and V. This equilibrium is always stable except when the fitness equals 1.

2. All genotypes. See Maple Code for expressions for the genotype fractions at equilibrium. This equilibrium is unstable. The Maple code shows this by plotting the modulus of the eigenvalues for all possible fitnesses.

3. No non-*Medea* individuals

$$D_{MM} = -\frac{V_{E,Het}}{V_{E,Het} - 2}$$
$$D_{M+} = \frac{2V_{E,Het} - 1}{V_{E,Het} - 2}$$
$$D_{++} = 0$$
$$S_{++} = 0$$

This equilibrium only exists for fitness values greater than or equal to 0.5. The eigenvalues are 0 and $2V_{E,Het}$. This equilibrium is stable when it exists, except at the boundaries where the analysis is inconclusive.

4. No non-*Medea* alleles.

$$D_{MM} = \frac{V_{E,Het}}{V_{E,Het} + 1}$$
$$D_{M+} = 0$$
$$D_{++} = 0$$
$$S_{++} = 0$$

The eigenvalues are 0 and $\frac{1}{2V_{E,Het}}$. Therefore this equilibrium is stable for fitnesses greater than 0.5, and unstable for lower fitnesses; stability at the equality is inconclusive.

Chapter 4:

Some strains of *Drosophila* show resistance to synthetic *Medea*

Introduction:

Medea successfully drives in the Hay lab stock (w^{1118}) *Drosophila melanogaster*. There is at least one line ($Medea^{myd88\text{-}1}$) that drives perfectly. That is, in over 12,000 progeny, a heterozygous female has never produced a fly without *Medea* (Chen, et al, 2007). However, the identical construct inserted at other locations in the genome is not as successful. A large fraction of the offspring are *Medea*-bearing, but the drive is not 100%. Here we use 2 such lines (CC and FF). Finally, the Hay lab has generated other versions of *Medea* with different miRNA toxins including *dah* (discontinuous actin hexagons) (unpublished data).

While *Medea* elements do not have to have the property that heterozygous females produce only *Medea* offspring. In the absence of fitness costs, if the *Medea* heterozygotes kill even a small fraction of their non-*Medea* progeny, *Medea* will spread (Wade and Beeman 1994, Ward and Su, unpublished). However, fitness costs lead to introduction thresholds (Wade and Beeman 1994) and residual non-Medea individuals in the population (Ward and Su, unpublished). Genes that confer resistance to *Medea*, these will be selected for, and if fitness costs are high and/or resistance is high, Medea can be lost from the population (Ward and Su, unpublished).

If *Medea* elements are going to be used in mosquitoes, they will need to function in a variety of genetic backgrounds. Because we have synthetic *Medea* in *Drosophila*, we can test whether our *Medea* lines can convert other drosophila strains. This can give us confidence that the *Medea* elements are not specific to a particular genome. However if our *Medea* lines do not drive in other genetic backgrounds, we need to revise our toxin-antidote strategy.

Materials and Methods.

We obtained 20 strains of *Drosophila* from the Tucson Stock Center (now at UCSD). These are listed in Table 1. These strains are globally diverse and hopefully represent geographic diversity. We then crossed 6 individual male flies from these lines to our *Medea* lines and backcrossed 10 hybrid females individually into the wild strain. Male progeny from this cross were scored by eye color. Because the white gene (causes eyes to be red) is located on the X chromosome, and males get their X chromosome from their mother, half the males should get an X chromosome without the white gene. The *Medea* element is marked with a white gene that does not express as strongly as the endogenous white gene, leading to orange eyes. Because the *Mede*a element is not on the X chromosome, it segregates independently the endogenous white gene. If *Medea* functions, all progeny should have red or orange eyes. If *Medea* does not drive, half the progeny will have red eyes, one quarter will have white and one quarter will have orange. See Figure 1 for a schematic.

Table 1: Location and time of collection for each fly line. The number refers to the UCSD stock center number of the form 14021-0231.xx where xx is the number in the stock column. Collection location is where the stock was taken and the date of collection, if known, is noted.

Stock	Collection Location	Date
0	**Hawaii**	**1955**
1	Peru	1956
2	Micronesia	
3	Australia	
4	Malaysia	1962
5	Guam	1968
6	India	1968
7	Taiwan	
14	Florida	
15	Brazil	
22	Mexico	2002
23	Greece (Crete)	2002
24	Congo	2003
34	St. Kitts (Caribbean)	2005
37	Arizona	2005
47	Catalina Island	2006
51	South Africa	2007
53	Mauritius (East of Madagascar)	2006
55	Arkansas	2007
56	Connecticut	2007

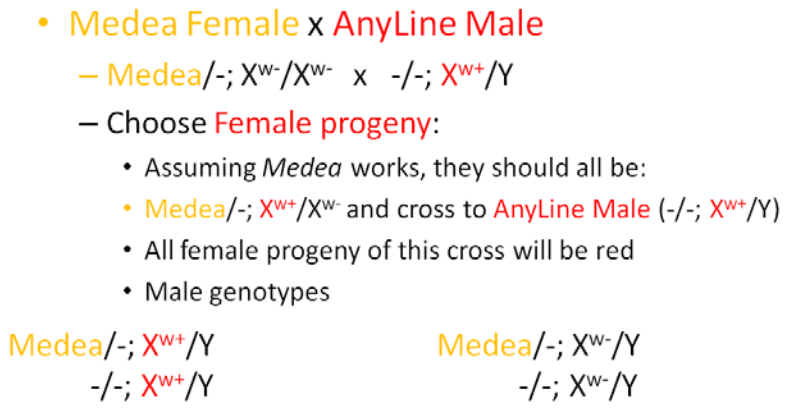

Figure 1: *Medea* crosses. This figure shows the crossing scheme to determine if *Medea Medea* works other strains. Notice that if *Medea* fails to drive, the males will be at 2:1:1 red to orange to white ratio, but if it drives, the males will be at a 1:1 ratio.

Results and discussion:

Several lines had to be excluded from our study. Lines 0, 2 and 56 did not grow well in my hands. Line 5 displayed evidence of hybrid digenesis and did not survive crosses. Line 23 died whenever it was placed on the carbon dioxide pad, possibly as a result of a sigma virus.

Between ~300 and ~2500 flies were counted for each *Medea* and strain combination. The distribution of eye colors in the males is shown in Figure 2. In *dah*, nearly all the lines show high levels of *Medea* drive, however, several show at least some resistance. Most of the wild strains show some resistance to *myd88* based *Medeas*. However, the position of the construct is important. The *myd88*-1 has less resistance than the other two positions. This is not surprising as CC and FF are known to not drive completely in w^{1118}. Some strains (such as 24) show relatively high levels of resistance in all lines. Others show resistance to cc and FF but not myd88-1.

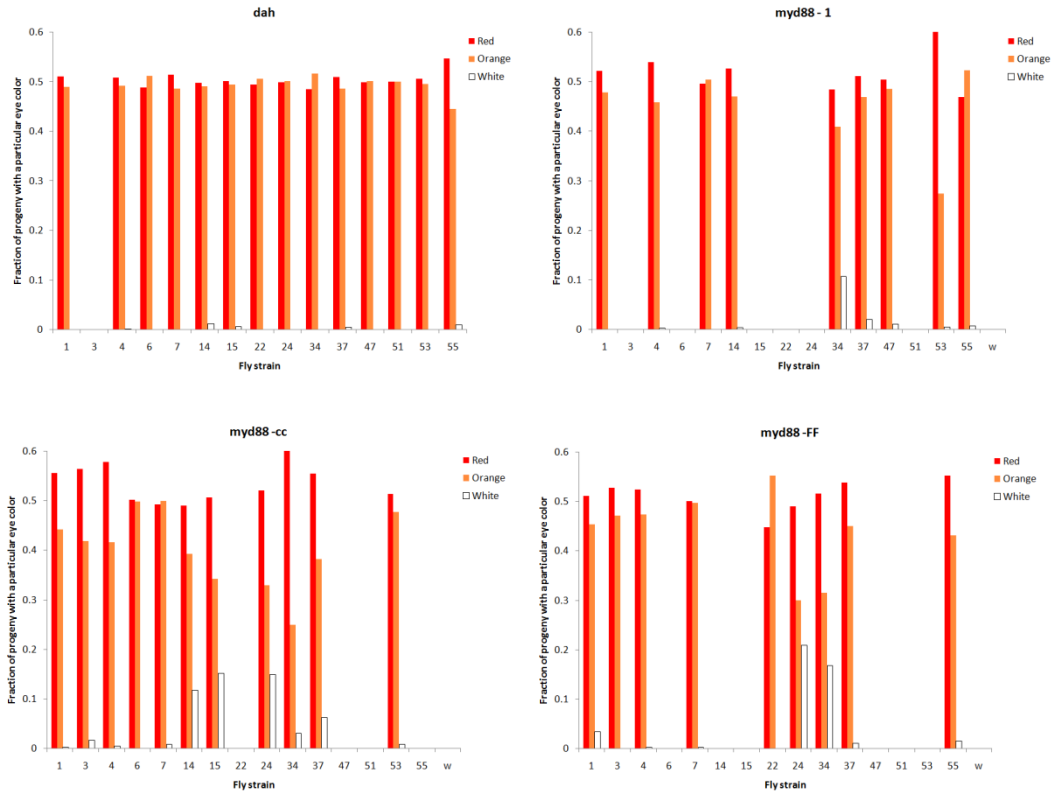

Figure 2. *Medea*'s ability to drive in various genetic backgrounds. If *Medea* works perfectly, the expected ratio of eye color (red:orange:white) is 1:1:0, if it fails, it will be at 2:1:1. The data that are missing simply mean those crosses were not set up due to time.

Because all crosses were set up with single females, we looked for variation in resistance to *Medea* within strains. If the resistance mechanism was not fixed in a strain, we would expect to see some crosses showing resistance to the *Medea* allele and other showing no resistance in a single strain. Uniform low level resistance will often allow *Medea* to spread. However, the presence, even at low levels of highly resistant alleles may cause *Medea* to be driven from a population. Because of the large number of crosses, finding statistically significant variation is difficult. However, in several cases one cross from line 34 would show no resistance and in other cases it showed resistance. However, the

number of flies that did not show resistance was not high enough to exclude 1) contamination of the sample or 2) statistical chance.

Only one mechanism of resistance was determined. The target sites for the miRNAs were sequenced and, in line 34, there is a single base pair deletion in the first target site. The loss of that base seems to confer resistance even with a second functioning target site. As the target sites are in the 5'UTR, it is not surprising that some polymorphisms occur. Other lines did not show mutations in the target sites.

Attempts to find other dominant resistant alleles were unsuccessful. We used balancer chromosomes to move a single chromosomes from the wild strains into w^{1118}. We then tested the resistance to *Medea* of these flies repeated the *Medea* crosses. The only chromosome that we isolated for *Medea* resistance was chromosome 2 of line 34, the target site mutation.

Resistance to *Medea* is common in wild strains of flies. However, these were first generation elements, containing only 2 miRNAs. Simply altering the gene that miRNA targets reduced the amount of natural resistance substantially; in fact in about 2/3 of the trials no resistance to $Medea^{dah}$ was noted. Knowledge of $Medea^{myd88}$ resistance did not correlate with $Medea^{dah}$ resistance, indicating that use both *dah* and *myd88* could be an effective strategy. Resistance can be due to mutations in target sites, but other mechanisms must also be at work. Both genomic position and identity of the target play roles in amount of resistance to *Medea*. When releasing *Medeas* in the field, ensuring

that all miRNAs can function independently can help decrease resistance. It will also be important to test *Medea* elements in local populations before releasing them into the wild.

Literature cited

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

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

Chapter 5: Conclusions

Medea is a maternal effect selfish genetic element with attractive properties for use as a synthetic gene drive system for the control of insect-borne illness. First, *Medea* elements can spread even if they confer a fitness cost to carriers. Second, under at least some conditions when *Medea* spreads it eliminates the non-*Medea* genotype from the population. Third, the synthetic form of *Medea* is the only gene drive mechanism that is both well understood at the molecular level, because it was designed with components of known behavior, and that has been demonstrated to drive population replacement. Finally, design considerations discussed by regarding ways to prevent recombinational separation of drive and disease refractoriness functions, to prevent selfish element spread in non-target species, and to carry out multiple cycles of population replacement, provide reasons to believe that the population genetic behavior of synthetic *Medea* elements can to some extent be controlled. Therefore, *Medea* is a logical target for concerted development efforts.

Our data show de novo synthesis of a selfish genetic element able to drive itself into a population in a manner consistant with our modeling efforts without need for any data correction. We extend out anaylsis with a deterministic model to show that *Medea* selfish genetic elements can drive rapid population replacement under a wide range of conditions, provided that they are introduced above a critical introduction frequency (or critical male introduction ratio), determined by the fitness costs associated with *Medea*. Our analysis highlights the importance of examining genotype frequencies as well as allele frequencies – an important point when working in systems that violate assumptions

of Hardy-Weinberg. In this case, the eggs retain a "memory" of their maternal genotype. Other gene drive systems will have unique challenges in their models.

A critical feature of *Medea*'s potential as a drive mechanism, highlighted throughout this work, is that under all conditions in which spread occurs, even when *Medea* carries a fitness cost and non-*Medea* alleles remain in the population, non-*Medea* individuals are permanently eliminated from the population. In some cases, when autosomal *Medea* elements have a $t_1>0$, or when *Medea* is located on the X, the non-*Medea* allele can also be eliminated from the population. The rate of *Medea* spread is a function of introduction ratio, fitness costs, and number of elements. Low fitness costs allow rapid spread at relatively low *Medea*/non-*Medea* male introduction ratios, while high fitness costs require higher introduction ratios in order for spread to occur quickly, or at all. The use of introduction strains that carry multiple copies of a particular element can further increase the rate of spread and allow the use of lower introduction ratios, though one element is likely to ultimately predominate within the population. In summary, the key to rapid population replacement with *Medea* is to have a high enough introduction ratio so that *Medea*-dependent killing of non-*Medea* alleles on a generation-by-generation basis, particularly for the early generations, is substantially greater than the coincident loss of fitness-compromised *Medea* alleles through natural selection.

Practical population replacement requires that transgenic individuals be refractory to disease transmission for many generations. *Medea* elements that are autosomally linked with $t_1>0$, or that are X-linked in a male heterogametic species (such as the malaria vector, *Anopheles*

gambiae) could be useful in this regard since having no non-*Medea* alleles in the population serves to maximize the number of genes for disease refractoriness in individual females in the population. These genes, because they are proposed not to confer a fitness benefit to carriers, will eventually undergo mutational decay. If such mutations result in a fitness increase to carriers (a loss of a fitness cost associated with their anti-disease function), and the non-*Medea* allele has a significant equilibrium frequency in the population, then insects permissive for disease transmission will appear. This appearance is delayed if each individual necessarily carries two *Medea* elements.

The ability to eliminate a specific allele from the population also provides a basis from which to carry out modifications of a transgenic population. For example, the toxin component of the drive mechanism also does not provide a fitness benefit to the organisms in whose genomes it resides. Therefore, it too will eventually undergo mutational decay to inactivity, resulting in the appearance of antidote-only alleles. Pre-existing diversity and mutation within the pathogen population may also contribute to the emergence of pathogen populations resistant to first generation effectors. It is also possible, though probably unlikely, that the presence of specific effectors will facilitate the emergence of new pathogens and forms of disease. For all of these reasons, it is essential that strategies be available to remove first-generation elements from the population, replacing them with second-generation elements, if desired. As discussed in Chen et al. (2007), second generation *Medea* elements can be generated that will spread at the expense of first generation elements, when both elements are located at the same chromosomal position. If second generation autosomal elements having $t_1>0$ are used, or the elements are X-linked, first generation elements can be eliminated from the population during this process. The

use of such second generation elements carries a price in that somewhat higher introduction frequencies are needed than with an autosomal *Medea* element having t_1=0. But given the importance that control over the fate of released transgenes is likely to have for the acceptance of population replacement as a viable strategy for disease prevention, this may be a small price to pay. This strategy does not restore the population to its pre-transgenic state, but it does provide a method for removing specific transgenes from the population.

What are the contexts in which area-wide population replacement with *Medea* can realistically be carried out? Our results suggest that in order for *Medea* to drive rapid population replacement within 10-20 generations (roughly 1-2 years), *Medea*/non-*Medea* male introduction ratios of between 1:10 and 1:1 are needed, depending on the nature and size of fitness costs, and the number of elements carried in the introduction strain. These numbers represent optimistic estimates because they assume that *Medea* males are competitive with wild males. However, *Medea* male mating competitiveness may be decreased (thereby lowering the effective *Medea*/non-*Medea* male introduction ratio) through several mechanisms. For example, factory rearing can select for distinct mating and other life history traits that are maladaptive in the wild. The factory strain may also be partially reproductively isolated from the wild population of interest. It should be possible to remove some of these genetic barriers to male mating competitiveness and offspring fitness by introgressing *Medea* into the wild genetic background prior to, or during factory rearing. However, non-genetic fitness costs associated with factory rearing (food, stress associated with handling and distribution to sites of release) may still result in loss of male mating competitiveness. The importance of each of

these variables will need to be evaluated, ideally in field tests, before accurate estimates of *Medea*'s ability to penetrate a specific population can be obtained.

Wild populations of *Aedes aegypti* and some *Anopheles* species have been estimated to range from 10,000-20,000 adults per village. These sizes are small compared with those associated with classical sterile male release in other insects; 68,000 per week in the case of the screw worm fly, and ~10^9 in the case of ongoing Mediterranean fruit fly suppression programs. With respect to mosquitoes, weekly factory production of 1,000,000 *Aedes aegypti* could be achieved routinely in the 1960s. Large numbers of *Anopheles* males have also been produced in factory environments using mid-twentieth century technologies. In some contexts it may also be possible to take advantage of naturally-occurring changes in mosquito population size to provide an environment in which *Medea* can more easily gain a foothold within a population. For example, while wet season populations of *Anopheles* adults per village in Mali can reach ~15,000, in the dry season these populations consist of only 1,000-3,000 adults. These encouraging points notwithstanding, it is important to emphasize that area-wide population replacement remains a daunting task. Disease-endemic regions can be very large (thousands of square miles), and consist of many villages, requiring that the number of *Medea* males to be released be scaled accordingly. Modest migration rates can result in rapid spread of *Medea* in space (J. M. Marshall, C. M. Ward and J. T. Su, unpublished observations). However, if inter-village distances and other environmental obstacles are such that mosquito migration between villages is minimal, *Medea* will need to be introduced at many sites in order to create area-wide protection within a reasonable timeframe. These facts, coupled with the issues associated with factory rearing and distribution noted above, argue that a detailed

analysis of the local mosquito population, the environment, and the amount of resources available to raise and distribute *Medea*-bearing insects will need to be carried out before population replacement is attempted in any given area. Finally, we note that the models examined here make a number of assumptions: infinite population size, non-overlapping generations, no age structure within the population, random mating, and no migration. This kind of model is often used to gain basic insights into population genetic processes. However, it provides only a qualitative snapshot of the conditions under which *Medea* can succeed in driving population replacement. It will be important to carry out more detailed modeling that takes account of the biology of specific pest species, stochasticity, as well as other variables that can influence rate of spread and functional lifetime in the wild.

Finally, we note that while models are critical for understanding the dynamics of spread and estimating population sizes, they are not a substitute ofr laboratory efforts. First, for pests such as mosquito species, there is little genetic or molecular information regarding genes and promoters used during oogenesis and early embryogenesis. This information is necessary for the creation of *Medea* elements and straightforward to generate, with the use of transcriptional profiling to identify appropriately expressed genes and transgenesis and RNA interference in adult females to identify those required for embryonic development, but it remains to be acquired. Finally, we must test any strains we develop in the lab with a variety of insects from the wild populations to ensure that those populations are not pre-adapted to be resistant to our engineered elements. From rapid screening it appears that populations will have variable amounts of resistance to alleles. Encouragingly, even our first generation elements drove in many, mnay populations. By increasing the number of target sites and selecting those targets to be in conserved

regions, we can create better candidate elements. Ulitmately, field tests will be required. Although an understanding of the above issues is critical for the success of any population-replacement strategy, the problems are not intractable, as evidenced by past successes in controlling pests by means of sterile-male release and as implied by our growing understanding of mosquito population genetics, immunity, and ecology.