UNIVERSITY OF CALIFORNIA

Los Angeles

The Dynamics of Transposable Elements
in Genetically Modified Mosquito Vectors

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biomathematics

by

John Macky Marshall

2008
The dissertation of John Macky Marshall is approved.

_____________________________
Ken Lange

_____________________________
Janet S. Sinsheimer

_____________________________
Charles E. Taylor, Committee Co-chair

_____________________________
Tom Chou, Committee Co-chair

University of California, Los Angeles

2008
# TABLE OF CONTENTS

LIST OF FIGURES ......................................................................................................... vi

LIST OF TABLES ........................................................................................................ viii

ACKNOWLEDGMENTS .............................................................................................. ix

VITA ................................................................................................................................ xiii

ABSTRACT OF THE DISSERTATION .................................................................... xv

1 CHAPTER 1 .............................................................................................................. 1
   1.1 Objectives and hypotheses ............................................................................ 1
      1.1.1 General hypotheses ............................................................................. 6
      1.1.2 Hypotheses examined in each chapter ................................................. 6
   1.2 Biology of transposable elements .............................................................. 7
      1.2.1 Regulation of transposition rate with copy number ......................... 10
      1.2.2 Fitness costs associated with transposable elements ...................... 11
      1.2.3 Internal deletion within transposable elements ............................... 13
   1.3 Ecology of *Anopheles gambiae* .............................................................. 14
      1.3.1 Chromosomal forms ......................................................................... 15
      1.3.2 Population structure .......................................................................... 17
   1.4 Other gene drive systems ......................................................................... 19
      1.4.1 Homing endonuclease genes ............................................................ 20
      1.4.2 Meiotic drive ..................................................................................... 20
      1.4.3 Medea .............................................................................................. 20
      1.4.4 Wolbachia ....................................................................................... 21
      1.4.5 Engineered underdominance ........................................................... 21
   1.5 Malaria eradication and control ............................................................... 22

2 CHAPTER 2 ............................................................................................................ 24
   2.1 Introduction ................................................................................................. 24

   2.2 Materials and methods ............................................................................ 29
      2.2.1 Mathematical model ........................................................................... 31
      2.2.2 Parameter values .............................................................................. 34
      2.2.3 Numerical analysis ............................................................................ 39
2.3 Results

2.3.1 Dissociation rate

2.3.2 Transpositional handicap

2.3.3 Fitness consequences

2.4 Discussion

2.4.1 Model limitations

2.4.2 Future directions

3 CHAPTER 3

3.1 Introduction

3.2 Model formulation

3.2.1 Reproduction

3.2.2 Transposition and deletion

3.2.3 Host fitness

3.2.4 Gamete formation

3.2.5 Population size changes

3.3 Parameter values

3.3.1 Transposition

3.3.2 Transpositional regulation

3.3.3 Deletion

3.3.4 Host fitness

3.3.5 Population size changes

3.3.6 Population size

3.3.7 Release strategy

3.4 Analysis

3.4.1 Critical parameter values for TE spread

3.4.2 TE loss in a natural population

3.4.3 TE loss following a transgenic release

3.4.4 Rate of TE spread following a transgenic release

3.4.5 Rate of TE spread beyond a prevalence of 10%

3.4.6 Sensitivity to model parameters

3.5 Discussion

3.5.1 Transpositional bursts

3.5.2 Population growth

3.5.3 Implications for a transgenic release

3.5.4 Model limitations

3.5.5 Future directions
# Chapter 4

## 4.1 Introduction

## 4.2 Analysis

- **4.2.1** Homing endonuclease genes
- **4.2.2** Transposable elements
- **4.2.3** Meiotic drive
- **4.2.4** Medea
- **4.2.5** Wolbachia
- **4.2.6** Engineered underdominance

## 4.3 Discussion

# Chapter 5

## 5.1 Review of hypotheses

- **5.1.1** The impact of dissociation on TE-mediated disease control
- **5.1.2** Molecular and ecological parameters for TE spread
- **5.1.3** The effect of gene drive on transgene containment

## 5.2 Genetically modified mosquitoes and the Concorde

- **5.2.1** Multiple vectors of malaria
- **5.2.2** The Concorde

## 5.3 Future directions

- **5.3.1** Barriers to the establishment of refractoriness for other gene drive systems
- **5.3.2** The effect of local population structure on the early spread of Wolbachia infections and Medea alleles
- **5.3.3** Modeling mosquito dispersal and the spatial spread of gene drive systems

# References
LIST OF FIGURES

Figure 1.1 Schematic representation of the two mechanisms by which Class II elements can be replicated. ................................................................. 9

Figure 1.2 Distribution of paracentric inversions on the right arm of chromosome 2 in An. gambiae. .................................................................................. 16

Figure 1.3 Graph showing the number of An. arabiensis and An. gambiae M and S molecular forms typed from collections taken in Banambani, Mali during 2005. .................................................. 18

Figure 2.1 Abortive gap repair mechanism of internal deletion. ......................... 26

Figure 2.2 Schematic for the spread of an element construct incorporating dissociation of the resistance gene. .......................................................... 30

Figure 2.3 Element spread with dissociation. ...................................................... 40

Figure 2.4 Maximum prevalence of disease-resistant vectors. ................................. 44

Figure 3.1 Graph showing the number of An. arabiensis and An. gambiae M and S molecular forms typed from collections taken in Banambani, Mali during 2005. .................................................. 68

Figure 3.2 Critical parameter values for TE spread. ............................................. 74

Figure 3.3 Asymptotic probabilities of TE loss as a function of replicative transposition rate. .................................................................................. 78

Figure 3.4 Probabilities of TE loss one year following release. ........................... 84

Figure 3.5 Rate of TE spread for the first year following a transgenic release. ....... 88

Figure 3.6 Rate of TE spread for the first 10 years following a transgenic release. ... 92

Figure 4.1 Schematic for the early spread of a HEG through a randomly mating mosquito population. ........................................................... 110

Figure 4.2 Asymptotic probabilities of HEG loss as a function of fitness cost. ....... 116

Figure 4.3 Schematic for the early spread of a TE through a randomly mating mosquito population. ........................................................... 121

Figure 4.4 Asymptotic probabilities of TE loss as a function of fitness cost. ....... 126
Figure 4.5 Schematic for the early spread of a Y-linked meiotic drive gene ($Y^D$) and X-linked response allele ($X^u$) through a randomly mating mosquito population. ................................................................. 132

Figure 4.6 Asymptotic probabilities of meiotic drive loss as a function of fitness cost. .................................................................................................................. 139

Figure 4.7 Schematic for the early spread of a Medea allele through a randomly mating mosquito population. .............................................................................. 144

Figure 4.8 Asymptotic probabilities of Medea loss as a function of fitness cost. ........ 147

Figure 4.9 Schematic for the early spread of a Wolbachia bacterium through a randomly mating mosquito population. ................................................................. 152

Figure 4.10 Asymptotic probabilities of Wolbachia loss as a function of fitness cost. .................................................................................................................. 157

Figure 4.11 Schematic for the early spread of a pair of engineered underdominance alleles through a randomly mating mosquito population. ......................... 164
LIST OF TABLES

Table 2.1 Parameter values, ranges and source references for models used in Chapter 2................................................................................................................................. 35

Table 3.1 Parameter values, ranges and source references for models used in Chapter 3........................................................................................................................... 64

Table 3.2 Sensitivity of loss probability and rate of TE spread to model parameters. ..... 95

Table 5.1 Hypotheses examined in the dissertation.......................................................... 173
ACKNOWLEDGMENTS

I wish to thank my advisor Charles Taylor for providing me with a fascinating and relevant thesis project, for suggesting important research questions, and for introducing me to many of the leaders in the field of malaria control. I would also like to thank Professor Taylor for his friendship, for providing me the opportunity to travel as part of my research, and for taking an interest in my extracurricular as well as academic activities. Section 5.2 of this dissertation is based on a paper in progress co-authored with Professor Taylor entitled “Genetically Modified Mosquitoes and the Concorde.”

I would like to thank the other members of my committee in the Department of Biomathematics. I thank Tom Chou for his sharp mathematical insight and for passionately insisting that my biological problems be more mathematically interesting. I thank Ken Lange for always having the answers to my questions and for always having the time to explain them to me. I thank Janet Sinsheimer for her kindness and for taking an interest in my work. Also, although not on my committee, I thank Sally Blower for welcoming me into her research group, and for teaching me many useful skills in mathematical epidemiology. I am also grateful to Robert Weiss in the Department of Biostatistics for advising me on a statistical project unrelated to my dissertation.

None of this research would have been possible without substantial financial support. I would like to acknowledge and express my gratitude for a two-year full scholarship from the Department of Biomathematics, the Edward and Isabel Kidson Scholarship from the New Zealand Vice Chancellors’ Committee, and a Dissertation
Year Fellowship from the UCLA Graduate Division. This research has also been supported by grant numbers R01AI51633 and 1R56AI072549 from the National Institutes of Health.

All of the work presented in this dissertation has benefited from discussions with leaders in the field of genetically modified mosquitoes. I am particularly grateful to David O’Brochta from the University of Maryland for many fruitful discussions on the molecular biology of transposable elements. I am also grateful to Marcelo Jacobs-Lorena and Kenneth Olson for discussions on the molecular details of refractory alleles, and to William Engels for discussions on $P$ elements and the accuracy of DNA repair. My understanding of other gene drive systems has been greatly improved thanks to discussions with Jason Rasgon, Fred Gould, Steve Sinkins, Michael Turelli, Bruce Hay, Austin Burt and Yunxin Huang.

In the Taylor lab of the Department of Ecology and Evolutionary Biology, I am particularly grateful to Mahamadou Touré for teaching me about the ecology of mosquitoes in Mali. I am also grateful to Mahamadou Touré and Nicholas Manoukis for allowing the reproduction of Figure 1.3 in this manuscript, and to Charles Taylor for allowing the reproduction of Figure 1.2. I thank the many other members of the Taylor lab – Sigrid Rian, Saul Lozano, Paul Levine, Travis Collier, Yoosook Lee, Changde Chen and Dave Yuan – for their friendship and willingness to help with any academic endeavour. In the Department of Biomathematics, I thank the many graduate students who have befriended me and helped me to understand the mathematical techniques needed to solve my biological problems.
In Mali, I am grateful to the staff and medical students at the Malaria Research and Training Center (MRTC) in Bamako for giving me hands-on experience identifying the different species of malaria vectors and for allowing me an opportunity to understand the applications of my mathematical models.

In Italy, I am grateful to Mario Coluzzi and Fabio Candura at the University of Rome (La Sapienza). Much of my understanding of the chromosomal forms of *Anopheles gambiae* stems from a summer spent in the lab of Professor Coluzzi, who essentially created this field; while much of my understanding of Italian culture stems from the fantastic hospitality of my Italian colleague Dr Candura.

In New Zealand, I am grateful to Peter Wills who mentored me as a physics undergraduate at the University of Auckland. Professor Wills introduced me to the disciplines of mathematical and theoretical biology, and has continued to take an active interest in my work for more than five years since I left the University of Auckland.

I wish to thank my friends in New Zealand, many of whom have visited me in Los Angeles, and others who have always set aside the time to catch up with me when I have returned home. I thank my friends in Los Angeles who have made me feel very at home in this city. I also thank my wonderful roommates at the Real World Westwood who have made life in Los Angeles so much more entertaining. There has never been a dull moment.

Finally, I would like to thank my family for their immeasurable support and encouragement throughout my entire life. I thank my two sisters – Annabel and Clementine – for their unconditional love and companionship; and I thank my parents for
always encouraging me to pursue my interests, and for supporting me in whatever I choose to do.
<table>
<thead>
<tr>
<th>Year</th>
<th>Degree/Program</th>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997-1999</td>
<td>BSc (Biological sciences)</td>
<td>University of Auckland</td>
<td>Auckland, New Zealand</td>
</tr>
<tr>
<td>2000-2002</td>
<td>BTech (Hons) (Optoelectronics)</td>
<td>University of Auckland</td>
<td>Auckland, New Zealand</td>
</tr>
<tr>
<td>2003-2004</td>
<td>MS (Biomathematics)</td>
<td>University of California at Los Angeles</td>
<td>Los Angeles, California</td>
</tr>
<tr>
<td>2003-2005</td>
<td></td>
<td>Department of Biomathematics Scholarship</td>
<td></td>
</tr>
<tr>
<td>2003-2006</td>
<td></td>
<td>Edward and Isabel Kidson Scholarship</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td>Complex Systems Summer School</td>
<td>Santa Fe Institute</td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td>Teaching Preceptorship</td>
<td>Department of Ecology and Evolutionary Biology, UCLA</td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
<td>UCLA Dissertation Year Fellowship</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td>William C. Reeves New Investigator Award</td>
<td></td>
</tr>
</tbody>
</table>
PUBLICATIONS AND PRESENTATIONS


ABSTRACT OF THE DISSERTATION

The Dynamics of Transposable Elements
in Genetically Modified Mosquito Vectors

by

John Macky Marshall
Doctor of Philosophy in Biomathematics
University of California, Los Angeles, 2008
Professor Charles E. Taylor, Co-Chair
Professor Tom Chou, Co-Chair

The possibility of using genetically modified mosquitoes (GMMs) to control malaria has been discussed for several decades. Concrete proposals for using such methods are under development and empirical measurements permit some evaluation. Using the most recent experimental data, we model the use of transposable elements (TEs) to drive refractory
genes into mosquito populations. Our models recommend a transgenic release at the beginning of the seasonal population growth phase and a transposition rate of 0.1 per TE per generation to satisfy public health goals. Concern is raised regarding the loss of linkage between the TE and refractory gene, and is heightened by the observation that TEs such as *Himar1* transpose significantly more frequently once free of exogenous DNA. This has the potential to significantly diminish the success of the disease control strategy. Concerns are also raised regarding the possibility that a few transgenic mosquitoes will escape from field cages during the course of preliminary studies to test their efficacy. If TEs are used as the drive mechanism, modeling suggests that fewer than ten escapees are sufficient to establish the transgene in the population. We also analyze loss probabilities for other gene drive systems following an accidental release. Our models suggest that homing endonuclease genes and TEs are among the most invasive of the gene drive systems currently being considered. For meiotic drive strategies consisting of a drive gene and response allele, the drive gene is very capable of spreading; however the response allele requires a fitness benefit in order to spread. Drive strategies that rely on decreasing the number of wild-type mosquitoes in order to increase their own prevalence – such as *Medea, Wolbachia* and engineered underdominance – require either a fitness benefit or minimum release size in order to spread. Research into other drive strategies is encouraged; however all drive strategies are ultimately confronted by the economic infeasibility of transforming at least five different vectors of human malaria in order to achieve some semblance of control. Parallels are suggested between GMMs and the ill-fated Concorde supersonic airplane.
1 CHAPTER 1
INTRODUCTION

1.1 OBJECTIVES AND HYPOTHESES

Malaria is one of the most serious health problems facing the developing world. It is estimated that there are 250-450 million clinical cases and more than one million deaths due to the disease in Sub-Saharan Africa ever year (Lindsay et al. 1998; Greenwood and Mutabingwa 2002). Furthermore, despite ambitious goals such as those of the Roll Back Malaria initiative to halve malaria deaths by 2010, malaria deaths had actually risen halfway into the program (Yamey 2004). Several methods of disease control have proven ineffective in recent years. Widespread usage of insecticides to suppress mosquito populations has resulted in the emergence and spread of insecticide-resistance genes (Weill et al. 2003). Similarly, resistance to anti-malarial drugs such as chloroquine has quickly spread through the Plasmodium population (Sidhu et al. 2002). Low technology solutions such as bed nets, raised housing and removal of standing water offer promise; however they currently lack the investment and infrastructure to substantially reduce the malaria burden. Consequently, there is interest in new and innovative methods of disease control.

One of the new methods currently being considered to reduce the prevalence of malaria is the use of genetically modified mosquitoes (GMMs) to drive refractory genes into the
disease vector population (Alphey et al. 2002). This idea has been discussed for decades (Craig 1963; Curtis 1968); although it has only been during the last decade that advancements in molecular biology have allowed development efforts to proceed in earnest. While most mosquitoes transmit malaria, some have been identified that carry alleles of immune system genes that inhibit development of the *Plasmodium* parasite – the infections agent of malaria (Osta et al. 2004). Alleles have also been engineered that reduce the ability of mosquitoes to transmit malaria or dengue fever (Olson et al. 1996; Ito et al. 2002). It is hoped that these alleles can be linked to a gene drive system that will drive them to fixation or near-fixation in the mosquito population, and that this population will no longer transmit malaria or dengue fever.

Several gene drive systems exist in nature. Some of the most promising systems currently being investigated include homing endonuclease genes, transposable elements, Medea elements, the intracellular bacterium *Wolbachia*, engineered underdominance genes, and meiotic drive (Sinkins and Gould 2006). Transposable elements (TEs) were one of the first drive systems to receive widespread attention, and they will form the focus of this dissertation.

TEs are particularly interesting genomic components due to their ability to transpose replicatively within a genome and hence spread throughout a population despite a fitness cost (Charlesworth et al. 1994). The idea of using TEs to drive refractory genes into mosquito populations was proposed decades ago (Curtis 1968) and gained substantial
support following the observation that $P$ elements spread through most of the wild-type *Drosophila melanogaster* population within the second half of the 20th century (Engels 1997). As currently envisioned, a TE construct for genetic modification will contain a transposase gene to facilitate its replication, a promoter, a marker gene, and several refractory genes to neutralize the disease agent (O’Brochta 2003).

The objective of this dissertation is to examine, given the data presently available, the promise that TE constructs, and GMMs in general, have for the control of malaria and other vector-borne diseases. A substantial amount of money is currently being invested in the development of transgenic mosquitoes in the hope that their eventual release will reduce the burden of vector-borne disease. As data becomes available, it is important that we assess the feasibility of this project, and whether funding from the already small malaria budget is being well-spent. In studying foreseeable problems of the technology, modeling can also suggest directions for future experimental and engineering studies, and offer some insight into the safe use and testing of transgenic technologies.

In Chapter 2, we investigate one of the primary concerns of transposon-based genetic control strategies – the possibility that the refractory gene will become dissociated from the TE construct as it spreads into the vector population. The TE-based control strategy is particularly vulnerable to dissociation occurring between the drive system and refractory gene. This is thought to occur during transposition when a gap is left by a newly-transposed element. If gap repair is prematurely aborted, then internal sequences of the...
construct are lost, and a shortened refractory allele may become nonfunctional. In this case, the mosquito will still be genetically modified, but will lack the ability to halt transmission of the parasite. While dissociation has been proposed by some researchers as a reason for the TE-mediated control strategy being doomed to failure (e.g. Curtis 2003), this is primarily a question of rate. We propose a differential equations model to assess the rate of dissociation that a TE-mediated control strategy can tolerate.

In Chapter 3, we propose a branching process model to study the early spread of a TE in a diploid mosquito population. This model is primarily motivated by the question of what molecular and ecological parameters will satisfy the requirements of a TE-mediated disease control strategy – a small probability of TE loss, and a fast rate of TE spread. *Anopheles gambiae*, the main vector of malaria in tropical Africa, serves as a case study for hosting a TE for the purpose of malaria control. A major feature of the demography of *An. gambiae* is the existence of large population size changes between the dry season and the peak of the rainy season (Taylor *et al.* 2001; Manoukis 2006). We therefore study the effect of these population size changes on the success of a TE-mediated disease control strategy within the framework of a branching process model.

In Chapter 4, we compare probabilities that each of the six main gene drive systems will be lost from a wild population following an accidental release. The motivation for this study is that, once transgenic strains of mosquitoes have been engineered, it is planned that they will be studied in large outdoor cages in a region where they might eventually
be released (Alphey et al. 2002; Scott et al. 2002). The catch-22 of these cage studies is that, while they are absolutely necessary to assess potential outcomes prior to a transgenic release, it is impossible to guarantee that no mosquitoes will ever escape while this assessment is taking place (Benedict et al. 2008). Gene drive systems are of particular concern because they are designed to enhance the invasiveness of introduced genes. We therefore propose branching process models for each of the six main gene drive systems and compare their extinction probabilities under ideal conditions. This allows us to suggest which system is safest in the event of an accidental release.

Chapters 2, 3 and 4 were originally formatted for publication as individual manuscripts and are therefore occasionally redundant in content. While the issues discussed in these chapters represent only a fraction of the problems associated with genetically modified mosquitoes and TE-mediated disease control, they are clearly posed and amenable to study with data that is currently available.
1.1.1 General hypotheses

- Dissociation has a significant effect on the efficacy of transposon-mediated disease control strategies (Chapter 2).
- There exists a set of molecular and ecological parameters for which TEs will spread through a randomly mating population (Chapter 3).
- The mechanism of gene drive has an effect on the containment of transgenic mosquitoes in the event of an accidental release (Chapter 4).

1.1.2 Hypotheses examined in each chapter

Chapter 2:

- The prevalence of disease-resistant vectors depends on the rate of dissociation between the TE and refractory gene (Hypothesis 2.1).
- A transpositional handicap due to the refractory gene decreases the prevalence of disease-resistant vectors (Hypothesis 2.2).
- A fitness benefit due to the refractory trait increases the prevalence of disease-resistant vectors (Hypothesis 2.3).
Chapter 3:

- There exists a set of molecular and ecological parameters for which TEs are unlikely to be lost from a randomly mating population (Hypothesis 3.1).
- There exists a set of molecular and ecological parameters for which TEs will spread at a rate acceptable to public health goals (Hypothesis 3.2).
- A TE released at the beginning of the population growth phase has a greater chance of persisting in the wild (Hypothesis 3.3).

Chapter 4:

- Gene drive systems enhance the invasiveness of introduced genes (Hypothesis 4.1).
- Certain gene drive systems are more likely to persist than others following an accidental release (Hypothesis 4.2).
- Gene drive systems are more likely to persist during conditions of population growth (Hypothesis 4.3).

1.2 **Biology of Transposable Elements**

A TE is a DNA sequence capable of inserting itself at a new location in the genome. The TEs being considered for genetic modification belong to Class II and contain repeat sequences which mark their boundaries, and their own transposase gene which catalyzes
transposition. Class II TEs move by a cut-and-paste mechanism (Plasterk and van Luenen 2002). This is a conservative mechanism of transposition that does not involve replication of elements directly. Despite the conservative nature of Class II element transposition, these transposons can be replicated as a consequence of the transposition reaction, leading to an increase in their copy number.

There are two ways Class II elements can be replicated (Figure 1.1). First, element excision and transposition temporarily leaves a double-stranded gap in chromosomal DNA that must be repaired or the cell will die. This gap might simply be sealed resulting in the loss of the element from that site and a relocation of the element to a new site. In some cases, however, the gap is filled by copying the information contained on the homologous chromosome. If the cell was originally homozygous for the element at that site, then the homologous chromosome will contain a copy of the element which can be copied into the gap. This results in a net gain of one element in the genome (Figure 1.1A). P elements and Mos1 are examples of TEs that can display this type of repair mechanism following element excision and transposition (Rio 2002; Plasterk and van Luenen 2002). Second, some TEs transpose during the S phase of the cell cycle. If a recently replicated element transposes to an unreplicated region of the genome, it will be replicated a second time resulting in a net gain of one element within the genome (Figure 1.1B). Mos1 in A. aegypti and the Ac/Ds element in corn time some of their transpositions to exploit S phase DNA replication (Chen et al. 1992; Wilson et al. 2003).
Figure 1.1 Schematic representation of the two mechanisms by which Class II elements can be replicated. (A) In templated gap repair, excision and transposition leaves a gap that is sometimes sealed by copying information on the homologous chromosome. (B) In S-phase transposition, a replicated element transposes to an unreplicated part of the genome and is replicated again.
Replication of a TE is essential for the element to increase in copy number within a host genome and hence to increase in frequency within a population. Under the cut-and-paste mechanism of transposition, most element remobilization events are non-replicative, resulting in the element simply changing position. This is relevant to inferring potential fitness costs of transposition and movement of elements to sites in the genome unlinked to other elements.

1.2.1 Regulation of transposition rate with copy number

A key detail of TE biology is that, for many families of TEs, transposition rate tends to decrease as the number of elements in the host genome increases – a phenomenon known as transpositional autoregulation. Autoregulation can sometimes be attributed to host factors involved in gene silencing (Wu and Morris 1999) or entirely to post-transcriptional regulation of the TE itself (Galun 2003; Townsend and Hartl 2000). It is thought that different families of TEs have evolved their own mechanisms to achieve a balance between selection for high element copy number and selection for carriers with fewer deleterious mutations (Charlesworth and Langley 1986). The relevance of these observations to this dissertation is that transposition rate should be modeled as a decreasing function of element copy number.

Insight into the form of this function can be obtained through considering the molecular kinetic mechanisms that could be causing the observed autoregulation. Townsend and Hartl (2000) considered a number of potential mechanisms for the mariner element. They
deduced that the only mechanism kinetically consistent with the observed autoregulation describes a transposition event which occurs when DNA-bound monomers meet, dimerize and bind to the outside ends of a TE thus leading to element mobilization.

The mechanism of regulation will differ from element to element; however for mariner at least, these kinetic equations will predict a form for the transposition rate function. In the absence of such insight, generic functions have been proposed by others. Charlesworth and Charlesworth (1983) proposed a hyperbolic form for replicative transposition as a function of copy number. Additionally, Le Rouzic and Capy (2005) proposed a threshold regulation function whereby transposition rate drops abruptly at a critical element copy number, and a continuous regulation function whereby transposition rate decreases exponentially with copy number.

1.2.2 Fitness costs associated with transposable elements

Host fitness is another important parameter that tends to decrease as the number of elements in the host genome increases. This relationship is of great importance to models of transposon population dynamics because reduction of host fitness is one of the main forces enabling the attainment of equilibrium element copy numbers in many species (Charlesworth and Charlesworth 1983). Given the lack of reliable experimental fitness measurements, most literature on fitness costs has a theoretical basis, and there is an ongoing debate as to how to model the effects of natural selection on TE copy number (Brookfield and Badge 1997).
First, due to insertional mutagenesis, each additional copy of a TE has the potential to disrupt a gene contributing to the fitness of the organism (Mackay 1989). This should result in a fitness cost that increases linearly with element copy number. Second, Montgomery et al. (1987) have proposed that recombination can occur between TEs of the same family that are present at different sites, resulting in deleterious chromosomal rearrangements and a fitness cost that increases with the square of TE copy number. Langley et al. (1988) and Bartolome et al. (2002) support this hypothesis by observing elevated numbers of TEs in areas of reduced recombination. The role of ectopic recombination in modulating element copy number has also been studied by Petrov et al. (2003). Third, Brookfield (1991, 1996) has proposed that the fitness costs of TEs arise primarily through the act of transposition, resulting in a fitness cost that is a function of the transposition rate.

While literature on fitness costs is primarily theoretical, experimental measurements do seem to corroborate the decrease in fitness with increasing copy number. Pasyukova et al. (2004) observed that isogenic lines of D. melanogaster that had diverged by 90 TE copies per haploid genome displayed a negative correlation between element copy number and both total fitness and egg hatchability. In the absence of reliable fitness measurements applicable to field releases of transgenic mosquitoes, we rely on the few measurements and large body of theory available to explore large regions of parameter space.
1.2.3 Internal deletion within transposable elements

A property of some TEs of concern to the success of TE-mediated control strategies is the deletion of DNA sequences within TEs during transposition. This is known as “internal deletion” and could potentially lead to loss of the refractory gene in the TEs being considered for release. Internal deletion has been observed for $P$ elements in $D. melanogaster$ in small cage experiments where insects containing ~30 autonomous $P$ elements were mixed with insects free of $P$ elements. The $P$ elements rapidly spread into the naïve genomes; however coincident with this spread was the accumulation of numerous internally deleted non-autonomous elements (Engels 1989). This deletion is thought to occur due to imperfect gap repair following element excision. Returning to Figure 1.1A, a double-stranded gap is introduced into chromosomal DNA following excision and transposition. This gap is sometimes filled by copying information from the homologous chromosome still containing the element; however if this process is interrupted before the element has been completely copied then the result is an internal deletion. Other elements have also been shown to accumulate internal deletions, presumably by a similar mechanism (Lohe et al. 2000; Rubin and Levy 1997).

Despite the observation of internal deletion in $P$ and other elements, there are still other elements that do not seem to have a propensity for losing internal sequences. The $Herves$ element in $An. gambiae$ has a long history in this species, is shared by sister species in the $An. gambiae$ complex, and can exist in natural populations in an almost exclusively intact
form (Subramanian et al. 2007). The implication of these observations is explored in Chapter 2 of this dissertation.

1.3 ECOLOGY OF ANOPHELES GAMBIAE

An. gambiae is the main vector of malaria in tropical Africa and consequently it is the primary species being considered to host a TE and effector gene for the purpose of malaria control. The origins of the species have been traced back to the last 4000 years when extensive agriculture in Africa began to penetrate the forest (Ayala and Coluzzi 2005). Following its origin, the species diversified into a variety of chromosomal forms making up the An. gambiae complex, each of which has adapted to its own particular niche.

The range of An. gambiae currently extends through most of Sub-Saharan Africa. This is a very diverse area, and the genetic diversity and population structure of the species is correspondingly complex. Much of this complexity has been described in western Africa, and in particular Mali – a region where three chromosomal forms of the species complex coexist (Touré et al. 1998). Eventually, research should be extended to other sites throughout Africa, particularly around central and east Africa; however in the discussion that follows, we will use the ecology of An. gambiae in Mali as a case study for the African continent.
1.3.1 Chromosomal forms

One of the major factors in the ecology of *An. gambiae* is the existence of at least five chromosomal forms that may be partially or totally reproductively isolated (Figure 1.2). These chromosomal forms are distinguished on the basis of six paracentric inversions on chromosome 2 – the *j, b, c, u* and *d* inversions on the right arm and the *a* inversion on the left arm (Coluzzi and Sabatini 1967; Bryan *et al.* 1982; Touré *et al.* 1983; Coluzzi *et al.* 1985) – although the Forest chromosomal form cannot be distinguished in this way (Charles Taylor, personal communication). Each of these chromosomal forms has its own distinct geographical distribution, and in the cases where these distributions overlap, gene flow between forms is limited. As an additional level of complication, there are also two molecular forms of *An. gambiae* whose distributions often coincide with those of the chromosomal forms (Favia *et al.* 1997; 2001); however these distributions do not always coincide (Della Torre *et al.* 2001; 2002).
Figure 1.2 Distribution of paracentric inversions on the right arm of chromosome 2 in *An. gambiae*. Data is based on a sample of 1422 female *An. gambiae* mosquitoes from Banambani, Mali reported by Touré *et al.* (1998) and summarized by Taylor *et al.* (2001). Clusters of individuals having similar combinations of inversions correspond to distinct chromosomal forms of *An. gambiae*. The Forest and Bissau chromosomal forms cannot be distinguished on the basis of these four inversions. The Bamako and Savanna chromosomal forms tend to coincide with the S molecular form of *An. gambiae*, while the Mopti chromosomal form tends to coincide with the M molecular form. The small number of hybrid forms suggests that the populations are partially reproductively isolated.
1.3.2 Population structure

A significant understanding of the population structure of An. gambiae is critical to the success of a transgenic release. At the well-studied village of Banambani, Mali and its surrounding areas, the picture of population structure is one of local An. gambiae populations concentrated within human villages and consisting of three chromosomal forms (Bamako, Savanna and Mopti) alongside Anopheles arabiensis populations (Taylor et al. 2001). The resulting population is structured in three distinct ways. First, the An. gambiae population is structured temporally since its population size changes within and between years. Mark-release-recapture studies in the village of Banambani indicate that the annual village abundance of An. gambiae varies by a factor of four; however the most dramatic population size changes occur within a single year when population sizes at the peak of the wet season are on the order of 100 times those during the dry season (Taylor et al. 2001). The fluctuation in population size is also illustrated in the weekly collections taken in Mali throughout 2005 (Figure 1.3). Both M and S molecular forms show peak abundances during the wet season from June to October (Manoukis 2006). The implication of these observations is explored in Chapter 3 of this dissertation.
Figure 1.3 Graph showing the number of *An. arabiensis* and *An. gambiae* M and S molecular forms typed from collections taken in Banambani, Mali during 2005. The difference in population size between the wet and dry seasons is clearly evident for both M and S molecular forms (from Manoukis 2006).
Second, restricted gene flow between chromosomal forms adds an additional level of structure to the *An. gambiae* population. The amount of gene flow between chromosomal forms is significantly higher than if they were distinct species, but is still much smaller than within forms in the same population. Third, the *An. gambiae* population is structured geographically by its concentration in discrete patches corresponding to villages. The different chromosomal forms also have distinct geographical distributions as described by Rian *et al.* (2007). The populations of each chromosomal form exchange migrants with neighboring villages (Touré *et al.* 1998; Dolo *et al.* 1999); however there appears to be more gene flow among villages that are close together than among villages that are further apart (Johnson 1969).

### 1.4 Other Gene Drive Systems

While TEs were one of the first mechanisms of gene drive to gain widespread attention; the rapid spread of *P* elements in *D. melanogaster* has not been repeated for any TE artificially introduced into a disease vector species (Atkinson and James 2002). In recent years, several other gene drive systems have been discovered (Sinkins and Gould 2006), and experiments are currently underway to study the ability of these systems to drive refractory genes into mosquito populations (Chen *et al.* 2007; Cha *et al.* 2006a). Some of the most promising gene drive systems are discussed in the following paragraphs. The ability of these systems to persist in the wild following an accidental release is explored in Chapter 4 of this dissertation.
1.4.1 Homing endonuclease genes

Homing endonuclease genes (HEGs) are a class of highly-specific DNA endonucleases found in some viruses, bacteria and eukaryotes (Windbichler et al. 2007). HEGs are able to spread through a population despite a fitness cost due to their overrepresentation in the gametes of a heterozygote. They achieve this by expressing an endonuclease which creates a double-stranded break at a highly-specific site that lacks the HEG. Homologous DNA repair then copies the HEG to the cut chromosome (Rong and Golic 2003).

1.4.2 Meiotic drive

Meiotic drive refers to any mechanism by which a heterozygous locus segregates at a greater-than-Mendelian frequency by destroying or disabling the homologous chromosome (Little 1991). Various mechanisms are known to result in meiotic drive (Hickey and Craig 1966; Lyttle 1977). Some of these reduce the quantity of functional sperm; however they do not necessarily result in reduced fertility (Sinkins and Gould 2006). Alleles that promote meiotic drive are able to spread through a population despite a fitness cost as a consequence of their increased inheritance.

1.4.3 Medea

Medea, also known as maternal-effect-dominant embryonic arrest, is another form of meiotic drive. Its dynamics have been studied in Tribolium beetles (Wade and Beeman 1994); however it has attracted much recent attention since an engineered Medea element
has been observed to rapidly spread through *Drosophila* populations in the laboratory (Chen *et al.* 2007). *Medea* is able to spread through a population despite a fitness cost through its ability to cause the death of all offspring of heterozygous females that do not inherit the *Medea* allele. This distorts the offspring ratio in favor of the *Medea* allele.

### 1.4.4 Wolbachia

*Wolbachia* is a maternally-inherited, intracellular bacterium found in a wide variety of invertebrate taxa. *Wolbachia* infections are associated with several host reproductive alterations including cytoplasmic incompatibility (Stouthammer *et al.* 1999), in which offspring of matings between infected males and uninfected females are completely or partially sterilized, while matings involving infected females always produce infected offspring. This favors the offspring ratio in favor of the *Wolbachia* infection, and *Wolbachia* is therefore able to spread rapidly through a population despite a fitness cost (Turelli and Hoffmann 1999).

### 1.4.5 Engineered underdominance

The simplest case of underdominance is when a trait is determined by two alleles at a single locus and the fitness of a heterozygote is less than that of either homozygote (Hartl and Clark 1989). The dynamics of underdominant traits are generally unstable and, depending on the initial frequency of the two alleles, one will tend to be lost while the other will reach fixation in the population (Crow and Kimura 1970; Spiess 1977). A
novel form of engineered underdominance has been suggested by Davis et al. (2001) that requires a much smaller release size in order to spread into a naïve population.

1.5 MALARIA ERADICATION AND CONTROL

It is appropriate to distinguish between “malaria eradication” and “malaria control.” “Malaria eradication” implies the complete cessation of malaria transmission and a basic reproductive number of the malarial infection less than one. This is a particularly lofty goal for the continent of Africa, and hence we will focus on the more modest goal of “malaria control” in this dissertation. “Malaria control” implies a reduction in the prevalence of malaria to a point where the disease is no longer a major public health problem.

There is a close relationship between the entomological inoculation rate and malaria prevalence. Hay et al. (2000) surveyed the existing literature on entomological inoculation rates across Africa from 1980 onwards and compared this data to measurements of malaria prevalence. Fitting the data to a number of mathematical models, Hay et al. (2000) found that the best fit was given by the model,

\[
P = 1 - \left(1 + \frac{bE}{rk}\right)^{-k}.
\]

Here, malaria prevalence, \( P \), is defined as the proportion of children in a given area that are infected with malaria, and the entomological inoculation rate, \( E \), is defined as the
average number of infective bites than an individual receives in one year. Additionally, \( b \) represents the probability that a bite from an infectious mosquito results in a new infection or superinfection, \( 1/r \) represents the expected time to clear each infection, and \( 1/k \) represents the variance of the gamma distribution of relative infection rates. It should be noted that the entomological inoculation rate can also be written as,

\[
E = mas, \quad \text{(1.2)}
\]

where \( m \) represents the number of mosquito vectors per human, \( a \) represents the average number of bites on a human per mosquito per year, and \( s \) represents the proportion of mosquito vectors with malaria sporozoites in their salivary glands.

It is apparent from Equation 1.1 that reducing the entomological inoculation rate will lead to a monotonic decrease in malaria prevalence. Genetically modifying mosquitoes with genes that render them refractory to malaria will reduce the number of disease vectors per human by a proportion equal to the fraction of refractory mosquitoes in the population. Equations 1.1 and 1.2 can then be used to give us some indication of the reduction in malaria prevalence we can expect for a given population frequency of refractory mosquitoes. The implications of these equations are explored in Chapter 5 of this dissertation.
2  CHAPTER 2

THE IMPACT OF DISSOCIATION ON TRANSPOSON-MEDIATED DISEASE
CONTROL STRATEGIES

2.1  INTRODUCTION

The creation of transgenic mosquitoes that have a fitness advantage when feeding on Plasmodium-infected blood (Marrelli et al. 2007) has greatly renewed interest in the use of genetically modified vectors as a means of disease control. Resistance genes are unlikely to reach fixation in a wild vector population on their own, partly because disease prevalence tends to be relatively low in vector populations (Beier et al. 1999), and partly because the selective advantage of a resistance gene diminishes as the disease becomes less prevalent (Boëte and Koella 2003). Consequently, a series of gene drive systems have been proposed to bias the disease-resistance gene in favor of fixation (Craig 1963; Curtis 1968). One such drive system is a transposable element (TE), which is able to spread through a population by virtue of its ability to replicate within a genome (Charlesworth et al. 1994).

Any successful gene drive strategy requires tight linkage between the drive system and effector gene (James 2005). A general concern of drive systems is that rare recombination events can lead to loss of linkage between the drive system and effector gene (Riehle et al. 2003; Knols and Scott 2003; Curtis 2003). For TEs, there is an additional concern that
internal deletion of DNA sequences within elements can occur during transposition. In the class of transposons being considered for genetic modification, this is thought to occur by an abortive gap repair mechanism (Rubin and Levy 1997). Following excision or transposition, a double-stranded gap is introduced into the host chromosomal DNA. This gap is sometimes filled by copying information from a homologous chromosome, sister chromatid or ectopic chromosomal site still containing the TE. If this process is interrupted, then the central portions of the element will not be copied, leading to an internal deletion event (Figure 2.1).
Figure 2.1 Abortive gap repair mechanism of internal deletion. (A) An intact TE construct consists, at the very least, of a transposase gene (Tr), a disease resistance gene (R) and a pair of inverted repeats (IR) marking its boundaries. (B) Following transposition or deletion, a double-stranded gap is introduced into the host chromosome DNA. (C) This gap is sometimes filled by copying information from a homologous chromosome, sister chromatid or ectopic chromosomal site containing the intact construct. (D) If this process is interrupted, then the central portions of the TE will not be copied. In some cases, this produces a dissociated construct not containing the resistance gene.
The rate of internal deletion is largely dependent on the TE and host species being considered. In *Drosophila melanogaster*, *P* (Engels 1989) and *hobo* (Gelbart and Blackman 1989) elements produce deletion derivatives at a significant frequency due to abortive gap repair. The same mechanism is primarily responsible for the formation of nonautonomous *Ds* elements in maize (Rubin and Levy 1997) and the prevalence of nonautonomous *mariner* elements in many natural populations (Lohe et al. 2000). Despite this, there are other TEs, such as the *Herves* element in *Anopholes gambiae*, that remain in an almost exclusively intact form throughout evolutionary history (Subramanian et al. 2007).

Variability in the rate of internal deletion poses the question: what dissociation rate can be tolerated by an effective disease control strategy? If the dissociation rate is exceptionally slow, then the TE will drive the resistance gene into the vector population before the resistance gene has a chance to become detached. However, if the dissociation rate is exceptionally fast, then the resistance gene will be lost from the construct as the transposon spreads through the population, preventing the resistance trait from reaching a sufficiently high prevalence for effective disease control.

The situation is complicated by the tendency of TEs to transpose less frequently as their size is increased. Experiments in which *Himar1* mariner elements were increased in size by adding various fragments of exogenous DNA (Lampe et al. 1998) found that the mobility of these elements is inversely proportional to their size. A similar relationship
was observed for $P$ elements in *D. melanogaster* (Spradling 1986). These results are particularly relevant to the use of genetically modified vectors since it is projected that any useful gene drive system will carry at least two effector genes, a marker gene and regulatory elements (Marcelo Jacobs-Lorena, personal communication). Of concern is the possibility that elements no longer carrying their transgenic load will spread more quickly than their intact counterparts.

Another complicating factor is the impact of disease-resistance genes on host fitness. Mounting an immune response is generally associated with an evolutionary cost in insects (Kraaijeveld and Godfray 1997) and hence any resistance gene is likely to be associated with a selective disadvantage. Despite this, transgenic mosquitoes have been created that have no noticeable fitness cost when fed on *Plasmodium*-free blood (Moreira et al. 2004) and are in fact more fit when fed on *Plasmodium*-infected blood (Marrelli et al. 2007). Whether the resistance gene confers a fitness benefit or cost to the host will significantly influence the fate of the resistance gene in the vector population. Successful disease control requires that the disease-resistance trait reaches a high prevalence in the vector population within a reasonable timeframe. Throughout our modeling, we make recommendations regarding the dissociation rates, transpositional handicaps and fitness costs or benefits required to achieve this goal.
2.2 **MATERIALS AND METHODS**

We use a system of ordinary differential equations to describe element spread through a randomly mating population incorporating dissociation (Figure 2.2). To simulate a transgenic release, we initiate the model with a proportion of the vector population having one copy of the intact TE construct and the rest having none. Element spread occurs as long as replicative transposition exerts a greater influence on element copy number than element deletion and host fitness costs. Dissociation leads to the number of intact constructs decreasing by one and the number of elements free of their transgenic load increasing by one. This reduces the number of TEs having a copy of the disease-resistance gene.
Figure 2.2 Schematic for the spread of an element construct incorporating dissociation of the resistance gene. An individual having two copies of the intact construct (TER) and two copies of the dissociated construct (TE) can gain a copy of either through transposition and lose a copy of either through deletion. The proportion of individuals having this genotype can also increase or decrease as a result of transposition or deletion. Dissociation is a one-way process leading to a TER becoming a TE. Analogous schematics can be drawn for every other genotype.
2.2.1 Mathematical model

Particles in the model may have between 0 and $T$ copies of the TE, where $T$ is the maximum element copy number attainable during the period of spread. Of these copies, $m$ are attached to the resistance gene and $n$ are dissociated from their transgenic load.

We keep track of the proportion of vectors having each genotype, $x_{(m,n)}$, using a separate differential equation for each one. The process of element spread can then be modeled by a system of $((T+1)^2 + T + 1)/2$ simultaneous ordinary differential equations:

$$\frac{dx_{(m,n)}(t)}{dt} = \theta(t) \sum_{i,j,k,l} p_{(i,j),(k,l),(m,n)} x_{(i,j)}(t) x_{(k,l)}(t) - \mu_{(m,n)}(t) x_{(m,n)}(t)$$

$$+ (m-1)(1-\delta) u_{m+n-1} x_{(m-1,n)}(t) + (n-1) u_{m+n-1} x_{(m,n-1)}(t)$$

$$+ (m+1) v x_{(m+1,n)}(t) + (n+1) v x_{(m,n+1)}(t)$$

$$+ (m+1) w_{(m+1,n-1)} x_{(m+1,n-1)}(t)$$

$$- \left( m(1-\delta) u_{m+n} + nu_{m+n} + mv + nv + mw_{(m,n)} \right) \vec{x}_{(m,n)}(t)$$

where $n,m \geq 0$ and $n+m \leq T$.

The first two terms in Equation 2.1 describe the birth and death rates of a vector having the genotype $(m,n)$. The death rate of a type-$(m,n)$ vector, $\mu_{(m,n)}$, is given by the equation:

$$\mu_{(m,n)}(t) = 1 + d(t)(m+n)^p + m\sigma - 1_{\{m>0\}} \sigma \sum_{n=0}^{T} x_{(0,n)}(t)$$

(2.2)

The first term in this equation represents the death rate of a type-$(0,0)$ vector, the second term represents the fitness cost of $m+n$ TEss, the third term represents the fitness cost of
$m$ resistance genes, and the fourth term represents the selective advantage of the 
resistance trait. The fitness cost of a single TE, $d(t)$, is assumed to decrease over time 
according to the equation:

$$d(t) = (d_0 - d_\infty)2^{-t/t_{1/2}} + d_\infty.$$  \hspace{1cm} (2.3)

Here, $d_0$ represents the mean fitness cost of a new TE insert, $d_\infty$ represents the mean 
fitness cost of a TE segregating in the population, and $t_{1/2}$ represents the number of 
vector generations after which the mean fitness cost of a TE is $(d_0 + d_\infty)/2$. Following 
Charlesworth and Charlesworth (1983), we assume that the fitness cost of the TE is a log-
concave function of element copy number. The power $p$ determines the rate at which 
this fitness cost increases with total TE copy number (Charlesworth and Charlesworth 
1983). The fitness cost of the resistance gene, $\sigma_-$, is assumed to increase linearly with 
dosage, $m$; while the fitness benefit of the resistance trait, $\sigma_+$, is conditional upon the 
presence of one or more resistance genes, and is proportional to the fraction of the vector 
population that is susceptible to the disease at time $t$.

The total birth rate of disease vectors, $\theta(t)$, is equal to the total death rate, normalizing 
for the sum of $x_{(i,j)}(t)x_{(k,l)}(t)$ terms in the birth rate term. This leads to the equation:

$$\theta(t) = \frac{\sum_{m,n} H_{(m,n)}(t)x_{(m,n)}(t)}{\sum_{i,j,k,l} x_{(i,j)}(t)x_{(k,l)}(t)}.$$ \hspace{1cm} (2.3)
The birth rate of a type- \((m,n)\) vector is then given by summing over every pair of genotypes in the population and weighting each pair by the probability that, when a type- \((i,j)\) vector mates with a type- \((k,l)\) vector, the offspring will be of type- \((m,n)\).

Assuming that all TEs are fairly well-spaced and hence segregate independently, this probability is given by:

\[
P_{(i,j),(k,l),(m,n)} = \sum_{x=0}^{m} \sum_{y=0}^{n} \frac{1}{2^{i+j+k+l}} \binom{j}{y} \binom{l}{n-y} \binom{i}{x} \binom{k}{m-x}.
\]  

(2.4)

The tendency of some elements to transpose locally (Tower et al. 1993; Zhang and Spradling 1993; Guimond et al. 2003) and display homing behavior (Guimond et al. 2003), though important, is beyond the scope of this study.

The third and fourth terms in Equation 2.1 describe the rate at which replicative transposition increases the proportion of type- \((m,n)\) vectors in the population. This rate is proportional to the rate of replicative transposition and, due to the presence of repression of transposition (Charlesworth and Charlesworth 1983; Townsend and Hartl 2000; Le Rouzic and Capy 2005), is also a function of total element copy number. In a genome having a total of \(i\) element copies, the replicative transposition rate can be described as,

\[
u_i = u_1 2^{-c(i-1)},
\]

(2.5)

where \(c\) determines the rate at which the replicative transposition rate falls off with additional element copies, and \(u_1\) is the replicative transposition rate when there is only a single element in the genome. Transposition is slowed down by the presence of a
resistance gene, and so the replicative transposition rate of an intact TE construct is given by 
\[(1 - \delta)u_i\], where \(\delta\) represents the fractional decrease in transposition rate due to the 
effector gene.

The fifth and sixth terms in Equation 2.1 describe the rate at which element deletion 
increases the proportion of type-\((m,n)\) vectors in the population. The rate of element 
deletion, \(\nu\), is very small and we assume this to be constant, independent of the total TE 
copy number or transgenic load. The seventh term describes the rate at which dissociation 
increases the proportion of type-\((m,n)\) vectors in the population. Internal deletion is 
hypothesized to occur during the act of transposition, and hence its rate is modeled as 
proportional to the rate of replicative transposition:

\[
w_{(m,n)} = \frac{u_{m+n} \cdot w^*}{u_1}.
\] 

(2.6)

Here, \(w^*\) is the dissociation rate when there is only a single intact construct in the 
genome. Finally, the eighth term describes the rate at which replicative transposition, 
element deletion and dissociation decrease the proportion of type-\((m,n)\) vectors in the 
host population.

2.2.2 Parameter values

Table 2.1 contains the parameter values and ranges used in the implementation of our 
model. At present there is little or no data regarding the post-integration behavior of the 
candidate elements in human disease vectors such as \textit{An. gambiae} and \textit{Aedes aegypti}, so
most of these values were taken from measurements in other species, in particular *D. melanogaster* (Nuzhdin et al. 1997; Maside et al. 2000; Guimond et al. 2003; Pasyukova et al. 2004).

### Table 2.1 Parameter values, ranges and source references for models used in Chapter 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Estimated value</th>
<th>Range</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_i$</td>
<td>Replicative transposition rate</td>
<td>0.1 TE$^{-1}$gen$^{-1}$</td>
<td>–</td>
<td>Seleme et al. (1999), Vasilyeva et al. (1999)</td>
</tr>
<tr>
<td>$c$</td>
<td>Transpositional regulation parameter</td>
<td>2.866 TE$^{-1}$</td>
<td>–</td>
<td>Subramanian et al. (2007), Townsend and Hartl (2000)</td>
</tr>
<tr>
<td>$v$</td>
<td>Element deletion rate</td>
<td>$4 \times 10^{-6}$ TE$^{-1}$gen$^{-1}$</td>
<td>–</td>
<td>Nuzhdin et al. (1997), Maside et al. (2000)</td>
</tr>
<tr>
<td>$d_0$</td>
<td>Initial TE fitness cost</td>
<td>0.02 TE$^{-1}$</td>
<td>[0.01, 0.1] TE$^{-1}$</td>
<td>Crow and Simmons (1983), Mackay et al. (1992)</td>
</tr>
<tr>
<td>$d_{\infty}$</td>
<td>Asymptotic TE fitness cost</td>
<td>$10^{-3}$ TE$^{-1}$</td>
<td>–</td>
<td>Nuzhdin et al. (1997), Pasyukova et al. (2004)</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>Halftime of TE fitness cost decline</td>
<td>5 gen</td>
<td>[2, 20] gen</td>
<td>Charlesworth (1991)</td>
</tr>
<tr>
<td>$p$</td>
<td>Rate of decrease in vector fitness</td>
<td>1.5</td>
<td>–</td>
<td>Charlesworth and Charlesworth (1983)</td>
</tr>
<tr>
<td>$w^*$</td>
<td>Dissociation rate</td>
<td>0.05 TE$^{-1}$gen$^{-1}$</td>
<td>[0, 0.1] TE$^{-1}$gen$^{-1}$</td>
<td>Levis et al. (1985), Robertson et al. (1988)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Transpositional handicap</td>
<td>0.95</td>
<td>[0, 0.99]</td>
<td>Lampe et al. (1998); Spradling (1986)</td>
</tr>
<tr>
<td>$\sigma_-$</td>
<td>Fitness cost of resistance gene</td>
<td>–</td>
<td>[0, 0.05] TE$^{-1}$</td>
<td>Ahmed et al. (2002), Beier et al. (1999)</td>
</tr>
<tr>
<td>$\sigma_+$</td>
<td>Fitness benefit of resistance gene</td>
<td>–</td>
<td>[0, 0.05]</td>
<td>Marrelli et al. (2007), Appawu et al. (2003)</td>
</tr>
<tr>
<td>$\chi_{(1,0)}$</td>
<td>Initial proportion of transgenic vectors</td>
<td>0.025</td>
<td>[0.01, 0.1]</td>
<td>Taylor et al. (2001)</td>
</tr>
</tbody>
</table>
We consider a baseline replicative transposition rate of \( u = 0.1 \) per element per generation (TE\(^{-1}\) gen\(^{-1}\)). Although this is a fairly high transposition rate, it is realistic (Seleme et al. 1999; Vasilyeva et al. 1999) and several modeling approaches have recommended it as a minimum requirement for gene drive to occur in a timeframe acceptable to public health goals (Rasgon and Gould 2005; Le Rouzic and Capy 2006). We chose the transpositional regulation parameter to be \( c = 2.866 \) per element (TE\(^{-1}\)) in order to achieve an average equilibrium element copy number of four. This was motivated by the observation that Herves, a functional Class II TE in An. gambiae, is present across many geographical locations at an average copy number of four (Subramanian et al. 2007). Herves is of particular interest to transposon-mediated gene drive strategies since it rarely undergoes internal deletion and maintains its structural integrity over long periods of time (Subramanian et al. 2007).

Element deletion events are so rare that they are often not observed in laboratory line experiments (Nuzhdin and Mackay 1994). We consider a deletion rate of \( v = 4 \times 10^{-6} \) TE\(^{-1}\) gen\(^{-1}\) as suggested by pooled estimates from several laboratory line experiments (Nuzhdin et al. 1997; Maside et al. 2000).

Following Mackay et al. (1992), we estimate the mean fitness cost of a new TE insertion by the average fitness cost of a spontaneous mutation. This has been estimated as \( d_o = 0.02 \) TE\(^{-1}\) (Mukai et al. 1972; Ohnishi 1977; Crow and Simmons 1983), and is a reasonable estimate in the early stages of TE spread when insertional mutagenesis is the
dominant fitness cost and selection has not yet eliminated TEs with higher fitness costs (Charlesworth 1991). For TEs already segregating in the population, the mean fitness cost has been estimated from simple models of population dynamics by sampling chromosomes from natural populations and assuming that selection and transposition have an equal and opposite influence on TE copy number at equilibrium (Nuzhdin et al. 1997; Nuzhdin 2000). This has led to an estimate of fitness cost per element between $10^{-5}$ and $10^{-3}$ (Nuzhdin et al. 1997). We consider a fitness cost of $d_w = 10^{-3}$ TE$^{-1}$ since studies on laboratory lines of *D. melanogaster* suggest a fitness cost toward the upper end of this range (Pasyukova et al. 2004).

Charlesworth (1991) has proposed models to understand the abundance of TEs in nature when chromosomal sites vary with respect to the effect of element insertions on host fitness. These models predict that TEs will become more highly represented at neutral or weakly selected sites, and that the mean fitness cost of a TE will decline over a time-scale “roughly equal to the reciprocal of the rate of transposition.” For a transposition rate of $u_1 = 0.1$ TE$^{-1}$ gen$^{-1}$, this corresponds to a fitness decline over a scale of ten generations. We therefore consider a halftime of $t_{1/2} = 5$ mosquito generations for the mean fitness cost of the TE to fall from $d_o$ to $(d_o + d_w)/2$. Following Charlesworth and Charlesworth (1983), we also consider a power of $p = 1.5$ to account for additional fitness costs at higher copy numbers, such as ectopic recombination (Langley et al. 1988).
A realistic range for the rate at which exogenous DNA becomes dissociated from an engineered TE construct can be deduced from experiments in which $P$ elements were screened for loss of a visible marker gene (Daniels et al. 1985; Levis et al. 1985; Robertson et al. 1988). Here, expression of the rosy and white marker genes is disrupted at a rate of $w' \approx 0.05$ TE$^{-1}$gen$^{-1}$. Data on the fractional decrease in replicative transposition rate due to the presence of exogenous DNA (Spradling 1986; Lampe et al. 1998), the fitness cost of mounting an immune response (Kraaijeveld and Godfray 1997; Moreira et al. 2004) and the selective advantage of being disease-resistant (Beier et al. 1999; Marrelli et al. 2007) can be used to form realistic ranges for the parameters $\delta$, $\sigma_-$ and $\sigma_+$ respectively.

Limiting our considerations to a single village where a disease vector is present, we estimate the initial proportion of disease vectors having one copy of the intact TE construct. The village of Banambani in Mali, West Africa serves as a well-studied example for the malaria vector An. gambiae (Taylor et al. 2001). Here, there is an estimated abundance of 2,000 mosquitoes of any given chromosomal form at the peak of the dry season (Taylor et al. 2001). Considering a release of 50 transgenic mosquitoes, each having a single copy of the intact TE construct, leads to an initial release proportion of $x_{(1,0)}(0) = 0.025$. 

38
2.2.3 Numerical analysis

As a measure of the spread of disease-resistance, we calculate the proportion of vectors having one of more copies of the intact construct over time. This quantity can be calculated numerically using the fourth order Runge-Kutta method. For the parameterization described above, the maximum number of TEs in the genome is $T = 16$. This can be used to reduce the number of equations being solved simultaneously.

2.3 Results

2.3.1 Dissociation rate

Using this system of differential equations, we consider the impact of varying the dissociation rate on the spread of the disease-resistance gene through a randomly mating vector population. In the simplest case (Figure 2.3), the resistance gene has no fitness consequences for the host and confers no transpositional handicap on the TE. The total genomic copy number reaches an equilibrium of four within 20 years and is very close to equilibrium after 10 years (Figure 2.3A-C). Dissociation is hypothesized to occur during the act of transposition, and hence the proportion of dissociated TEs increases most rapidly during the early stages of element spread. Once equilibrium has been reached, transposition still occurs at a slower rate to counteract the forces of selection, deletion and the random effects of gamete sampling. Consequently, the proportion of intact constructs decreases slowly until the resistance gene is lost from the population.
Figure 2.3 Element spread with dissociation. The resistance gene has no impact on host fitness or transposition rate. (A-C) Total element copy number reaches an equilibrium of four within 20 years, while the proportion of intact element copies decrease monotonically over time. Lines correspond to numerical solutions of the differential equation model for three different dissociation rates. (D) The prevalence of the disease-resistant vectors reaches a maximum within ten years and then decreases gradually. The maximum prevalence and duration of presence depend largely on dissociation rate.
The proportion of element copies that have an intact copy of the disease-resistance gene is largely dependent on the dissociation rate. For a dissociation rate of 0.001 TE\(^{-1}\)gen\(^{-1}\), almost all element copies contain an intact resistance gene, and the average number of intact gene copies per individual is significantly greater than three even after 40 years of spread (Figure 2.3A). This suggests that, for default parameters, a dissociation rate of 0.001 TE\(^{-1}\)gen\(^{-1}\) is acceptable for public health goals. A dissociation rate of 0.01 TE\(^{-1}\)gen\(^{-1}\) may also be tolerable since, in this case, intact and dissociated TE constructs are present at fairly equal numbers up to 40 years following release (Figure 2.3B). However, for a dissociation rate of 0.05 TE\(^{-1}\)gen\(^{-1}\), the vast majority of elements are devoid of an intact resistance gene well before the element construct has had a chance to reach equilibrium in the vector population (Figure 2.3C).

One indicator of an effective disease control strategy is the presence of a high number of disease-resistant vectors over a significant time span. Assuming that only a single copy of the resistance gene is necessary for its expression, the prevalence of disease-resistant vectors is equal to the proportion of vectors having at least one copy of the resistance gene. This quantity is plotted over time in Figure 2.3D for four different dissociation rates. For default parameters, disease-resistant vectors reach a maximum prevalence greater than 80% for dissociation rates less than 0.013 TE\(^{-1}\)gen\(^{-1}\) and always reach a maximum prevalence within the first decade of spread.
To put this into perspective, dissociation rates have been measured for $P$ elements by screening for the loss of a visible marker gene within the element (Daniels et al. 1985; Levis et al. 1985; Robertson et al. 1988) and are typically on the order of 0.05 TE$^{-1}$gen$^{-1}$. According to our model, disease-resistant vectors reach a maximum prevalence of 18% at this rate, again suggesting that a dissociation rate of 0.05 TE$^{-1}$gen$^{-1}$ is unacceptable for public health goals. On the other hand, for a dissociation rate of 0.01 TE$^{-1}$gen$^{-1}$, disease-resistant vectors reach a maximum prevalence close to 87% and remain at a high prevalence for several decades following release. This confirms that a dissociation rate of 0.01 TE$^{-1}$gen$^{-1}$ is potentially acceptable for public health goals.

An interesting conclusion that follows from Figure 2.3D is that, if the prevalence of disease-resistant vectors reaches a significantly high proportion in the vector population, then this prevalence will remain high for several decades following release. This implies that a high maximum prevalence of disease-resistant vectors will also tend to satisfy the requirement of a long duration of prevalence. For the remainder of this paper, we will use the maximum prevalence of disease-resistant vectors as an indicator of an effective disease-control strategy.

The qualitative predictions of this model are relatively insensitive to changes in lesser-known parameters such as the fitness cost of a new TE insert, $d_0$, the halftime for TE fitness cost to fall, $t_{1/2}$, and the proportion of transgenic mosquitoes released, $x_{(1,0)}(0)$. The release proportion of 2.5%, although motivated by the use of 50 transgenic
mosquitoes in a small village, also applies, for example, to the release of 5,000 transgenic mosquitoes in an endemic region with 200,000 wild-type mosquitoes. Increasing the release proportion to 10% decreases the time taken for the resistance trait to reach near-fixation and moderately increases the maximum prevalence that the resistance trait reaches in the population. Decreasing the release proportion to 1% has the converse effect (results not shown).

2.3.2 Transpositional handicap

We extend this simple model of dissociation to the case where the effector gene compromises the rate of replicative transposition (Figure 2.4A). Using a parameter, $\delta$, to account for the fractional reduction in transposition rate due to the presence of an effector gene, we model the impact of a transpositional handicap on the prevalence of disease-resistant vectors. Even a modest transpositional handicap can significantly reduce the maximum prevalence of disease-resistance. As seen in Figure 2.4A for a transpositional handicap of $\delta = 0.25$, the dissociation rate must be less than $0.005 \text{ TE}^{-1}\text{gen}^{-1}$ for the maximum prevalence of disease-resistant vectors to exceed 80%. When the transpositional handicap is $\delta = 0.5$ then the dissociation rate must be less than $0.001 \text{ TE}^{-1}\text{gen}^{-1}$ to achieve the same maximum prevalence; and when the transpositional handicap is $\delta = 0.95$ then dissociation can essentially not be tolerated.
Figure 2.4 Maximum prevalence of disease-resistant vectors. The resistance gene influences both transposition rate and host fitness. (A) The replicative transposition rate of an intact element construct is reduced by a fraction, $\delta$. Increasing the value of $\delta$ decreases the maximum prevalence of resistant vectors and makes dissociation requirements more demanding. (B) The fitness cost conferred by the resistance gene, $\sigma_-$, has a large impact on its maximum prevalence. (C) Dissociation requirements are considerably relaxed when the resistance gene confers a fitness benefit, $\sigma_+$. (D) A highly advantageous resistance gene is able to spread irrespective of a daunting dissociation rate or transpositional handicap.
To put these results into perspective, the RNAi-based strategy for resistance to multiple dengue serotypes currently consists of an effector gene, marker gene and regulatory elements amounting to ~6.5 kb (Kenneth Olson, personal communication). A Himar1 mariner element containing a construct of this size would suffer a reduction in transposition rate of $\delta = 0.97$ (Lampe et al. 1998). Typical constructs now used in laboratory experiments are on the order of ~5-6 kb (Marcelo Jacobs-Lorena, personal communication), and would suffer a transpositional handicap of $\delta = 0.95$ if attached to a Himar1 mariner transposase gene (Lampe et al. 1998). Any construct acceptable for release in the field will need to have at least two or three effector genes in order to counteract evolution of the disease agent and to increase effectiveness of the intervention (Marcelo Jacobs-Lorena, personal communication). These field release constructs will be on the order of ~8-12 kb and, contained within a Himar1 mariner transposon, would cause a transpositional handicap of $\delta = 0.99$ or more (Lampe et al. 1998).

Although these transposition rate calculations specifically apply to the Himar1 mariner element, they clearly make the point that the transpositional handicap of an effector gene is of great concern to any transposon-mediated control strategy. The transpositional properties of any candidate element must be measured in both the presence and absence of a transgene construct, and the size of the construct should be minimized wherever possible, while still satisfying the requirements for effective intervention. As seen in Figure 2.4A, a transpositional handicap of $\delta = 0.1$ is tolerable for dissociation rates less than $0.01 \text{ TE}^{-1}\text{gen}^{-1}$; and higher transpositional handicaps are tolerable for smaller
dissociation rates. An additional concern, however, is that gap repair becomes significantly less efficient for segments larger than 11 kb (Johnson-Schlitz and Engels 2006). This provides an additional incentive for minimizing the size of the transgene construct.

2.3.3 Fitness consequences

We consider the impact of a fitness cost associated with the effector gene using a parameter, $\sigma_-$, to account for the increase in vector death rate due to the effector gene. As with the case of a transpositional handicap, even a small fitness cost to the host can significantly reduce the maximum prevalence of disease-resistant vectors. As seen in Figure 2.4B, when the fitness cost of the effector gene is $\sigma_- = 0.005 \text{ TE}^{-1}$ then the dissociation rate must be less than $0.005 \text{ TE}^{-1}\text{gen}^{-1}$ for the maximum prevalence of disease-resistant vectors to exceed 80%. When the fitness cost of the effector gene is $\sigma_- = 0.01 \text{ TE}^{-1}$, the maximum tolerable dissociation rate to reach the same prevalence is $0.001 \text{ TE}^{-1}\text{gen}^{-1}$; and when the fitness cost is $\sigma_- = 0.05 \text{ TE}^{-1}$ then effective disease control is virtually impossible.

To account for a selective advantage associated with the effector gene, we introduce a parameter, $\sigma_+$, to model the mean increase in mosquito longevity due to the resistance trait. According to our model results (Figure 2.4C), when the fitness benefit of the effector gene is small ($\sigma_+ = 0.005 – 0.01$) then the dissociation rate required for effective
disease control is moderately relaxed. However, when the fitness benefit is large ($\sigma_+ = 0.05$) then disease control becomes feasible for a large range of dissociation rates – the maximum prevalence of disease-resistant vectors exceeds 80% for dissociation rates less than 0.05 TE$^{-1}$gen$^{-1}$ and exceeds 75% for dissociation rates less than 0.1 TE$^{-1}$gen$^{-1}$.

It is difficult to put these results into perspective due to the lack of reliable comparative fitness measurements in vector populations; however fitness costs have been documented in several insect species due to both mounting an immune response (Moret and Schmid-Hempel 2000; Ahmed et al. 2002) and maintaining the physiological machinery necessary to do so (Kraaijeveld and Godfray 1997; Koella and Boete 2002). Ahmed et al. (2002) found that egg production was reduced by 18.6% in An. gambiae mosquitoes whose immune system was artificially stimulated with lipopolysaccharides. Given that the proportion of mosquitoes infected with malaria parasites varies between 1-20% in Africa (Beier et al. 1999), this corresponds to a mean fitness cost of $\sigma_- = 0.002 – 0.04$ TE$^{-1}$, depending on the region of Africa. Measurements of the fitness cost of immune surveillance are not as well-documented as they are for immune deployment.

These considerations are meaningless if the corresponding selective advantage of the resistance trait is not also considered. Indeed, Hurd et al. (2005) found that the fitness cost of resistance to malaria is equal to the selective advantage of avoiding infection, and hence the two factors effectively cancel each other out. If this were the case, then fitness effects could be completely ignored.
A recent encouraging result is that transgenic mosquitoes have been engineered that exhibit no measurable fitness cost when fed on *Plasmodium*-free blood (Moreira *et al.* 2004) and exhibit a fitness benefit when fed on *Plasmodium*-infected blood (Marrelli *et al.* 2007). Although these results were obtained from cage experiments using the rodent malaria parasite *Plasmodium berghei*; it is conceivable that similar results could be obtained for the human malaria parasite *Plasmodium falciparum* and vector *An. gambiae*. Marelli *et al.* (2007) measured a 50% fitness benefit of being transgenic and a 35% fitness cost of being homozygous for the transgene; however this fitness measurement was on a population of mosquitoes that were all exposed to *Plasmodium*. In western Kenya, ~6.3% of *An. gambiae* mosquitoes are infected with *Plasmodium* (Shililu *et al.* 1998); while in northern Ghana, ~11% of *An. gambiae* mosquitoes are infected (Appawu *et al.* 2003). Assuming that a 50% fitness benefit is possible without overdominance, then these mosquito infection rates correspond to a mean selective advantage of $\sigma_+ = 0.03$ and $\sigma_+ = 0.05$ respectively.

Thus far, the effects of transpositional handicaps and fitness costs and benefits have been considered in isolation. However, of considerable interest is whether the selective advantage of resistance could counteract the detrimental effects of a transpositional handicap. In Figure 2.4D, we consider the impact of a fitness benefit on the spread of refractoriness when there is a transpositional handicap of $\delta = 0.95$. Interestingly, the maximum prevalence of disease-resistant vectors depends more on the fitness benefit
than on the dissociation rate. For a fitness benefit of $\sigma_+ = 0.05$, the maximum prevalence of disease-resistant vectors varies between 76% for a dissociation rate of 0.001 TE$^{-1}$gen$^{-1}$ and 67% for a dissociation rate of 0.1 TE$^{-1}$gen$^{-1}$. For fitness benefits of $\sigma_+ = 0.01$ or less, disease control is not feasible. The implication of this result is that a highly advantageous resistance gene ($\sigma_+ > 0.05$) can spread to a high prevalence in the vector population irrespective of a daunting dissociation rate or transpositional handicap.

2.4 DISCUSSION

Several commentaries on the use of TEs to drive resistance genes into vector populations have expressed concern that dissociation will ultimately lead to the resistance gene being lost from the population (Curtis 2003; Hahn and Nuzhdin 2004). One of the key conclusions of this modeling effort is that the epidemiological outcome of the gene drive strategy critically depends on the rate at which the resistance gene dissociates from the drive system. For slow rates of dissociation (0.01 TE$^{-1}$gen$^{-1}$) disease control is expected to occur within decades, while for intermediate rates of dissociation (0.05 TE$^{-1}$gen$^{-1}$) disease-resistant vectors are predicted to reach a much smaller maximum prevalence (less than 20%) and to be lost within decades. These contrasting scenarios highlight the importance of having good estimates of dissociation rates for the TEs and host species being considered for transgenic release. They also lead to the interesting prediction that, if dissociation inhibits the success of the project, then the resistance gene could be lost from the vector population within a human time frame.
Another recommendation of the model is that the size of the effector gene and associated regulatory elements should be restricted in order to limit the transpositional handicap of a TE attached to its transgenic load. While small transpositional handicaps may be tolerable, experiments on the Himar1 mariner element indicate that the effector gene constructs currently being developed in the laboratory would suffer a reduction in transposition rate of at least 95% if attached to this element. It is therefore important that, in addition to measuring the dissociation rate of the candidate elements, we also measure their transposition rates both in the presence and absence of the transgene construct. It is not sufficient that a transposable element replicates quickly within a genome – its transposition rate must also be relatively unaffected by exogenous DNA.

The fitness impact of the refractory gene is also of great importance in determining its fate in the vector population. Although a selective advantage may not be sufficient on its own to achieve fixation (Boéte and Koella 2003), when combined with a drive system it greatly improves the chances of success. A large selective advantage of refractoriness is a crucial factor in outweighing the daunting effects of dissociation and reduction in transposition rate. It is therefore essential to have comparative measurements of the fitness of both transgenic and wild-type mosquitoes both exposed and unexposed to the disease agent. If a transgenic release becomes more feasible, it will also be essential to make these measurements under taxing field conditions using the human vector and disease agent, and to develop innovative ways for differentiating these effects from
inbreeding depression and the process of transgenesis. Such measurements are not currently available.

2.4.1 Model limitations
Symptomatic of any mathematical analysis, simplifications have been made in this model that may compromise its predictions. Firstly, the choice of \textit{Herves} as a model element that has an average copy number of four (Subramanian \textit{et al.} 2007) is not representative of the vast number of TEs that are present at much higher copy numbers (Charlesworth \textit{et al.} 1992). This choice was motivated in part by the fact that \textit{Herves} rarely undergoes internal deletion, and also by the fact that model complexity greatly increases with element copy number. To assess the dynamics of TEs with higher copy numbers, another modeling framework should be proposed.

Secondly, in the model formulation all genomic elements are assumed to act independently during gamete formation, implying that all element copies are at least 50 centimorgans apart. This contradicts the tendency for TEs to jump locally rather than distally (Tower \textit{et al.} 1993; Zhang and Spradling 1993; Guimond \textit{et al.} 2003) and to home in on certain genomic regions (Guimond \textit{et al.} 2003). This is an important consideration since clustering of element copies can reduce the rate of element spread (Rasgon and Gould 2005); however in order to focus on the impact of dissociation, this consideration is beyond the scope of the present study.
2.4.2 Future directions

The model presented here is of special relevance to diseases such as malaria and dengue fever that are currently being considered for transposon-mediated control strategies (Scott et al. 2002). The present study could be supplemented by imbedding the spread of the resistance gene within a model of the epidemiology of malaria or dengue fever (Boëte and Koella 2002). This would give us a better idea of the exposure rate of vectors to the malaria parasite or dengue virus. It would also allow us to better predict the maximum prevalence of disease-resistant vectors and hence the rate of dissociation that is required for effective disease control.
3 CHAPTER 3
A BRANCHING PROCESS FOR THE EARLY SPREAD OF A TRANSPOSABLE ELEMENT IN A DIPLOID POPULATION

3.1 INTRODUCTION

Transposable elements (TEs) are particularly interesting genomic components due to their ability to transpose replicatively within a genome and hence spread throughout a population despite a fitness cost. The replicative ability of TEs has led to their widespread prevalence in the genomes of many taxa to the extent that various families of TEs account for ~90% of the Salamander genome (Marracci et al. 1996) and ~45% of the human genome (Biemont and Vieira 2004). In some cases, studies have shown TE spread to be very rapid. The P element is perhaps the best example of this, having spread through most of the wild-type Drosophila melanogaster population in the span of a few decades (Engels 1989).

Observations like these have inspired the idea of using TEs as drive mechanisms for spreading disease resistance genes into vector populations (Craig 1963; Curtis 1968). In recent years, advances in molecular biology (Atkinson and James 2002; Marrelli et al. 2007) and ecology (Scott et al. 2002) have allowed this idea to become a feasible control strategy for vector-borne disease. Hence it is of great epidemiological interest to have
some idea of the molecular and ecological conditions under which a TE will spread through a host population.

Models of TE dynamics have tended to focus on factors affecting element spread at the level of the genome and the host individual. Here, the replicative ability of the TE is weighed against its fitness cost to the host, its excision rate, and the amount by which transposition is suppressed with increasing copy number. Accounting for these factors, a number of models have been proposed to study the evolution and distribution of TEs in general (Charlesworth and Charlesworth 1983; Charlesworth and Langley 1986; Sawyer and Hartl 1986; Moody 1988; Brookfield and Badge 1997). These analyses have tended to focus on equilibrium distributions, comparing them against the distribution of TE copy number in nature.

The conditions necessary for TE spread have been investigated in a number of studies that have used a meiotic drive parameter to characterize the departure from Hardy-Weinberg equilibrium due to the presence of a TE (Ribeiro and Kidwell 1994; Kiszewski and Spielman 1998; Boete and Koella 2003). Other properties of TEs, such as the maximum transposition rate and the various mechanisms of transpositional regulation, have been investigated by Le Rouzic and Capy (2005) and Struchiner et al. (2005). Additionally, the tendency for TEs to transpose locally rather than distally has been addressed in simulations by Rasgon and Gould (2005).
In this paper, we apply a branching process model to examine the early stages of a TE invasion. Although our model has similar structural details to previously-published branching process models (Sawyer and Hartl 1986; Moody 1988; Basten and Moody 1991), it differs in that it is applicable to diploid host populations and explicitly examines the effects of population size changes on TE spread. The motivation for this extension is the planned use of a TE to drive an antimalarial gene into an *Anopheles gambiae* mosquito population in Africa. Despite this focus, the predictions of the model apply generally and similar strategies are currently being considered for the control of dengue in *Aedes aegypti* (Scott et al. 2002).

*An. gambiae* is the main vector of malaria in tropical Africa, and consequently it is the primary species being considered to host a TE for the purpose of malaria control. A major feature of the demography of *An. gambiae* is the existence of population size changes within and between years (Taylor et al. 2001; Manoukis 2006). The most dramatic of these occur between the dry season and the peak of the rainy season when the population size may change by several orders of magnitude within six months (Manoukis 2006). The population is also structured chromosomally by the existence of up to five chromosomal forms that may be partially or totally reproductively isolated (Touré et al. 1983; Coluzzi et al. 1985); and geographically by its concentration in discrete patches corresponding to villages (Touré et al. 1998).
We will restrict our attention to the spread of a TE through a single chromosomal form of *An. gambiae* in a single village. This will allow us to focus on the effects of temporal population structure on the release strategy. A successful transgenic release on the village scale must have two basic properties. Firstly, it must be efficient in establishing the effector gene in the host population (James 2005); and secondly, the drive system must work within a time frame acceptable to public health goals (Braig and Yan 2001). We will address these requirements by investigating the conditions that minimize the probability of TE loss and maximize the rate of TE spread. Within this context, we will make recommendations regarding the parameters that will be required in order to effectively control vector-borne disease.

### 3.2 Model formulation

We use a continuous-time multi-type branching process to model the early stages of TE spread through a randomly mating host population. Particles in the model are of *T* types corresponding to hosts infected with *i* copies of the TE, where *i* ∈ {1, 2,..., *T*} and *T* ≥ 1. Here, *T* can be approximated as the number of sites that will be occupied in the early stages of TE spread. A host having *i* TE copies may also be referred to as a “type-*i*” host. Uninfected hosts are not kept track of in this model because the majority of individuals belong to what may be thought of as a reservoir of uninfected hosts.
3.2.1 Reproduction

Mating between organisms cannot be modeled explicitly within the confines of a branching process model due to the requirement that the particles in a branching process must be independent (Lange 2003). However, since the vast majority of individuals are uninfected in the early stages of spread, we can imagine that all mating events involving infected hosts will be with individuals from the reservoir of uninfected hosts. This approximation is valid up to a prevalence of about 10% since, in a randomly mating population, less than 1% of all matings will involve two infected hosts at these prevalences. Mating between two infected hosts is modelled in Section 3.4.5 for prevalences greater than 10%.

We consider a budding model in which hosts do not die when they have offspring. This enables us to separate the birth and death rates for each host type. Here, hosts of type- $i$ are assumed to mate with uninfected host organisms from the reservoir to give rise to offspring at a constant rate $\theta$. The number of TEs in the offspring’s genome is then determined by: (a) whether a replicative transposition or element deletion event occurred in the cell that gave rise to the gamete contributed by the infected host; and (b) the number of TEs in this diploid cell that are passed on to the haploid gamete during meiosis.
3.2.2 Transposition and deletion

Transposition and deletion are modeled by assuming that a proportion $\alpha_i$ of gametes are derived from cells in which a replicative transposition event has occurred, while a proportion $\beta_i$ of gametes are derived from cells in which an element deletion event has occurred. The replicative transposition rate for a type-$i$ host, $\alpha_i$, is equal to the replicative transposition rate per TE in a type-$i$ host, $u_i$, multiplied by the number of TEs in the host genome, $i$ (i.e. $\alpha_i = iu_i$). Here, $u_i$ is generally a decreasing function of $i$ to account for suppression of transposition with increasing copy number (Weinreich et al. 1994; Wu and Morris 1999; Townsend and Hartl 2000). Similarly, the deletion rate for a type-$i$ host, $\beta_i$, is equal to the deletion rate per element, $v$, multiplied by the number of elements in the host genome, $i$ (i.e. $\beta_i = iv$), where $v$ is generally considered a constant.

A number of alternative models for the relationship between transposition rate and TE copy number were considered to assess their influence on the early spread of the TE. In each of these models, $a$ represents the transposition rate in a genome containing a single TE (i.e. $u_1 = a$). Our default model describes the replicative transposition rate, $u_i$, as a linear function of $i$,

$$u_i = a(1 - b(i-1)),$$

where $b$ is the fraction by which transposition rate falls off with each additional TE copy. The no transpositional regulation model corresponds to the linear transpositional regulation model with the parameter $b$ set to zero (i.e. $u_i = a$). The threshold regulation
model (Le Rouzic and Capy 2005) describes the case where no regulation occurs until a certain critical TE copy number is reached, \( i_r \), after which transposition is suppressed by a constant amount, \( a - a_r \), i.e.

\[
 u_i = \begin{cases} 
 a, & i \leq i_r \\
 a_r, & i > i_r 
\end{cases}, \quad a > a_r . \tag{3.2}
\]

Finally, the continuous regulation model (Le Rouzic and Capy 2005) describes the case where each additional TE copy reduces the transposition rate by a smaller increment, reaching a transposition rate of \( (a - a_r)/2 \) for a TE copy number of \( i_r \), and finally converging to a minimum transposition rate of \( a_r \) as copy number becomes very large, i.e.

\[
 u_i = (a - a_r)2^{-(i-1)/(i_r-1)} + a_r . \tag{3.3}
\]

Equation 3 is qualitatively similar to the hyperbolic equation for transposition rate as a function of TE copy number proposed by Charlesworth and Charlesworth (1983). The continuous regulation model is the commonly used when self-regulation of transposition is considered (Labrador and Corces 1997) and could be a consequence of the production of repressors (Lemaitre et al. 1993) or overproduction inhibition (Lohe and Hartl 1996). The threshold regulation model is less commonly used, although a number of molecular mechanisms have been proposed for its existence (Le Rouzic and Capy 2005).
3.2.3 Host fitness

The fitness cost associated with additional TE copies is modeled by varying the death rate, $\mu_i$, according to the number of TE copies, $i$, that the host genome contains. Here, $\mu_i$ is an arbitrary increasing function of $i$. A number of alternative models for the relationship between death rate and TE copy number were considered to assess their influence on the early spread of the TE. In each of these models, the death rate of a host containing no copies of the TE is set to 1, and $d$ represents the increase in death rate for a host containing a single TE in its genome (i.e. $\mu_i = 1 + d$). Our default model describes the death rate, $\mu_i$, as a linear function of $i$,

$$\mu_i = 1 + di.$$  

(3.4)

This model suggests that each additional genomic insertion has an equal and additive effect on host fitness during the early stages of TE spread, possibly due to the effects of insertional mutagenesis (Mackay 1989). The neutral insertion model is a special case of this model with the parameter $d$ set to zero, i.e. $\mu_i = 1$. The independent fitness cost model (Charlesworth and Charlesworth 1983) describes the case where each additional TE copy increases the host death rate by an identical factor, $1 + d$, i.e.

$$\mu_i = (1 + d)^i.$$  

(3.5)

Finally, the log-concave fitness cost model (Charlesworth and Charlesworth 1983) is similar to the linear fitness cost model with the exception that TE copy number is raised to the power $p$, where $1 < p < 2$, i.e.

$$\mu_i = 1 + di^p.$$  

(3.6)
In both the independent fitness cost model and the log-concave fitness cost model, the increase in host death rate rises with each additional TE copy. This is desirable since additional fitness costs arise at higher copy numbers, for example Montgomery et al. (1987) have proposed that ectopic recombination can occur between TEs of the same family present at different sites, and Brookfield (1991; 1996) has proposed that fitness costs arise primarily through the act of transposition at higher copy numbers. The log-concave fitness cost model is favored by Charlesworth and Charlesworth (1983) since it allows an equilibrium distribution of copy number to be obtained in their model of the population dynamics of TEs.

### 3.2.4 Gamete formation

The number of TEs in the haploid gamete is determined by the number of TEs in the diploid cell that are passed on during meiosis. For a diploid cell with \( i \) copies of the TE we assume, to a first approximation, that all of these TEs are far enough apart from each other that they segregate independently. Under this assumption, the probability of having \( j \) copies in a gamete is proportional to the number of ways of choosing \( j \) elements from a total of \( i \). Similarly, if a replicative transposition event has occurred in the diploid cell, then the probability of having \( j \) copies in the gamete is proportional to the number of ways of choosing \( j \) elements from \( i + 1 \), or from \( i - 1 \) if a deletion event has occurred.

The number of TEs in the offspring’s genome is equal to the number of TEs in the haploid gamete from the infected host because the other parent is uninfected by the TE so contributes no elements to the offspring’s genome.
Putting this all together within the framework of a multi-type continuous-time branching process (Lange 2003; Dorman et al. 2004), we have a finite number of independently acting infected hosts of \( T \) types that reproduce and die. Each host having \( i \) element copies lives an exponentially distributed length of time with death intensity \( \lambda_i = \mu_i + \theta \), and at the end of its life produces on average \( f_{ij} \) hosts with \( j \) element copies according to the equation,

\[
f_{ij} = \frac{1}{2^{i-1}} \binom{i}{j} \theta \beta_i + \frac{1}{2^i} \binom{i}{j} \theta (1 - \alpha_i - \beta_i) + \frac{1}{2^{i+1}} \binom{i+1}{j} \theta \alpha_i + \theta \lambda_i^{1,(i=j)}, \tag{3.7}
\]

where \( i, j \in \{1, 2, \ldots, T\} \).

### 3.2.5 Population size changes

Growth and decline in the total population size can be incorporated within the framework of the branching process simply by altering the value of the birth rate parameter, \( \theta \). This is related to the population growth rate of the reservoir of uninfected hosts, \( r \), by the following equation,

\[
\theta = 2(1 + r). \tag{3.8}
\]

The population size of uninfected hosts is then modelled using the exponential growth model,

\[
N_t = N_0 e^{rt}, \tag{3.9}
\]

where \( N_0 \) is the initial population size, \( N_t \) is the population size at time \( t \), and time is measured in generations. For exponential population growth, \( r \) should be positive; while
for exponential population decline, $r$ should be negative. For a fluctuating population size, the sign of $r$ should be time-dependent and oscillate in sign.

### 3.3 Parameter values

Table 3.1 contains the default parameter values for our model. At present there is little or no data regarding the post-integration behavior of the candidate TEs in *An. gambiae*, so most of these values are taken from measurements in other species, in particular *D. melanogaster* (Mackay *et al.* 1992; Nuzhdin *et al.* 1997; Maside *et al.* 2000; Guimond *et al.* 2003; Pasyukova *et al.* 2004).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Estimated value</th>
<th>Range</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>Replicative transposition rate</td>
<td>0.1 TE$^{-1}$ gen$^{-1}$</td>
<td>[10$^{-2}$, 0.3]</td>
<td>Charlesworth et al. (1992), Seleme et al. (1999)</td>
</tr>
<tr>
<td>$a_r$</td>
<td>Repressed transposition rate</td>
<td>10$^{-4}$ TE$^{-1}$ gen$^{-1}$</td>
<td>[10$^{-5}$, 10$^{-3}$]</td>
<td>Le Rouzic and Capy (2005), Nuzhdin (2000)</td>
</tr>
<tr>
<td>$b$</td>
<td>Transpositional regulation parameter</td>
<td>0.1</td>
<td>[-0.1, 0.1]</td>
<td>Biémont (1994), Pasyukova et al. (1998)</td>
</tr>
<tr>
<td>$i_t$</td>
<td>Threshold TE copy number</td>
<td>5</td>
<td>[2, 5]</td>
<td>Le Rouzic and Capy (2005), Biémont (1994)</td>
</tr>
<tr>
<td>$i_c$</td>
<td>TE copy number when transpositional repression is 50%</td>
<td>3</td>
<td>[2, 5]</td>
<td>Le Rouzic and Capy (2005), Biémont (1994)</td>
</tr>
<tr>
<td>$v$</td>
<td>TE deletion rate</td>
<td>$4 \times 10^{-6}$ TE$^{-1}$ gen$^{-1}$</td>
<td>–</td>
<td>Nuzhdin et al. (1997), Maside et al. (2000)</td>
</tr>
<tr>
<td>$d$</td>
<td>Fitness cost of TE</td>
<td>0.02 TE$^{-1}$</td>
<td>[10$^{-3}$, 0.4]</td>
<td>Mackay et al. (1992), Irvin et al. (2004)</td>
</tr>
<tr>
<td>$p$</td>
<td>Power determining rate of decrease in host fitness</td>
<td>1.5</td>
<td>[1, 2]</td>
<td>Charlesworth and Charlesworth (1983)</td>
</tr>
<tr>
<td>$r$</td>
<td>Population growth rate</td>
<td>0.2 gen$^{-1}$</td>
<td>[-0.4, 0.4]</td>
<td>Taylor et al. (2001), Manoukis (2006)</td>
</tr>
<tr>
<td>$N_0$</td>
<td>Initial population size</td>
<td>$2 \times 10^3$</td>
<td>[2 $\times 10^3$, 2 $\times 10^6$]</td>
<td>Taylor et al. (2001), Li et al. (1999)</td>
</tr>
<tr>
<td>$n_r$</td>
<td>Release size</td>
<td>50</td>
<td>[1, 10$^7$]</td>
<td>Charles Taylor (pers. comm.)</td>
</tr>
<tr>
<td>$i_r$</td>
<td>Copy number at release</td>
<td>1</td>
<td>[1, 4]</td>
<td>-</td>
</tr>
<tr>
<td>$t_r$</td>
<td>Time of release</td>
<td>0 gen</td>
<td>[0, 22.8]</td>
<td>Mahamoudou Touré (pers. comm.)</td>
</tr>
</tbody>
</table>

Table 3.1 Parameter values, ranges and source references for models used in Chapter 3.
3.3.1 Transposition

The replicative transposition rate with only a single TE in the genome, $a$, has been estimated indirectly by sampling chromosomes from natural populations (Nuzhdin et al. 1997; Charlesworth et al. 1992; Biemont et al. 1997), and directly by scoring insertion sites in laboratory populations and rescoring after many generations (Maside et al. 2000). This has led to estimated replicative transposition rates ranging from $\sim 10^{-5}$ per element per generation for many TEs in *D. melanogaster* (Pasyukova et al. 2004; Charlesworth et al. 1992) and *D. simulans* (Nuzhdin 1995; Vieira and Biemont 1997) up to transposition rates greater than 0.1 per element per generation for very active individual *I* elements (Seleme et al. 1999; Vasilyeva et al. 1999).

The upper bound of the replicative transposition rate, $a$, was inflated in our analysis to account for the fact that most replicative transposition rates in the literature are estimated from host genomes with copy numbers much greater than one. This underestimates the transposition rate in genomes having a single TE since transposition tends to be suppressed at higher copy numbers and some TEs tend to show bursts of transpositional activity following introduction into a new host population (Yang et al. 2006).

3.3.2 Transpositional regulation

For the linear model of transpositional regulation, we consider both positive and negative values of the fractional reduction in transposition rate with increasing copy number, $b$. While most studies show that replicative transposition rate is negatively correlated with
copy number (Biemont 1994; Jensen et al. 1999; Biemont et al. 1999), studies of copia and Doc retrotransposons in D. melanogaster suggest that replicative transposition rate is positively associated with copy number (Nuzhdin et al. 1996; Pasyukova et al. 1998).

For the threshold and continuous regulation models of transposition, we need an estimate of the repressed rate of transposition, \( a_r \), as well as the copy number at which this repression occurs, denoted by \( i \), for the threshold model and \( i_c \) for the continuous regulation model. Since we are interested in the early stages of TE spread when copy numbers tend to be very small, we choose values of these parameters that are consistent with the value of parameter \( b \).

### 3.3.3 Deletion

Element deletion events are so rare that they are often not observed in laboratory line experiments (Nuzhdin and Mackay 1994), and rates of element deletion, \( v \), are thought to be at least two orders of magnitude less than the TE’s baseline transposition rate (Nuzhdin et al. 1997; Nuzhdin 2000). Maside et al. (2000) calculated a pooled excision rate of \( 3.95 \times 10^{-6} \) per element per generation from a laboratory study of 11 families of TEs in D. melanogaster. This is in good agreement with the pooled estimate of \( 4.05 \times 10^{-6} \) per element per generation calculated by Nuzhdin et al. (1997) from other laboratory line experiments (Eggleston et al. 1988; Harris 1963; Nuzhdin and Mackay 1995).
3.3.4 Host fitness

The parameter $d$ represents the increase in host death rate with each additional TE copy in the genome and is approximately equal to the decrease in host fitness due to each additional TE copy. Following Mackay et al. (1992), we estimated the fitness cost of a new genomic insertion by the average fitness cost of a spontaneous mutation. This has been estimated as $\sim 0.02$ per element (Mukai et al. 1972; Ohnishi 1977; Crow and Simmons 1983) and is a reasonable estimate in the early stages of TE spread when insertional mutagenesis is the dominant fitness cost and selection has not yet eliminated TEs with higher fitness costs (Charlesworth 1991). For the log-concave model of host fitness, we consider a power of $p = 1.5$ following Charlesworth and Charlesworth (1983).

3.3.5 Population size changes

We are interested in both long-term population growth, and the seasonal population size changes that occur in An. gambiae in Banambani, Mali (Figure 3.1). We have chosen to focus on the village of Banambani as a case study for the African continent since the ecology of An. gambiae has been intensively studied there (Taylor et al. 2001; Lanzaro et al. 1998; Tripet et al. 2005; Touré et al. 1994).
Figure 3.1 Graph showing the number of *An. arabiensis* and *An. gambiae* M and S molecular forms typed from collections taken in Banambani, Mali during 2005 (Manoukis 2006). The difference in population sizes between wet and dry seasons is clearly evident in both M and S molecular forms and can be modeled approximately by alternating exponential growth and decline (overlaid curve).
Recent collections taken in Banambani suggest that peak population densities during the wet season are around 100 times those during the dry season (Manoukis 2006), while conservative estimates are that seasonal population densities differ by a factor of ten (Taylor et al. 2001). The generation time of mosquitoes is around 16 days, following observations that mosquito maturation time is 12 days and the first eggs are laid four days following maturation (Mahamoudou Touré, personal communication). Assuming a ten-fold population size increase over six months (11.4 mosquito generations), this yields a population growth rate of \( r = 0.2 \) per generation, and a growth rate of \( r = -0.2 \) per generation for the corresponding population decline (Figure 3.1).

These seasonal population size changes are superimposed over historical population expansions. Evidence from microsatellite allele size data suggests a recent population expansion in the malaria vectors An. gambiae and An. arabiensis (Donnelly et al. 2001). Statistical tests on this data suggest an expansion that occurred on the order of \( 10N_e \) generations ago (where \( N_e \) is the effective population size prior to the population expansion) and may be contemporaneous with excessive penetration of agriculture into the African forest 4000 years ago (Coluzzi et al. 2002; Ayala and Coluzzi 2005). Similar population expansions have been inferred for D. melanogaster populations. For example, the expansion of D. melanogaster into northern Africa and Eurasia \( \sim 10^4 \) years ago (Baudry et al. 2004; Thornton and Andolfatto 2006) is consistent with nucleotide data and coalescent simulations suggesting a \( 10^2 \) to \( 10^6 \)-fold population size increase occurring around the same time (Pool et al. 2006).
3.3.6 Population size

Mark-release-recapture experiments on *An. gambiae* conducted in Banambani over five years found an average peak abundance of $\sim 6 \times 10^4$ (Taylor *et al.* 2001; Taylor and Manoukis 2006). Dividing this by three to account for the three chromosomal forms in Banambani yields a peak abundance of $\sim 2 \times 10^4$ per chromosomal form, or a population size prior to growth of $N_0 = 2 \times 10^3$. These population sizes are relatively small compared to the effective ancestral population size of *D. melanogaster*, which has been shown to be $\sim 2.0 \times 10^6$ (Li *et al.* 1999).

3.3.7 Release strategy

In analyses where we considered the natural invasion of a TE following a horizontal transfer event, we initialized the branching process with a single host having a single copy of the TE. In our analyses of the genetic control strategy, we initialized the branching process with $n_r$ transgenic hosts each having $i_r$ TE copies released $t_r$ generations following the beginning of the population growth phase. The dynamics of TE spread were then monitored in continuous time up to a prevalence of 10%.
3.4 **ANALYSIS**

3.4.1 **Critical parameter values for TE spread**

One of the features of an ideal drive system listed by Braig and Yan (2001) is the requirement that the drive must be strong enough to compensate for its inherent fitness cost. If the drive of a TE outweighs its fitness cost, then the TE has a chance to spread through the population and possibly reach fixation; however if the fitness cost outweighs the drive, then the TE will eventually be lost from the population. Within the framework of a branching process model, the possibility of TE spread corresponds to the case where the branching process is supercritical. A supercritical process has a basic reproductive number greater than one (Lange 2003).

The basic reproductive number is equal to the average number of offspring of the same type that a particle eventually generates (Van den Driessche and Watmough 2002). If we restrict ourselves to the very early stages of TE spread when the vast majority of infected hosts have only one or two copies of the TE, then the reproductive number of a type-1 host, $R_{1,1}$, can be defined recursively as,

$$R_{1,1} = f_{1,1} + f_{1,2} R_{2,1}.$$  \hspace{1cm} (3.10)

Here, $f_{1,1}$ is the average number of offspring of a type-1 host that are of type-1, $f_{1,2}$ is the average number of offspring of a type-1 host that are of type-2, and $R_{2,1}$ is the expected number of type-1 hosts that a type-2 host eventually generates. Following similar reasoning, $R_{2,1}$ can be defined recursively as,
\[ R_{2,1} = f_{2,1} + f_{2,2}R_{2,1}, \]  

(3.11)

where \( f_{2,1} \) and \( f_{2,2} \) are similarly defined. Equation 3.11 can then be rearranged and substituted into Equation 3.10 to obtain the basic reproductive number for a type-1 host,

\[ R_{1,1} = f_{1,1} + \frac{f_{1,2}f_{2,1}}{1-f_{2,2}}. \]  

(3.12)

Assuming linear models of transpositional regulation and host fitness, we can calculate the basic reproductive number by substituting Equations 3.1, 3.4, 3.7 and 3.8 into Equation 3.12. Straightforward algebra then shows that for a type-1 host we have the basic reproductive number,

\[
R_{1,1} = \frac{(r+1)\left(a^2(b-1)(r+1) - 2a(r+1)(b(v-3) - v + 2) \right)}{2(d+2r+3)(4d-r + a(b-1)(r+1)+2(r+1)+1)}.
\]  

(3.13)

The drive of the TE outweighs its fitness cost when \( R_{1,1} > 1 \).

Before analyzing the full implications of this equation, first let us consider a simplified scenario in which transpositional regulation and element deletions are negligible \( (b = v = 0) \) and the population size is constant \( (r = 0) \). This leads to the basic reproductive number,

\[
R_{1,1} = \frac{4a + a^2 - 24d - 6}{2(d+3)(a-4d-1)},
\]  

(3.14)

which is greater than one when,

\[
a > d + 1 - \sqrt{1-7d^2}.
\]  

(3.15)
For small fitness costs \((d < 0.01\) per element), Equation 3.15 leads to the well known result that TE drive outweighs fitness cost when,

\[
a > d ,
\]

and hence selection and transposition have an equal and opposite influence on TE copy number at equilibrium (Nuzhdin et al. 1997). For larger fitness costs \((d > 0.01\) per element), the magnitude of the transposition rate must be slightly greater than the magnitude of the fitness cost in order for TE spread to be critical.

When transpositional regulation and element deletions are not negligible and the population size is not constant, we must use Equation 3.13 to determine the conditions for TE spread to be critical (Figure 3.2). The critical balance between transposition rate and fitness cost is highly sensitive to changes in the population growth rate. This sensitivity is most visible when the transposition rate and fitness cost are small. For moderate population growth \((r = 0.005\) per generation), TE spread is always supercritical for fitness costs less than 0.005 per element. For moderate population decline \((r = -0.005\) per generation), TE spread is always subcritical for transposition rates less than 0.005 per element per generation. The criticality of TE spread is relatively insensitive to the degree of transpositional regulation (results not shown), the model of transpositional regulation, and the model of host fitness with increasing copy number (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).
Figure 3.2 Critical parameter values for TE spread. The default parameters from Table 3.1 are used. Curves correspond to values of transposition rate and fitness cost for which TE spread is critical. For the region of parameter space to the left of each curve TE spread is supercritical; while to the right of each curve TE spread is subcritical. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*. 
The case where \( r = 0.2 \) per generation corresponds to the seasonal population growth phase of \textit{An. gambiae}. During this phase, TE spread is supercritical for all transposition rates when the fitness cost is less than 0.2 per element. The implication of this result is that the conditions for supercritical TE spread are less restrictive during periods of population growth. This may enable the TE to become established in its host population prior to a subsequent population decline.

It should be noted that, in the case of \textit{An. gambiae}, population growth is seasonal; while the above calculation applies to a period of continuous population growth. The case of seasonal population growth will be treated in Sections 3.4.3 through 3.4.5. Additionally, knowledge of the conditions necessary for TE spread does not, on its own, confer knowledge of the rate of TE spread and the probability of TE loss. These quantities will be calculated in Sections 3.4.2 through 3.4.4.

**3.4.2 TE loss in a natural population**

When TE spread is supercritical the eventual extinction of the TE is uncertain, however the probability that the TE is lost from the population may still be very high. Following a horizontal transfer a TE will be present in its new host population as a single copy in only a single host. At such a low prevalence, the probability of a recently transferred TE being lost through genetic drift is considerable, even if it has a very high transposition rate.
Considering the simplified scenario in which transpositional regulation and element deletion are negligible and population size is constant, Kaplan et al. (1985) calculated the probability that a single TE is lost from its host population, $e_1$, to be the smallest solution of the equation,

$$
e_1 = \exp(-1 - e_1 \exp(-(1 - e_1) u_1)) .$$

Here, $u_1$ is the transposition rate in a host carrying only a single element. This is in good agreement with simulation results of Le Rouzic and Capy (2005) for transposition rates of $u_1 < 1$ per element per generation.

To determine the probability that the TE is lost from the host population when the population size is not constant, we use the branching process model described by Equations 3.1-3.9. We begin by defining the probability generating function for the process (Dorman et al. 2004). This is a function of the vector $s = (s_1, s_2, ..., s_T)$ and is defined as,

$$P_i(s) = \sum_{j} p_j s^j = \sum_{j} p_j s_1^{j_1} s_2^{j_2} ... s_T^{j_T} ,$$

where $p_j$ is the probability that a type-$i$ particle gives rise to $j_1$ type-1 particles, $j_2$ type-2 particles, and so on, and $j$ is defined as the vector $j = (j_1, j_2, ..., j_T)$. Substituting the $p_j$ terms from the branching process into Equation 3.18, we have the probability generating function,

$$P_i(s) = \frac{\mu_i}{\lambda_i} + (f_{i,0} + \sum_{j=1}^{T} f_{i,j-1}) s_i + \sum_{j=1}^{T} f_{i,j} s_i s_j - \frac{2(1+r)}{\lambda_i} s_i^2 ,$$

76
where \( i \in \{1, 2, \ldots, T\} \). The probability that the TE is eventually lost from the population beginning with a single type-1 host, \( e_1 \), is then given by the smallest solution of the system of \( T \) simultaneous equations (Harris 1963),

\[
P_i(e) = e_i \forall i,
\]

(3.20)

where \( e_i \) is the probability that a type-\( i \) host is eventually lost from the population. Note that we have calculated the asymptotic loss probabilities here, although the model is only valid up to a prevalence of 10%. This paradox is averted by considering that beyond a prevalence of 10% the TE has a high enough presence in the population that it is very unlikely to go extinct.

For a population of constant size, TE loss is uncertain when the transposition rate for a single TE copy is greater than the fitness cost of a single genomic insertion (Figure 3.3A). This is in agreement with the prediction of equation 16. Loss probability falls as transposition rate increases and the fitness cost decreases, but even in the generous case of a neutral insertion with a transposition rate of 0.1 per element per generation, the loss probability is still as high as 92% (Figure 3.3A). This suggests that the majority of TEs will become extinct following their introduction and many introductions are likely to occur before one of these TEs colonizes a new species.
Figure 3.3 Asymptotic probabilities of TE loss as a function of replicative transposition rate. The default parameters from Table 1 are used. (A) The fitness cost of a genomic insertion is varied. The case where $d = 0$ per element is comparable to the solution of Kaplan et al. (1985). (B) Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of An. gambiae.
To test for consistency with the framework of Kaplan *et al.* (1985), the smallest solution to Equation 3.17 is shown (Figure 3.3A). Since Kaplan *et al.* (1985) considered a neutral insertion; this is best compared to the branching process when $d = 0$ per element. These curves are in good qualitative agreement with each other, although the solution of Kaplan *et al.* (1985) produces slightly smaller loss probabilities. Incidentally, the loss probabilities of Kaplan *et al.* (1985) are also slightly smaller than the simulated results of Le Rouzic and Capy (2005), suggesting that the solution of Kaplan *et al.* (1985) may slightly underestimate the probability of TE loss.

The prospects for a TE to colonize a new host species are dramatically improved during conditions of population growth (Figure 3.3B). The impact of population growth on reducing TE loss probability is most visible for population growth on the order of $r = 0.1$ per generation. At this growth rate, TE loss probability is ~7.5% lower than for the case of a constant population size over the entire range of transposition rates inferred from nature. When the rate of population growth is $r = 0.2$ per generation, as is the case during the seasonal population growth of *An. gambiae*, TE loss probability is lowered by an additional 7.5%. Combining a high transposition rate with conditions of population growth will further improve the odds of the TE colonizing a new host species. For a comparable rate of population decline ($r = −0.1$ per generation) TE loss probability is ~7.5% higher than for the case of a constant population size, and TE loss is certain for transposition rates less than 0.11 per element per generation.
The sensitivity of TE loss probability to the model of transpositional regulation and host fitness (Equations 3.1-3.6) was also tested under a variety of parameterizations (supplemental Appendices at http://johnmm.bol.ucla.edu/te/). The implication of these results is that whether the TE is lost from the population or increases in prevalence exponentially is relatively independent of the way in which transposition rate and fitness cost change with increasing copy number. What happens at higher copy numbers will no doubt affect the distribution of element copy number in the later stages of spread, but whether the TE spreads or not seems to be primarily determined by what happens when the genomic copy number is one.

3.4.3 TE loss following a transgenic release

Given the impact of population size changes on the eventual loss of a newly transferred TE, the implications this has for genetic control strategies in An. gambiae populations is of great interest. Indeed, one of the features of an ideal drive system listed by James (2005) is the requirement that the drive system be efficient in establishing an effector gene in the population. Therefore the TE loss probability following a transgenic release should be exceptionally small. To account for the seasonal population size changes of An. gambiae, we calculate the probability that the TE is lost from the An. gambiae population one year following its release. This allows us to account for a full cycle of population growth and decline.
We consider the release of a single transgenic mosquito \( t_r \) generations following the beginning of the population growth phase having \( i \) copies of the TE. To calculate the probability that this TE is lost from the population one year following release, we derive a differential equation for the loss probability as a function of time, \( e_i(t) \), and solve this at a time one year following release subject to the initial condition \( e_i(0) = 0 \forall i \). During the first season, the loss probabilities are characterized by the system of nonlinear ordinary differential equations (Lange 2003),

\[
\frac{de_i(t)}{dt} = -\lambda_i e_i(t) + \lambda_i P_i(e(t)) , \tag{3.21}
\]

where \( e(t) \) is defined as the vector \( (e_1(t), e_2(t), ..., e_T(t)) \).

During the second season, the loss probabilities can be calculated by characterizing the distribution of copy numbers at the end of the first season at time \( t_1 \), and calculating the probability that each of these lineages of infected individuals becomes extinct during a time \( t_2 = t - t_1 \) into the second season. This is encapsulated by the multivariate generating function,

\[
Q_i(t_1, e(t_2)) = \sum_{k} \Pr(Z_{i_1} = k \mid Z_0 = u_i) e(t_2)^k , \tag{3.22}
\]

where \( Z_i = (Z_{i,1}, Z_{i,2}, ..., Z_{i,T}) \) is the vector of particle counts at time \( t \geq 0 \), \( u_i \) is the standard unit vector whose \( i \)th entry is equal to one, and \( e(t_2) \) is the vector of loss probabilities whose entries are calculated using the population growth rate parameter for
the second season. This generating function is characterized by the system of nonlinear
ordinary differential equations (Lange 2003),
\[
\frac{dQ_i(t,e(t_2))}{dt} = -\lambda_i Q_i(t,e(t_2)) + \lambda_i P_i(Q(t,e(t_2))) , \tag{3.23}
\]
where \(Q(t,e(t_2))\) is defined as the vector \((Q_1(t,e(t_2)), Q_2(t,e(t_2)), \ldots, Q_r(t,e(t_2)))\).

Solving this at a time \(t_2\) into the second season subject to the initial condition
\[Q_i(0,e(t_2)) = e_i(t_2) \forall i\] gives the probability of TE loss at time \(t\), where \(t = t_1 + t_2\). This
procedure can be iterated to calculate the TE loss probabilities during subsequent
seasons.

Next, having calculated the TE loss probability for a single transgenic mosquito one year
following release, this result can be extended to a transgenic release consisting of \(n\)
mosquitoes, where \(n\) is defined as the vector \((n_1, n_2, \ldots, n_r)\) and \(n_i\) represents the number
of mosquitoes having \(i\) copies of the TE at the time of release. This follows simply from
the independence of particles so that the collective extinction probability is \(e(t)^n\). For a
release consisting of \(n\) transgenic hosts each having \(i\) copies of the TE, the extinction
probability is \(e_i(t)^n\).

One of the main results of this calculation concerning the release strategy is that the
model favors a transgenic release immediately following the dry season when the \textit{An.}
gambiae population begins to grow (Figure 3.4A). A release at this time reduces the
probability that the TE is lost from the population in the first year following its release.

The model also suggests that, in the event that knowledge of the population growth rate is imprecise, the control strategy should bias towards a release after the beginning of the population growth phase. This follows from observations that TE loss probabilities are smaller for a transgenic release at the beginning of the rainy season than at the end of the dry season (Figure 3.4A). A release under these conditions in a population of fluctuating size has a better chance of persisting than an identical release in a population of constant size.
Figure 3.4 Probabilities of TE loss one year following release. The default parameters from Table 3.1 are used, with the exception that $n_r = 1$. (A) The *An. gambiae* host population is assumed to grow for six months, and then decline for six months. Release time is measured in generations following the beginning of the population growth phase. (B) The replicative transposition rate is varied. (C) Release size is increased to 50. (D) Copy number per individual at release is increased to four.
Assuming a release at the beginning of the rainy season, Figure 3.4B then depicts the TE loss probability as a function of time for a variety of transposition rates. Clearly, a TE with a high transposition rate is less likely to be lost from the host population. At the end of the year, a TE with a transposition rate of 0.01 per element per generation has a loss probability of 91% while a TE with a transposition rate of 0.1 per element per generation has a loss probability of 83%. TE loss probabilities one year following release are relatively insensitive to the model of transpositional regulation and host fitness (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).

The simplest and most effective way to improve the efficiency of a TE drive strategy is to release more transgenic mosquitoes (Figure 3.4C). According to model predictions, increasing the release size to 10 reduces the loss probability to 15% one year following release. For a release size of 25 the loss probability falls below 1%. Another way to increase the number of TEs released into the population is to increase the copy number of the introduced mosquitoes (Figure 3.4D). This reduces the loss probability of the TE, but to a lesser extent than increasing the release size by the same amount. Therefore we favor the release of a greater number of transgenic mosquitoes each having a single copy of the TE.

3.4.4 Rate of TE spread following a transgenic release

To satisfy the second requirement of Braig and Yan (2001) that the drive system must work within a public health time frame, we consider the influence of population growth
and various model parameters on reducing the time taken for the TE to reach a proportion of 10% in the *An. gambiae* population.

The most intuitive and complete information derivable from the branching process is the mean and variance in the number of disease vectors having different TE copy numbers over time. For an initial release consisting of *n* mosquitoes, this is given by the matrix exponential (Dorman *et al.* 2004),

\[ E(Z_t) = ne^{\Theta t}, \quad (3.24) \]

where \( Z_t \) is the random vector of particle counts whose entries \( Z_{i,j} \) represent the number of disease vectors having \( i \) copies of the TE at time \( t \), and \( \Theta \) is the branching process matrix whose entries are given by \( \lambda_i(1_{i=j}) \). The above equation is valid for the first season, however during the second season the particle counts are given by,

\[ E(Z_t) = ne^{\Theta_1 t}e^{\Theta_2 t}, \quad (3.25) \]

where \( \Theta_1 \) and \( \Theta_2 \) are the branching process matrices calculated using the first and second season population growth rates respectively. This procedure can be iterated to obtain the vector of particle counts during subsequent seasons.

The variance of the particle count vector during the first season can be calculated following the methodology in the Appendix of Dorman *et al.* (2004). This applies to the variance matrix \( V_i(Z_t) \) for a transgenic release involving a single transgenic mosquito.
having \( i \) copies of the TE. For a release consisting of \( n \) transgenic mosquitoes, the variance matrix follows from the independence of clans (Lange 2003), and is given by,

\[
Var(Z_i) = \sum_{i=1}^{T} n_i V_i(Z_i) .
\] (3.26)

During the second season, the variance matrix can be calculated by conditioning on the state of the particle count vector at the end of the first season. Simple algebraic manipulation then gives,

\[
Var(Z_i) = \sum_{i=1}^{T} E(Z_{t,i}) V_i(Z_{t,i}) + \sum_{i=1}^{T} Var(Z_{t,i}) E_i(Z_{t,i}) E_i(Z_{t,i})^*,
\] (3.27)

where \( E_i(Z_i) \) represents the mean particle count vector at time \( t \) given a single transgenic mosquito having \( i \) copies of the TE at time 0. This approach can be iterated to calculate the variance in particle counts over subsequent seasons.

Applying this approach to a release of 50 transgenic mosquitoes, we calculated the mean and variance in the number of each host type over time. The mean number of transgenic hosts increases exponentially during the rainy season and then decreases exponentially during the subsequent dry season, reflecting the dynamics of the uninfected \( An. gambiae \) population (Figure 3.5A). The replicative behavior of the TE also leads to transgenic \( An. gambiae \) being over five times more prevalent one year following release. Of the 260 expected transgenic \( An. gambiae \) at the end of the dry season, 236 have a single copy of the TE, 23 have two TE copies, and 1 has three TE copies (Figure 3.5A). This Poisson-like distribution of copy number is largely a consequence of the increase in copy number
due to replicative transposition being counteracted by the reduction of copy number due to gamete formation and mating with uninfected hosts.

Figure 3.5 Rate of TE spread for the first year following a transgenic release. The default parameters from Table 1 are used. (A) The number of transgenic *An. gambiae* with different TE copy numbers are tracked over time. Mean numbers of hosts are shown as bold lines. The variance of host numbers having a single TE is illustrated by distributing points about the mean according to a normal distribution with variance equal to the variance in host numbers (note that the distribution of host numbers does not necessarily obey a normal distribution). (B) The proportion of transgenic *An. gambiae* is tracked over time. Population growth rate is varied. (C) Replicative transposition rate is varied. (D) The waiting time for the TE to reach a proportion of 10% in the host population is calculated as a function of replicative transposition rate. Release size is varied.
In contrast to the results for TE loss probabilities, the rate of population growth has no influence on the proportion of hosts having the TE one year following its release (Figure 3.5B). For higher population growth rates, the proportion of individuals having the TE increases more quickly during the rainy season and more slowly during the dry season, eventually ending at the same proportion regardless of population growth rate over the period of a year. Despite this independence, a release at the beginning of the population growth phase is recommended since it results in the TE having the highest initial proportion in the population.

Whether the TE increases in proportion in the first year following its release is largely determined by the difference between its replicative transposition rate and fitness cost (Figure 3.5C). The TE increases its proportion in the host population when the magnitude of its transposition rate exceeds its fitness cost, which is an interesting parallel to Equation 3.16. The requirement of a rapid spread (Lange 2003) is most effectively achieved by ensuring that the transposition rate is sufficiently high. For example, the TE reaches a proportion of 10% within a year of its release when its transposition rate is greater than 0.087 per element per generation (Figure 3.5C). The rate of TE spread one year following release is relatively independent of the model of transpositional regulation and host fitness (supplemental Appendices at http://johnmm.bol.ucla.edu/te/).

Finally, the waiting time for the mean number of transgenic An. gambiae to reach a proportion of 10% in the population is predominantly a function of replicative
transposition rate and release size (Figure 3.5D). While the waiting times starting with release sizes of 50 and 100 are relatively similar, the waiting times beginning with a single transgenic host are significantly longer. Similar comparisons over a smaller range of release sizes are shown in the supplemental Appendices (http://johnmm.bol.ucla.edu/te/). The relevance of these comparisons is that a more effective strategy than releasing hundreds of mosquitoes into one population might be to release fewer mosquitoes into multiple villages or into multiple chromosomal forms in the same village. This should reduce the fixation time across a larger geographical area.

3.4.5 Rate of TE spread beyond a prevalence of 10%

One major restriction of the branching process model is the assumption that host organisms do not interact during mating. This is not a good approximation for prevalences of transgenic hosts above 10%; hence another modeling framework is required to make inferences regarding the ultimate fate of the TE in the host population.

We consider a system of ordinary differential equations (ODEs) in which the proportion of hosts having \( k \) copies of the TE is kept track of by its own differential equation,

\[
\frac{dx_k(t)}{dt} = \theta(t) \sum_{i=0}^{T} \sum_{j=0}^{T} p_{ij} x_i(t) x_j(t) - \mu_k x_k(t) .
\]

Here, \( x_k(t) \) represents the proportion of hosts having \( k \) TE copies, where \( k \in \{0,1,\ldots,T\} \) and \( T \) can be approximated as the number of genomic sites that will be occupied when the TE reaches equilibrium in the host population. This equation simply describes the
difference in birth and death rates for hosts having $k$ TE copies – $\theta(t)$ is the overall birth rate parameter; $p_{ijk}$ is the probability that, when a host having $i$ TE copies mates with a host having $j$ TE copies, the offspring has $k$ TE copies; and $\mu_k$ is the death rate of a host having $k$ TE copies, as previously described.

The form of the birth rate term $\theta(t)$ is given by setting the total birth rate equal to the total death rate. Then, to calculate the probability that an offspring has $k$ TE copies, we first consider the probability, $p_{im}$, that a diploid host having $i$ TE copies produces a haploid gamete having $m$ TE copies. This is almost identical to the quantity $f_{im}$ for the branching process and is given by,

$$p_{im} = \frac{1}{2^{i-1}} \binom{i-1}{m} \beta_i + \frac{1}{2^i} \binom{i}{m} (1 - \alpha_i - \beta_i) + \frac{1}{2^{i+1}} \binom{i+1}{m} \alpha_i . \quad (3.29)$$

The probability that, when a type-$i$ host mates with a type-$j$ host, the offspring is of type-$k$ is then given by,

$$p_{ijk} = \sum_{m=0}^{k} p_{im} p_{j,(k-m)} , \quad (3.30)$$

where $i, j, k \in \{0,1,\ldots,T\}$, and both $p_{im}$ and $p_{j,(k-m)}$ are given by Equation 3.29.

In order to test the consistency of the ODE model with the branching process model of TE spread, we calculated the proportion of transgenic hosts over time for a variety of transposition rates (Figure 3.6). For both modeling frameworks, we used the continuous model of transpositional regulation, the linear model of host fitness, and the default
parameters from Table 3.1 with the exception that $r = 0$. The results suggest that the two modeling frameworks are consistent up to a prevalence of 10%, and that the major assumption of the branching process – that each mating only involves a single transgenic host – is indeed satisfied when the prevalence of transgenic hosts is low.

Figure 3.6 Rate of TE spread for the first 10 years following a transgenic release. Predictions of the branching process model are shown as thick, solid curves for prevalences less than 10%. Predictions of the ODE model are shown as narrow, patterned curves for prevalences up to 100%. The default parameters from Table 1 are used, with the exception that $r = 0$. Transposition rate is varied.
Figure 3.6 also implies that the early stages of TE spread provide a good indication of the ultimate fate of the TE in the population. When the transposition rate is 0.02 per element per generation, the TE declines in prevalence in the early stages and then continues to decline. When the transposition rate is 0.1 per element per generation, the TE quickly reaches a prevalence of 10% and then continues quickly to near-fixation. When the transposition rate is 0.06 per element per generation, the TE reaches 10% prevalence slowly and approaches near-fixation slowly.

The TE reaches a high maximum prevalence in the host population (Figure 3.6), but does not reach complete fixation because there will always be a proportion of offspring that do not inherit the TE. Fixation is preferable to near-fixation; however a high prevalence of the effector gene may be sufficient to decrease the reproductive number of the pathogen to below one. A parallel epidemiological model is required for an accurate assessment of disease control.

3.4.6 Sensitivity to model parameters

The sensitivity of the TE loss probability to any continuous model parameter $\gamma$ can be calculated by taking its partial derivative, $\frac{\partial e(t)}{\partial \gamma}$, and scaling this by the range of values observed in nature, $\Delta \gamma$. During the first season, this can be calculated by taking the partial derivative of Equation 3.21 with respect to $\gamma$ and integrating up to time $t_i$ subject to the initial condition that $\frac{\partial e_i(0)}{\partial \gamma} = 0 \forall i$. During the second season, the sensitivity of the TE loss probability can be calculated by taking the partial derivative of
Equation 3.22 with respect to $\gamma$ and integrating up to time $t_1$ subject to the initial condition that $\partial Q_i(0, e(t_2)) / \partial \gamma = \partial e_i(t_2) / \partial \gamma di$. Here, the second season population growth rate is used in the calculation of the initial condition, and the first season population growth rate is used in the integration.

The sensitivity of the rate of TE spread can be inferred from the sensitivity of the dominant eigenvalue, $(\partial \rho / \partial \gamma) \Delta \gamma$, when the population growth rate is set to 0. This follows from the observation that, on the scale of years, the rate at which the proportion of TEs in the population increases is independent of $r$ (Figure 3.5C). The sensitivity of the dominant eigenvalue can be calculated following the methodology in Section 11 of Dorman et al. (2004).

Table 3.2 shows the sensitivities of the TE loss probability and rate of TE spread to the continuous parameters of our model. Sensitivities for the alternative models of transpositional regulation and host fitness are shown in the supplemental Appendices (http://johnmm.bol.ucla.edu/te/). These sensitivity analyses do not account for discrete parameters such as the release size and copy number at release; however a number of important implications can be deduced for prioritizing parameters that should be accurately known for a transgenic release. Since we have more control over and better estimates of transposition rate and release size, then we should focus on these parameters in engineering efforts. However, it is also important that we have good estimates of less easily manipulated parameters such as the population growth rate and fitness cost of a
genomic insertion. These are some of the most influential factors in the spread of the TE and are essential to ensure that our model predictions are accurate.

<table>
<thead>
<tr>
<th>Model parameter $\gamma$</th>
<th>Parameter range</th>
<th>Sensitivity of asymptotic loss probability $(\partial e_i / \partial \gamma)\Delta \gamma$</th>
<th>Sensitivity of one year loss probability $(\partial e_i(t) / \partial \gamma)\Delta \gamma$</th>
<th>Sensitivity of rate of element spread $(\partial \rho / \partial \gamma)\Delta \gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.3 TE$^{-1}$ gen$^{-1}$</td>
<td>-0.20</td>
<td>-0.19</td>
<td>0.28</td>
</tr>
<tr>
<td>$b$</td>
<td>0.2</td>
<td>0.0016</td>
<td>0.00094</td>
<td>-0.0035</td>
</tr>
<tr>
<td>$d$</td>
<td>0.4 TE$^{-1}$</td>
<td>0.37</td>
<td>0.37</td>
<td>-0.47</td>
</tr>
<tr>
<td>$r$</td>
<td>0.8 gen$^{-1}$</td>
<td>-0.68</td>
<td>-0.56</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3.2 Sensitivity of loss probability and rate of TE spread to model parameters.
3.5 DISCUSSION

The proposed model provides us with a succinct framework to explore the early stages of TE spread through a diploid population. We can also make some crude recommendations regarding the implications this has for the use of TEs as drive mechanisms in the control of vector-borne disease. We chose a branching process to model this phenomenon since it has the benefit of being analytically tractable while still having the stochasticity of a Markov chain (Lange 2003). This has enabled us to investigate the probability that a TE is lost from its host population as a result of genetic drift, and how this probability is affected during periods of population growth and decline.

TEs face significant challenges upon being transferred into a new host population, thus making the consideration of genetic drift particularly relevant (Le Rouzic and Capy 2005). Whether a TE begins its invasion through horizontal transfer mediated by a vector (Silva and Kidwell 2000; Robertson et al. 1998), introgression (Labrador et al. 1999) or the de novo appearance of a new TE within the host species, the process invariably begins with only a single element. The fate of the element in its new host population is then a product of its internal properties, the population dynamics of the host species, and chance.
3.5.1 Transpositional bursts

There is a large body of evidence supporting the existence of heightened transposition rates at low copy numbers, which may enable TEs to prosper in the early stages of spread. Several mechanisms may contribute to this, including hybrid dysgenesis (Bregliano and Kidwell 1983; Bucheton 1990), overproduction inhibition (Lohe and Hartl 1996), and gene silencing (Jensen et al. 2002). Furthermore, evidence for a recent transpositional burst in the DINE-1 element of Drosophila yakuba has been presented by Yang et al. (2006). Predictions of this and other models (Le Rouzic and Capy 2005; Kaplan et al. 1985) suggest that a heightened transposition rate following introduction into a novel species will reduce the loss probability and speed up the rate of TE spread in the new population. However, even under generous conditions (a neutral TE and heightened transposition rate of 0.1 per element per generation), the loss probability can still be as high as 92% (Figure 3.3A).

3.5.2 Population growth

The prospects for a TE to colonize a new host species are dramatically improved during conditions of population growth. According to model predictions, when the rate of population growth is \( r = 0.2 \) per generation, as is the case during the seasonal population growth phase of An. gambiae, the probability of TE loss is lowered by \( \sim 15\% \) compared to the case of a constant population size (Figure 3.3B). Although seasonal population growth occurs on a time scale of six months, growth can also occur over a much longer time period when a species colonizes a new geographical location. The recent population
expansion of the malaria vectors *An. gambiae* and *An. arabiensis* (Donnelly *et al.* 2001) and the expansion of *D. melanogaster* into northern Africa and Eurasia (Baudry *et al.* 2004; Thornton and Andolfatto 2006) may be examples of this. The prediction of this model is that such a period of growth is a ripe time for a TE to colonize a new species.

### 3.5.3 Implications for a transgenic release

Any plan to release a TE into a novel host population for the purpose of disease control should aim to exploit the existence of both heightened transposition rates and population size changes to improve its likelihood of success. In the case of *An. gambiae*, dramatic population size changes between the dry season and the peak of the rainy season (Figure 3.1) should be used to the advantage of the release strategy. To this end, a transgenic release is recommended immediately following the dry season when the *An. gambiae* population begins to grow (Figure 3.4A). This allows the TE the best chance to establish itself in the population prior to the subsequent population decline. A less intuitive prediction is that a release slightly after the beginning of the growth phase is more efficient than a release prior to the growth phase. This suggests that, in the event that knowledge of population size changes is imprecise, the release strategy should bias towards a late release.

In addition to manipulating the release time, the model makes a number of other suggestions to lower the loss probability and increase the rate of TE spread through the *An. gambiae* population. TE loss probability is most effectively reduced by increasing the
release size of transgenic mosquitoes to more than 25, while increasing the TE copy number of the released mosquitoes is less effective since the TEs are more likely to be lost at once (Figure 3.4C and D). This is a larger release size than that suggested by Struchiner et al. (2005) who suggested a release of more than eight transgenic hosts; however Struchiner et al. (2005) also found that the TE copy number of the released hosts is relatively inconsequential.

The most effective way to ensure that the TE spreads at a rate acceptable to public health goals is by ensuring that the introduced TE has a high replicative transposition rate upon release. Here, the model recommends that a TE with a replicative transposition rate of 0.1 per element per generation is realistic (Seleme et al. 1999; Vasilyeva et al. 1999) and will reach a proportion of 10% in a village population in an acceptable time frame (Figure 3.4C). This is in agreement with Rasgon and Gould (2005) for the case where transposition is biased towards unlinked sites, and Le Rouzic and Capy (2006) who suggest that realistically useful TEs will have a transposition rate as high as 0.1 per element per generation.

A number of parameters that are less easily engineered also have important implications for the success of the disease control strategy. The fitness cost of a genomic insertion is the most influential parameter on the rate of TE spread, and the rate of population growth is one of the most influential parameters on TE loss probability (Table 3.2). It is important that we seek accurate estimates of these parameters.
3.5.4 Model limitations

Symptomatic of any mathematical analysis, simplifications have been made that may compromise the model predictions. Firstly, in the model formulation all TEs are assumed to act independently during gamete formation, implying that all element copies are at least 50 centimorgans apart. This contradicts the tendency for TEs to jump locally rather than distally (Tower et al. 1993; Newfield and Takaesu 1999) and to home in on certain genomic regions (Guimond et al. 2003; Hudson et al. 1995). Clustering of element copies can reduce the rate of TE spread since tightly linked element copies rarely segregate during meiosis and are effectively inherited as a single unit (Rasgon and Gould 2005). This also threatens the result that the majority of hosts only have a single element copy in the early stages of TE spread, since the replicative effect of transposition is counteracted less by the dilutive effect of sampling during gamete formation when elements are clustered. These are clearly important considerations and worthy of much further study; however we have neglected them in the present analysis in order to focus on the implications of population size changes.

Secondly, the present model is restrictive since it only applies to TE spread up to a prevalence of 10%. The assumption that all matings involve at least one uninfected host is required by the independence of particles in a branching process, but doesn’t allow us to ask questions about the waiting time for a TE to reach fixation. Despite this, modeling TE spread up to a prevalence of 10% captures the impact of population size changes.
fairly well, since the main effects of population growth and decline are stochastic and most relevant when the TE is present in small numbers. Furthermore, the mean number of transgenic hosts for prevalences above 10% can be calculated using the differential equations model (Equations 3.28-3.30).

Finally, of relevance to a transgenic release, another molecular detail that has not been accounted for in this paper is the potential for internal deletion to lead to loss of the effector gene from the TE construct (Engels 1989; Rubin and Levy 1997; Lohe et al. 2000). A related concern is that TEs that have undergone internal deletion may spread through the population more quickly than intact introduced TEs that still carry their transgenic load. Future modeling should assess the severity of this concern and experimental studies should be conducted on candidate TEs in An. gambiae.

3.5.5 Future directions
To gain a deeper insight into the impact of population structure on TE spread, we should consider the spatial as well as temporal population structure of a species, and how these interact together. In the case of An. gambiae, this would involve accounting for the existence of multiple chromosomal forms (Touré et al. 1983; Coluzzi et al. 1985; Coluzzi and Sabatini 1967) and how their relative abundances change throughout the year (Figure 3.1). This should be superimposed over the geographical distribution of mosquito populations in discrete patches corresponding to villages (Touré et al. 1983) and the relative abundance of each chromosomal form from village to village (Touré et al. 1998).
Clearly there is a relationship between the spatial and temporal dimensions of population structure, and an understanding of this will better enable us to assess the potential benefits of TE-mediated strategies for the control of vector-borne disease.
4 CHAPTER 4

THE EFFECT OF GENE DRIVE ON CONTAINMENT OF TRANSGENIC MOSQUITOES

4.1 INTRODUCTION

Mosquito-borne diseases such as malaria and dengue fever continue to pose a major health problem through much of the world. In the absence of any single effective disease control strategy, much interest has been directed at the use of gene drive mechanisms to spread anti-pathogen genes through mosquito populations (Craig 1963; Curtis 1968; Alphey et al. 2002). Several gene drive systems exist in nature, and it is hoped that refractory genes will be associated with these systems and driven into mosquito populations within a timeframe acceptable to public health goals (James 2005). Some of the most promising gene drive systems currently being investigated include homing endonuclease genes, transposable elements, Medea elements, the intracellular bacterium Wolbachia, engineered underdominance genes, and meiotic drive (Sinkins and Gould 2006).

Any transgenic mosquito project is expected to involve several stages of testing – first in the laboratory, then in indoor cages, and then in outdoor cages exposed to the ambient environment in a region where transgenic mosquitoes might eventually be released (Alphey et al. 2002; Scott et al. 2002). Laboratory studies will investigate the efficacy of
the transgene at preventing disease as well as testing for unintentional adverse effects. The Core Working Group on Guidance for Contained Field Trials (Benedict et al. 2008) has identified several potential adverse effects of transgenic mosquitoes that must be assessed prior to a release. These include an enhanced vectorial capacity for non-target pathogens, increased mosquito longevity or reproductive capacity, behavioral changes that lead to a higher biting rate, and a decreased susceptibility to other control measures such as insecticides. The Working Group also expressed the need to investigate the rate of horizontal DNA transfer between mosquitoes and non-target organisms, since other species may also acquire an increased capacity to transmit disease or disrupt an essential ecological function.

Much can be studied in the laboratory; however there are some potential adverse effects of transgenic mosquitoes that can only be assessed in outdoor cages (Benedict et al. 2008). A realistic assessment of mosquito longevity and reproductive capacity must be carried out under more natural conditions of climate and light variation. Ambient cages are also necessary to assess the population growth rate and the carrying capacity of the environment. Furthermore, a female bias in the wild sex ratio is problematic since only female mosquitoes transmit disease. Realistic assessments of such a bias can only be studied in an ambient cage.

By segregating transgenic organisms from the field, ambient cages provide a useful intermediate research stage between the laboratory and the environment; however
complete physical containment can never be guaranteed. The Working Group (Benedict et al. 2008) has outlined several possible breaches of containment – some of which can be avoided, but some of which are very difficult to protect against. These include unpredictable environmental damage due to earthquakes or lightning, leakage of water containing eggs or larvae, breaches of containment due to sabotage or burglary, and just simple human error. Therefore, there is a possibility that transgenic mosquitoes will be accidentally released into an environment that is conducive to their survival before the effectiveness and safety of the gene drive strategy has been ascertained.

Most genes introduced in small numbers are very likely to be lost from the environment, even in the presence of a selective advantage (Fisher 1922; Haldane 1927; Wright 1931). However gene drive mechanisms enhance the invasiveness of introduced genes, and therefore introduced transgenes may be less likely to be lost than ordinary genes following an accidental release. Given that the organism currently being considered for genetic alteration is a vector of human disease, it is particularly important that the invasiveness of selfish DNA be accounted for in the risk management of ambient cage trials.

Here, we analyze the probability that transgenic DNA consisting of an anti-pathogen gene and drive system is lost from a mosquito population following an accidental release. Several gene drive mechanisms are currently being considered to spread anti-pathogen
genes into mosquito populations, each having its own unique dynamics. We therefore analyze the loss probability associated with each system separately.

For an initial comparison of gene drive strategies, we calculate the asymptotic extinction probability for each system. A major feature of the demography of *Anopheles gambiae*, the main vector of malaria in tropical Africa, is the existence of population size changes within and between years (Taylor *et al.* 2001; Manoukis 2006). Given the influence of population size changes on gene loss, we also calculate the extinction probabilities under conditions of population growth and decline. Although there are several other factors that will influence the loss or persistence of transgenic DNA following an accidental release, it is hoped that these calculations will inform the risk management of planned ambient cage trials involving selfish DNA.

### 4.2 Analysis

#### 4.2.1 Homing endonuclease genes

Homing endonuclease genes (HEGs) are a class of highly-specific DNA endonucleases found in some viruses, bacteria and eukaryotes (Windbichler *et al.* 2007). HEGs are able to spread through a population despite a fitness cost due to their overrepresentation in the gametes of a heterozygote. They achieve this by expressing an endonuclease which creates a double-stranded break at a highly-specific site that lacks the HEG. Homologous DNA repair then copies the HEG to the cut chromosome (Rong and Golic 2003).
To calculate the basic reproductive number of the HEG allele following an accidental release, we consider a two-type continuous-time branching process in which type-1 particles are heterozygous and type-2 particles are homozygous for the HEG. We consider a budding model in which the host does not die following reproduction. The death rate of all mosquitoes is set to $\mu = 1 \text{ gen}^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have a fecundity of $\theta$, where:

$$\theta = 2(1 + r) \quad (4.1)$$

and $r$ represents the population growth rate.

In the early stages of spread, almost all matings involve at least one wild-type mosquito. All matings between wild-type mosquitoes and mosquitoes homozygous for the HEG will produce heterozygotes. Matings between wild-type mosquitoes and heterozygotes for the HEG will produce heterozygotes with probability $(1 + t)/2$ and wild-types with probability $(1 - t)/2$. Here, the homing rate, $t$, represents the fraction by which the HEG allele is overrepresented in the gametes of a heterozygote.

Since homozygotes will be lost from the population after one generation, we are interested in the basic reproductive number of mosquitoes that are heterozygous for the HEG. The basic reproductive number is equal to the average number of offspring that a particle eventually generates (Van den Driessche and Watmough 2002). For a
heterozygote, this is simply given by the average number of heterozygotes that it produces at the end of its life, \( f_{1,1} \).

In order to calculate this quantity, we first need to consider the fitness cost associated with the HEG allele. We assume that a female homozygous for the HEG has a reduced fecundity of \( \theta(1 - s) \). Male mosquitoes that mate with wild-type females do not suffer from reduced fecundity, and so the mean fecundity of a homozygote is \( \theta(1 - s/2) \). This reduction in fecundity is assumed to have a dominance factor of \( h \in [0,1] \), where \( h = 1 \) represents a dominant fitness cost, \( h = 0 \) represents a recessive fitness cost, and \( h = 1/2 \) represents an additive fitness cost. The mean fecundity of a heterozygote is therefore \( \theta(1 - hs/2) \).

According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in Figure 4.1. The death intensity for heterozygotes is equal to

\[
\lambda_1 = \mu + \theta(1 - hs/2),
\]

and the death intensity for homozygotes is equal to

\[
\lambda_2 = \mu + \theta(1 - s/2).
\]

The basic reproductive number for a mosquito heterozygous for the HEG is then given by:

\[
R_{1,1} = f_{1,1} = \frac{\theta}{\lambda_1} \left( 1 - \frac{hs}{2} \right) \left( 1 + \frac{1 + t}{2} \right).
\]
When the basic reproductive number is greater than one, HEG spread is supercritical, and the HEG has a nonzero probability of spreading through the mosquito population.

Simplifying Equation 4.4, this means that an accidentally released HEG has some chance of spreading into a wild population when:

$$t > \frac{hs - (2 - hs)r}{(2 - hs)(1 + r)}.$$  \hspace{1cm} (4.5)

When the population size is constant, then the condition for HEG spread simplifies to:

$$t > \frac{hs}{2 - hs}.$$ \hspace{1cm} (4.6)

One implication of this result is that only the fitness cost to the heterozygote determines whether the HEG has a chance of spreading or not. The magnitude of fitness cost that is tolerable is given by Equations 4.5 and 4.6.
Figure 4.1 Schematic for the early spread of a homing endonuclease gene (HEG) through a randomly mating mosquito population. Mosquitoes homozygous for the HEG ($HH$) have a reduced fecundity of $\theta(1 - s/2)$ and mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for the HEG ($Hh$). Heterozygotes have a reduced fecundity of $\theta(1 - hs/2)$ and mate with wild-type mosquitoes to give birth to heterozygotes with probability $(1 + t)/2$ and wild-types with probability $(1 - t)/2$. All mosquitoes have a death rate of $\mu$. 

\[ \theta(1 - \frac{hs}{2}) \frac{1 + t}{2} \] 

\[ \theta(1 - \frac{s}{2}) \frac{1}{2} \]
When the basic reproductive number is greater than one, HEG spread is possible; however the probability of HEG loss may still be very high. To determine the extinction probability of the HEG, we first define the probability generating function for the HEG branching process. This is a function of the vector $\mathbf{z} = (z_1, z_2)$ and is defined as,

$$ P_i(z) = \sum_{j=1}^{2} p_{ij} z^j, \quad (4.7) $$

where $p_{ij}$ is the probability that a type-$i$ particle gives rise to $j_1$ type-1 particles and $j_2$ type-2 particles, and $j$ is defined as the vector $j = (j_1, j_2)$. Substituting the $p_{ij}$ terms from the HEG branching process into this equation, we have the probability generating function,

$$ P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1} \left(1 - \frac{hs}{2}\right) \frac{1-t}{2} z_1 + \frac{\theta}{\lambda_1} \left(1 - \frac{hs}{2}\right) \frac{1+t}{2} z_1^2, \quad (4.8) $$

$$ P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2} \left(1 - \frac{s}{2}\right) z_1 z_2. \quad (4.9) $$

The probability that the HEG allele is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations (Harris 1963),

$$ P_i(e) = e_i \forall i, \quad (4.10) $$

where $e_1$ is the loss probability beginning with a heterozygote for the HEG, and $e_2$ is the loss probability beginning with a homozygote for the HEG. For the HEG branching process, this system of equations has the solution,

$$ e_1 = \min \left\{ \frac{2}{(1+r)(2-hs)(1+t)} \right\}, \quad (4.11) $$
\[ e_2 = \frac{1}{1 + (1 + r)(2 - s)(1 - e_1)} \quad (4.12) \]

The extinction probability beginning with a heterozygote, \( e_1 \), is a decreasing function of homing rate and population growth rate and an increasing function of heterozygote fitness cost. The extinction probability beginning with a homozygote, \( e_2 \), is a function of \( e_1 \) and is less than one when \( e_1 \) is less than one.

Note that we have calculated the asymptotic loss probabilities here, although the model is only valid in the early stages of spread when the reservoir of wild-types is particularly large. This paradox is averted by considering that, when mating events begin to occur between two mosquitoes having the HEG allele, then the HEG allele has already reached a high enough presence in the population that it is very unlikely to go extinct. This assumption similarly applies to each of the drive systems discussed in this paper.

To make sense of these equations, it helps to have some idea of the parameter values and ranges that exist in nature. For HEGs, the most important parameter is the homing rate, \( t \), which describes the fractional increase in representation of HEGs in the gametes of a heterozygote. An estimate of homing rate has been hinted at for the transfer of a Tet-resistance transcription unit between two plasmids (Windbichler et al. 2007). The act of homing requires both DNA cleavage and repair using the HEG-carrying site as a template. In this case, the complete HEG allele was transferred to \( \sim 10\% \) of cleaved sites, suggesting a range of homing rates between 0 and 10% since not all sites will be cleaved.
This estimate may not be particularly insightful (Austin Burt, personal communication) since it is likely that HEGs will behave differently in the mosquito germ line.

Rong and Golic (2003) have studied the dynamics of HEG alleles in D. melanogaster; however their studies have been directed more at molecular dynamics and less at the rates at which these processes occur. In the absence of good parameter estimates, Derecet et al. (2007) have investigated the full range of homing rates, $t \in [0,1]$.

The fitness cost associated with the HEG allele depends on the disease control strategy being employed. A HEG designed to drive a refractory gene into the mosquito population will have a relatively small fitness cost equal to the fitness cost of the refractory gene. This is difficult to estimate due to the lack of reliable comparative fitness measurements in vector populations; however fitness costs have been documented in several insect species due to both mounting an immune response (Moret and Schmid-Hempel 2000; Ahmed et al. 2002) and maintaining the physiological machinery necessary to do so (Kraaijeveld and Godfray 1997; Koella and Boete 2002).

A recent encouraging result is that transgenic mosquitoes have been engineered that exhibit no measurable fitness cost when fed on Plasmodium-free blood (Moreira et al. 2004) and exhibit a fitness benefit when fed on Plasmodium-infected blood (Marrelli et al. 2007). Accounting for the proportion of mosquitoes infected with malaria parasites (Beier et a. 1999), these results correspond to a mean homozygote fitness cost in the
range \( s \in [-0.05, 0.04] \) (Marshall 2008). There is little data on the homozygosity of a fitness cost associated with a refractory allele and so this is best assumed as being additive \( (h = 0.5) \).

A HEG designed to induce a genetic load or bias the sex ratio in order to reduce the size of the vector population will have a much larger fitness cost. For this strategy, it is hoped that females homozygous for the HEG will be sterile and males will suffer minimal fitness cost (Austin Burt, personal communication). This suggests an average homozygous fitness cost of \( s \leq 1 \). The best estimate for the homozygosity of this fitness cost comes from classic data for \( D. \ melanogaster \) (Austin Burt, personal communication) in which recessive lethal alleles have a homozygosity of \( h = 0.02 \).

A third strategy of disease control is to use a HEG that disrupts a gene regulating the ability of mosquitoes to function as efficient vectors for the malaria parasite. This strategy is likely to confer a very small fitness cost on the mosquito, and may in fact confer a fitness benefit similar to that observed by Marelli et al. (2007) for the refractory gene approach.

The population dynamics of the wild-type mosquito population are equally relevant to an assessment of each of the gene drive strategies. The \( An. \ gambiae \) population of Banambani, Mali, serves as a well-studied example (Taylor et al. 2001; Lanzaro et al. 1998; Tripet et al. 2005; Touré et al. 1994), with recent collections suggesting that peak
population densities during the wet season are at least ten times those during the dry season (Taylor et al. 2001). Assuming a mosquito generation time of ~16 days (Mahamadou Touré, personal communication) and a population change over a period of six months, this yields a population growth rate of $r = 0.2$ per generation, and a corresponding rate of decline of $r = -0.2$ per generation.

HEGs are one of the most invasive gene drive systems available; and hence are one of the most likely to spread following an accidental release. For both HEG strategies – refractory gene (Figure 4.2A) and genetic load (Figure 4.2B) – a single homozygous mosquito having a HEG with a relatively modest homing rate of $t = 0.1$ has a loss probability of less than 90% over the entire range of likely fitness costs. This leads to a persistence probability of more than 10% for all realistic parameterizations, which is very high for a single escapee. These loss probabilities are noticeably reduced during periods of population growth (Figure 4.2C).
Figure 4.2 Asymptotic probabilities of HEG loss as a function of fitness cost. (A) Homing rate is varied for the refractory gene HEG strategy. (B) Homing rate is varied for the genetic load HEG strategy. (C) Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of An. gambiae. (D) The escapee genotype is varied and the homozygote release size is increased to 10.
The probability that the HEG becomes established in the population is most influenced by the number of escapees during an accidental release (Figure 4.2D). Provided that the condition for HEG spread is satisfied (Equation 4.5), then the loss probability declines with the power of the number of escapees. Homozygotes are slightly more likely to persist in the population essentially because they bring two HEGs into the wild population; while heterozygotes only bring one. According to model predictions under default parameters, an escape of five homozygotes reduces the loss probability to 43%, while an escape of 25 homozygotes reduces the loss probability to 1.5%. HEGs are therefore able to spread under the full range of realistic conditions, and are more likely to persist than not for escape sizes greater than four.

4.2.2 Transposable elements

Transposable elements (TEs) are particularly interesting genomic components due to their ability to transpose replicatively and hence spread throughout a population despite a fitness cost (Charlesworth et al. 1994). The observation that P elements spread through most of the wild-type D. melanogaster population within a few decades (Engels 1989) has inspired the idea of using TEs as drive mechanisms for spreading anti-pathogen genes into mosquito populations (Craig 1963, Curtis 1968).

To model the early stages of TE spread following an accidental release, we consider a $T$-type continuous-time branching process in which a type-$i$ particle has $i$ copies of the TE, where $i \in \{1,2,...,T\}$ and $T \geq 1$. Here, $T$ can be approximated as the number of sites that
will be occupied in the early stages of TE spread. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1$ gen$^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity $\theta$ as described by Equation 4.1.

Transposition and deletion are modelled by assuming that a proportion $\alpha_i$ of gametes are derived from cells in which a replicative transposition event has occurred, while a proportion $\beta_i$ of gametes are derived from cells in which an element deletion event has occurred. The replicative transposition rate for a type-$i$ host, $\alpha_i$, is equal to the replicative transposition rate per TE in a type-$i$ host, $u_i$, multiplied by the number of TEs in the host genome, $i$ (i.e. $\alpha_i = iu_i$). Here, $u_i$ is generally a decreasing function of $i$ to account for suppression of transposition with increasing copy number (Weinreich et al. 1994; Wu and Morris 1999; Townsend and Hartl 2000). We model transposition rate as an exponentially decreasing function of copy number,

$$u_i = u_i 2^{-c(i-1)},$$

(4.13)

where $c$ determines the rate at which the replicative transposition rate falls off with additional element copies (Marshall 2008). Similarly, the deletion rate for a type-$i$ host, $\beta_i$, is equal to the deletion rate per element, $\nu$, multiplied by the number of elements in the host genome, $i$ (i.e. $\beta_i = iv$), where $\nu$ is generally considered a constant.
The fitness cost associated with additional TE copies is modelled by assuming that a female having \(i\) TE copies has a reduced fecundity of \(\theta(1 - s_i)\), where \(s_i\) is an increasing function of \(i\). For the early stages of TE spread, we describe fitness cost as a linear function of \(i\) (i.e. \(s_i = id\)), where \(d\) represents the fractional decrease in female fecundity associated with each additional TE copy in the genome. Male mosquitoes that mate with wild-type females do not suffer from reduced fecundity, and so the mean fecundity of a type-\(i\) host is \(\theta(1 - s_i / 2)\).

The number of TEs in the haploid gamete is determined by the number of TEs in the diploid cell that are passed on during meiosis. For a diploid cell with \(i\) copies of the TE we assume, to a first approximation, that all of these TEs are far enough apart from each other that they segregate independently. Under this assumption, the probability of having \(j\) copies in a gamete is proportional to the number of ways of choosing \(j\) elements from a total of \(i\). Similarly, if a replicative transposition event has occurred in the diploid cell, then the probability of having \(j\) copies in the gamete is proportional to the number of ways of choosing \(j\) elements from \(i + 1\), or from \(i - 1\) if a deletion event has occurred.

In the early stages of spread, almost all matings involve at least one wild-type mosquito, and so the number of TEs in the offspring’s genome is equal to the number of TEs in the gamete contributed by the parent infected with the TE. According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in
Figure 4.3. In this process, a type- $i$ host lives an exponentially distributed length of time with death intensity,

$$
\lambda_i = \mu + \theta(1 - s_i / 2) ,
$$

(4.14)

and at the end of its life produces on average $f_{ij}$ hosts with $j$ element copies according to the equation:

$$
f_{ij} = \frac{\theta}{\lambda_i} \left( 1 - \frac{s_i}{2} \right) \left[ 1 + \frac{1}{2^{i-1}} \binom{i-1}{j} \beta_j + \frac{1}{2^i} \binom{i}{j} (1 - \alpha_i - \beta_j) + \frac{1}{2^j} \binom{i+1}{j} \alpha_i \right] ,
$$

(4.15)

where $i, j \in \{1, 2, ..., T\}$.
Figure 4.3 Schematic for the early spread of a transposable element (TE) through a randomly mating mosquito population. Mosquitoes having $i$ copies of the TE have a reduced fecundity of $\theta(1 - s_i / 2)$ and mate with wild-type mosquitoes to give birth to an average of $\mu$ mosquitoes having $j$ TE copies, where $\mu$ is defined in Equation 14 and $j \in \{0, ..., i + 1\}$. All mosquitoes have a death rate of $\mu$. 

\[
\theta \left(1 - \frac{s_i}{2}\right) \left[ \frac{1}{2^{i+1}} \left( \begin{array}{c} i-1 \\ i+1 \end{array} \right) \beta_i + \frac{1}{2^i} \frac{1}{i+1} \left(1 - \alpha_i - \beta_i\right) + \frac{1}{2^{i+1}} \frac{1}{i+1} \alpha_i \right]
\]
The basic reproductive number of a TE following an accidental release can by studied by assuming that, in the early stages of TE spread, the vast majority of infected hosts have only one or two TE copies. The basic reproductive number for a mosquito having a single TE copy, \( R_{1,1} \), can then be defined recursively as,

\[
R_{1,1} = f_{1,1} + f_{1,2}R_{2,1} .
\]  

(4.16)

Here, \( f_{1,1} \) is the average number of offspring of a type-1 host that are of type-1, \( f_{1,2} \) is the average number of offspring of a type-1 host that are of type-2, and \( R_{2,1} \) is the expected number of type-1 hosts that a type-2 host eventually generates. Following similar reasoning, \( R_{2,1} \) can be defined recursively as,

\[
R_{2,1} = f_{2,1} + f_{2,2}R_{2,1} ,
\]  

(4.17)

where \( f_{2,1} \) and \( f_{2,2} \) are similarly defined. Equation 4.17 can then be rearranged and substituted into Equation 4.16 to obtain the basic reproductive number for a type-1 host,

\[
R_{1,1} = f_{1,1} + \frac{f_{1,2}f_{2,1}}{1 - f_{2,2}} .
\]  

(4.18)

There is no simple solution to this equation; however if we consider the simplified scenario in which transpositional regulation and element deletion are negligible \( (c = v = 0) \), this leads to the basic reproductive number,

\[
R_{1,1} = \frac{(d - 2)(r + 1)(6 - u_1(u_1 + 4) - r(6 + u_1(4 + u_1)) + d(r + 1)(54 + u_1(u_1 + 4))))}{4(d + 3 + r(d + 2))(r - 1 + (r + 1)(u_1 - d(9 + u_1)))} .
\]  

(4.19)
When the basic reproductive number is greater than one, TE spread is supercritical, and the TE has a nonzero probability of spreading through the mosquito population. When transpositional regulation and element deletion are not negligible, we must use Equation 4.18 to determine the conditions for TE spread to be supercritical.

When TE spread is supercritical the eventual extinction of the TE is uncertain, however the probability that the TE is lost from the population may still be very high. To determine the extinction probability of the TE, we first define the probability generating function for the TE branching process. This is a function of the vector \( z = (z_1, \ldots, z_T) \) and is defined as,

\[
P_i(z) = \sum_j p_y z^j = \sum_j p_y z_1^j \cdots z_T^j ,
\]

where \( p_y \) is the probability that a type-\( i \) particle gives rise to \( j_1 \) type-1 particles, \( j_2 \) type-2 particles, and so on, and \( j \) is defined as the vector \( j = (j_1, \ldots, j_T) \). Substituting the \( p_y \) terms from the TE branching process into this equation, we have the probability generating function,

\[
P_i(z) = \frac{H}{\lambda_i} + (f_{i,0} + f_{i,T})z_i + \sum_{j=1}^{T} f_{ij} z_j - \frac{\theta}{\lambda_i} \left( 1 - \frac{s_i}{2} \right) z_i^2 ,
\]

The probability that the TE is eventually lost from the population is then given by the smallest solution of the system of \( T \) simultaneous equations described by Equation 4.10. There is no simple analytic solution in this case, and so the system is best solved by numerically iterating:
\[ e_i = \frac{\mu}{\lambda_i} + (f_{i,0} + f_{i,T+1})e_i + \sum_{j=1}^{T} f_{ij}e_i e_j - \theta \left( \frac{1}{\lambda_i} - \frac{s_i}{2} \right) e_i^2, \]  

(4.22)

for \( i \in \{1, \ldots, T\} \), where \( e_1 \) is the loss probability beginning with an individual having a single TE copy, \( e_2 \) is the loss probability beginning with an individual having two TE copies, and so on.

To make sense of these equations, it helps to have some idea of the parameter values and ranges that exist in nature. Currently, there is little or no data regarding the behavior of the candidate TEs in human disease vectors such as \( An. gambiae \) and \( Aedes aegypti \), so most of these estimates have been taken from measurements in \( D. melanogaster \).

We consider a baseline replicative transposition rate of \( u_1 = 0.1 \) per element per generation (TE\(^{-1}\)gen\(^{-1}\)). Although this is a fairly high transposition rate, it is realistic (Seleme \textit{et al.} 1999; Vasilyeva \textit{et al.} 1999) and several modeling approaches have recommended it as a minimum requirement for gene drive to occur in a timeframe acceptable to public health goals (Rasgon and Gould 2005; Le Rouzic and Capy 2006). Following Marshall (2008), we consider a transpositional regulation parameter of \( c = 2.9 \) per element (TE\(^{-1}\)). This value was chosen since it produces an equilibrium TE copy number consistent with the \textit{Herves} element in \( An. gambiae \) (Subramanian \textit{et al.} 2007).

We also consider a deletion rate of \( v = 4 \times 10^{-6} \) TE\(^{-1}\)gen\(^{-1}\) as suggested by pooled estimates from several laboratory line experiments (Nuzhdin \textit{et al.} 1997; Maside \textit{et al.} 2000).
The fitness cost associated with a TE can originate from both the mutagenic nature of a new genomic insertion and the effects of being associated with a refractory gene. Following Mackay et al. (1992), we estimate the fitness cost of a new genomic insertion by the average fitness cost of a spontaneous mutation. This has been estimated as ~0.02 per element (Mukai et al. 1972; Ohnishi 1977; Crow and Simmons 1983) and is a reasonable estimate in the early stages of TE spread when insertional mutagenesis is the dominant fitness cost and selection has not yet eliminated TEs with higher fitness costs (Charlesworth 1991). The fitness cost of being associated with a refractory gene has been discussed for the case of HEGs, and has been estimated by the range $[-0.05,0.04]$. Combining these two estimates suggests a range for fitness costs on the order of $d \in [-0.03,0.06]$ per element.

While TEs are not as invasive as HEGs, they are very able to spread through a mosquito population following an accidental release. A TE having a transposition rate greater than $u_1 = 0.05 \ TE^{-1}\ gen^{-1}$ has a small chance of persisting in the population over the entire range of likely fitness costs (Figure 4.4A). For an ambitious yet feasible transposition rate of $u_1 = 0.1 \ TE^{-1}\ gen^{-1}$, the persistence probability of a single escapee is between 4% and 9% for all realistic parameterizations. The prospects for a TE to colonize a wild population are dramatically improved during conditions of population growth (Figure 4.4B).
Figure 4.4 Asymptotic probabilities of TE loss as a function of fitness cost. (A) Replicative transposition rate is varied. (B) Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*. (C) Copy number per individual at release is increased to five. (D) Release size is increased to 25.
The probability that the TE is lost from the population is decreased by the number of TEs that are present in the escapee (Figure 4.4C); however it is decreased even more by the number of escapees during an accidental release (Figure 4.4D). According to model predictions, an escape of five mosquitoes each infected with a single TE decreases the loss probability to 72%. For 10 escapees, the loss probability decreases to 51%; while for 25 escapees, the loss probability decreases to 19%. TEs are therefore less invasive than HEGs; however they are still very likely to establish themselves in a population following an accidental release. Under default conditions, a TE is more likely to persist than not for escape sizes greater than 10.

4.2.3 Meiotic drive

Meiotic drive refers to any mechanism by which a heterozygous locus segregates at a greater-than-Mendelian frequency by destroying or disabling the homologous chromosome (Little 1991). Various mechanisms are known to result in meiotic drive (Hickey and Craig 1966; Lyttle 1977). Some of these reduce the quantity of functional sperm; however they do not necessarily result in reduced fertility (Sinkins and Gould 2006). Alleles that promote meiotic drive are able to spread through a population despite a fitness cost as a consequence of their increased inheritance.

We consider a meiotic drive strategy which has been recently modeled (Huang et al. 2007) utilizing a Y-linked meiotic drive gene ($Y^D$) to drive an X-linked drive-insensitive response allele ($X^{it}$) into the mosquito population. The Y-linked meiotic drive gene is able
to spread into the population by virtue of its overrepresentation in the gametes of \(X^{sn}D\)
males, where \(X^{sn}\) is a wild-type drive-sensitive allele at the same locus as the \(X^{di}\) allele. As
the meiotic drive gene increases its prevalence in the population, it selects for the drive-
insensitive \(X^{di}\) allele. This selection occurs because, while the \(X^{sn}\) allele is
underrepresented in the gametes of \(X^{sn}D\) males, the \(X^{di}\) allele is represented at normal
Mendelian frequencies. If an anti-pathogen gene is linked to the \(X^{di}\) allele, it will therefore
be driven into the population.

To model the spread of the meiotic drive gene (\(Y^{D}\)) and drive-insensitive response allele
(\(X^{di}\)), we consider a 5-type continuous-time branching process in which type-1 particles
are males having the X-linked drive-sensitive allele and Y-linked meiotic drive gene
(\(X^{sn}Y^{D}\)), type-2 particles are males having the X-linked drive-insensitive allele and Y-
linked meiotic drive gene (\(X^{di}Y^{D}\)), type-3 particles are males having the X-linked drive-
insensitive allele (\(X^{di}Y^{d}\)), type-4 particles are females homozygous for the drive-
insensitive allele (\(X^{di}X^{di}\)), and type-5 particles are heterozygous females (\(X^{di}X^{sn}\)).

We consider a budding model in which the death rate of all mosquitoes is set to \(\mu = 1\)
gen\(^{-1}\), and fitness differences between particle-types are accounted for by differences in
female fecundity. Wild-type mosquitoes are assumed to have fecundity \(\theta\) as described by
Equation 4.1.
In the early stages of spread, almost all transgenic males will mate with wild-type females \((X^{en} X^{en})\) while almost all transgenic females will mate with wild-type males \((X^{en} Y^{d})\). This means that, one generation following an accidental release, there will be no mosquitoes having both transgenic alleles because all offspring will inherit a wild-type allele from their wild-type parent. Matings involving type-1 males \((X^{en} Y^{D})\) will produce more type-1 males with probability \((1 + t) / 2\) and will produce wild-type females with probability \((1 − t) / 2\). Here, the meiotic drive parameter, \(t\), represents the fraction by which the \(Y^{D}\) allele is overrepresented in the gametes of \(X^{en} Y^{D}\) males.

Since the two transgenic alleles, \(Y^{D}\) and \(X^{it}\), become separated in the early stages of spread, we are actually interested in calculating the basic reproductive numbers of the two alleles separately. The \(Y^{D}\) allele will spread when the basic reproductive number of type-1 \((X^{en} Y^{D})\) individuals is greater than one; and the \(X^{it}\) allele will spread when the basic reproductive numbers of both type-3 \((X^{it} Y^{d})\) and type-5 \((X^{en} X^{it})\) individuals are greater than one.

In order to calculate these quantities, we first need to consider the fitness costs associated with each transgenic allele. We assume that a female homozygous for the \(X^{it}\) allele has a reduced fecundity of \(\theta(1 − s_{y})\). This reduction in fecundity is assumed to have a dominance factor of \(h \in [0,1]\) and hence male \((X^{it} Y^{d})\) and female \((X^{en} X^{it})\) mosquitoes that only have one copy of the \(X^{it}\) allele have a reduced fecundity of \(\theta(1 − h s_{y})\). Male mosquitoes that have the \(Y^{D}\) allele \((X^{en} Y^{D})\) are assumed to have a reduced fecundity of
\( \theta(1-s_D) \); and finally, male mosquitoes that have both the \( X^\alpha \) allele and the \( Y^D \) allele (\( X^\alpha Y^D \)) are assumed to have a reduced fecundity of \( \theta(1 - h s_u - s_D) \).

According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in Figure 4.5. The death intensities for each particle type are given by:

\[
\begin{align*}
\lambda_1 &= \mu + \theta(1-s_D), \\
\lambda_2 &= \mu + \theta(1-h s_u - s_D), \\
\lambda_3 &= \lambda_5 = \mu + \theta(1-h s_u), \\
\lambda_4 &= \mu + \theta(1-s_u).
\end{align*}
\] (4.23) (4.24) (4.25) (4.26)

The process is also characterized by the following offspring counts:

\[
\begin{align*}
f_{1,1} &= \frac{\theta(1-s_D)}{\lambda_1} \left(1 + \frac{1+t}{2}\right), \quad (4.27) \\
f_{3,3} &= \frac{\theta(1-h s_u)}{\lambda_3}, \quad (4.28) \\
f_{3,5} &= \frac{\theta(1-h s_u)}{\lambda_3} \left(\frac{1}{2}\right), \quad (4.29) \\
f_{5,3} &= \frac{\theta(1-h s_u)}{\lambda_5} \left(\frac{1}{4}\right), \quad (4.30) \\
f_{5,5} &= \frac{\theta(1-h s_u)}{\lambda_5} \left(1 + \frac{1}{4}\right). \quad (4.31)
\end{align*}
\]
Upon mating with wild-type females, type-1 males \((X^{s}\nu Y^{D})\) can only give rise to type-1 males or wild-type females. Their basic reproductive number is therefore given by:

\[
R_{1,1} = f_{1,1} = \frac{\theta(1-s_D)(3+t)}{2\mu + 2\theta(1-s_D)}.
\]  

(4.32)

When this basic reproductive number is greater than one, the meiotic drive gene has a nonzero probability of spreading through the mosquito population. Simplifying Equation 4.32, this means that an accidentally released meiotic drive gene has a chance of spreading into a wild population when:

\[
t > \frac{s_D - r(1-s_D)}{1-s_D + r(1-s_D)}.
\]  

(4.33)

When the population size is constant, then the condition for spread of the meiotic drive gene simplifies to:

\[
t > \frac{s_D}{1-s_D}.
\]  

(4.34)

This result describes the maximum fitness cost that a meiotic drive gene of given strength, \(t\), can tolerate when the population size is constant. Equation 4.33 shows that these conditions are relaxed during periods of population growth.
Figure 4.5 Schematic for the early spread of a Y-linked meiotic drive gene \((Y^D)\) and X-linked response allele \((X^u)\) through a randomly mating mosquito population. Mosquitoes having both transgenic alleles \((X^uY^D)\) have a reduced fecundity of \(\theta(1-hs_u-s_D)\) and mate with wild-type mosquitoes to give birth to \(X^uX^u\) females with probability \(1/2\) and \(X^uY^D\) males with probability \(1/2\). \(X^uY^D\) males have a reduced fecundity of \(\theta(1-s_D)\) and mate with wild-type mosquitoes to give birth to \(X^uY^D\) males with probability \((1+t)/2\) and wild-type females with probability \((1-t)/2\). The remaining transitions are described in the text. All mosquitoes have a death rate of \(\mu\).
Calculating the basic reproductive numbers for type-3 \( (X^dY^d) \) and type-5 \( (X^dX^m) \) mosquitoes is a little more complicated since type-3 males can mate with wild-type females to produce type-5 females. Additionally, type-5 females can mate with wild-type males to produce type-3 males and more type-5 females. The basic reproductive numbers of type-3 and type-5 mosquitoes are therefore defined recursively analogous to Equation 4.18. The basic reproductive number for a type-3 male is,

\[
R_{3,3} = f_{3,3} + \frac{f_{3,5}f_{5,3}}{1 - f_{5,3}} = \frac{1}{2} + \frac{4}{5} \frac{1}{hs_u (1 + r) - r + 1} + \frac{9}{10} \frac{1}{hs_u (1 + r) - 2r - 3}, \quad (4.35)
\]

and the basic reproductive number for a type-5 female is,

\[
R_{5,5} = f_{5,5} + \frac{f_{5,3}f_{3,5}}{1 - f_{3,3}} = \frac{(1 + r)(1 - hs_u)(6 + r - hs_u (1 + r))}{6 - 4hs_u + 4r(1 - hs_u)}. \quad (4.36)
\]

When these basic reproductive numbers are greater than one, the X-linked drive-insensitive allele has a nonzero probability of spreading through the mosquito population. Simplifying Equations 4.35 and 4.36, this means that an accidentally released drive-insensitive allele has a chance of spreading into a wild population when:

\[
hs_u < \frac{r}{1 + r}. \quad (4.37)
\]

When the population size is constant, then the condition for spread of the drive-insensitive allele becomes:

\[
s_u < 0. \quad (4.38)
\]

This result suggests that, if the initial release size is very small, then the drive-insensitive allele will only spread through the population if it confers a fitness benefit to the mosquito. Under conditions of population growth, the allele will be able to tolerate a
fitness cost as described by Equation 4.37; however it is only the fitness cost of having a single \( \lambda^{it} \) allele that matters.

Interestingly, the strength of the meiotic drive gene, \( t \), does not relax the conditions for spread of the \( \lambda^{it} \) allele. This is because, in the early stages of spread, almost all matings are with the reservoir of wild-types, and so the increase in the proportion of individuals having the \( \lambda^{it} \) allele is insignificant. The condition for the spread of the meiotic drive gene is more relaxed than the condition for the spread of the drive-insensitive allele. This suggests that, following an accidental release, a very feasible possibility is that the \( \lambda^{it} \) allele will be lost while the \( YD \) allele will continue to spread through the mosquito population.

When meiotic drive is supercritical, eventual extinction is uncertain; however the extinction probability may still be very high. To determine this probability when the population size is not constant, we first define the probability generating function for the meiotic drive branching process. This is a function of the vector \( z = (z_1, \ldots, z_5) \) and is defined as,

\[
P_i(z) = \sum_{j=1}^{5} p_y z^j = \sum_{j=1}^{5} p_y z_1^{j_1} z_2^{j_2} z_3^{j_3} z_4^{j_4} z_5^{j_5},
\]

where \( p_y \) is the probability that a type-\( i \) particle gives rise to \( j_1 \) type-1 particles, \( j_2 \) type-2 particles and so on, and \( j \) is defined as the vector \( j = (j_1, \ldots, j_5) \). Substituting the
\( p_y \) terms from the branching process into this equation, we have the probability generating functions,

\[
P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta(1-s_D)}{\lambda_1} \frac{1}{2} (1-t)z_1 + \frac{\theta(1-s_D)}{\lambda_1} \frac{1}{2} (1+t)z_1^2 ,
\]

(4.40)

\[
P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta(1-hs_{it} - s_D)}{\lambda_2} \frac{1}{2} z_2 z_1 + \frac{\theta(1-hs_{it} - s_D)}{\lambda_2} \frac{1}{2} z_2 z_3 ,
\]

(4.41)

\[
P_3(z) = \frac{\mu}{\lambda_3} + \frac{\theta(1-hs_{it})}{\lambda_3} \frac{1}{2} z_3 + \frac{\theta(1-hs_{it})}{\lambda_3} \frac{1}{2} z_3 z_5 ,
\]

(4.42)

\[
P_4(z) = \frac{\mu}{\lambda_4} + \frac{\theta(1-s_{it})}{\lambda_4} \frac{1}{2} z_5 z_4 + \frac{\theta(1-s_{it})}{\lambda_4} \frac{1}{2} z_5 z_5 ,
\]

(4.43)

\[
P_5(z) = \frac{\mu}{\lambda_5} + \frac{\theta(1-hs_{it})}{\lambda_5} \frac{1}{2} z_5 + \frac{\theta(1-hs_{it})}{\lambda_5} \frac{1}{4} z_5 z_5 + \frac{\theta(1-hs_{it})}{\lambda_5} \frac{1}{4} z_5^2 .
\]

(4.44)

The probability that the meiotic drive gene and drive-insensitive allele are both eventually lost from the population is then given by the smallest solution of the system of five simultaneous equations described by Equation 4.10. For the meiotic drive branching process, this system of equations has the solution:

\[
e_1 = \min \left\{ \frac{1}{(1+r)(1-s_D)(1+t)} \right\} ,
\]

(4.45)

\[
e_3 = e_5 = \min \left\{ \frac{1}{(1+r)(1-hs_{it})} \right\} ,
\]

(4.46)

\[
e_2 = \frac{1}{1 + (2 - e_1 - e_3)(1 + r)(1 - s_D - hs_{it})} ,
\]

(4.47)

\[
e_4 = \frac{1}{1 + 2(1-e_3)(1 + r)(1 - s_{it})} .
\]

(4.48)
Here, $e_1$ is the loss probability beginning with a single type-1 male, $e_2$ is the loss probability beginning with a single type-2 male, and so on. When a type-1 male is released, these equations calculate the extinction probability of the $Y^D$ allele; when a type-3, type-4 or type-5 female is released, the equations calculate the extinction probability of the $X^{it}$ allele; and when a type-2 male is released, the equations calculate the probability that both alleles are eventually lost from the wild mosquito population.

The extinction probability beginning with a type-1 male is a decreasing function of meiotic drive strength and population growth rate and an increasing function of fitness cost due to the meiotic drive gene. Its form is similar to the extinction probability of a heterozygote for the HEG allele. The extinction probability beginning with a type-3 or type-5 female is a decreasing function of population growth rate and an increasing function of the fitness cost due to a single drive-insensitive allele. As explained earlier, the meiotic drive parameter does not influence the extinction probability of the $X^{it}$ allele. The extinction probability beginning with a type-4 female is a function of $e_3$ and is less than one when $e_3$ is less than one. Finally, the extinction probability beginning with a type-2 male is a function of both $e_1$ and $e_3$ and is less than one when both $e_1$ and $e_3$ are less than one.

To make sense of these equations, we need to have some idea of the parameter values and ranges that exist in nature. One of the most important parameters for meiotic drive is the parameter $t$ which represents the fraction by which the $Y^D$ allele is overrepresented in the
gametes of $X^a Y^D$ males. There is currently very little data available for this parameter, and so will rely somewhat on values used in the modeling literature. Although Cha et al. (2006a; 2006b) have studied the effects of the meiotic drive gene $M^D$ on the population dynamics of *Aedes aegypti*, these studies have not yielded any estimates for the parameter $t$. Researchers with experience modeling meiotic drive suggest that the degree of sex distortion should be as high as $t = 0.9$ (Yunxin Huang, personal communication); while the goal of some molecular biologists is to have no females produced at all (Fred Gould, personal communication). Huang et al. (2007) have investigated the full range of meiotic drive parameter values, $t \in [0,1]$, with particular emphasis on the value $t = 0.8$.

The fitness cost associated with the meiotic drive system can originate from both the meiotic drive gene and the X-linked drive-insensitive allele. As for the meiotic drive parameter, $t$, very little data is available for these fitness costs, and so we will rely somewhat on values used in the modeling literature. Researchers with expertise in modeling meiotic drive suggest that a fitness cost on the order of $s_D \approx s_\mu \approx 0.05$ is desirable (Yunxin Huang, personal communication). Huang et al. (2007) have modeled the spread of meiotic drive for fitness costs of 0.05 and 0.2, and have found that a fitness cost on the order of $s_D = s_\mu = 0.05$ is necessary for gene drive to occur over default conditions. We will investigate a fitness cost in the range $s_D, s_\mu \in [0,0.2]$.

For the meiotic drive strategy modeled by Huang et al. (2007), there is an additional fitness cost due to the $X^d$ allele being associated with a refractory gene. This fitness cost
has been discussed for the case of HEGs, and has been estimated by the range
\([-0.05, 0.04]\). Combining these two fitness costs for the \(X^{it}\) allele suggests a fitness cost in
the range \(s_{it} \in [-0.05, 0.24]\). We assume that the fitness cost of the \(X^{it}\) allele is additive
and hence choose \(h = 0.5\).

The meiotic drive strategy by which a Y-linked meiotic drive gene (\(Y^D\)) is used to drive
an X-linked drive-insensitive allele (\(X^{it}\)) into the population is interesting in that, before
the Y-linked drive gene has been driven into the population, the X-linked response allele
is very likely to be lost from the population. In the absence of population growth, the Y-
linked drive gene requires a fitness benefit (\(s_{it} < 0\)) in order to spread (Equation 4.38).
This means that in Figure 4.6A, the loss probability is equal to the loss probability of the
\(Y^D\) allele, since it is certain that the \(X^{it}\) allele and refractory gene will eventually go
extinct. The \(Y^D\) allele is able to persist in the population despite a fitness cost by virtue of
its overrepresentation in the gametes of \(Y^D X^{it}\) males. For a high yet feasible meiotic drive
strength of \(t = 0.8\), the persistence probability of a single escapee is between 19% and
30% for all realistic fitness costs (Figures 4.6A and 4.6B), and the \(Y^D\) allele has a chance
of persisting even during a dramatic population decline (Figure 4.6C).
Figure 4.6 Asymptotic probabilities of meiotic drive loss as a function of fitness cost. (A) Meiotic drive strength and the fitness cost of the meiotic drive allele are varied. (B) Meiotic drive strength and the fitness cost of the drive-insensitive response allele are varied. (C) Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*. (D) The escapee genotype is varied and the $X'^{D} Y^{D}$ release size is increased to five.
The probability of transgenes persisting following an accidental release depends largely on the genotype of the escaped mosquito. If the escapee only carries the transgenic $X^{it}$ allele (as is the case for genotypes $Y^{d}X^{it}$, $X^{sn}X^{it}$ and $X^{it}X^{it}$), then this allele must confer a fitness benefit in order to spread (Figure 4.6D). If the escapee only carries the transgenic $Y^{D}$ allele (as is the case for genotype $Y^{D}X^{it}$), then the escapee has a persistence probability greater than 27% over the entire range of realistic fitness costs (Figure 4.6D).

Interestingly, an escapee having both transgenic alleles ($Y^{D}X^{it}$) has a slightly smaller persistence probability than an escapee only having the transgenic $Y^{D}$ allele. This is because it is usually only the $Y^{D}$ allele that persists following an accidental release, and the transgenic $X^{it}$ allele confers a fitness cost to the escapee.

Under default conditions, for five escapees having both transgenic alleles the loss probability decreases to 20%; while for 10 escapees the loss probability decreases to 4%. The meiotic drive allele, $Y^{D}$, is therefore more invasive than a HEG or TE; however the X-linked response allele, $X^{it}$, is less invasive than either. The meiotic drive allele is more likely to persist than not for escapes consisting of more than two mosquitoes having the $Y^{D}$ allele; while the drive-insensitive response allele requires a fitness benefit in order to have any chance of spreading, and even then is very likely to go extinct.

### 4.2.4 Medea

*Medea*, also known as maternal-effect-dominant embryonic arrest, is another form of meiotic drive. Its dynamics have been studied in *Tribolium* beetles (Wade and Beeman
1994); however it has attracted much recent attention since an engineered Medea element has been observed to rapidly spread through Drosophila populations in the laboratory (Chen et al. 2007). Medea is able to spread through a population despite a fitness cost through its ability to cause the death of all offspring of heterozygous females that do not inherit the Medea allele. This distorts the offspring ratio in favor of the Medea allele.

In the early stages of spread, the vast majority of mosquitoes belong to what may be thought of as a reservoir of wild-types. One implication of this is that the death of wild-type offspring of female heterozygotes will have very little influence on the spread of the Medea allele following an accidental release because the vast majority of mosquitoes will still be wild-types. The early spread of the Medea allele will therefore be very similar to the spread of a new mutation.

Taking this consideration into account, we model the spread of the Medea allele following an accidental release using a two-type continuous-time branching process in which type-1 particles are heterozygous and type-2 particles are homozygous for the Medea allele. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1$ gen$^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity $\theta$ as described by Equation 4.1.
Matings between wild-type mosquitoes and mosquitoes homozygous for the *Medea* allele will produce heterozygotes; while matings between wild-type mosquitoes and heterozygotes for the *Medea* allele will produce offspring that are half heterozygotes and half wild-types (wild-types will die if their heterozygote parent is female). Since homozygotes will be lost from the population after one generation, we are interested in the basic reproductive number of mosquitoes that are heterozygous for the *Medea* allele. For a heterozygote, this is simply given by the average number of heterozygotes that it produces at the end of its life, $f_{1,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with the *Medea* allele. We assume that a female homozygous for *Medea* has a reduced fecundity of $\theta(1 - s)$. Male mosquitoes that mate with wild-type females do not suffer from reduced fecundity, and so the mean fecundity of a homozygote is $\theta(1 - s/2)$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0,1]$, and hence the mean fecundity of a heterozygote is $\theta(1 - hs/2)$.

According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in Figure 4.7. The death intensity for a heterozygote is equal to:

\[
\lambda_1 = \mu + \theta(1 - hs/2),
\]

and the death intensity for a homozygote is equal to:
\[ \lambda_2 = \mu + \theta(1 - s/2) \]  \hspace{1cm} (4.50)

The basic reproductive number for a mosquito heterozygous for the *Medea* allele is then given by:

\[ R_{1,1} = f_{1,1} = \theta \left( \frac{1 - hs}{2} \right) \left( \frac{1 + \frac{1}{2}}{2} \right) \]  \hspace{1cm} (4.51)

When the basic reproductive number is greater than one, *Medea* spread is supercritical, and the *Medea* allele has a nonzero probability of spreading through the mosquito population. Simplifying Equation 4.51, this means that an accidentally released *Medea* element has some chance of spreading into a wild population when:

\[ hs < \frac{2r}{1 + r} \]  \hspace{1cm} (4.52)

When the population size is constant, then the condition for *Medea* spread simplifies to:

\[ hs < 0 \]  \hspace{1cm} (4.53)

One implication of this result is that only the fitness cost to the heterozygote determines whether the *Medea* allele has a chance of spreading or not. If the population size is constant, the *Medea* allele must confer a fitness benefit in order to have some chance of spreading. During periods of population growth, the magnitude of fitness cost that is tolerable is given by Equation 4.52.
Figure 4.7 Schematic for the early spread of a Medea allele through a randomly mating mosquito population. Mosquitoes homozygous for Medea (\(MM\)) have a reduced fecundity of \(\theta(1 - s/2)\) and mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for Medea (\(Mm\)). Heterozygotes have a reduced fecundity of \(\theta(1 - hs/2)\) and mate with wild-type mosquitoes to give birth to heterozygotes with probability 1/2 and wild-types with probability 1/2. All mosquitoes have a death rate of \(\mu\).
When the basic reproductive number is greater than one, Medea spread is possible; however the probability of Medea loss may still be very high. To determine the extinction probability of the Medea allele, we first define the probability generating function for the Medea branching process. This is a function of the vector $z = (z_1, z_2)$ and is defined in Equation 4.7. Substituting the $p_{ij}$ terms from the Medea branching process into this equation, we have the probability generating function,

$$P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1} \left(1 - \frac{hs}{2}\right) z_1 + \frac{\theta}{\lambda_1} \left(1 - \frac{hs}{2}\right) \frac{1}{2} z_1^2,$$

$$P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2} \left(1 - \frac{s}{2}\right) z_1 z_2.$$

The probability that the Medea allele is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Equation 4.10. For the Medea branching process, this system of equations has the solution,

$$e_1 = \min \left\{ \frac{2}{(1+r)(2-hs)}, 1 \right\},$$

$$e_2 = \frac{1}{1 + (1+r)(2-s)(1-e_1)}.$$

The extinction probability beginning with a heterozygote, $e_1$, is a decreasing function of population growth rate and an increasing function of heterozygote fitness cost. The extinction probability beginning with a homozygote, $e_2$, is a function of $e_1$ and is less than one when $e_1$ is less than one.
To make sense of these equations, we need to have some idea of the fitness consequences of the Medea allele. While very few measurements exist for this parameter, Chen et al. (2007) have fitted data from their synthetic Medea element in Drosophila to a model of Medea spread. They found that their data is most consistent with an element that has no fitness cost; however their confidence interval for this parameter, $s \in [-0.23, 0.1]$, is large. The fitness cost of being associated with a refractory gene has been discussed for the case of HEGs, and has been estimated by the range $[-0.05, 0.04]$. Combining these two estimates suggests a range for the fitness cost of a Medea allele on the order of $s \in [-0.28, 0.14]$.

Following Chen et al. (2007), we assume that the fitness cost of the Medea allele is additive and hence choose a degree of homozygosity of $h = 0.5$. The proportion of wild-type embryos that die because they do not produce an antidote to the toxin produced by a female Medea-infected parent, $t$, although important, is not relevant to the early spread of the Medea allele.

Medea is an ideal drive system for the needs of transgene containment since it will spread very quickly following an intentional release (Chen et al. 2007); however it requires either a fitness benefit or period of population growth in order to spread following an accidental release (Figure 4.8). During a period of population decline, the Medea allele has no chance of spreading; however for a population growth rate of $r = 0.1$ gen$^{-1}$, the
persistence probability for a single homozygous escapee is between 11% and 17% over
the full range of realistic parameterizations (Figure 4.8A).

Figure 4.8 Asymptotic probabilities of Medea loss as a function of fitness cost. (A) Population growth rate is varied. The case where \( r = 0.2 \) per generation corresponds to the seasonal population growth phase of An. gambiae. (B) The escapee genotype is varied and the homozygote release size is increased to 10.
The probability that a beneficial Medea allele is lost from the population following an accidental release is most affected by the number of escapees during an accidental release (Figure 4.8B). For a realistic fitness benefit of $s = -0.05$, an escape of 10 mosquitoes homozygous for the Medea allele decreases the loss probability to 78%. For 25 escapees, the loss probability decreases to 54%. Under default conditions, the Medea allele is more likely to persist than not for homozygous escape sizes greater than 27.

Medea is much less prone to spread following an accidental release than HEGs, TEs or meiotic drive genes; however it should still be kept in mind that an escape of ~30 transgenic mosquitoes is entirely feasible. If these mosquitoes are refractory to malaria, it is also feasible that they may possess a slight fitness advantage over wild mosquitoes. Additionally, the observation of Medea alleles in nature suggests that naturally-occurring Medea alleles may actually confer an innate selective advantage to their host; as hinted at by the confidence interval for fitness effects of the synthetic Medea allele studied by Chen et al. (2007). Therefore, while Medea is significantly safer than HEGs, TEs and meiotic drive genes; we should still be weary of its ability to spread following an accidental release.

4.2.5 Wolbachia

Wolbachia is a maternally-inherited, intracellular bacterium found in a wide variety of invertebrate taxa. Wolbachia infections are associated with several host reproductive alterations including cytoplasmic incompatibility (Stouthammer et al. 1999), in which
offspring of matings between infected males and uninfected females are completely or partially sterilized, while matings involving infected females always produce infected offspring. This favors the offspring ratio in favor of the \textit{Wolbachia} infection, and \textit{Wolbachia} is therefore able to spread rapidly through a population despite a fitness cost (Turelli and Hoffmann 1999).

To calculate the basic reproductive number of a \textit{Wolbachia} bacterium following an accidental release, we consider a two-type continuous-time branching process in which type-1 particles are females infected with the \textit{Wolbachia} bacterium and type-2 particles are \textit{Wolbachia}-infected males. We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1$ gen$^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity $\theta$ as described by Equation 4.1.

In the early stages of spread, almost all matings involve at least one mosquito uninfected by the \textit{Wolbachia} bacterium. Crosses between \textit{Wolbachia}-infected females and uninfected males tend to produce \textit{Wolbachia}-infected males and females; however maternal transmission is imperfect such that a proportion, $u$, of their offspring are uninfected and a proportion, $1-u$, of their offspring are infected by \textit{Wolbachia}. Crosses between \textit{Wolbachia}-infected males and uninfected females tend to be infertile with a proportion, $t$, of their offspring suffering from CI-induced sterility and a proportion,
1−t, of their offspring surviving. The viable offspring of this cross, however, are uninfected and may be thought of as belonging to the reservoir of uninfected hosts.

Since Wolbachia is maternally inherited, we are only interested in the basic reproductive number of Wolbachia-infected females. This is equal to the average number of Wolbachia-infected females that it produces at the end of its life, \( f_{i,1} \).

In order to calculate this quantity, we first need to consider the fitness cost associated with a Wolbachia infection. We assume that a female infected with the Wolbachia bacterium has a reduced fecundity of \( \theta(1 − s) \). Infected males that mate with uninfected females do not suffer from reduced fecundity, and so their fecundity is \( \theta \).

According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in Figure 4.9. The death intensity for a Wolbachia-infected female is equal to:

\[
\lambda_1 = \mu + \theta(1 − s) ,
\]

and the death intensity for a Wolbachia-infected male is equal to:

\[
\lambda_2 = \mu + \theta .
\]

The basic reproductive number for a Wolbachia-infected female is then given by:

\[
R_{1,1} = f_{i,1} = \frac{\theta(1 − s)}{\lambda_1} \left( 1 + \frac{1−u}{2} \right) .
\]
When the basic reproductive number is greater than one, *Wolbachia* spread is supercritical, and the *Wolbachia* bacterium has a nonzero probability of spreading through the mosquito population. Simplifying Equation 4.60, this means that an accidentally released *Wolbachia*-infected female has some chance of spreading into a wild population when:

\[
(1 + r)(1 - s)(1 - u) > 1.
\]  

(4.61)

This equation implies that population growth can somewhat compensate for imperfect maternal transmission and a fitness cost associated with the *Wolbachia* bacterium. When the population size is constant \((r = 0)\), this equation is consistent with the condition derived by Turelli *et al.* (1992) for *Wolbachia* spread. According to Turelli *et al.* (1992), since maternal transmission is always imperfect \((u > 0)\) then, when the population size is constant, the *Wolbachia* bacterium must confer a fitness benefit to the mosquito in order for it to spread from a low prevalence. The magnitude of the fitness benefit, \(-s\), required for *Wolbachia* spread is then given by:

\[
-s > \frac{1}{1-u} - 1.
\]  

(4.62)
Figure 4.9 Schematic for the early spread of a *Wolbachia* bacterium through a randomly mating mosquito population. *Wolbachia*-infected females ($W_f$) have a reduced fecundity of $\theta(1 - s)$ and mate with wild-type mosquitoes to give birth to *Wolbachia*-infected males and females, each with probability $\frac{1}{2}(1 - u)$, and wild-types with probability $u$. All mosquitoes have a death rate of $\mu$. 
When the basic reproductive number is greater than one, *Wolbachia* spread is possible; however the probability of *Wolbachia* loss may still be very high. To determine the extinction probability of the *Wolbachia* bacterium, we first define the probability generating function for the *Wolbachia* branching process. This is a function of the vector \( z = (z_1, z_2) \) and is defined in Equation 4.7. Substituting the terms from the *Wolbachia* branching process into this equation, we have the probability generating function,

\[
P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta(1-s)}{\lambda_1} uz_1 + \frac{\theta(1-s) 1-u}{2 \lambda_1} z_1^2 + \frac{\theta(1-s) 1-u}{2 \lambda_1} z_1 z_2 ,
\]

\[
P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2} z_2 .
\]

The probability that *Wolbachia* is eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Equation 4.10. For the *Wolbachia* branching process, this system of equations has the solution,

\[
e_1 = \min \left\{ \frac{1}{(1+r)(1-s)(1-u)} \right\} ,
\]

\[
e_2 = 1 .
\]

The extinction probability beginning with a *Wolbachia*-infected female, \( e_1 \), is a decreasing function of population growth rate and an increasing function of fitness cost. It is also a decreasing function of the reduction in maternal transmission. An accidental release beginning with a *Wolbachia*-infected male is certain to go extinct because *Wolbachia* is maternally inherited.
To make sense of these equations, we need to have some idea of the parameter values and ranges that exist in nature. One of the most important parameters for the spread of a *Wolbachia* infection is the proportion, $t$, of uninfected embryos from a cross involving an uninfected female and an infected male that die due to cytoplasmic incompatibility. This parameter is not so relevant in the early stages of spread, however, since surviving offspring belong to what may be thought of as a reservoir of uninfected hosts.

Another important parameter is the proportion, $u$, of offspring from *Wolbachia*-infected females that are uninfected by *Wolbachia*. Maternal transmission is high, so this parameter tends to be relatively small. Charlat *et al.* (2004) have estimated a 95% confidence interval for this parameter for *D. melanogaster* in the range $u \in [0.01, 0.17]$. Point estimates and ranges estimated for other species are within this range – for example, $u = 0.01$ for *Culex pipiens* in California (Rasgon and Scott 2003), $u = 0.025$ for *D. melanogaster* in Australia (Hoffmann *et al.* 1994, 1998), and $u \in [0.03, 0.15]$ for *D. simulans* in California (Weeks *et al.* 2007).

The fitness consequences of a *Wolbachia* infection are particularly relevant to its ability to spread. Recent measurements suggest that a *Wolbachia* infection can induce either a fitness benefit or cost, depending on the host species, genetic background (Dean 2006), and the amount of time that the infection has been present in the host species (Weeks *et al.* 2007). In terms of fitness costs, Hoffmann *et al.* (1990) have measured a reduction in
fecundity of $s \in [0.1,0.2]$ for *Wolbachia*-infected *D. melanogaster*, and Stevens and Wade (1990) have measured a reduction in overall fitness of $s = 0.37$ for *Wolbachia*-infected *Tribolium* beetles. Neutral *Wolbachia* infections ($s = 0$) have been observed in *Drosophila yakuba* (Charlat et al. 2004), *C. pipiens* (Rasgon and Scott 2003), and in natural populations of *D. simulans* (Turelli and Hoffmann 1995, 1999).

It is becoming increasingly recognized that *Wolbachia* may not just be a reproductive parasite but may also be a mutualist, and hence confer a selective advantage to its host (Steve Sinkins, personal communication). Weeks *et al.* (2007) have noted that, although 20 years ago California *D. simulans* were shown to have a reduced fecundity due to *Wolbachia* infection in the laboratory, *Wolbachia*-infected females now exhibit an increased fecundity of $s = -0.1$ in the laboratory. Dobson *et al.* (2002) have also observed a fecundity advantage due to *Wolbachia* superinfection in the mosquito *Aedes albopictus*; however this is yet to be validated in field populations (Jason Rasgon, personal communication). Combining these measurements and accounting for the potential effects of a refractory gene suggests a range of fecundity effects due to *Wolbachia* infection on the order of $s \in [-0.1,0.37]$.

Like *Medea*, *Wolbachia* is an ideal drive system for the needs of transgenic containment since it will spread following an intentional release (Turelli and Hoffmann 1999); however it requires either a fitness benefit or period of population growth in order to spread following an accidental release (Figure 4.10). The conditions for spread are a little
more restrictive for Wolbachia than they are for Medea, primarily because maternal transmission of Wolbachia is imperfect. As the efficiency of maternal transmission decreases, then the fitness benefit required for Wolbachia spread becomes larger (Figure 4.10A). Population growth can also compensate for a fitness cost and imperfect maternal transmission. For a population growth rate of $r = 0.1 \text{ gen}^{-1}$ and a maternal transmission rate of 97%, the Wolbachia bacterium will spread for fitness costs less than 6.5% (Figure 4.10B).
Figure 4.10 Asymptotic probabilities of *Wolbachia* loss as a function of fitness cost. (A) The proportion of offspring from *Wolbachia*-infected females that are uninfected by *Wolbachia* is varied. (B) Population growth rate is varied. The case where $r = 0.2$ per generation corresponds to the seasonal population growth phase of *An. gambiae*. (C) Release size is increased to 25.
For a *Wolbachia* bacterium that is able to spread following an accidental release, its loss probability is most affected by the number of escapees during an accidental release (Figure 4.10C). For a realistic fitness benefit of \( s = -0.05 \), an escape of 10 *Wolbachia*-infected females decreases the loss probability to 83%. For 25 infected females, the loss probability decreases to 63%. Under default conditions, the *Wolbachia* bacterium is more likely to persist than not for infected female escape sizes greater than 37. It is likely that less than half of the escapees from an accidental release will be infected females, and so this corresponds to a required escape size greater than 74 in order for it to be more likely than not that the *Wolbachia* bacterium will spread following an accidental release.

*Wolbachia* is less prone to spread than *Medea* following an accidental release; however, as for *Medea*, it should be kept in mind that an escape of \(~75\) infected mosquitoes is possible. Recent measurements suggest that, in addition to a fitness benefit due to malaria refractoriness, the *Wolbachia* infection may in fact confer a selective advantage to the escapees over wild mosquitoes (Weeks *et al.* 2007). Additionally, the observation that *Wolbachia* bacteria exist in nature suggests that *Wolbachia* infections have been able to spread from very low prevalences, in contradiction to the prediction by Turelli and Hoffmann (1999) that the initial frequency of *Wolbachia* must exceed \(~10\%\) in order to spread. The fact that real populations are locally structured may assist in the emergence of a new *Wolbachia* infection (Steve Sinkins and Jason Rasgon, personal communication). Therefore, while *Wolbachia* is one of the safest drive systems; there is
still a possibility that a transgenic *Wolbachia* strain will persist in the wild mosquito population following an accidental release.

It should also be noted that the extinction probabilities calculated here are for the *Wolbachia* drive strategy in which the refractory gene is engineered directly into the *Wolbachia* genome. Other drive strategies utilizing *Wolbachia* have been proposed by Turrelli and Hoffmann (1999) and Sinkins and Godfray (2004) and may be associated with slightly different loss probabilities.

### 4.2.6 Engineered underdominance

The simplest case of underdominance is when a trait is determined by two alleles at a single locus and the fitness of a heterozygote is less than that of either homozygote (Hartl and Clark 1989). The dynamics of underdominant traits are generally unstable and, depending on the initial frequency of the two alleles, one will tend to be lost while the other will reach fixation in the population (Crow and Kimura 1970; Spiess 1977). However, the problem with single-locus underdominance as a form of gene drive is that, in order for an introduced allele to reach fixation, it must be introduced at a very high frequency in the population which is often not feasible for gene drive strategies (Curtis 1968).

A novel form of engineered underdominance has been suggested by Davis *et al.* (2001) that doesn’t require such a high release size in order to spread into a naïve population. In
this system there are two transgenic constructs, $\alpha$ and $\beta$, each of which possesses a lethal gene and a suppressor gene that down-regulates the expression of the lethal gene on the other construct. This system is most efficient when the two transgenic constructs are inserted at loci on nonhomologous chromosomes, and when there are anti-pathogen genes associated with each construct.

Since both loci can be homozygous for the wild-type allele, homozygous for the transgenic allele, or heterozygous, then there are nine possible genotypes for the engineered underdominance system. However, individuals that only possess one transgenic construct are unviable because they express a lethal gene while lacking its suppressor. This means that there are only five viable genotypes – four of which possess both transgenic constructs, and one which belongs to the reservoir of wild-types.

Following an accidental release, we are interested in tracking the four viable genotypes possessing both transgenic constructs. Let us denote the transgenic allele at the first locus as $\alpha$, the wild-type allele at the first locus as $A$, the transgenic allele at the second locus as $\beta$, and the wild-type allele at the second locus as $B$. Utilizing these symbols, we consider a four-type continuous-time branching process in which type-1 particles are heterozygous at both loci ($\alpha A \beta B$), type-2 particles are heterozygous at the first locus ($\alpha A \beta \beta$), type-3 particles are heterozygous at the second locus ($\alpha a \beta B$), and type-4 particles are homozygous at both loci ($\alpha a \beta \beta$).
We consider a budding model in which the death rate of all mosquitoes is set to $\mu = 1$ gen$^{-1}$, and fitness differences between particle-types are accounted for by differences in female fecundity. Wild-type mosquitoes are assumed to have fecundity $\theta$ as described by Equation 4.1.

In the early stages following an accidental release, almost all matings involve at least one wild-type mosquito. This means that the only viable transgenic offspring will be heterozygous at both loci, since all offspring will inherit a wild-type allele from their wild-type parent at both loci. In calculating the basic reproductive number of the engineered underdominance constructs, we are therefore interested in the average number of type-1 offspring that a type-1 mosquito produces at the end of its life, $f_{i,1}$.

In order to calculate this quantity, we first need to consider the fitness cost associated with the engineered underdominance constructs. It is likely that the two constructs will each confer a fitness cost to the host, since binding of a suppressor to its corresponding lethal gene may be imperfect. We assume that a female homozygous for the $\alpha$ allele has their fecundity reduced by a fraction, $s$. Male mosquitoes that mate with wild-type mosquitoes do not suffer from reduced fecundity, and so the mean reduction in fecundity due to being homozygous for the $\alpha$ allele is $s/2$. This reduction in fecundity is assumed to have a dominance factor of $h \in [0,1]$, and hence the mean reduction in fecundity due to being heterozygous for the $\alpha$ allele is $hs/2$. The $\beta$ allele is expected to cause a similar reduction in fecundity, and so type-1 particles are expected to have a fecundity of
\( \theta(1 - hs), \) type-2 and type-3 particles have an expected fecundity of \( \theta(1 - (s + hs)/2), \) and type-4 particles have an expected fecundity of \( \theta(1 - s). \)

According to standard branching process theory (Lange 2002), these assumptions lead to the branching process shown in Figure 4.11. The death intensities for each particle type are equal to:

\[
\lambda_1 = \mu + \theta(1 - hs),
\]

(4.67)

\[
\lambda_2 = \lambda_3 = \mu + \theta \left( 1 - \frac{s + hs}{2} \right),
\]

(4.68)

\[
\lambda_4 = \mu + \theta(1 - s).
\]

(4.69)

The basic reproductive number for a mosquito heterozygous at both loci is then given by:

\[
R_{1,1} = \frac{f_{1,1}}{\lambda_1} = \frac{\theta(1 - hs) \left( 1 + \frac{1}{4} \right)}{\lambda_1}.
\]

(4.70)

When the basic reproductive number is greater than one, spread of the engineered underdominance alleles is supercritical, and the engineered constructs have a nonzero probability of spreading through the mosquito population. Simplifying Equation 4.70, this means that an accidentally released mosquito having both underdominance constructs has some chance of spreading into a wild population when:

\[
- hs > \frac{1-r}{1+r}.
\]

(4.71)

When the population size is constant, then the condition for engineered underdominance spread simplifies to:

\[
- hs > 1.
\]

(4.72)
Noting that the fecundity of a type-1 particle is $\theta(1 - hs)$, this condition means that, when the population size is constant, the underdominance alleles have a chance of spreading when the fecundity of a type-1 particle is greater than $2\theta$. The fecundity of a wild-type mosquito is equal to $\theta$, and so this implies that a heterozygote for both underdominance alleles must be twice as fecund as a wild-type mosquito in order for the underdominance alleles to have some chance of spreading into a population of constant size. This condition is moderately relaxed during periods of population growth, as described by Equation 4.71.
Figure 4.11 Schematic for the early spread of a pair of engineered underdominance alleles through a randomly mating mosquito population. Mosquitoes homozygous for one or both alleles (\(\alpha A\beta\beta\), \(\alpha\alpha\beta\beta\), \(\alpha\alpha\beta\beta\) mate with wild-type mosquitoes to give birth to mosquitoes heterozygous for both alleles (\(\alpha\alpha\beta\beta\)). Heterozygotes for both alleles have a reduced fecundity of \(\theta(1-hs)\) and mate with wild-type mosquitoes to give birth to heterozygotes for both alleles with probability \(1/4\), wild-types with probability \(1/4\) and unviable offspring with probability \(1/2\). All mosquitoes have a death rate of \(\mu\).
When the basic reproductive number is greater than one, spread of the underdominance alleles is possible; however the probability of loss of these alleles may still be very high. To determine the extinction probability of the underdominance alleles, we first define the probability generating function for the engineered underdominance branching process. This is a function of the vector \( z = (z_1, z_2, z_3, z_4) \) and is defined as,

\[
P_i(z) = \sum_{j=1}^{4} p_{ij} z^j = \sum_{j=1}^{4} p_{ij} z_1^{j_1} z_2^{j_2} z_3^{j_3} z_4^{j_4}, \tag{4.73}
\]

where \( p_{ij} \) is the probability that a type-\( i \) particle gives rise to \( j_1 \) type-1 particles, \( j_2 \) type-2 particles, and so on, and \( j \) is defined as the vector \( j = (j_1, ..., j_4) \). Substituting the \( p_{ij} \) terms from the engineered underdominance branching process into this equation, we have the probability generating function,

\[
P_1(z) = \frac{\mu}{\lambda_1} + \frac{\theta}{\lambda_1} (1 - hs) \frac{3}{4} z_1 + \frac{\theta}{\lambda_1} (1 - hs) \frac{1}{4} z_1^2, \tag{4.74}
\]

\[
P_2(z) = \frac{\mu}{\lambda_2} + \frac{\theta}{\lambda_2} \left( 1 - \frac{s + hs}{2} \right) \frac{1}{2} z_2 + \frac{\theta}{\lambda_2} \left( 1 - \frac{s + hs}{2} \right) \frac{1}{2} z_2 z_1, \tag{4.75}
\]

\[
P_3(z) = \frac{\mu}{\lambda_3} + \frac{\theta}{\lambda_3} \left( 1 - \frac{s + hs}{2} \right) \frac{1}{2} z_3 + \frac{\theta}{\lambda_3} \left( 1 - \frac{s + hs}{2} \right) \frac{1}{2} z_3 z_1, \tag{4.76}
\]

\[
P_4(z) = \frac{\mu}{\lambda_4} + \frac{\theta}{\lambda_4} (1 - s) z_1 z_4, \tag{4.77}
\]

The probability that engineered underdominance alleles are eventually lost from the population is then given by the smallest solution of the system of two simultaneous equations described by Equation 4.10. For the engineered underdominance branching process, this system of equations has the solution,
The extinction probability beginning with a heterozygote for both alleles, \( e_1 \), is a decreasing function of population growth rate and an increasing function of heterozygote fitness cost. The extinction probabilities beginning with any other viable genotype are all functions of \( e_1 \), and are less than one when \( e_1 \) is less than one.

To make sense of these equations, we need to have some idea of the fitness consequences that engineered underdominance constructs have on their hosts. The two allele, two locus strategy for engineered underdominance has only recently been proposed as a mechanism of gene drive and so there are no experimental measurements of its effects on female fecundity. We will therefore rely on fitness costs used in the modeling literature.

Magori and Gould (2006) have investigated a large range of fitness costs due to underdominance constructs, \( s \in [0,0.5] \). While they note that a neutral underdominance construct is unrealistic; their model predicts that a refractory gene could be driven into a population in the presence of a fitness cost of \( s = 0.05 \) provided a sufficient release ratio is achieved. Assuming that a beneficial refractory gene may be associated with each engineered construct, we explore an expanded range of fecundity effects on the order of
$s \in [-0.05,0.5]$. For completeness, we also consider a full range for the degree of homozygosity of $h \in [0,1]$.

Engineered underdominance constructs are the safest gene drive system for the needs of transgenic containment. Their safety in the event of an accidental release primarily arises from the high release ratio required for them to spread. Davis et al. (2001) showed that, even under ideal conditions in which there are no fitness costs associated with the constructs, transgenic insects must exceed a frequency of ~27% in order to have some chance of spreading through the population. The required release ratio increases as fitness costs are accounted for (Magori and Gould 2006); and fecundity effects due to underdominance constructs are unlikely to be positive given the constructs are engineered to express lethal genes down-regulated by suppressors.

According to Equation 4.71, engineered underdominance constructs face certain extinction following an accidental release for the entire range of realistic parameters. Even for the most generous estimate of population growth rate ($r = 0.2 \ \text{gen}^{-1}$), a heterozygote for one of the constructs must have a fitness benefit of at least 67% in order to have some chance of spreading from a low prevalence. This is unachievable for even the most generous estimates of the parameters $h$ and $s$. In order to spread following an escape from a field cage, the escape size must represent a significant fraction of the wild population, as predicted by Davis et al. (2001) and Magori and Gould (2006).
It should be noted that the two allele, two locus strategy for engineered underdominance is not the only possible strategy for spreading underdominance constructs into a wild population; however for the time being it is reasonable to model this since more complicated strategies are further down the road (Fred Gould, personal communication). Variants of this strategy have been proposed by Magori and Gould (2006); and although they are also very likely to be lost following an accidental release, their extinction probabilities and required release ratios may be smaller than for the two allele, two locus strategy.

4.3 DISCUSSION

In the event of an actual escape of transgenic mosquitoes from an ambient field cage, there will be many factors influencing the loss or persistence of transgenic DNA. Escapees from field cages will likely suffer from some degree of inbreeding depression and may be slightly maladapted to conditions in the wild. Control measures may also be put in place to reduce the spread of transgenic DNA – for example, the use of vegetation-free zones or “trap crops” surrounding the cage to restrict mosquito dispersal. Consequently, the loss probabilities calculated here contain systematic errors and are mainly of comparative interest for studying the possible outcomes under a variety of gene drive strategies.

Comparison of the predictions of the six branching processes in this paper suggests that engineered underdominance constructs provide the safest gene drive strategy for the
needs of transgenic containment. According to model predictions, this system will face certain extinction for the entire range of realistic parameters provided that the release proportion is less than ~27% of the wild population (Davis et al. 2001). The release proportion required for transgenic spread is even higher if there is a fitness cost associated with the constructs.

The catch-22 of this result is that engineered underdominance constructs are also not very invasive following an intentional release. Any gene drive strategy with the aim of controlling disease over a large geographic area must have the ability to spread between several partially-isolated subpopulations of disease vectors. The problem with engineered underdominance constructs is that, even under the most generous parameterizations, they require a migration rate consistently above 3% per generation in order to spread into neighbouring populations (Davis et al. 2001). This required migration rate is higher than those observed between subpopulations of An. gambiae mosquitoes in Mali, West Africa (Taylor et al. 2001).

TEs and HEGs, on the other hand, are very capable of spreading between subpopulations of disease vectors following an intentional release (Taylor and Manoukis 2003; Deredec et al. 2007); but are also very capable of spreading following an accidental release. According to model predictions with default parameters, a TE is more than 50% likely to persist for escape sizes greater than 10; while a HEG is more than 50% likely to persist for homozygote escape sizes greater than four. These escape sizes are entirely within the
realm of possibility; and hence the invasiveness of TEs and HEGs is offset by the risk of establishment following an accidental release.

Meiotic drive systems are also of concern following an accidental release. For the meiotic drive strategy analyzed in which a Y-linked meiotic drive gene is used to drive an X-linked response allele into the population, the Y-linked drive gene is very capable of spreading following an accidental release, while the X-linked response allele requires population growth or a fitness benefit in order to spread. If we are concerned about the unknown effects of an attached refractory allele, then the configuration of this meiotic drive strategy is good since the refractory allele is to be attached to the X-linked response allele, and hence is likely to be lost following an accidental release. However, if we are also concerned about the unknown effects of the drive gene on the mosquito population, then this strategy is worrying since the drive gene is more than 50% likely to persist for escapes of more than two mosquitoes carrying the drive gene.

Medea and Wolbachia provide a good compromise between invasiveness following an intentional release and containment following an accidental release. Both systems will spread very quickly following an intentional release (Chen et al. 2006; Turelli and Hoffmann 1999); however they require either a fitness benefit or population growth in order to spread following an accidental release. The conditions for spread are a little more restrictive for Wolbachia, primarily because maternal transmission of Wolbachia is imperfect.
Despite this, we should still be wary of the ability of Medea and Wolbachia to persist in a wild mosquito population following an accidental release. It is feasible that Medea or Wolbachia-infected mosquitoes could possess a slight fitness advantage over wild-type mosquitoes. Under such conditions, an advantageous Medea allele is more than 50% likely to persist for homozygote escape sizes greater than 27, while an advantageous Wolbachia bacterium is more than 50% likely to persist for escape sizes greater than 74. The concern that we should be wary of is that the existence of a fitness advantage under natural conditions will not be known until experiments have been carried out in ambient field cages. During these experiments, an escape of this size is entirely within the realm of possibility.

An additional concern is that Medea and Wolbachia may be more invasive following an accidental release than predicted by simple models of population dynamics. The observation of Medea alleles and Wolbachia bacteria in nature (Wade and Beeman 1994; Stevens and Wade 1990) suggests that both have been able to spread from very low prevalences. The existence of local population structure has not been accounted for in the branching processes in this paper; however it may help to enable the establishment of a new Wolbachia infection, Medea allele or drive-insensitive response allele for a meiotic drive system from a very low initial prevalence.
Despite these concerns, Medea and Wolbachia still represent the best compromise between invasiveness and containment for the six gene drive systems currently being considered (Sinkins and Gould 2006). The mechanism of gene drive should be thought of as a form of biological containment in the design and implementation of ambient cage trials. This is only one of many considerations in the design of such experiments; however it is an important additional consideration to the comprehensive list compiled by Benedict et al. (2008). Given the lack of knowledge of the outcomes of such technology, all efforts should be taken to prevent the release of transgenic strains before their efficacy and side-effects have been adequately studied.
5 CHAPTER 5
SUMMARY AND CONCLUSIONS

5.1 REVIEW OF HYPOTHESES

Throughout this dissertation, we have used available data to examine the promise that genetically modified mosquitoes (GMMs) have for the control of malaria and other vector-borne diseases. Focusing on transgenic TE constructs, we have examined the impact of dissociation between the drive system and refractory gene. We have also analyzed the molecular and ecological conditions for which TEs will spread through a randomly mating population, and the probabilities that each of the six main gene drive systems will persist in a wild population following an accidental release. The corresponding hypotheses are listed in Table 5.1. Our analyses suggest that TE-mediated gene drive is unlikely to be successful as a sole or primary means of malaria control.

<table>
<thead>
<tr>
<th>Index</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The prevalence of disease-resistant vectors depends on the rate of dissociation between the TE and refractory gene.</td>
</tr>
<tr>
<td>2.2</td>
<td>A transpositional handicap due to the refractory gene decreases the prevalence of disease-resistant vectors.</td>
</tr>
<tr>
<td>2.3</td>
<td>A fitness benefit due to the refractory trait increases the prevalence of disease-resistant vectors.</td>
</tr>
<tr>
<td>3.1</td>
<td>There exists a set of molecular and ecological parameters for which TEs are unlikely to be lost from a randomly mating population.</td>
</tr>
<tr>
<td>3.2</td>
<td>There exists a set of molecular and ecological parameters for which TEs will spread at a rate acceptable to public health goals.</td>
</tr>
<tr>
<td>3.3</td>
<td>A TE released at the beginning of the population growth phase has a greater chance of persisting in the wild.</td>
</tr>
<tr>
<td>4.1</td>
<td>Gene drive systems enhance the invasiveness of introduced genes.</td>
</tr>
<tr>
<td>4.2</td>
<td>Certain gene drive systems are more likely to persist than others following an accidental release.</td>
</tr>
<tr>
<td>4.3</td>
<td>Gene drive systems are more likely to persist during conditions of population growth.</td>
</tr>
</tbody>
</table>

Table 5.1 Hypotheses examined in the dissertation.
5.1.1 The impact of dissociation on TE-mediated disease control

In agreement with Hypothesis 2.1, dissociation rate has a large effect on the population frequency of refractory mosquitoes. The best estimates of dissociation rate have been measured for P elements in D. melanogaster. For P elements engineered with a functional gene between their inverted repeats, this gene has lost its function at an average rate of 0.05 per element per generation (TE<sup>-1</sup> gen<sup>-1</sup>) (Levis et al. 1985; Robertson et al. 1988). At this rate, model predictions are that the refractory allele will reach a maximum frequency of ~20% in the mosquito population (Chapter 2). According to Equation 1.2, this corresponds to a ~20% reduction in the Entomological Inoculation Rate (EIR). The Africa-wide average of EIR has been reported as 134.2 infective bites per person per year (Smith et al. 2005), with a corresponding malaria prevalence of 73.3%. According to Equation 1.1, reducing the EIR by 20% will reduce the Africa-wide malaria prevalence by a mere 1.5% to 71.8%.

The P element of D. melanogaster is fairly unique, and there is evidence to suggest that other TEs may have much smaller dissociation rates (Submaranian et al. 2007). However even for dissociation rates on the order of 0.001 TE<sup>-1</sup> gen<sup>-1</sup>, model predictions are that the refractory allele will reach a maximum frequency of 97.6% in the mosquito population (Chapter 2), leading to an Africa-wide malaria prevalence of 37%. In all cases these temporary gains are eventually lost and, unless novel and effective transgenic strains can be released repeatedly, the mosquito population will become susceptible to malaria again.
These predictions are all made neglecting the massive impact of transpositional handicaps on the spread of intact TE constructs. Given the current insert sizes (Kenneth Olson and Marcelo Jacobs-Lorena, personal communication), a Himar1 mariner transposon would have its transposition rate decreased by at least 95% (Lampe et al. 1998). Under these circumstances, in agreement with Hypothesis 2.2, model predictions are that the refractory gene would barely spread at all (Chapter 2). A fitness benefit can partially make up for this handicap, in agreement with Hypothesis 2.3. However, even for the best case scenario of a 5% fitness benefit and a dissociation rate of 0.001 TE\(^{-1}\) gen\(^{-1}\), model predictions are that the refractory allele will reach a maximum frequency of 76% in the mosquito population (Chapter 2), leading to an Africa-wide malaria prevalence of 63%. This is nowhere near the level of malaria control envisioned.

5.1.2 Molecular and ecological parameters for TE spread

TEs are more invasive than most genes following introduction into a novel population due to their ability to spread despite a fitness cost. Therefore, in agreement with Hypothesis 3.1, there is a particularly generous subset of parameter space for which a TE is able to spread. For small fitness costs and a constant population size, our model predicts that TE spread is supercritical when the magnitude of the replicative transposition rate is greater than the magnitude of the fitness cost. This condition is relaxed during periods of population growth (Chapter 3).
As with any introduced gene, the probability of loss is very high for small release sizes. A TE with a high yet realistic transposition rate of 0.1 TE\(^{-1}\) gen\(^{-1}\) (Seleme et al. 1999; Vasilyeva et al. 1999) has a loss probability of 83% following a year of population growth and decline; however this probability falls off quickly as the release size increases. For a release or ten transgenic mosquitoes under default conditions, the loss probability falls to 15% one year following release. This probability falls below 1% for a release consisting of 25 transgenic mosquitoes engineered with the TE construct (Chapter 3). These results are promising for the efficiency of an intentional release; however they cause concern for the possibility of an accidental release. Given the impossibility of absolute containment (Benedict et al. 2008), this leads us to question the acceptability of planned outdoor cage trials that are necessary for the technology to proceed.

Of interest to an intentional release is the rate at which a TE will spread through a population. A requirement of any successful gene drive strategy is that the system must work within a public health time frame (Braig and Yan 2001). In agreement with Hypothesis 3.2, there is also a generous subset of parameter space for which a TE will satisfy this requirement. Rate of TE spread is most influenced by the difference between transposition rate and fitness cost. Assuming a fitness cost equal to that of a spontaneous mutation and a transposition rate on the order of 0.1 TE\(^{-1}\) gen\(^{-1}\), a TE will spread from a prevalence of 2.5% to near fixation in a randomly mating population in a less than four years (Chapter 3). The optimal release time is at the beginning of the population growth phase, in agreement with Hypothesis 3.3 (Chapter 3).
5.1.3 The effect of gene drive on transgene containment

It is worth noting that, while TEs have a high probability of being established in a wild mosquito population, not all of the drive mechanisms under consideration are this invasive. Hypothesis 4.2 is correct in stating that certain gene drive systems are more likely to persist than others following an accidental release; however the statement of Hypothesis 4.1 that “gene drive systems enhance the invasiveness of introduced genes” is conditional upon both the drive mechanism and the release size.

Homing endonuclease genes (HEGs) and TEs are among the most invasive of the gene drive strategies currently being discussed and are capable of spreading regardless of the release size. For the meiotic drive strategy in which a Y-linked meiotic drive gene is used to drive an X-linked response allele into the population, the drive gene is very capable of spreading following an accidental release; however the response allele requires population growth or a fitness benefit in order to spread. The response allele will spread in the absence of these two factors on the condition of a high release size in which the Y-linked drive gene is prevalent enough to drive the response allele into the population before it is lost (Chapter 4).

Drive strategies that rely on decreasing the number of wild-type mosquitoes in order to increase their own prevalence, such as Medea, Wolbachia and engineered underdominance, are unlikely to spread following an accidental release but are capable of
spreading following an intentional release. All three of these systems require either a fitness benefit, population growth or a large release size in order to spread following an accidental release (Chapter 4). For engineered underdominance, the required release size is particularly large – a population frequency greater than ~27% – however for Medea and Wolbachia, the required release size is significantly smaller.

Despite this, we should still be wary of the ability of Medea and Wolbachia to persist following an accidental release. All drive systems, with the possible exception of engineered underdominance constructs, are prone to spread under conditions of a fitness benefit, and whether or not they have such a benefit can only be properly assessed in a cage from which they may escape. This concern is confounded by the observation that, in agreement with Hypothesis 4.3, all gene drive systems are more likely to persist during conditions of population growth (Chapter 4), and such conditions exist for African mosquito populations for much of the year.

5.2 GENETICALLY MODIFIED MOSQUITOES AND THE CONCORDE

Results from this dissertation bring to mind problems encountered by another ambitious high-technology challenge – the British/French Concorde supersonic airplane. We explore parallels between the two endeavors; but first highlight an additional problem for any transgenic control strategy – the existence of multiple malaria vectors in Africa with multiply overlapping geographical distributions (Coetzee et al. 2000).
5.2.1 Multiple vectors of malaria

The likelihood of dissociation occurring during a TE-mediated control strategy implies that Africa-wide malaria prevalence will be at best halved by transgenic TE constructs. If we assume that another gene drive system is capable of driving a refractory gene to fixation in a single population, then this strategy will still be confronted by the economic infeasibility of modifying at least five different vectors of human malaria in order to achieve some semblance of control.

To illustrate this point, let us again consider Africa as a whole. For African countries in which the EIR has been estimated by species, the total EIR is on average 144 infective bites per person per year (Smith et al. 2005; Coetzee et al. 2000). About one third of these bites are from *Anopheles arabiensis*, another third is contributed by *An. gambiae s.s.*, a tenth is contributed by *Anopheles funestus*, and a variety of other species contribute the remaining bites.

Interpreting these fractions using Equations 1.1 and 1.2, we see that even in the event that we are able to completely modify up to six species of malaria vectors in Africa, malaria will continue to pose a significant public health burden on the continent. In the absence of any modification, the average child prevalence of malaria throughout Africa is ~74%. Modifying the entire populations of *An. arabiensis* and *An. gambiae s.s.* will reduce this to ~60%; while modifying an additional four species of malaria vectors will reduce malaria prevalence to ~40%. Modifying up to six species of malaria vectors will be
extremely expensive, and even this will only reduce the child malaria prevalence to ~40%, which is not even close to the level of disease control desired.

These results clearly show that the projected impact of GMMs on the malaria burden in Africa is modest at best, and that GMMs are unlikely to be successful as a sole or primary means of malaria control. It is possible that they may compliment other strategies of malaria control; however this must be weighed against their substantial cost. This leads us to consider parallels between the GMM effort and another bold and expensive adventure with high technology - the Concorde supersonic airplane.

5.2.2 The Concorde
Planning for commercial supersonic transport (SST) began in 1957. The goal of SST was to provide an aircraft that would carry passengers routinely and profitably between major cities (Owen 1997). Three groups carried forward plans for such airplanes - the USSR, the US, and a British-French consortium. The USSR was the first to complete such an airplane; however it crashed in an early demo and never really recovered. The American SST was budgeted carefully with the view that “If it doesn't make money, it doesn't fly.” Projected fuel costs implied that such a flight would be far more expensive than a regular flight, and environmental concerns, in addition to noise associated with supersonic flight, eventually let to its abandonment in 1999 before design even began.
Despite this, the British-French consortium did proceed with their plans and eventually built 20 Concorde SST planes, flying 13 of them commercially. As could have been expected, Concorde flight was particularly costly, with a seat costing approximately ten times the amount of a seat on a regular flight. Recognizing this, the British tried to get out of their contract, but they were unable to do so because the contract had been written in such a manner that neither side could pull out without the other and the French did not want to terminate. Eventually, following a fatal accident in 2000, all of the Concorde SST planes were grounded, and the project was abandoned entirely in 2003. Most of the planes are now in museums.

After much review, the general consensus has been that “the Concorde proved at once a technological triumph and an economic disaster” (Owen 1997). The project made many technical advances and demonstrated that international cooperation was possible on extremely demanding, technical projects. However the Concorde itself was very expensive and only a seventh of the costs were ultimately recovered (Arnold 2003).

There are many parallels between GMMs and the Concorde – both involve high levels of new technology, there are serious reasons for economic and environmental concern, there is a need for international cooperation, and there has been surprisingly little analysis of the anticipated costs and benefits. A consolation of the Concorde was that its technology and infrastructure eventually led to the design of the Airbus, which is one of the most profitable commercial airliners in flight today. However, if spin-off technology is to be a
motivation for the GMM project, then we should seriously consider the outcome of funds taken from the already small malaria budget. As Laurie Garrett argued in a recent issue of *Foreign Affairs* regarding the explosion of interest in global health, “Billions of dollars ought to buy better” (Farmer and Garrett 2007).

We are not claiming that the GMM effort will necessarily become the economic or environmental disaster that the Concorde became; but rather that the parallels between the GMM effort and the Concorde are too stark to be ignored. Consequently, we call for a hard analysis of these parallels. We call for a wide engagement of public discussion, and for a comprehensive administrative infrastructure to be put in place. Finally, we call for a continuing analysis of the potential benefits and side-effects of the use of GMM technology. Without this, we are concerned that the GMM effort could produce “a technological triumph and a public health disaster.”

5.3 **Future directions**

5.3.1 **Barriers to the establishment of refractoriness for other gene drive systems**

Dissociation is a particular concern for TE-mediated disease control strategies; however there are several drive strategies for which is it is irrelevant and for which there are other more important concerns. Chen *et al.* (2007) have proposed a strategy for reducing the rate of dissociation between the refractory and drive genes of a *Medea* element. This essentially involves locating the *Medea* toxin and refractory genes within an intron of the
antidote. Dissociation is likely to be a concern for the refractory gene HEG strategy; however the genetic load HEG strategy is immune to such concerns. A concern for the genetic load HEG strategy is that mutant HEGs that are rendered less deleterious to their host will have a selective advantage over functional deleterious HEGs. This may undermine the population suppression goals of the genetic load strategy. Engineered underdominance constructs are also relatively immune to the effects of dissociation; however their low degree of invasiveness means that they are unable to spread between many partially-isolated subpopulations of disease vectors (Davis et al. 2001).

In order to study the efficacy of other drive systems at spreading refractory genes into mosquito populations, it is important that we identify the potential barriers to their success. Mathematical modeling can then be used to assess the severity of these concerns.

5.3.2 The effect of local population structure on the early spread of Wolbachia infections and Medea alleles

Wolbachia and Medea provide the best compromise between invasiveness following an intentional release and containment following an accidental release. Models assuming randomly mating populations predict that both systems require either a fitness benefit or population growth in order to spread following an accidental release (Chapter 4). Despite this, observations of Medea alleles and Wolbachia bacteria in nature (Wade and Beeman 1994; Stevens and Wade 1990) suggest that both have been able to spread from very low prevalences, possibly even in the presence of a fitness cost.
Given the necessity of effective containment during outdoor cage trials, it is important that we have a full understanding of the persistence probability of *Medea* alleles and *Wolbachia* bacteria following an accidental release. Steve Sinkins and Jason Rasgon (personal communication) have suggested that local population structure may assist in the emergence of a new *Wolbachia* infection. The same may apply to a new *Medea* allele. Alternative modeling frameworks should be used to assess these hypotheses, and to check whether predictions made under the assumption of a randomly mating population are overly restrictive.

### 5.3.3 Modeling mosquito dispersal and the spatial spread of gene drive systems

The fundamental requirement of any gene drive strategy is that the transgenic construct disperses over a large geographic area on a time scale acceptable to public health goals (Braig and Yan 2001). To date, there has been little detailed study of the dispersal of mosquito vectors and how this relates to the distribution of huts, villages and egg-laying sites (Kiszewski and Spielman 1998; Taylor and Manoukis 2003). As recognized by Kiszewski and Spielman (1998), a “realistic analysis of the conditions for an antimalarial transgenic release requires a model that accounts for the effects of the discontinuous distribution of hosts in space and time.”

Reaction-diffusion models (Murray 2004) are particularly well-suited to studying the geographic spread of a TE or other drive system through a heterogeneous environment.
Realistic ranges on the parameters determining the distribution of mosquito travel distances can be obtained from mark release recapture experiments conducted for *An. gambiae* in and around the village of Banambani, Mali (Touré *et al.* 1998). The time taken for TEs and other gene drive systems to spread over large geographical areas can then be calculated, and it may be determined how this is affected by the existence of restricted gene flow between chromosomal forms and neighboring villages.

There are a great many problems to be addressed regarding the use of GMMs to control malaria and other vector-borne diseases. In this dissertation, we have addressed some of the problems that we considered to be important and tractable at the time of writing. As research into GMMs advances, there will undoubtedly be other problems that will become relevant and tractable. We encourage a continuing analysis of the potential problems and rewards of GMM technology, and an ongoing assessment of the ability of GMMs to control one of the developing world’s most devastating public health problems.
REFERENCES


Arnold, J., 2003 Why economists don’t fly Concorde. BBC News Online.


Deredec, A., A. Burt, and C. Godfray, 2007 Homing endonuclease genes: New tools for population engineering and control. Selfish DNA Meeting, NESCent, Durham, North Carolina, USA.


Farmer, P., and L. Garrett, 2007 From “marvelous momentum” to health care for all: Success is possible with the right programs. Foreign Affairs 86 (2).

Favia, G., A. Della Torre, M. Bagayoko, A. Lanfrancotti, N. Sagnon et al., 1997 Molecular identifications of sympatric chromosomal forms of An. gambiae and further evidence of their reproductive isolation. Insect Mol. Biol. 10: 19-23.


Manoukis, N. C., 2006 Studies on the ecology and adaptation of Anopholes gambiae in Mali and their impacts on malaria transmission and control (PhD thesis). University of California, Los Angeles.


Pool, J. E., V. Bauer DuMont, J. L. Mueller, C. F. Aquadro, 2006 A scan of molecular variation leads to narrow localization of a selective sweep affecting both


Shililu, J. I., W. A. Maier, H. M. Seitz, and A. S. Orago, 1998 Seasonal density, sporozoite rates and entomological inoculation rates of *Anopheles gambiae* and *Anopheles funestus* in a high-altitude sugarcane growing zone in western Kenya. Tropical Medicine and International Health **3**: 706-710.


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


Weinreich, M. D., A. Gasch, W. S. Reznikoff, 1994 Evidence that the *cis* preference of the *Tn5* transposase is caused by nonproductive multimerization. Genet. Dev. **8**: 2363-2374.


Wright, S., 1931 Evolution in Mendelian populations. Genetics **16**: 97-159.


