MAKING GENE BIO-DEGRADABLE: PARAMETER ESTIMATION AND MATHEMATICAL MODELING OF SPREAD OF ANTI-PATHOGEN GENE IN MOSQUITOES

A Thesis

by

NEDA NAJMITABRIZI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Mark Lawley
Co-Chair of Committee,	Lewis Ntaimo
Committee Member,	Zachary Adelman
Head of Department,	Mark Lawley

December 2017

Major Subject: Industrial Engineering

ABSTRACT

Mosquitoes can transfer a lot of dangerous infections which place a massive burden on human health. Scientists have achieved significant progress in the last decades in the detection and management of mosquito-borne diseases; however, these diseases continue to threaten human life.

In the last decades, gene drive systems have been considered as a biological conversion method to prevent transmission of vector-borne disease. Through this approach, scientists have been trying to spread a trait in mosquitoes' population by altering their genetics so that mosquitoes become resistant to the disease. In this case, mosquitoes cannot get infected and hence, we do not need to worry about the disease transmission by mosquitoes. However, genetic alteration has risks and scientists do not know how the new genetic will influence the ecosystem; what will be the potential mutation and if the new genetic does not work as expected, how it is reversible? Hence, none of the gene drive methods has been conducted in nature yet.

In this thesis, we studied the biodegradability of an engineered gene that can get spread through the population by a gene drive method and then eliminate itself returning the genotype into the wild-type genetic structure. A mathematical model was constructed to study this feature for five different gene drive systems. Based on simulation results, three of them succeeded while one succeeded after modification in mechanism and one failed to achieve the biodegradability goal.

To validate our model, we put two parameters equal to zero that provided the results equivalent to present studies and got the results matched. Future laboratory-based experiments will be conducted to test the validity of our results.

DEDICATION

To my husband and best friend, for his unconditional support and motivation.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Dr. Mark Lawley, my committee members Dr. Zach Adelman and Dr. Lewis Ntaimo for their continuous guidance and support of my research, as well as their patience, motivation and immense knowledge.

I would like to thank Dr.Madhav Erraguntla, who was always willing to help and give his best suggestions.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professors Dr. Mark Lawley and Dr. Lewis Ntaimo of the Department of Industrial and Systems Engineering and Professor Dr. Zach Adelman of the Department of Entomology

All work for the thesis was completed by the student, under the advisement of Dr. Zach Adelman of the Department of Entomology at Texas A&M University and the committee members.

Funding Sources

Graduate study was supported by DARPA project from Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

The first chapter of this thesis provides information about the mosquito-borne disease, related complications, and gene drive methods. In the last two sections of this chapter, the motivation, goal, and selected methodology are discussed.

I.1 Mosquito-borne disease

Mosquitoes are dangerous creatures. They transmit serious diseases that cause more than one million human deaths per year [1]. The most prevalent mosquito-borne protozoan diseases and viruses are malaria, zika, dengue, chikungunya, and yellow fever [1,2]. The disease can be transmitted by the bite of an infected female mosquito. While the majority of them occurs in underdeveloped countries, however, they are on the move across the world due to factors like traveling and global change.

I.1.1 Protozoan disease

Malaria is a famous mosquito-borne vector disease. Parasitic protozoans, Plasmodium, which cause malaria, can be transmitted by infected female Anopheles mosquitoes.

In 2015, about 85% of malaria incidents were reported from sub-Saharan Africa [3], and currently about half of the world population is at the risk of infection of this disease [4,5]. While there is a treatment for malaria, however, it can be lethal if it causes severe side effects such as cerebral malaria, pulmonary edema, liver, kidney or spleen failure, severe anemia, and low blood sugar [6]. Another reason for a high mortality rate of malaria is

that the majority of incidents occur in underdevelopment countries with lack of doctors and medical resources [3].

I.1.2 Virus disease

The bite of an infected female Aedes mosquito can transfer zika, dengue, chikungunya and yellow fever depending on the type of virus in the mosquito's body. for which there is no treatment or vaccine except for yellow fever.

Zika infection usually has mild symptoms which last for 2-7 days, but if the infection occurs during pregnancy, it causes microcephaly or other fetal brain abnormalities in the newborn baby. It also is a cause of Guillain-Barré syndrome, a movement disorder that leads to paralysis or death [7].

The symptoms of mild dengue are very similar to flu, however, if it does not get controlled, it develops into severe dengue with the symptoms such as hemorrhage, shock, vomiting blood and difficult breathing. At this state, the conditions are life-threatening in children and require critical medical care [7].

Like dengue, chikungunya has flu-like symptoms, but it can be misdiagnosed where dengue occurs. While most patients can recover, however, serious cases such as eye, neurological and heart complications can occur and may cause death in old people [7].

Yellow fever causes same symptoms as mentioned above. It can be difficult to diagnose and can be confused with dengue, chikungunya, and malaria. About 15% of yellow fever patients enter to the toxic phase of this illness at which severe symptoms such as hemorrhage and kidney failure occur and cause death in half of those patients at toxic phase.

Mosquito-borne diseases have spread in new areas due to traveling and changes in climate and micro-organisms. They have potential to cause significant mortality and morbidity as well as enormous economic damage. Even if they do not lead to death, they may affect the life quality of patients for the rest of their lives. Most people think that this type of disease is a matter for individuals and specialists in tropical areas, but in fact, they are an issue for doctors, veterinarians, economists and policy makers globally.

In the last decades, significant progress has been achieved in detection and management of vector-borne diseases. In this regard, gene drive methods can play a major role in preventing disease transmission. The next section of this chapter explains the gene drive mechanism and different corresponding methods.

I.2 Gene drive system

The mechanism of this approach is to spread a selfish gene which is inherited at the super-Mendelian rate. It is a revolutionary technique of altering the odds that gives scientists the power of changing or eliminating organisms and promoting the inheritance of a specific trait. The first gene drive system was proposed at 1968 in London School of Hygiene and Tropical medicine. It used translocations mechanism to drive anti-pathogen gene into wild-type populations [8]. Nowadays, gene drive techniques have progressed in their goal to drive anti-pathogen gene traits into wild-type populations to combat the vector-borne disease. Hence, a gene drive system can help an anti-pathogen gene linked gene spread through a vector population and become fixed, thus preventing parasite transmission. The anti-pathogen gene can be spread through vector population by just releasing some transgenic mosquitoes which is called

inundative release to achieve population replacement, or by a gene drive method such as (1) engineered Under-dominance system, (2) the Maternal Effect of Dominant Embryonic Arrests (MEDEA), (3) driving-Y method, and (4) Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). In next sub-sections, we provide definition and explanation for each method.

I.2.1 Inundative release

In this approach, transgene mosquitoes are used as a tool to control and extinguish the mosquito-borne disease. Male, parasite-resistant transgene mosquitoes are introduced into wild-type population [9]. The offspring will inherit the new gene based on the Mendelian rate which helps to spread the anti-pathogen gene into the wild population. However, inheritance of at least one allele of the new genotype will place fitness cost on the reproduction of those individuals.

I.2.2 Engineered under-dominance system

This mechanism was proposed the first time by Davis et al. [10]. It is classically identified as a genetic condition at which heterozygotes has a high fitness cost compared to homozygous individuals [11]. This system can drive the desired gene into the wild population by releasing transgenic individuals.

I.2.3 The Maternal Effect of Dominant Embryonic Arrests (MEDEA)

This mechanism is the product of maternally expressed toxin gene and zygotically expressed antidote gene. Hence, non-Medea bearing offsprings will be vulnerable and will be killed by the toxin gene, but Medea- bearing eggs will rescue themselves and will survive due to the antidote gene [12]. It means that if the mother

has at least one allele of transgenic genotype, its offspring will survive if they inherit at least one allele of the transgenic genotype, otherwise they will die.

I.2.4 Driving-Y

This method spreads the anti-pathogen gene which is linked to the Y-chromosome of an engineered male mosquito. The Y-chromosome shreds the X chromosome of the germline so that it distorts sex ratio, and as a result, the gamete will predominantly carry the Y chromosome [13]. Hence, if the father is transgenic mosquito, its offsprings will predominantly be male. This mechanism helps prevention of the spread of disease by the extinction of the vector population.

I.2.5 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

The concept of CRISPR/Cas9 gene drive was first proposed by Burt et al. [14]. This system works based on two key molecules: an enzyme which is called *Cas9* and a piece of RNA which is called guide RNA (gRNA). gRNA finds the spot on the DNA which needs to be cut and repair. Then the enzyme cuts that specific spot. At this point, gRNA can copy the new gene in the target place [15,16]. Hence, an anti-pathogen gene linked gene spreads through a vector population, thus preventing parasite transmission.

While the success of these mechanisms has been validated in simulation and laboratory-based studies [10,11-12,14-17], none have been tried in nature due to the risks. There are uncertainties about the potential risks which pose several questions for scientists. Will there be any genetic mutation of the converted population? If this is the case, how will it affect the environment? How will they become resistant to pesticides? If the wild population is converted or suppressed, how will the balance of the ecosystem be

affected? Moreover, what if the gene drive does not work as expected? In that case, will the scientists be able to reverse the drive? If so, how will the economic burden be handled? One of the recent ideas to minimize potential risks is to engineer a genotype, which first spreads an anti-pathogen gene throughout the population and then removes itself from the nature.

I.3 Research motivations and goals

As mentioned in previous sections, mosquito-borne vector diseases continue threatening global health. It is vital to prevent them rather than managing the symptoms and side effects after the incident. Significant progress has been achieved in this regard. However, there are risks and concerns about the implementation of those methods in nature preventing scientists from conducting them in the real world. It would be much safer to conduct a method that spreads a trait which can remove itself when the desired effect is achieved in the population. At present, there is no clear understanding of how these seemingly counterintuitive goals are achievable: spreading a gene into a population to fixation and then completely removing the gene from the wild.

Thus the primary purpose of this study is to show that these goals are not mutually exclusive and can be controlled based on the desired results. We simulated and studied different gene drive methods with the property of self-elimination to see whether they succeed to spread a bio-degradable anti-pathogen gene in wild-type mosquito's population. Furthermore, we provided an estimation of parameters which are required to run the experiments in the laboratory.

I.4 Mathematical modeling approach for this study

Nowadays, system science is considered an essential interdisciplinary tool to deal better with a complex system. The challenges and systems we face in the 21st century are highly interlinked and multidisciplinary. Hence, we need to formulate and consider general principals and different representations of a system to answer the questions [18]. System science is highly efficient in this regard. Methods such as agent-based modeling and mathematical modeling have been used in a variety of domains.

In this study, we have used mathematical modeling to simulate different gene drive methods to see whether they can succeed in spreading a bio-degradable anti-pathogen gene. Several studies have developed simulations to check how a gene drive can spread an anti-pathogen gene, but none has studied the self-elimination property of the anti-pathogen gene.

CHAPTER II

LITERATURE REVIEW

Engineering a biodegradable anti-pathogen gene is a novel idea that to the best of our knowledge is not studied in any published research. Hence, the literature review contains the result of previous studies regarding the success of gene drive systems in conversion or suppression of vector population.

Okamoto et al. have studied introgression of anti-pathogen gene in very large number into wild-type population [19]. They have used Skeeter Buster which is a stochastic spatially explicit model for Aedes mosquitoes to investigate whether releasing mosquitoes with only one anti-pathogen gene will cause population replacement. This method models the population dynamics of Aedes mosquitoes in an urban area. They found that releasing transgenic mosquitoes into wild-type population will result in population replacement which means that the anti-pathogen gene will spread through population and mosquitoes will become resistant to the disease, but their genotype will no longer be wild type.

Edgington et al. [20] have considered two-locus engineered under-dominance method and have conducted population genetics model which basically is the study of genetic variation among the population. They have assumed an infinite, closed, randomly mating population with non-overlapping generation. Furthermore, mutation and migration have been neglected in this study. They found that single release of transgenic mosquitoes will cause them to stay in population and found that this method does not need a high

initial number of transgenic mosquitoes to spread the anti-pathogen gene and hence, the release number of about 1/3 of the wild-type population will cause the trait to get fixed in the population.

Wade et al. [21] have also conducted population genetics to describe the expected population dynamics after conducting MEDEA mechanism. They found that in the absence of any fecundity fitness cost, MEDEA spreads the new genetic structure to fixation through the population. They also, have studied the properties and natural occurrence of MEDEA in Tribolium. Based on their study, this mechanism happens naturally in Tribolium population in Aisa, North and South America, Africa and Europe but it is almost absent in Indian and Australian population of Tribolium. In their laboratory experiments, they detected four genetic loci which could exhibit the mortality pattern of MEDEA. Chen et al. [22] have developed a synthetic MEDEA elements to explore the population replacement in Drosophila. In this approach, MEDEA encodes both maternally expressed toxin and zygotically expressed antidote. The toxin kills non-MEDEA bearing progeny and antidote rescue MEDEA bearing offspring.

Galizi et al. [23] have studied the distortion system for the control of malaria which is called driving-Y. They ran a laboratory-based experiment at which the goal was to suppress the population by releasing transgenic male mosquitoes. In this mechanism, Y-chromosome plays the selfish gene role which its inheritance rate is above mendelian inheritance rate, and the anti-pathogen gene is linked into it. Hence their progeny will dominantly be male. They found this mechanism to be two orders of magnitude more efficient than releasing sterile males for causing population extinction.

CRISPR was proposed by Esvelt et al. [15]. In that study, they have explained the way CRISPR method spreads the trait along with the role of the enzyme in cutting the chromosome and copying the trait. While the professional biological methods have been discussed in that study, Nobel et al. [24] have constructed and analyzed a mathematical model CRISPR gene drive mechanism. They have assumed an infinitely large population and random mating. In their study, they have included the effect of multiple cutting by multiple gRNA. Their results have shown that this mechanism can invade wild-type populations spreading the trait through it, but it could have limited utility due to the reproduction of resistance alleles.

As explained earlier, none of the current studies have considered self-elimination of the anti-pathogen gene. In this research, we have included bio-degradability of the anti-pathogen gene through gene drive methods.

CHAPTER III

GENOTYPES, PARAMETERS, AND MODEL STRUCTURE

This chapter explains the method and parameters we have used in our mathematical modeling. Our model simulates the spread of bio-degradable anti-pathogen gene into wild-type mosquito population through different gene drive methods.

III.1 Genotypes, alleles, fitness cost, and parameters

Genotype is the genetic structure of an organism which has two alleles for each gene. If cells of an organism contain two different alleles of a gene, it is called heterozygotes; otherwise, it is called homozygous. In this study, we use ww as the genotype of initial wild-type population, and gg and wg as the homozygous (e.g. gg) and heterozygotes (e.g. wg) genotypes of initial transgenic mosquitoes.

There may be a fitness cost for reproduction ability of a specific allele or genotype. It is basically considered as less reproductivity and higher mortality.

Number and types of genotypes, parameters, and initial population scenarios may differ from one mechanism to another, but there are three probabilities that are held in common by all mechanisms.

For all gene drive methods we have applied three probabilities for g allele transmission: (1) self-elimination, (2) self-generation, and (3) resistance, which are denoted by α, β and γ respectively such that: $\alpha + \beta + \gamma = 1$.

Self-elimination causes g allele to eliminate itself and turn to wild-type allele which depends on the mechanism of gene drive, thus g will be turned into w or w*.

Self-generation causes g allele to be inherited as g allele in next generation, and resistance probability, γ, will cause g allele to turn into g*, a crashed allele which cannot eliminate itself. In the process of self-elimination, g allele first gets crashed and then gets repaired into wild-type allele. If the repair process cannot succeed, g allele stays as its crashed form at which it cannot turn into wild-type anymore. Figure 1 shows the transmission of g allele based on the gene drive mechanism.

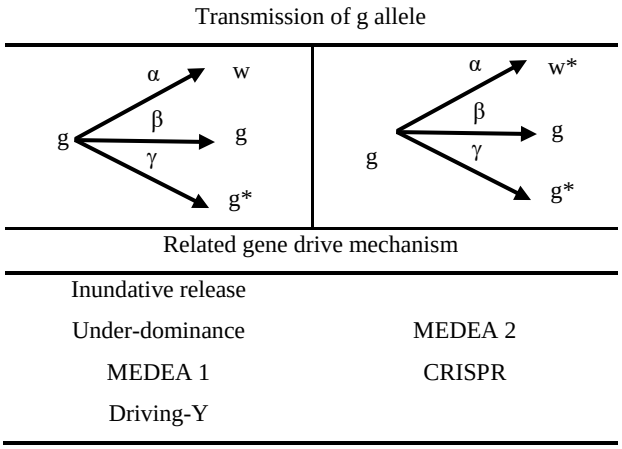

Figure 1. Transmission of g allele based on different gene drive mechanisms

In addition to those common probabilities, there are three other probabilities related to w allele transmission of wg and wg* genotypes in CRISPR method: (1) successful repair probability that turns w allele into g allele, (2) unsuccessful repair probability that transforms w into w', and (3) unsuccessful cleavage probability that makes w allele to be inherited as w allele by offspring, which are denoted by δ, η and ρ respectively. Figure 2 depicts the transmission of w allele of wg or wg* genotypes in

CRISPR mechanism and we will explain the nature of w' and w* in the corresponding section.

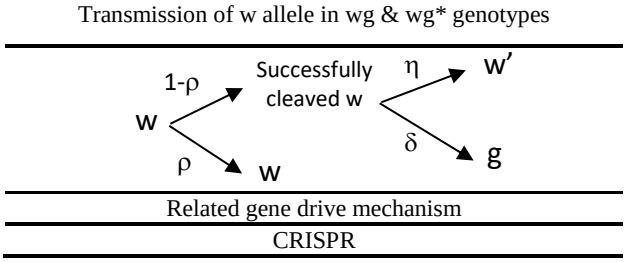

Figure 2. Transmission of w allele of wg & wg* genotypes in CRISPR mechanisms

Parameters we have included in our models are: (1) average daily reproduction rate, denoted by λ, is the average number of eggs laid by a female mosquito per day, (2) average mortality rate, μ_A, is the average number of mosquitoes that die per day, (3) fitness cost per g/g* allele, *f*, which is applied in all models except under-dominance system (4) fitness cost for heterozygotes in under-dominance method, f_U, (5) fitness cost for vulnerable wild-type mosquitoes in MEDEA mechanism, f_M, and, (6) fitness cost for female mosquitoes in driving-Y system, f_H. The average mortality rate has been applied for total population of each genotype. Total population of each genotype is sum of juvenile and adult mosquitoes. By juvenile, we mean the combination of eggs, larvae and pupae stages. At each stage, some mosquitos may die daily, hence we have considered an average mortality rate for total population. Table 1 summarizes the probabilities and parameters.

Parameter/Probability	Description	Value	Method
λ	Average daily reproduction rate	7	All methods
μ_A	Average daily mortality rate	0.3	All methods
f	Fitness cost per g/g* allele	5%	All methods except under-dominance
f_U	Fitness cost for heterozygotes	100%-90%-80%	Under-dominance
f_M	Fitness cost for vulnerable wild-type	100%-90%-80%	MEDEA
f_H	Fitness cost for female mosquitoes	99.9%-95%-90%	Driving-Y
α	Probability of self-elimination for g allele	0 – 0.8	All methods
β	Probability of self-generation for g allele	1-α-γ	All methods
γ	Probability of becoming resistant for g allele	0.05	All methods
ρ	Unsuccessful cleavage probability for w allele	0.005	CRISPR
δ	Successful repair probability for w allele	0.99	CRISPR
η	Unsuccessful repair probability for w allele	0.01	CRISPR

Table 1. List of parameters used in the study along with the corresponding method

In the next subsections, we explain the different possible outcome of genotypes and initial scenarios we have conducted in our mathematical modeling along with more details about different fitness cost and other parameters we have applied for each gene drive methods.

III.1.1 Inundative release

For this method, we have considered three different initial population scenarios: (1) 66% ww + 33% male gg, (2) 100% wg, and, (3) 100% gg. For each initial scenario, we have run three different simulations: (a) $\gamma = 0$ without any application of fitness cost into g/g* (g or g*) allele, (b) $\gamma = 0$ with application of 5% fitness cost into g/g* allele, and, (c) $\gamma = 5\%$ with application of 5% fitness cost into g/g* allele; in total, we have run 9

simulations, for each of nine levels of α. In this mechanism, self-elimination probability, α, turns g allele into w allele that is a wild-type allele resistant to the pathogen transmission.

Modeling by $\gamma = 0$ is an idealistic approach at which g allele may have only two states, w or g. However, in reality, there is a chance that g allele gets crashed.

The possible outcome of genotypes in this system is ww, wg and gg for $\gamma = 0$, however, when we include $\gamma = 5\%$, we will have other three genotypes as wg*, gg*, and g*g*. When the initial population is 66% ww + 33% male gg, the first generation will consist of ww, wg, and wg* (in the case of $\gamma=0$) and other three genotypes will appear from the second generation.

In contrast to the first scenario of the initial population, all six genotypes will appear from the first generation.

III.1.2 Under-dominance method

Likewise inundative release, we have studied three different initial population scenarios for under-dominance method as (1) 100% gg, (2) 25% ww + 75% gg, and, (3) 40% ww + 60% gg. For each scenario, we have studied six different combinations of γ and fitness cost as (a) $\gamma = 0$ and heterozygotes fitness cost = 100%, (b) $\gamma = 0$ and heterozygotes fitness cost = 90%, (c) $\gamma = 0$ and heterozygotes fitness cost = 80%, (d) $\gamma = 5\%$ and heterozygotes fitness cost = 100%, (e) $\gamma = 5\%$ and heterozygotes fitness cost = 90%, and, (f) $\gamma = 5\%$ and heterozygotes fitness cost = 80% which in total requires 18 simulation runs for each of nine levels of α. Similar to the inundative release, α turns g

allele into pathogen-resistant wild-type allele, w. If the $\gamma = 0$, the possible genotypes are ww, wg and gg but for non-zero γ we have wg*, gg* and g*g* as well and since in this system, the initial population includes female gg mosquitoes, all genotypes appears from first generation.

III.1.3 MEDEA

For this mechanism, we have studied two approaches: (1) self-elimination probability, α, turns g allele into wild-type, pathogen-resistant allele, w, which still makes ww vulnerable in MEDEA mechanism, (2) α, turns g allele into wild-type, pathogen-resistant allele, w* which can survive in MEDEA method. Hence, any wild-type genotype that includes w* is not vulnerable in this mechanism anymore. The only remaining vulnerable genotype is ww.

In the "result" part, we have explained the reason we have studied two approaches for MEDEA.

For both approaches, we have considered three initial population scenarios: (1) 100% gg, (2) 25% ww + 75% male gg, and, (3) 50% ww + 50% male gg. Likewise to the under-dominance method, for each scenario, we have studied six different combination of γ and fitness cost as (a) $\gamma = 0$ and vulnerable wild-type fitness cost = 100%, (b) $\gamma = 0$ and vulnerable wild-type fitness cost = 90%, (c) $\gamma = 0$ and vulnerable wild-type fitness cost = 80%, (d) $\gamma = 5\%$ and vulnerable wild-type fitness cost = 100%, (e) $\gamma = 5\%$ and vulnerable wild-type fitness cost = 90%, and, (f) $\gamma = 5\%$ and vulnerable wild-type fitness cost = 80% which in total it required 36 simulation runs for each level of α. In addition to fitness cost for vulnerable wild-type genotypes, we have applied 5% fitness cost into each g/g* allele.

The difference in the type of self-elimination between two MEDEA approaches, has caused different outcomes for genotypes. In the first one, the possible genotypes are ww, wg, wg*, gg, gg* and g*g*, that ww, wg, and wg* appear from the first generation and the others appear from the second generation (for non-zero γ) when the initial population is a combination of ww and gg, but all appear from the first generation if the initial population is 100% gg. However, In the second approach, there are ten possible outcomes: ww, wg, ww*, wg*, w*w*, w*g, w*g*, gg, gg*, g*g* such that the first four genotypes appear from the first generation, and the rest of them can be seen from the second generation (for non-zero γ) except when the initial population in 100% gg in the second approach of MEDEA. In that case, there are six possible genotypes as w*w*, w*g, w*g*, gg, gg* and g*g* which all of them appears from the first generation and are not vulnerable to the toxin in MEDEA mechanism.

III.1.4 Driving-Y

Nature of the driving-Y is to spreads Y-chromosome as the selfish element. To spread the anti-pathogen gene by this gene drive system, the g allele must be linked to the Y-chromosome denoted by g^Y. Since the g allele cannot be attached to X-chromosome, it implies that we cannot introduce $g^X g^Y$genotype male mosquitoes into the wild-type population. The released male mosquitoes have $w^X g^Y$genotypes at which g is attached to Y-chromosome. Wild-type mosquitoes genotype is denoted by $w^X w^Y$ as male mosquitoes and $w^X w^X$ as female mosquitoes. Furthermore, during the reproduction process, g allele cannot transfer from Y-chromosome to X-chromosome.

In this mechanism, we have studied one initial population scenario which is 66% ww + 33% $w^X g^Y$ in six different states: (1) γ = 0 and 99.9% fitness cost for vulnerable female mosquitoes, (2) γ = 0 and 95% fitness cost for vulnerable female mosquitoes, (3) γ = 0 and 90% fitness cost for vulnerable female mosquitoes, (4) γ = 5% and 99.9% fitness cost for vulnerable female mosquitoes, (5) γ = 5% and 95% fitness cost for vulnerable female mosquitoes, and (6) γ = 5% and 90% fitness cost for vulnerable female mosquitoes that in total it required six simulation runs for each level of α at which α turns the g allele into pathogen-resistant wild-type allele, w^Y.

The possible outcome of genotypes are $w^X w^Y$, $w^X w^X$, $w^X g^Y$ and $w^X g *^Y$ (for non-zero γ) and all of them appear from first generation.

III.1.5 CRISPR

For this method, we have studied only one initial population scenario which is 90% ww + 10% male gg mosquitoes, in two states (1) γ = 0, and (2) γ = 5% . In each state, we have considered only 5% fitness cost per g/g* allele which means two simulation runs for each level of α at which α turns g allele into pathogen-resistant wild-type allele, w*. However, the modeling of this system is more complicated compared to previous gene drive methods because of the presence of extra parameters and alleles. In this system, in addition to α, β and γ, there are three more probabilities that we define them here.

We explained earlier that in CRISPR mechanism, the enzyme *Cas9*, cuts the target point on chromosome and gRNA repairs the cut point by copying the g allele. The goal of this engineered system is a successful cleavage and a successful repair in w allele.

However, it may not be the case always. There is a chance that the enzyme cannot succeed to cleave the chromosome. In this case, w allele gets inherited by new offspring as w allele. We have denoted this probability by ρ. If the cleavage succeeds, the gRNA tries to repair the cleaved point by copying the g allele. If the gRNA succeed in its goal, w allele turns into g allele. We denote the corresponding probability by δ. However, there is a chance that gRNA cannot repair the cleaved point. In this case, we have a crashed w allele, denoted by w' which is cleaved by *Cas9* enzyme but could not get repaired by gRNA. We denote this probability by η. These occurrences all happen in wg and wg* alleles.

The possible alleles in this mechanism are w, w', w*, g, g* which cause fifteen different genotypes as ww, ww', ww*, w'w', w'w*, w*w*, wg, wg*, w'g, w'g*, w*g, w*g*, gg, gg*, g*g*. The first six genotypes are considered as wild-type genotypes. For non-zero γ, ww, ww*, wg, and wg* appear from the first generation while the others appear from the second generation.

Table 2 summarizes each gene drive system as well as studied initial population scenarios and possible outcomes of genotypes.

Method	Initial population	Fitness cost	Genotypes ($\gamma = 0.05$)
Inundative release	66% ww + 33% gg (male) 100% wg 100% gg	5% per g/g*	ww wg wg* gg gg* g*g*
Under-dominance	100% gg 25% ww + 75% gg 40% ww + 60% gg	100% cost (wg/wg*) 90% cost (wg/wg*) 80% cost (wg/wg*)	ww wg wg* gg gg* g*g*
Medea-1	100% gg 25% ww + 75% gg (male) 50% ww + 50% gg (male)	100% cost (vulnerable wild-type) 90% cost (vulnerable wild-type) 80% cost (vulnerable wild-type)	ww wg wg* gg gg* g*g*
Medea-2	100% gg	100% cost (vulnerable wild-type) 90% cost (vulnerable wild-type) 80% cost (vulnerable wild-type)	w*w* w*g w*g* gg gg* g*g*
	25% ww + 75% gg (male) 50% ww + 50% gg (male)	100% cost (vulnerable wild-type) 90% cost (vulnerable wild-type) 80% cost (vulnerable wild-type)	ww ww* w*w* wg wg* w*g w*g* gg gg* g*g*
Driving-Y	66% ww + 33% wg (male)	99.9% (vulnerable female) 95% (vulnerable female) 90% (vulnerable female)	$w^X w^Y$ $w^X w^X$ $w^X g^Y$ $w^X g*^Y$
	90% ww + 10% gg (male)	5% per g/g*	ww ww' ww* w'w' w'w* w*w* wg w'g w*g wg* w'g* w*g* gg gg* g*g*

Table 2. Gene drive methods, initial population scenarios, and genotype outcomes

III.2 Model description

We have used Malthusian population model as a growth model, and reaction model has been used to calculate the reproduction size for each genotype population which considers all possible interactions between genotypes that cause reproduction of a specific genotype.

$$[G_K]_{t+1} = [G_K]_t + \frac{d[G_K]_t}{dt} \tag{1.1}$$

s.t

$$\frac{d[G_k^M]_t}{dt} = \left(\sum_i 0.5 \times \lambda \times [G_i^F]_t \times \frac{\sum_j p_{ij}^k [G_j^M]_t}{[G_t^M]}\right) - \left(\mu_A \times [G_k^M]_t\right) \tag{1.2}$$

$$\frac{d[G_k^F]_t}{dt} = \left(\sum_i 0.5 \times \lambda \times [G_i^F]_t \times \frac{\sum_j p_{ij}^k [G_j^M]_t}{[G_t^M]}\right) - \left(\mu_A \times [G_k^F]_t\right) \tag{1.3}$$

$$\{i, j, k = ww, wg, wg*, gg, gg*, g*g*, ww*, ww', w*w*, w'w', w'w*, w'g, w'g*, w*g, w*g*$$

$[G^F]$ and $[G^M]$ denotes the number of unmated female and male mosquitoes with a specific genotype, $[G_t^M]$ is the total number of male mosquitoes at time t, p_{ij}^k is the probability that female mosquito i mates male mosquito j and reproduces mosquito k and, equations (1.2) and (1.3) are change of population of male and female mosquitoes with genotype k at time t, respectively.

We have assumed a homogeneous constant environment which means climatic conditions such as rainfall patterns, temperature, humidity, and seasonality are assumed to be constant. The interaction between humans and mosquitoes and other animals and mosquitoes are not included in our models.

We have studied self-elimination ability of anti-pathogen gene within two years in (1) inundative release of transgenic mosquitoes, (2) under-dominance method, (3) MEDEA, (4) driving-Y and, (5) CRISPR. In all models, we assume that initial population consists of unmated adult mosquitoes. At each generation, the model considers the number of unmated mosquitoes involved in the reproduction process. As we states we called egg, larvae and pupae stages as a juvenile stage and allocated 12 days to this juncture after which mosquitoes will emerge to the adult stage and get involved in the reproduction process. The total population of each genotype is considered as the summation of juvenile and adult mosquitoes. An average fixed daily mortality rate has been allocated for the total population. The next chapter provides results and finding from simulation of different gene drive systems.

CHAPTER IV

RESULTS

As stated earlier, we have studied self-elimination ability of an anti-pathogen gene within two years in five different gene drive systems. Self-elimination is successful when the total population can recover to the wild-type population within the studied time window.

For each mechanism, we have run the corresponding model and have provided the results in this chapter. In total, seventy-one simulations have been run for each level of α to provide the answers for the following question: (1) Can the engineered genotype be biodegradable? (2) What is the proper threshold for parameters to make a gene biodegradable?

All results have been explained based on plots. In each plot, X-axis shows the time in unit of days, and Y-axis indicates the proportion of wild-type population within two years. The "dotted" line in some plots indicates the state at which $\gamma = \alpha = 0$ which indicates there is no self-elimination, that is, the g allele is permanent.

IV.1 Results for inundative release

We start the results section with inundative release mechanism. Combination of 33% ww and 66% male gg mosquitoes has been considered as the first scenario of the initial population. Then we have tested two other scenarios as 100% wg and 100% gg for initial population, and for each scenario, we have provided the results considering a

combination of γ and fitness cost. The following subsections provide the results for this method.

IV.1.1 Initial population scenario: 67% ww + 33% gg (male)

For this initial population, first, we have considered $\gamma = 0$ which means that there is not any reproduction of g* allele. Also, we have not included fitness cost for g allele. We found that for $\alpha < 0.16$, the total population cannot recover to the wild-type which means that bio-degradability of the anti-pathogen gene is not achievable at these levels of α. However, for genes engineered for $\alpha \geq 0.16$, total population can recover to the wild type. Plot (a) in Figure 3 shows the results.

In the next step, we included 5% fitness cost in our model but still keeping $\gamma = 0$. The goal was to test the effect of fitness cost in the biodegradability of the anti-pathogen gene. It helped the total population to recover to the wild-type even in the absence of self-elimination probability. The result shows the importance of fitness cost. The reproduction and mortality rate of transgenic population has changed by only 5%, but it provides the confidence that even for $\alpha = 0$, the anti-pathogen gene can eliminate itself from the population within two years. The result is shown in plot (b) in Figure 3.

In the first two steps, we kept $\gamma = 0$. However, it is not realistic. In the real world γ is not zero and it is presumed to be between 1% and 5%. In this step, we have included $\gamma = 5\%$ and considered 5% fitness cost per g/g* allele. Plot (c) in Figure 3 shows the results. The "dotted" curve in the plot, belongs to $\alpha = \gamma = 0$. When γ is nonzero, α cannot be zero. Non-zero γ implies that there must have been self-elimination attempt which was not

successful and has caused a crashed form of g allele. Hence, for $\alpha = 0$ we have considered data corresponding to $\gamma = 0$.

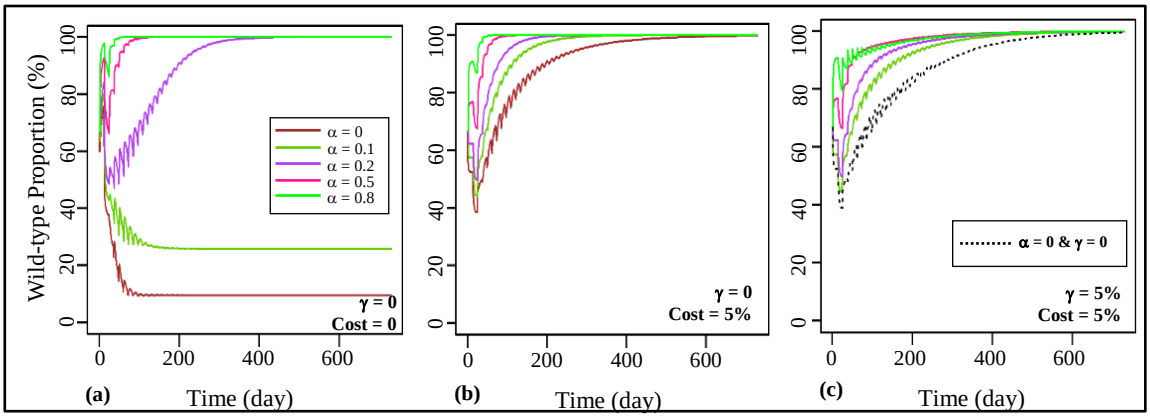

Figure 3. Inundative release, dynamic of wild-type population; Initial population = 67% ww + 33% gg (male)

IV.1.2 Initial population scenario: 100% wg

When we start with 100% wg as the original population, we expect to see all possible genotypes from the first generation. Likewise the first initial population scenario, we started the simulation of this population by keeping γ and fitness cost equal to zero. We found that total population can recover to the wild-type for all levels of α except $\alpha < 0.07$. Plot (a) in Figure 4 shows the results.

In the next step, we included the fitness cost. Likewise the first initial population, anti-pathogen gene can eliminate itself within two years even in the absence of self-elimination probability. The result is shown in plot (b) in Figure 4.

Increasing γ to 5%, makes delay in population recovery, however, likewise previous state, 5% fitness cost per g/g* allele helps the anti-pathogen gene remove itself from the population within two years. Plot (c) Figure 4 shows the results.

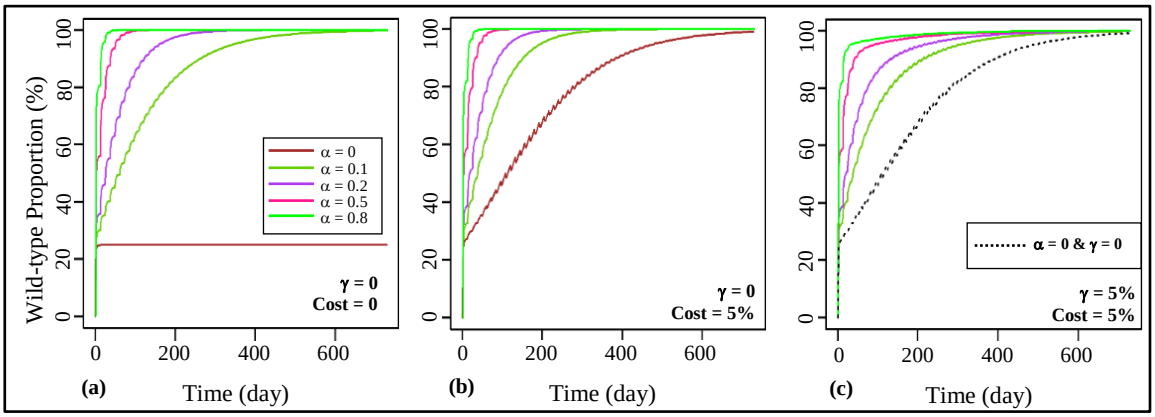

Figure 4. Inundative release, dynamic of wild-type population; Initial population = 100%wg

IV.1.3 Initial population scenario: 100% gg

We repeated the same processes when the initial population is 100% gg mosquitoes. For $\gamma = 0$ and f = 0, total population can recover to the wild-type for $\alpha \geq 0.08$ and it fails when $\alpha < 0.08$. When the self-elimination probability is zero, g allele cannot turn into w allele. Furthermore, the initial population does not include any wild-type mosquitoes. Hence, the proportion of wild-type is zero.

Including 5% fitness cost for g/g* does not make any difference for $\alpha = 0$ but if we compare the curves in the plot (a) with the equivalent curves in the plot (b) of Figure 5, we can see the improvement in the lower amount of α. The fitness cost helped the total

population to recover to the wild-type fast compared to state at which fitness cost is not included.

Then, we continued with a realistic form of this system. Beside f=5%, we included $\gamma = 5\%$. We found that anti-pathogen gene still can eliminate itself for $\alpha \geq 0.08$ but there is a delay in recovery time.

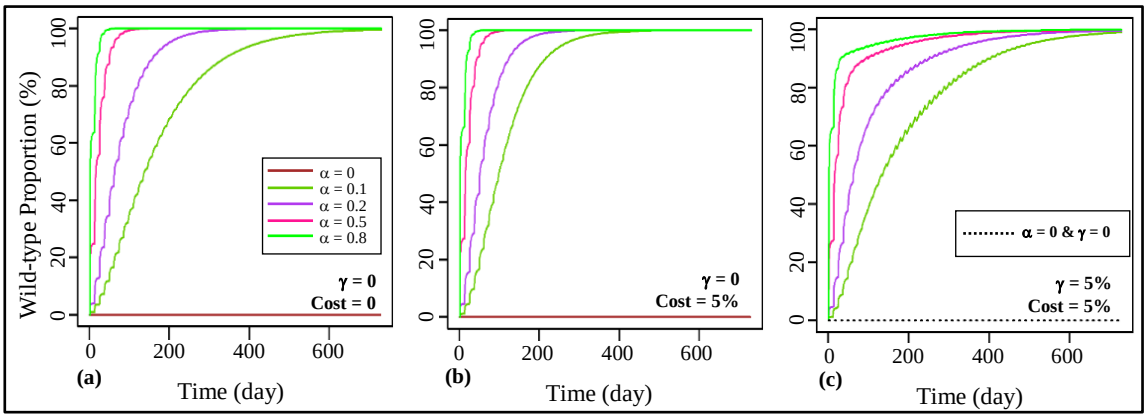

Figure 5. Inundative release, dynamic of wild-type population; Initial population = 100% gg

IV.2 Results for under-dominance method

This section provides the results for under-dominance method. We started with the initial population as 100% gg mosquitoes and tested three different fitness costs for heterozygotes genotypes considering $\gamma = 0$ and $\gamma = 5\%$. Then we tested two various combinations of ww and gg genotypes as initial population and studied the effect of corresponding fitness cost and γ in the recovery of the total population. The following sub-sections provide the results for this method.

IV.2.1 Initial population scenario: 100% gg

Starting with 100% gg mosquitoes causes all possible genotypes to appear from the first generation. Keeping $\gamma = 0$ prevents reproduction of g* allele, hence 100% heterozygotes fitness cost has been applied only for wg genotypes. We found that, self-elimination fails for $0 \leq \alpha < 0.18$ but for higher values of α, total population can recover to the wild-type within two years.

We simulated the same initial population scenario, changing heterozygotes fitness cost to 90% and 80% keeping $\gamma = 0$ and we got $\alpha \geq 0.16$ and $\alpha \geq 0.13$ respectively for population recovery. Plots (b) and (c) in Figure 6 show the results for f_U = 90% and 80% respectively.

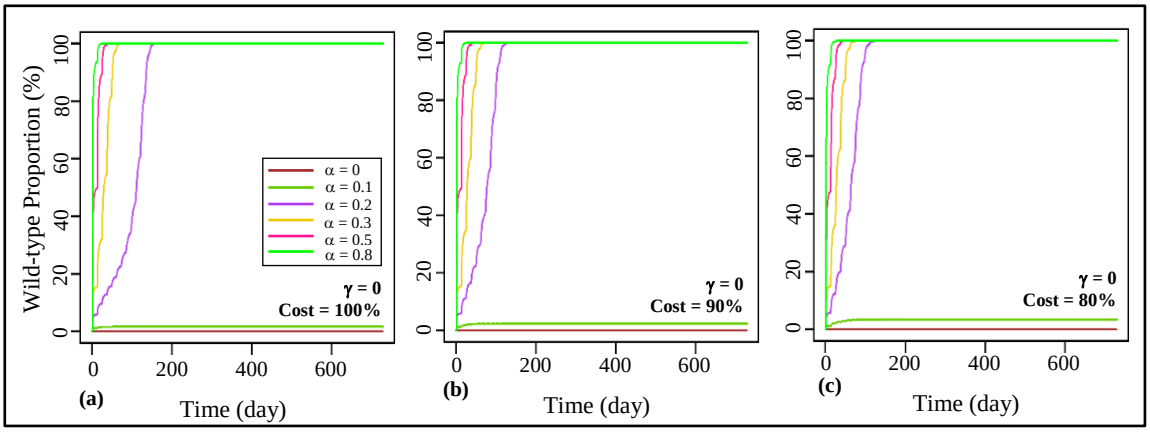

Figure 6. Under-dominance method, dynamic of wild-type population; Initial population=100% gg, $\gamma = 0$

In the next step, we increased γ to 5% and ran the same simulations. Plots (a), (b) and (c) in Figure 7, shows the results for f_U = 100%, 90% and 80% respectively when the

initial population is 100% gg mosquitoes. Since the γ is non-zero, the heterozygotes fitness cost applies for wg* as well. The results show that total population recover to the wild-type for $\alpha \geq 0.27$, $\alpha \geq 0.24$, and $\alpha \geq 0.22$ for 100%, 90% and 80% fitness costs respectively at which the lower limit of α has increased compared to state that γ was zero. It was expected, because γ introduce g* allele and it causes reproduction of three different non-wild-type genotypes and hence, the system needs higher α to eliminate anti-pathogen gene and return to the wild-type. Figure 7 summerizes the results.

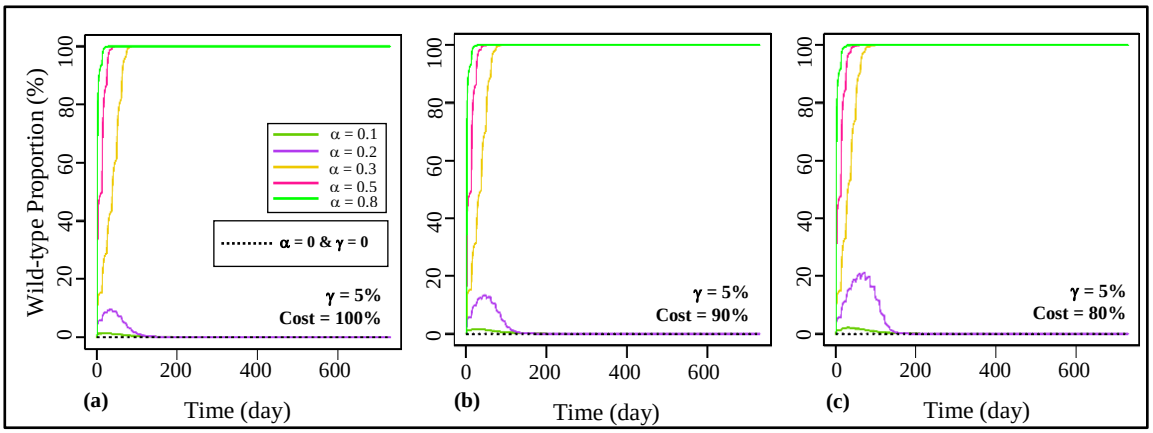

Figure 7. Under-dominance method, dynamic of wild-type population; Initial population= 100% gg, γ = 5%

IV.2.2 Initial population scenario: 25% ww + 75% gg

We modified the intial population and set it as 25% ww + 75% gg. Since the initial gg mosquitoes have been considered as male and female, hence, all possible genotypes appear from first generation. We tested the same fitness costs considering γ equal to zero and 5%. Plots (a), (b) and (c) in Figure 8 show the results for f_U = 100%, 90% and 80%

respectively considering $\gamma = 0$. We found that, self-elimination is achievable for $\alpha \geq 0.17$, $\alpha \geq 0.15$ and $\alpha \geq 0.14$ for 100%, 90% and 90% fitness costs respectively.

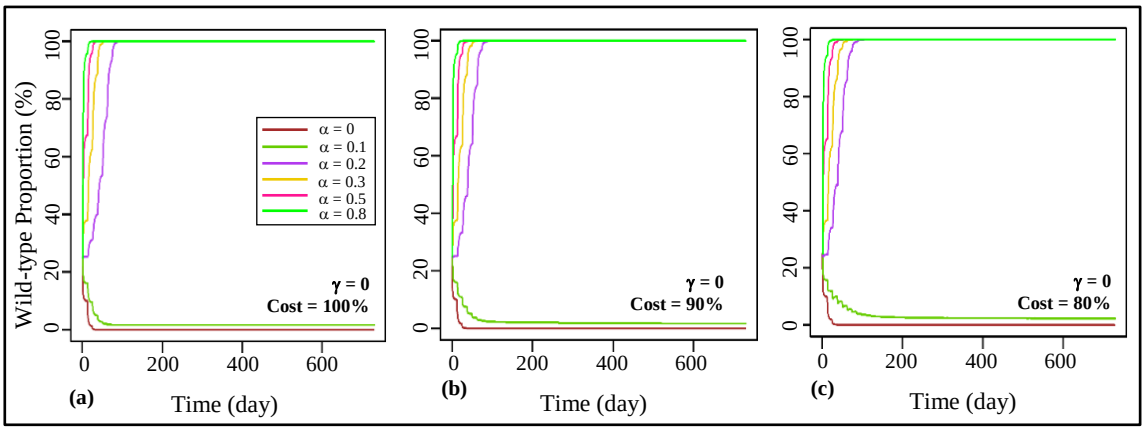

Figure 8. Under-dominance method, dynamic of wild-type population; Initial population= 25% ww + 75% gg, $\gamma = 0$

Plots (a), (b) and (c) in Figure 9 show the wild-type proportion for 100%, 90% and 80% heterozygotes fitness cost, respectively considering $\gamma = 5\%$ and same initial population. As result showed, anti-pathogen gene can eliminate itself when $\alpha \geq 0.2$, $\alpha \geq 0.18$, $\alpha \geq 0.16$ when fitness cost is 100%, 90% and 80% respectively within two years. Plots in Figure 9 show a time delay in recovery of total population to the wild-type.

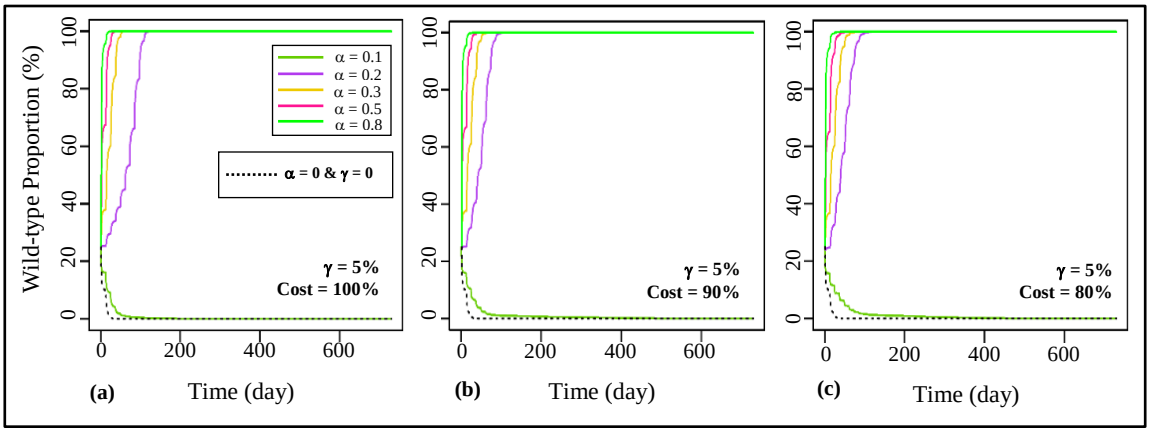

Figure 9. Under-dominance method, dynamic of wild-type population; Initial population= 25% ww + 75% gg, γ = 5%

IV.2.3 Initial population scenario: 40% ww + 60% gg

We repeated all simulation changing the initial population into 40% ww + 60% gg. All possible genotypes appear from the first generation because the system considers female and male gg mosquitoes in initial population scenario. First, we ran the model considering $\gamma = 0$ for three different heterozygotes fitness cost and the results are summarized in plots (a), (b) and (c) in Figure 10. We found that, total population can recover to the wild-type for $\alpha \geq 0.1$, $\alpha \geq 0.09$ and $\alpha \geq 0.08$ when fitness cost is 100%, 90% and 80% respectively.

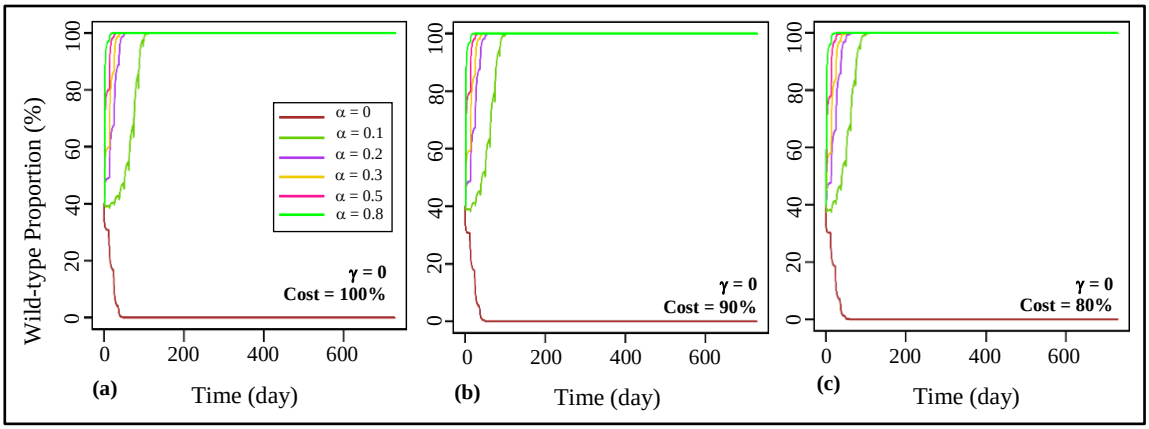

Figure 10. Under-dominance method, dynamic of wild-type population; Initial population= 40% ww + 60% gg, γ = 0

Including γ = 5% in model, does not change the level of α significantly at which anti-pathogen gene can eliminate itself from the population. It just made a delay in this process. Plots (a), (b) and (c) in Figure 11 show the wild-type proportion for 100%, 90% and 80% heterozygotes fitness cost respectively.

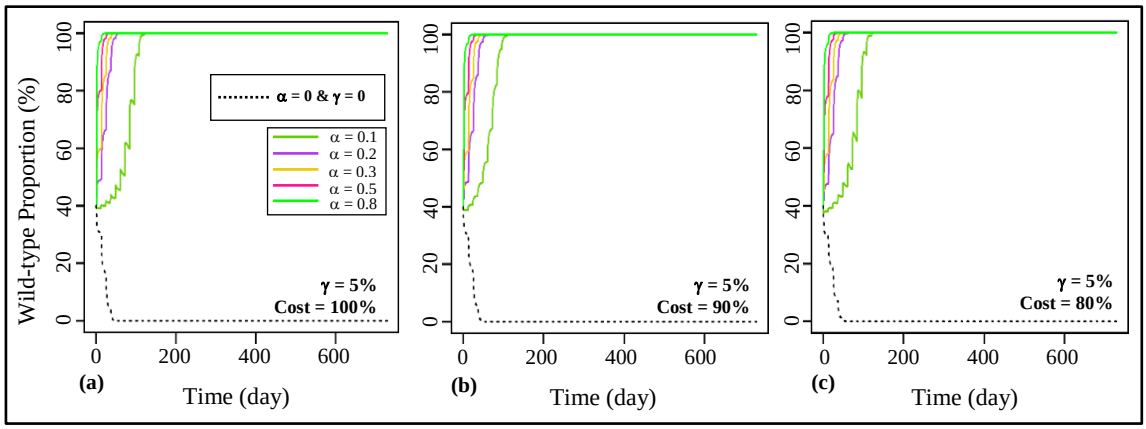

Figure 11. Under-dominance method, dynamic of wild-type population; Initial population= 40% ww + 60% gg, γ = 5%

Combination of 40% ww + 60% gg mosquitoes seems an ideal initial population scenario because it allows anti-pathogen gene to spread into the population and eventually eliminate itself within two years when $\alpha \geq 0.1$ for 100% and 90% fitness costs and $\alpha \geq 0.09$ for 80% fitness cost.

IV.3 Results for MEDEA (First approach)

This section provides the results for MEDEA method. As mentioned earlier, wild-type genotype offspring is vulnerable in this mechanism. It means that if the mother's genotype has at least one selfish allele, only the offspring which inherit at least one selfish allele survive and wild-type offspring may have f_M = 100%, 90% and 80% fitness cost.

In this system, we considered three different initial population as 100% gg, 25% ww + 75% gg and 50% ww + 50% gg mosquitoes. Unlike under-dominance method, the gg mosquitoes are only male in the second and third scenarios. For each initial population scenario, we tested three different fitness cost for vulnerable wild-type mosquitoes considering $\gamma = 0$ and 5%. In this system, in addition to wild-type fitness cost, additional 5% fitness cost has been applied for g/g* alleles.The following subsections provide the simulation results for each state.

IV.3.1 Initial population scenario: 100% gg

For this state, we modeled the MEDEA method with the initial population of 100% gg mosquitoes. Plots (a) and (b) in Figure 12 show the wild-type proportion when the wild-type fitness cost is 90% and 80% respectively. When the initial population is 100% gg, the only chance that we get wild-type genotype is when $\alpha > 0$, but on the other hand, we apply 100% to the wild-type mosquitoes which means that we do not have any ww

mosquitoes within two years. However, decreasing fitness cost to 90% and 80% gives 10% and 20% reproduction chance to the wild-type mosquitoes. This is the reason that for this state we have only two plots correspond to 90% and 80% fitness cost.

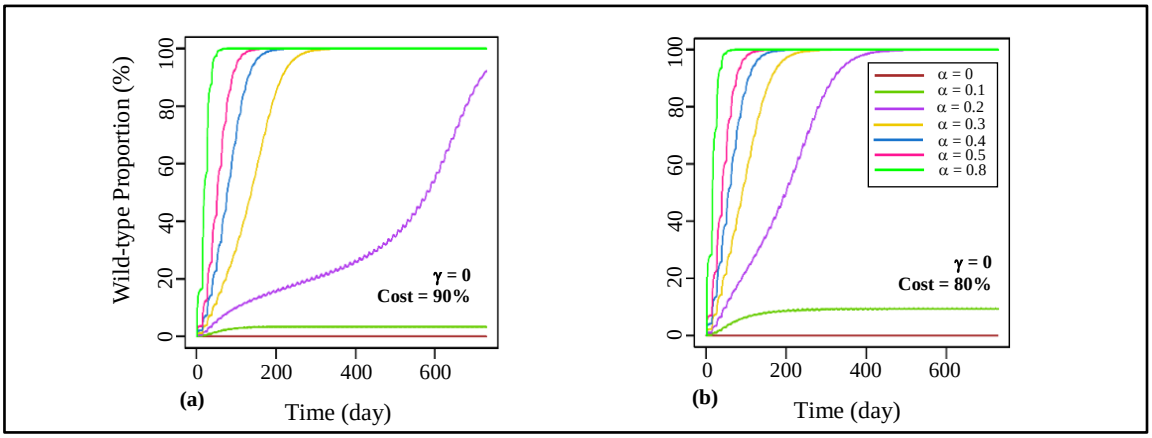

Figure 12. MEDEA-1 method, dynamic of wild-type population; Initial population= 100 % gg, $\gamma = 0$

The results showed that, when the fitness cost is 90%, total population can recover to the wild-type when $\alpha \geq 0.21$. The lower level of α decreases to the 0.16 when the fitness cost is 80%.

In the next step, we included $\gamma = 5\%$ and ran the model with same initial population and same fitness costs. The results are summarized in plots (a) and (b) in Figure 13. As soon as we increased the amount of γ, total population could not return to the wild-type even for the highest value of α. It was expected because there is a significantly high fitness cost applied to wild-type population. The mechanism kills the significant portion of wild-type. On the other hand, by introducing $\gamma = 5\%$, g* allele was introduced into a system that caused reproduction of wg*, gg* and gg* that not only increase the number of none-

wild-type mosquitoes in the population but kills more ww offspring. As a result, MEDEA mechanism with 100% gg mosquitoes as initial population, could not achieve biodegradability goal of anti-pathogen gene when $\gamma = 5\%$.

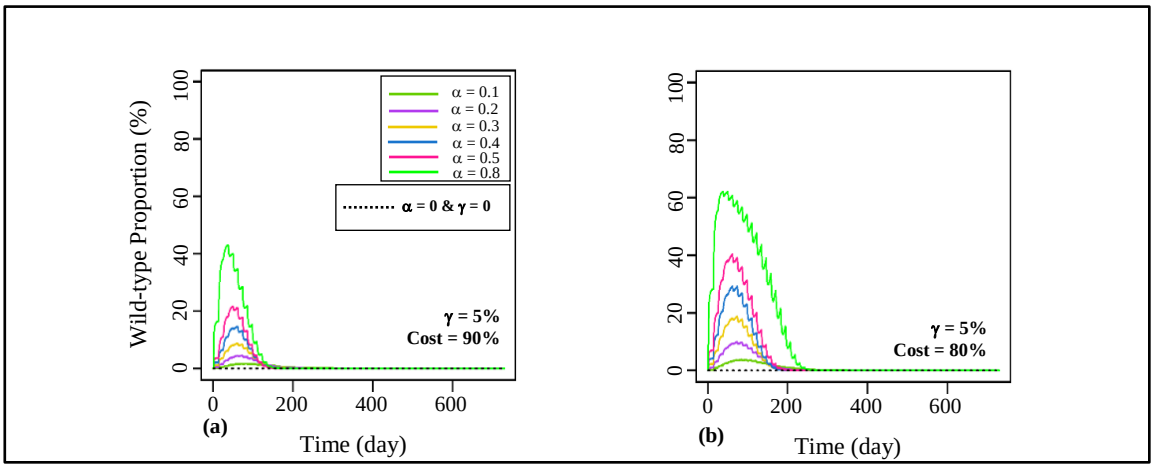

Figure 13. MEDEA-1 method, dynamic of wild-type population; Initial population = 100 % gg, $\gamma = 5\%$

IV.3.2 Initial population scenario: 25% ww + 75% gg (male)

We modified the initial population to 25% ww + 75% gg. Here the initial gg mosquitoes are only male. Hence, when $\gamma = 0$, ww and wg appear from the first generation, and reproduction of gg starts from the second generation. Plots (a), (b) and (c) in Figure 14 show the simulation results for 100%, 90% and 80% fitness cost respectively.

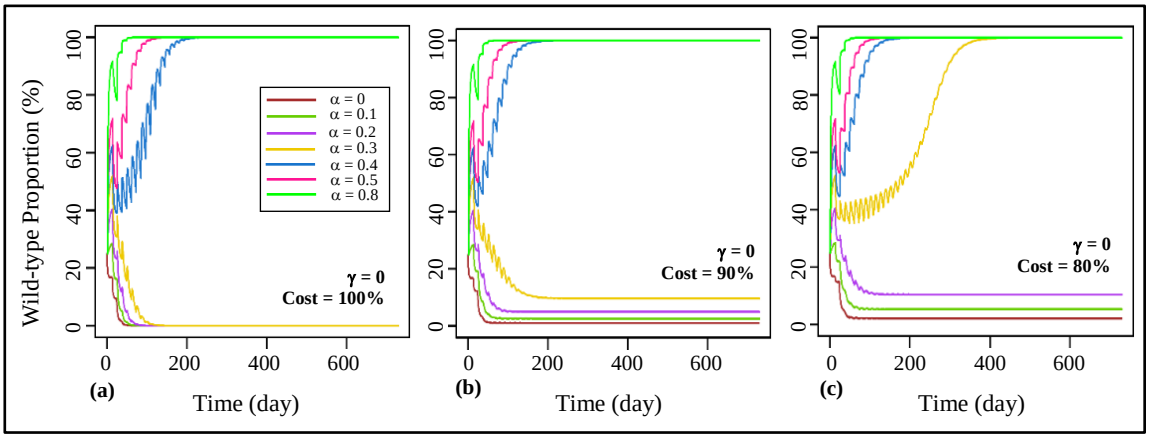

Figure 14. MEDEA-1 method, dynamic of wild-type population; Initial population = 25% ww + 75% gg, $\gamma = 0$

We found that, for 100%, 90% and 80% fitness costs, total population can recover to the wild-type when $\alpha \geq 0.38$, $\alpha \geq 0.34$ and $\alpha \geq 0.3$ respectively.

By including $\gamma = 5\%$, biodegradability is achievable when α is at least 0.8 and 0.6 for fitness equal to 100% and 90% respectively. By reducing the fitness cost to 80%, anti-pathogen gene can eliminate itself when $\alpha \geq 0.5$ but it is still concerning.

When the system can recover only when $\alpha \geq 0.8$, it means that the mechanism does not have enough time to spread the anti-pathogen gene. As soon as it is introduced into the population, it gets eliminated. While reducing the fitness cost reduced the lower level of α to 0.5, it still leaves a concern regarding biodegradability of the anti-pathogen gene. The lower level of self-elimination makes scientists confidence because it allows the anti-pathogen gene get spread through the population while gradually removes it from nature. Furthermore, scientists can revert the experiment if the mechanism does not work as expected. Figure 15 summerizes the results.

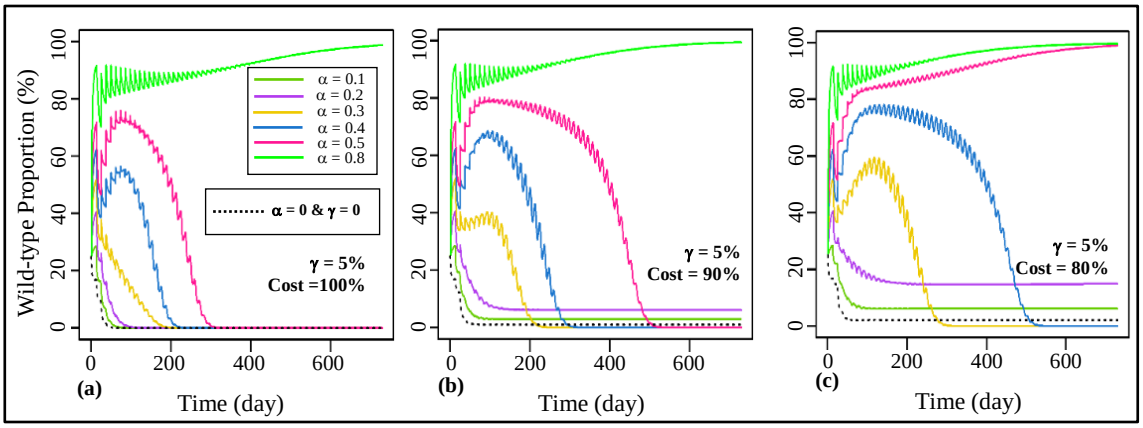

Figure 15. MEDEA-1 method, dynamic of wild-type population; Initial population = 25% ww + 75% gg, γ = 5%

IV.3.3 Initial population scenario: 50% ww + 50% gg (male)

Since two previous initial population scenarios did not provide the expectations, we modified the initial population to 50% ww + 50% gg (male) to check the level of α for biodegradability. Likewise previous states, we started with γ = 0. Plots (a), (b) and (c) in Figure 16 provides the results. We found that, when the wild-type fitness cost is 100% and 90%, total population can recover to the wild-type when $\alpha \geq 0.35$ and $\alpha \geq 0.31$ respectively. When the fitness cost reduced to 80% the lower level of α decreased to 0.28.

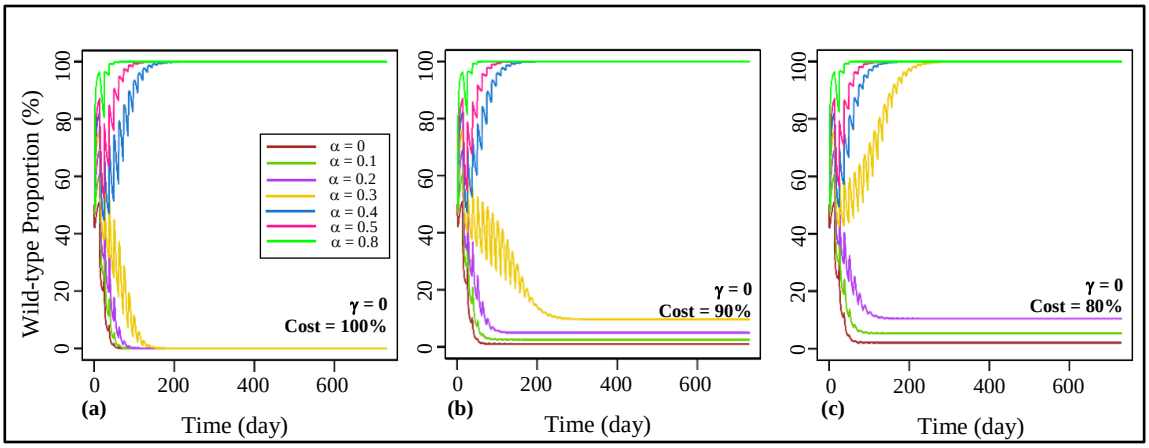

Figure 16. MEDEA-1 method, dynamic of wild-type population; Initial population = 50% ww + 50%gg, $\gamma = 0$

In the next step, we included $\gamma = 5\%$ in the model. Anti-pathogen gene could remove itself from the population for $\alpha \geq 0.5$ when the fitness cost is 100% and 90 %. The lower level of α decreases to 0.4 when the fitness cost is 80%. The results are summarized in plots (a), (b) and (c) in Figure 17. We mentioned earlier that realistic model includes non-zero γ. Hence, we evaluate the performance of the mechanism considering $\gamma = 5\%$.

Based on the results, the best possible outcome of this mechanism belongs to third initial population scenario, 50% ww + 50% gg (male) when the wild-type fitness cost is 80% and the corresponding interval for α is $\alpha \geq 0.4$.

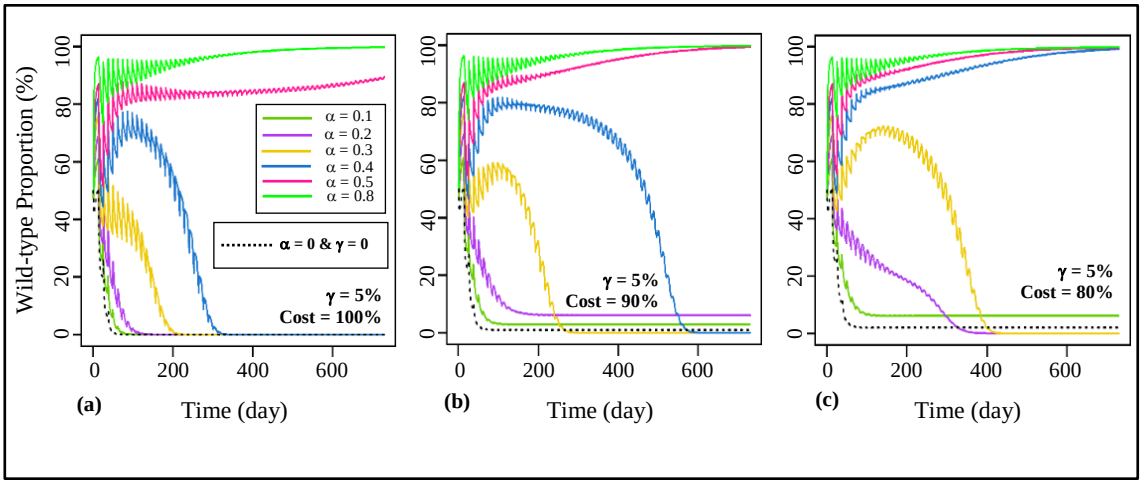

Figure 17. MEDEA-1 method, dynamic of wild-type population; Initial population = 50% ww + 50%gg, γ = 5%

As explained earlier, a model with a lower level of α provides more confidence in the sense of biodegradability and reversibility, and it is important to achieve the goal for various fitness cost and different initial population scenarios. In this approach of MEDEA, only one initial population scenario at a specific fitness cost relatively could meet the goal.

Insufficient results from the first approach of MEDEA from a self-elimination perspective, motivated us to improve the self-elimination mechanism for MEDEA to achieve biodegradability at a lower value of α. Hence the w* allele is introduced. The assumption is g allele can transform into w* allele with probability of α. While w* is considered as wild-type allele, since it is a transformation of the transgenic allele, g, it does not have MEDEA associated fitness cost. Hence, any wild-type genotype which at least has one w* allele will survive through MEDEA. The only vulnerable wild-type genotype will be ww which will be the offspring of ww male and ww female. In the next section, we model the second approach of MEDEA.

IV.4 Results for MEDEA (Second approach)

In this section, we have provided the results for the second method of MEDEA. Likewise the first approach we considered three different initial population scenario, and for each one, we tested three different wild-type fitness cost for $\gamma = 0$ and 5%. Additional 5% fitness cost per g/g* allele has been applied in the model.

IV.4.1 Initial population scenario: 100% gg

Here**,** we started with 100% gg as initial population scenario and keeping $\gamma = 0$, tried 100%, 90% and 80% wild-type fitness cost which is applied only for ww. Other wild-type genotypes, w*w* and ww* are not vulnerable since they have at least one w* allele. Plots (a), (b) and (c) in Figure 18 provides the results.

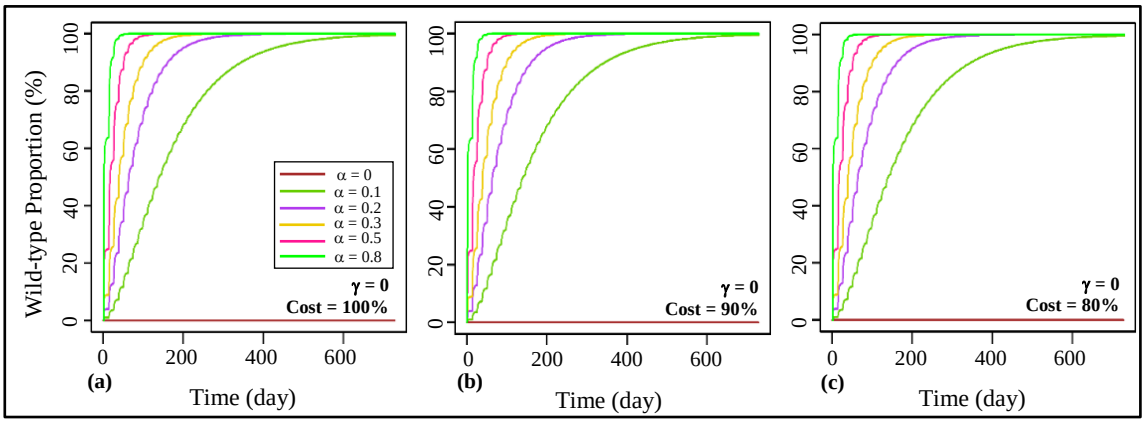

Figure 18. MEDEA-2 method, dynamic of wild-type population; Initial population = 100 % gg, $\gamma = 0$

In this case, since there are not any initial ww genotypes, all wild-type genotypes are w*w* and ww* which do not have ant fitness cost. We found that, total population can recover to the wild-type for $\alpha \geq 0.02$ for each fitness cost.

Moving to the next step, we increased γ to 5% and repeated the simulations. As we see in plots in Figure 19, we got the results similar to the state at which γ = 0. The only difference is a time delay in population recovery. This is the advantage of the second approach of MEDEA.

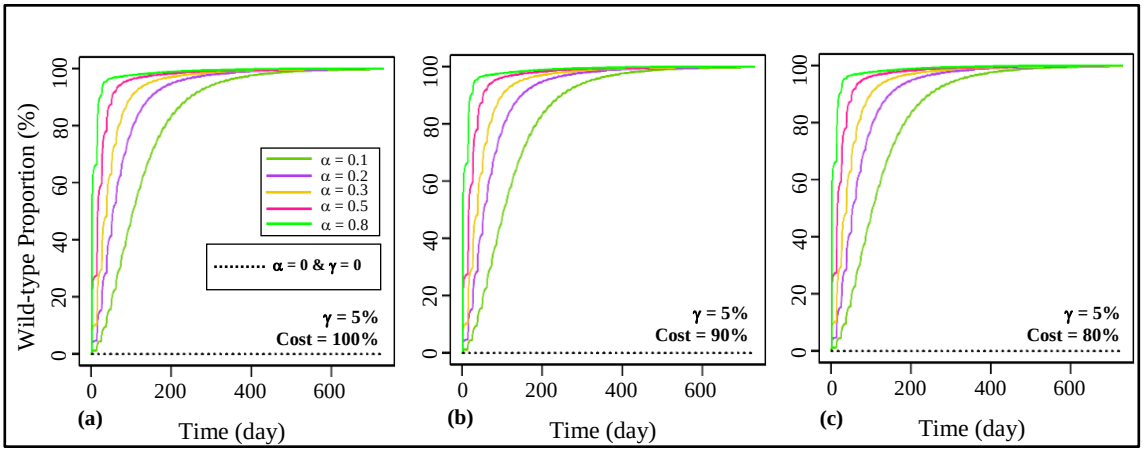

Figure 19. MEDEA-2 method, dynamic of wild-type population; Initial population = 100 % gg, γ = 5%

In the first approach of MEDEA, when γ was zero and initial population was only 100% gg, population could not return to the wild-type for f_M=100%. For the other two fitness costs, it could recover when α was 0.21 and 0.16 respectively. When we increased γ to 5%, the result got worst and total population could not revert to wild-type for any value of α. Hence, in comparison to the first approach of MEDEA, the result of second approach of MEDEA for 100% gg initial population has significantly improved and total population can recover to the wild-type for α ≥ 0.04.

IV.4.2 Initial population scenario: 25% ww + 75% gg (male)

We changed the initial population to 25% ww + 75% gg, but only male mosquitoes are considered as gg initial population. By keeping γ = 0, we ran the model for different fitness costs. Plots (a), (b) and (c) in Figure 20 show the simulation results for f_M =100%, 90% and 80% respectively. We found the same values for α as the first initial population scenario. We found that for α ≥ 0.02, anti-pathogen gene can eliminate itself from population for all values of fitness costs.

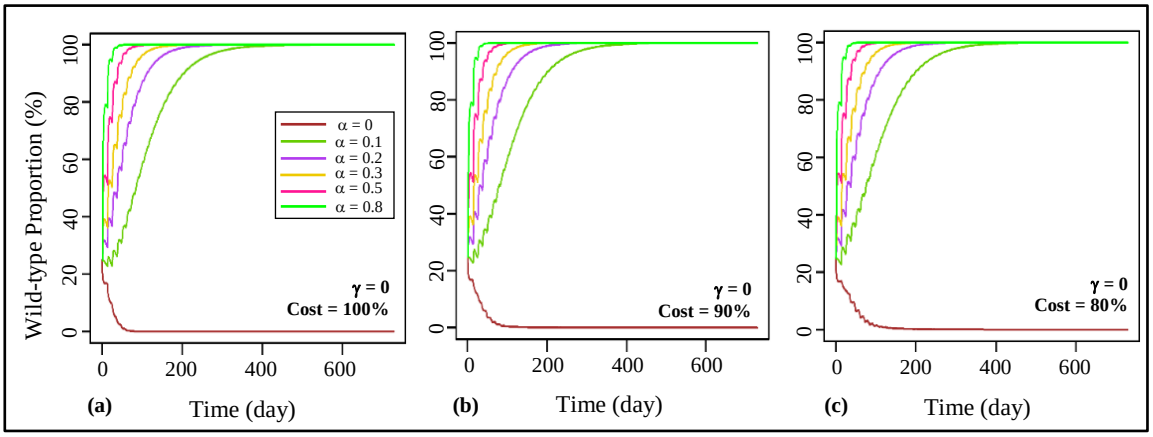

Figure 20. MEDEA-2 method, dynamic of wild-type population; Initial population = 25% ww + 75% gg, γ = 0

By increasing γ to 5% we found that for α ≥ 0.03 total population can recover to the wild-type. Furthermore, there is not a significant difference between three plot. It implies that reducing fitness cost from 100% to 90% and 80% do not affect the wild-type proportion. Because only a small portion of wild-type population is ww, and a major part of that belongs to ww* and w*w* which do not have a fitness cost, hence reducing the fitness cost, does not increase the portion ww significantly.

Moving to the realistic simulation, we increased γ to 5% and repeated the model with three different fitness costs. The results are summarized in Figure 21.

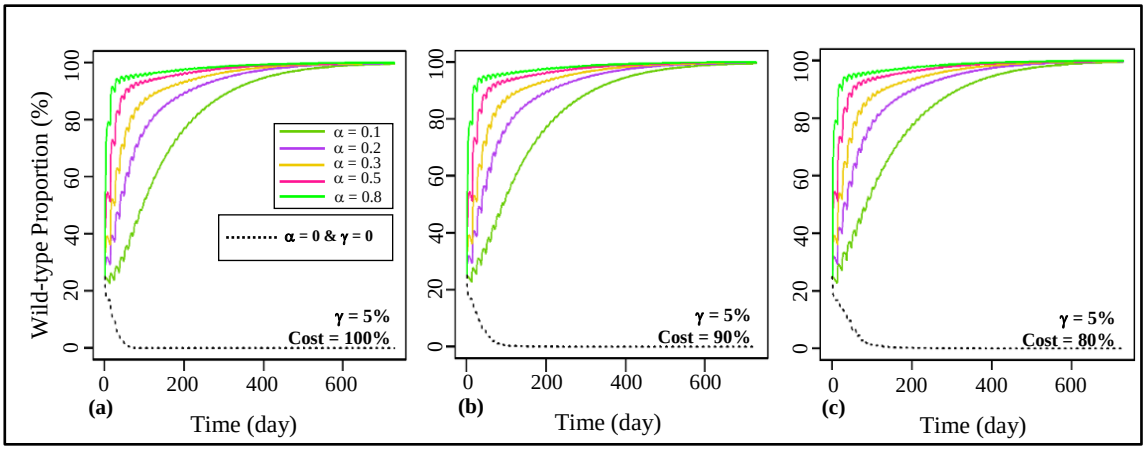

Figure 21. MEDEA-2 method, dynamic of wild-type population; Initial population = 25% ww + 75% gg, γ = 5%

IV.4.3 Initial population scenario: 50% ww + 50% gg (male)

The mechanism of the second approach of MEDEA provided the requirements of biodegradability based on two initial population scenarios. In this step, we modified the initial population to 50% ww + 50% gg (male) to check whether the mechanism still works for this scenario. As previously stated, we started with γ = 0 and tried three fitness costs. We got the same results. Anti-pathogen gene could eliminate itself from the population for $\alpha \geq 0.02$. The results are summarized in Figure 22.

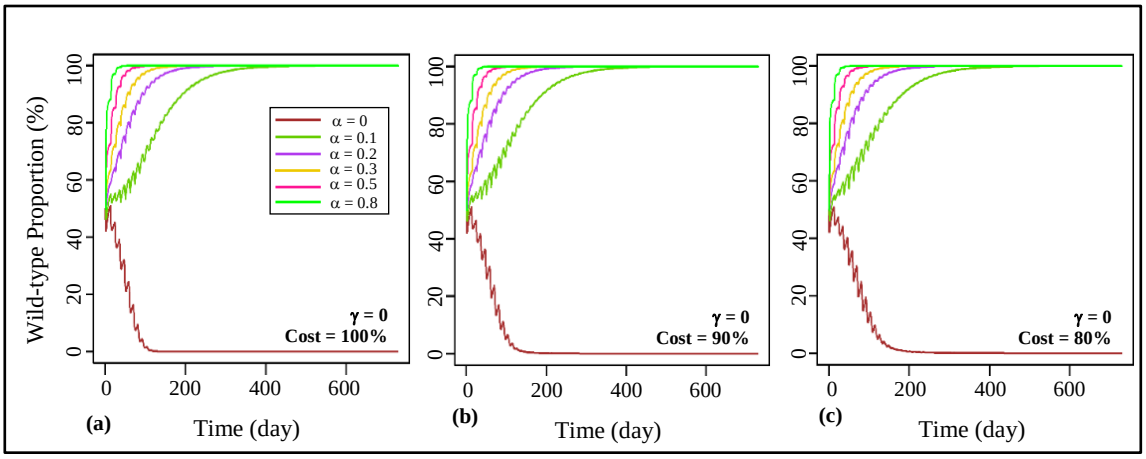

Figure 22. MEDEA-2 method, dynamic of wild-type population; Initial population = 50% ww + 50% gg, γ = 0

Increasing γ to 5% made a delay in recovery time of total population, but it did not change the values of α for different fitness costs. We found that anti-pathogen gene could eliminate itself from the population for α ≥ 0.03. Plots (a), (b) and (c) in Figure 23 show the results for f_M = 100%, 90% and 80% respectively. Likewise second scenario of initial population, there is not a significant difference between three plots that implies that fitness cost does not have a significant effect in portion of ww in wild-type population.

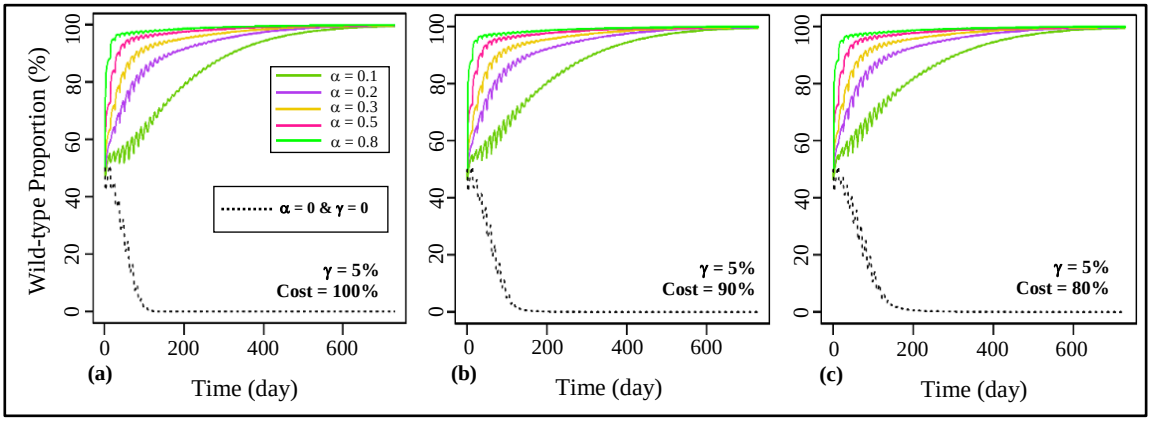

Figure 23. MEDEA-2 method, dynamic of wild-type population; Initial population = 50% ww + 50% gg, γ = 5%

IV.5 Results for driving-Y

In this section, we provided the results for driving-Y method. As we mentioned earlier, in this mechanism, anti-pathogen gene must be linked to the Y-chromosome. It means that we have g^Y allele but not g^X hence the initial population can not be $g^X g^Y$. Furthermore g allele can not transfer to X-chromosome during the reproduction process. In addition, this mechanism causes sex-bias in population which means that if the father carries at least one selfish allele, which is g^Y allele is our study, his offspring will inherit it with a significant high probability and hence, will be male mosquitoes which is equivalent to high fitness cost for female offspring.

For this mechanism, we considered 67% ww + 33% wg (male) as initial population scenario. The genotype constructs of wild-type can be $w^X w^Y$ and $w^X w^X$ but for transgenic mosquitoes it must, be $w^X g^Y$. For this scenario we included f_H = 99.9%, 95% and 90% as female fitness cost. In addition to f_H, we included 5% fitness cost per g/g* allele. We ran the simulation with $\gamma = 0$ and $\gamma = 5\%$. Plots (a), (b) and (c) in Figure 24 show the results for 99.9%, 95% and 90% fitness cost respectively when $\gamma = 0$.

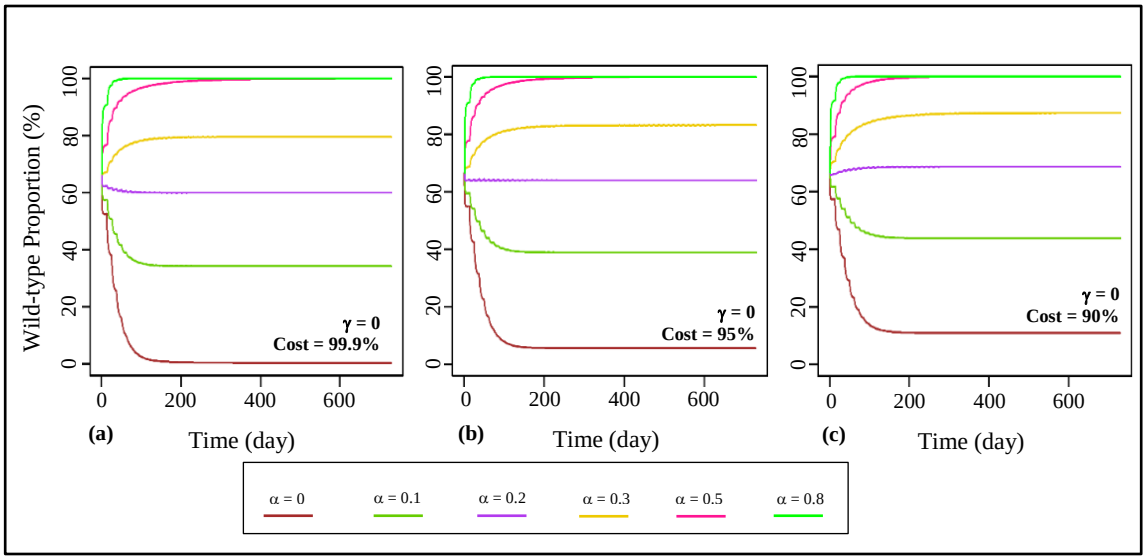

Figure 24. Driving-Y method, dynamic of wild-type population; Initial population = 67% ww + 33% wg, $\gamma = 0$

Total population can return to the wild-type when $\alpha \geq 0.47$, $\alpha \geq 0.45$ and $\alpha \geq 0.42$ for 99.9%, 95% and 90% fitness cost respectively. Another difference between plots is the time of recovery. Reducing the fitness cost, decrease the time of recovery. While the goal of biodegradability is achieved, however, it does not happen at lower levels of α. Hence we expected the worst results by increasing γ.

In next step, we increased γ to 5% to test our expectation. In this case, total population cannot recover to the wild-type and hence, the selfish gene spread through population for all three fitness costs. This mechanism not only causes the spread of anti-pathogen gene but spreads the Y-chromosome which increases the reproduction of male offspring. This process will reduce the overall replication in the system due to a decrease in the number of females. Hence, this gene drive method causes eradication of total population. Figure 25 shows dynamic of wild-type population in this system.

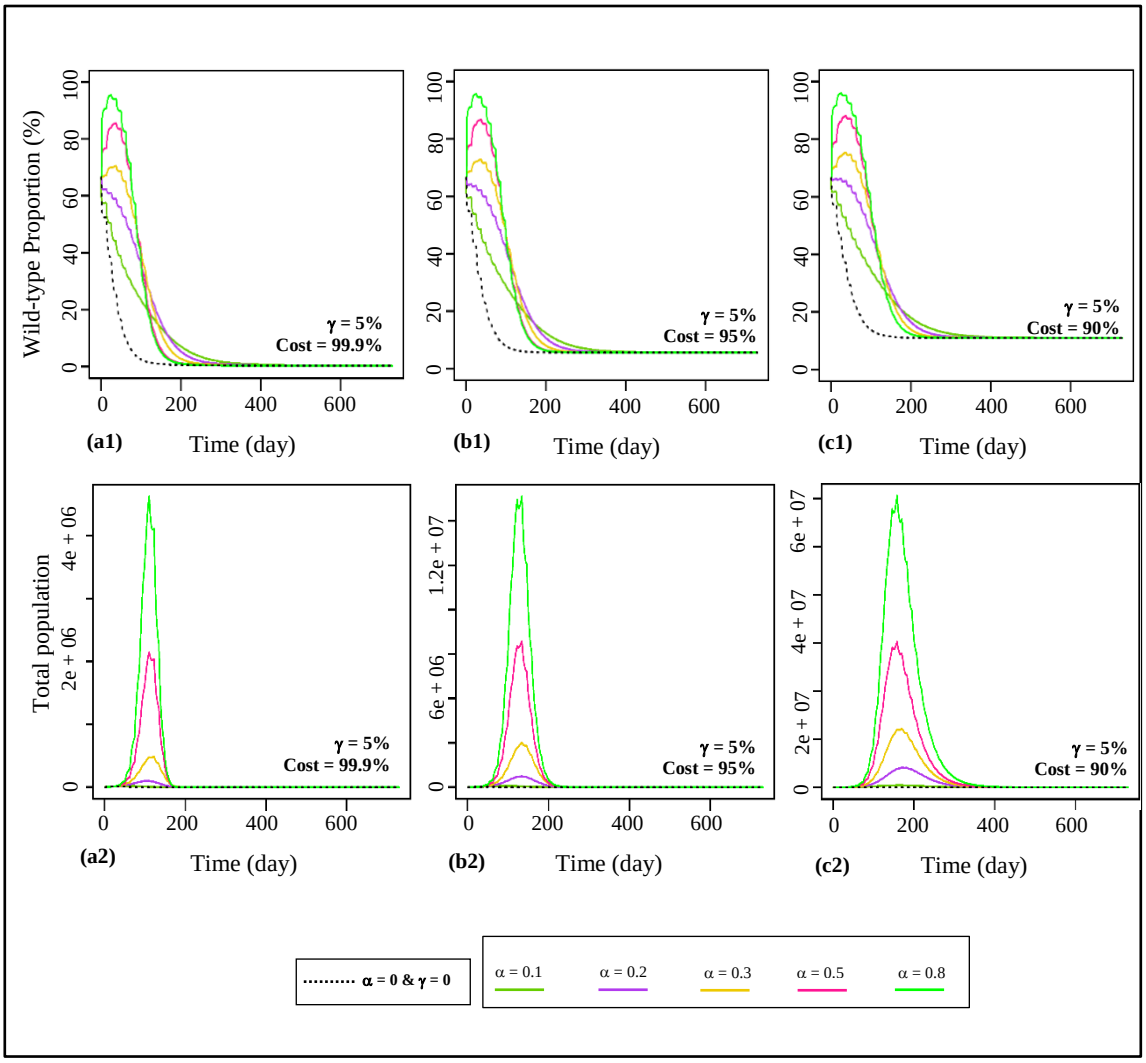

Figure 25. Driving-Y method, dynamic of wild-type population; Initial population = 67% ww + 33% wg, γ = 5%

IV.6 Results for CRISPR

This section provides the results for CRISPR method. We explained earlier that the goal of CRISPR mechanism is a successful cleavage of w allele by enzyme *Cas9* and successful repair of it by gRNA through copying the g allele at the cut point. Table 1 shows the successful cleavage and successful repair probabilities as 0.995 and 0.99 respectively. Hence, this mechanism allows the spread of anti-pathogen gene through

population, and self-elimination probability eliminates it gradually within two years. The only applied fitness is 5% per g/g* allele. For this mechanism we set the initial population to 90% ww + 10% gg (male) and ran the model for $\gamma = 0$ and 5%. Figure 26 shows the results for $\gamma = 0$.

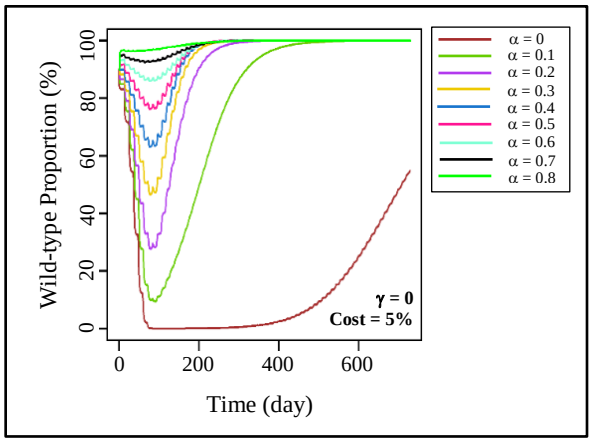

Figure 26. CRISPR method, dynamic of wild-type population; Initial population = 90% ww + 10% gg, $\gamma = 0$

As plot shows, this mechanism spreads the anti-pathogen gene through the population. Hence wild-type population starts decreasing but then after a while total population starts recovering to the wild-type depend on the value of α. Even for $\alpha = 0$, the plot shows the increase in the proportion of wild-type population. However, complete recovery may take more than two years which is out of the goal of our study. We found that for $\alpha \geq 0.05$ total population can recover to the wild-type population. Based on the represented results in Figure 24, we expected to see the same performance by increasing γ to %5. To test our expectation, we repeated the simulation with $\gamma = 5\%$. Figure 27 shows

the results, and it verifies our expectations. We found that, total population could revert to wild-type for $\alpha \geq 0.02$.

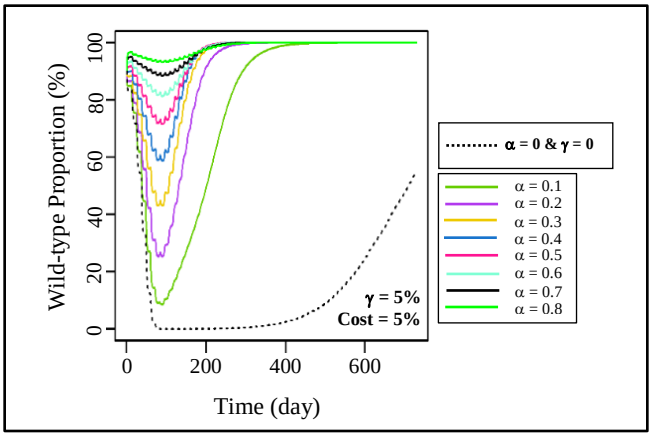

Figure 27. CRISPR method, dynamic of wild-type population; Initial population = 90% ww + 10% gg, $\gamma = 5\%$

IV.7 Validation of models

As mentioned earlier, modeling an engineered biodegradable anti-pathogen gene is a novel idea that to the best of our knowledge, there is not any valid published study about that. Future laboratory-based experiments will be developed to verify the validity of the result of the current study. However, there are previous studies related to the application of gene drive methods to spread an anti-pathogen gene in the wild-type population that was explained in chapter II. Those studies are equivalent to the state of our models at which α and γ is equal to zero. When α is equal to zero, it means that there is no biodegradability and when γ is equal to zero it means that there is not crashed form of transgenic allele because there is not self-elimination attempt. In inundative release

mechanism, the equivalent state belongs to the status at which in addition to α and γ, fitness cost per g/g* is also zero.

The study of Okamoto et al.[21] showed that releasing transgenic mosquitoes into wild-type population reduces the wild-type population so that after a while the number of wild-type population reaches the equilibrium state and stays on that at which the significant portion of total population contains transgenic mosquitoes and our modeling results based on three different initial population matched with this result.

For under-dominance mechanism, the study of Edgington et al. [21] showed that single release of transgenic mosquitoes would cause them to stay in population and only small portion of the population will contain wild-type mosquitoes. The results of our modeling based on three different initial population scenarios and three different heterozygotes fitness costs showed that when α and γ are zero, anti-pathogen gene spreads through the population and causes a significant reduction in wild-type proportion.

Wade et al. [24] studied the mechanism of MEDEA and found that in the absence of any fecundity fitness cost, MEDEA spreads the new genetic structure to fixation through the population. We studied two approaches of MEDEA, and for each one, we tested three different initial population scenarios. We got our results matched with the referenced study, and we showed that in the case of zero α and γ, anti-pathogen gene spread through the population and an only negligible portion of the total population belongs to wild-type mosquitoes.

For driving-Y mechanism, Galizi et al. [23] ran a laboratory-based experiment to show that driving-Y mechanism causes population suppression by releasing male

transgenic mosquitoes at with Y-chromosome plays selfish-gene role and their study verified population eradication. We simulated a population at which the initial population was the combination of 67% ww + 33% wg (male) and assumed that g allele in wg genotype is linked to Y-chromosome and we showed that not only driving-Y causes the spread of anti-pathogen gene in population but within two years causes population suppression.

For CRISPR mechanism, the proposed method by Esvelt et al. [15] was not modeled or conducted in the laboratory [24]. However, our simulation results matched with the expected outcome by the explained mechanism is the study by Esvelt et al. [15]. Future laboratory-based experiments will test the validity of our simulations.

CHAPTER V

DISCUSSION AND CONCLUSION

This chapter provides a summary of the development of the model, and the results obtained from the simulation. In the last part, potential areas of future work are presented.

V.1 Discussion

Gene drive systems are great methods of altering the odds; promising tools that give scientists the power to spread a new genetic construct in a target population or eliminate a harmful species. In last few decades, gene drive systems have been considered as a method to prevent the spread of mosquito-borne vector disease. However, due to uncertainties and unpredictable risks, none has been conducted in nature.

In this study, we introduced the engineered biodegradable anti-pathogen gene. The aim was to simulate mechanisms of different gene drive systems by which the anti-pathogen gene can spread through the population and then eliminate itself within two years. The success of this idea in simulation and laboratory-based experiments can help to pave the way of implementing the gene drive methods in the real world.

For this purpose, we build a mathematical model for different gene drive mechanisms. We started with inundative release of mosquitoes and forwarded to under-dominance, MEDEA, driving-Y and CRISPR methods. Modeling results showed the success of inundative release, under-dominance, and CRISPR mechanisms. However, we found that driving-Y mechanism suppresses the total population even at the high value of self-elimination probability due to the high female fitness cost. Furthermore, model results

showed that as long as we allocate MEDEA-related fitness cost to non-MEDEA bearing offspring, this mechanism will fail to achieve the goal of biodegradability. Only for a specific initial population scenario and at a relatively high value of self-elimination probability, total population can recover to the wild-type. However, we modify the mechanism of MEDEA so that wild-type allele, w*, which is transformed form for g allele, survives and makes a genotype with at least one w* allele survive. In this case, modeling results showed the success of this method regarding biodegradability of anti-pathogen gene based on different initial population scenarios and different MEDEA-related fitness costs.

V.2 Future works

Utilizing gene drive methods to combat vector-borne disease is a new domain that still requires experiments and studies to answer to several questions. Our study is not an exception, and we identified two ideas to expand the current study. The identified ideas are related to CRISPR method. First one is related to resistant alleles in CRISPR system that we indicated as w'. As we explained, resistant allele, w', is crashed form of w allele at which *Cas9* enzyme has tried to cleave it to copy the g allele, but it could not succeed. Hence, w is cleaved, but cannot get repaired. In this case w' is still a wild-type allele. However, it cannot be utilized to spread g allele anymore. Presence of resistance allele impacts the effectiveness of CRISPR mechanism regarding the spread of anti-pathogen gene. Hence, a fitness cost can be allocated to resistance alleles. However, it needs to be well-studied to identify a desired fitness cost for heterozygotes and homozygous resistant genotypes. The second idea is an answer to this question: How far mosquitoes can forward

in CRISPR mechanism before completely being removed from the population? This idea requires a spatial modeling which is a combination of continues and discrete spatial model and can provide an answer to that question.

CHAPTER VI

CONTRIBUTIONS

The thesis research made a significant contribution to the application of gene drive systems modeling to remove the vector-borne disease.

Previous research has sought to understand the mechanism of gene drive systems to replace the vulnerable wild-type population with the transgenic pathogen-resistance population. Our study offers an application of biodegradable anti-pathogen gene to replace a vulnerable wild-type population with wild-type pathogen-resistance population by modeling the spread of an engineered biodegradable anti-pathogen gene through different gene drive methods. Our research reported that biodegradability goal is achievable and can be utilized through gene drive systems such as inundative release mechanism, under-dominance, a modified form of MEDEA, and CRISPR methods. However, it is not applicable through driving-Y since it completely collapses total population, and CRISPR method needs more study and improvement regarding the presence of resistance alleles. To the best of our knowledge, there are no other studies on the self-elimination property of anti-pathogen gene. Thus, the implications of our model results provide valuable and insightful information for the development of engineered biodegradable anti-pathogen gene.

REFERENCES

[1] “Mosquito-Borne Diseases,” *Home*. [Online]. Available: https://amca.memberclicks.net/mosquito-borne-diseases. [Accessed: 01-Oct-2017].

[2] “The National Institute for Occupational Safety and Health (NIOSH),” *Centers for Disease Control and Prevention*, 22-Mar-2016. [Online]. Available: https://www.cdc.gov/niosh/topics/outdoor/mosquito-borne/other.html. [Accessed: 01-Oct-2017].

[3] J. Lyons, “Mosquito Borne Diseases,” *deBugged*, 31-Mar-2017. [Online]. Available: https://www.rentokil.com/blog/mosquito-borne-diseases. [Accessed: 01-Oct-2017].

[4] “Malaria,” *UNICEF DATA*. [Online]. Available: https://data.unicef.org/topic/child-health/malaria/. [Accessed: 01-Oct-2017].

[5] “10 Facts on Malaria,” *World Health Organization*. [Online]. Available: http://www.who.int/features/factfiles/malaria/en/. [Accessed: 20-Aug-2017].

[6] “Fact Sheet About Malaria,” *World Health Organization*. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs094/en/. [Accessed: 01-Oct-2017].

[7] “Mosquito-Borne Diseases,” *World Health Organization*. [Online]. Available: http://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/. [Accessed: 01-Oct-2017].

[8] “Gene Drive Overdrive,” *Nature Biotechnology*, vol. 33, no. 10, pp. 1019–1021, Aug. 2015.

[9] C. Boëte, F. Agusto, and R. Reeves, “Impact of Mating Behaviour on The Success of Malaria Control Through a Single Inundative Release of Transgenic Mosquitoes,” *Journal of Theoretical Biology*, vol. 347, pp. 33–43, 2014.

[10] Z. N. Adelman, *Genetic Control of Malaria and Dengue*. London Wall: Academic Press is an imprint of Elsevier, 2016.

[11] K. Magori, “Genetically Engineered Underdominance for Manipulation of Pest Populations: A Deterministic Model,” *Genetics*, vol. 172, no. 4, pp. 2613–2620, 2005.

[12] S. Davis, N. Bax, and P. Grewe, “Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations,” *Journal of Theoretical Biology*, vol. 212, no. 1, pp. 83–98, 2001.

[13] P. A. Eckhoff, E. A. Wenger, H. C. J. Godfray, and A. Burt, “Impact of Mosquito Gene Drive on Malaria Elimination in a Computational Model With Explicit Spatial and Temporal Dynamics,” *Proceedings of The National Academy of Sciences*, vol. 114, no. 2, pp. 255–267, 2016.

[14] A. Burt, “Site-Specific Selfish Genes as Tools for The Control and Genetic Engineering of Natural Populations,” *Proceedings of The Royal Society B: Biological Sciences*, vol. 270, no. 1518, pp. 921–928, Jul. 2003.

[15] T. H. Saey, “Explainer: How CRISPR Works,” Science News for Students, 31-Jul-2017. [Online]. Available: ttps://www.sciencenewsforstudents.org/article/explainer-how-crispr-works. [Accessed: 01-Oct-2017].

[16] "What is CRISPR-Cas9?" *Facts*, 19-Dec-2016. [Online]. Available: https://www.yourgenome.org/facts/what-is-crispr-cas9. [Accessed: 01-Oct-2017].

[17] C. F. Curtis, "Possible Use of Translocations to Fix Desirable Genes in Insect Pest Populations," *Nature*, vol. 218, no. 5139, pp. 368–369, 1968.

[18] A. Hieronymi, "Understanding Systems Science: A Visual and Integrative Approach," *Systems Research and Behavioral Science*, vol. 30, no. 5, pp. 580–595, 2013.

[19] K. W. Okamoto, M. A. Robert, F. Gould, and A. L. Lloyd, "Feasible Introgression of an Anti-pathogen Transgene into an Urban Mosquito Population Without Using Gene-Drive," *PLoS Neglected Tropical Diseases*, vol. 8, no. 7, Mar. 2014.

[20] M. P. Edgington and L. S. Alphey, "Conditions for Success of Engineered Underdominance Gene Drive Systems," *Journal of Theoretical Biology*, vol. 430, pp. 128–140, 2017.

[21] M. J. Wade and R. W. Beeman, "The Population Dynamics of Maternal-Effect Selfish Genes" *Genetics*, vol. 138, pp. 1309–1314, 1994.

[22] C.H. Chen, H. Huang, C. M. Ward, J. T. Su, L. V. Schaeffer, M. Guo, B. A. Hay, "A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in Drosophila," *Science*, vol. 316, no. 5824, pp. 597-600, 2007

[23] R. Galizi, L. A. Doyle, M. Menichelli, F. Bernardini, A. Deredec, A. Burt, B. L. Stoddard, N. Windbichler, and A. Crisanti, "A Synthetic Sex Ratio Distortion System for the Control of the Human Malaria Mosquito," *Nature Communications*, vol. 5, Oct. 2014.

[24] C. Noble, J. Olejarz, K. M. Esvelt, G. M. Church, and M. A. Nowak, “Evolutionary Dynamics of CRISPR Gene Drives,” *Science Advances*, vol. 3, no. 4, 2017.