CRISPR-based innovative genetic tools for control of Anopheles gambiae mosquitoes

Citation

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029729

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current/terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
CRISPR-based innovative genetic tools for control of *Anopheles gambiae* mosquitoes

A dissertation presented

by

Andrea L. Smidler

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

April 2019
CRISPR-based innovative genetic tools for control of Anopheles gambiae mosquitoes

ABSTRACT

Malaria and other mosquito-borne diseases pose an immense burden on mankind. Since the turn of the century, control campaigns have relied on the use of insecticide-impregnated bed nets and indoor residual sprays to stop Anopheles mosquitoes from transmitting the malaria parasite. Although these are our best strategies to control the spread of disease, wild mosquito populations are developing resistance to insecticides at an alarming rate, making disease control increasingly challenging. In the search for new powerful strategies aimed at controlling malaria-transmitting Anopheles populations, we can now exploit a suite of powerful genome engineering tools to control wild populations and monitor releases. In this dissertation we utilize CRISPR/Cas9 technology and other genetic engineering tools in Anopheles gambiae to generate genetically sterile males for population suppression, to assess the feasibility of developing evolutionarily stable gene drives for population replacement, and to expand the genetic toolkit for field releases of genetically modified mosquitoes.

We develop CRISPR/Cas in A. gambiae for male genetic sterilization and for basic biological study. Using this system we generate a line of mutant mosquitoes with deletions in Zero Population Growth (ZPG), a gene critical for germ cell development. Resulting male mutants show no sperm in the testes and sterilize the females with which they mate, demonstrating that similar systems could be adapted for use in Sterile-Insect Technique (SIT)-like release campaigns.

We also test whether CRISPR/Cas9 can facilitate the sustainable and stable spread of gene drives in natural mosquito populations. Specifically, we design gene drives that have the potential of being evolutionarily stable by insertion into haplolethal ribosomal genes. To facilitate this goal, we create gene drive docking lines via a novel knockin technology for insertion of complex DNA templates into genetically
intractable loci. We identify multiple challenges associated with such systems, including the occurrence of Minute-like mutant phenotypes that present severe fitness costs when targeting haploinsufficient genes, a general decay in gRNA function over time that has consequences for all gene drives designed to date, and the critical need for precisely controlling Cas9 expression to avoid large fitness costs.

Finally, we develop and validate a novel transgenic tool for monitoring GM field releases. We generate transgenic lines expressing a fusion of a fluorescent marker with a male seminal protein specifically in male reproductive tissues. Incorporation of this fluorescence fusion into the mating plug, which is transferred to the females during mating, allows the visual identification of successful mating events for a few hours after copulation. This transgenic tool enables effective monitoring of GM male mating competitiveness in field trials, overcoming current limitations.

The work outlined here significantly expands the genetic toolkit for the manipulation of Anopheles mosquitoes, facilitating the implementation of genetic control strategies aimed at malaria-transmitting vector populations.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES AND TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>GENOTYPE NOMENCLATURE</td>
<td>x</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>xi</td>
</tr>
</tbody>
</table>

## CHAPTER 1: Introduction

1.1 The burden of malaria and its current control ............................................. 2
1.2 A brief history of malaria and its eradication from Italy and the United States – a precedent for mosquito elimination to control malaria ............................................. 4
1.3 Mosquito genetic control – An introduction ............................................. 6
1.4 Population suppression systems ...................................................................... 8
   1.4.1 Sterile Insect Technique for mosquito population suppression ...................... 8
   1.4.2 RIDL-based transgenic population suppression systems ..................................... 9
1.5 Population replacement systems ................................................................... 10
   1.5.1 Endonuclease gene drives – Mode of action and history of development .............. 11
   1.5.2 TALEN and ZFN-mediated gene drives .............................................................. 12
   1.5.3 CRISPR-mediated gene drives ....................................................................... 13
   1.5.4 Safety considerations concerning gene drive ............................................. 14
   1.5.5 Anti-parasite cargoes for population replacement .................................... 15
1.6 The problem of drive-resistant alleles (DRAs) ............................................. 16
1.7 Learning from other CRISPR gene drives ..................................................... 18
1.8 Fitness considerations for gene drives ...................................................... 20
   1.8.1 CRISPR off-target mutagenesis .................................................................... 21
   1.8.2 Gene drive-imposed evolutionary selective pressures .................................. 22
1.9 Designing evolutionarily stable gene drives for population replacement ........... 25
1.10 Mosquito genetic control strategies need more tools .................................. 26
1.11 Summary of Aims ......................................................................................... 28
1.12 References ................................................................................................. 30

## CHAPTER 2: CRISPR-mediated germline knockouts for genetic sterilization of male Anopheles gambiae ............................................. 46

ABSTRACT ............................................................................................................. 47
INTRODUCTION ................................................................................................. 48
RESULTS ............................................................................................................ 49
   2.1 Generation of transgenic CRISPR-expression lines ......................................... 50
   2.2 ZPG mutants fail to develop normal testes .................................................. 51
2.3 Male ZPG mutants are highly sterile ................................................................................. 52
2.4 Characterization of fertile ZPG mutant males ................................................................. 53
2.5 Female ZPG mutants display severe ovarian atrophy ...................................................... 54
DISCUSSION ................................................................................................................................. 56
FUTURE DIRECTIONS ................................................................................................................. 58
METHODS ..................................................................................................................................... 61
REFERENCES ................................................................................................................................. 65

CHAPTER 3: Advancement towards developing evolutionarily stable gene drives
for Anopheles gambiae population replacement .................................................................................. 69

ABSTRACT .................................................................................................................................... 70

INTRODUCTION .............................................................................................................................. 71

RESULTS ........................................................................................................................................ 72

3.1 Gene drive design .................................................................................................................. 72
3.2 Gene drive docking site locus identification ......................................................................... 72
3.3 Designing two docking sites in Rpl11-Rpt1 –
benefits and drawbacks of each design ......................................................................................... 74
  3.3 A. Designing dRPLT – a gene drive docking site with Rpt1 knocked-out ............................ 76
  3.3 B. Designing dRPLTu – a gene drive docking site with Rpt1 function maintained....... 77
3.4 Knockin of dRPLT docking line into Rpl11-Rpt1 by HDR ................................................... 77
3.5 Developing transgenic Interlocus Gene Conversion (IGC)
for knockin into genetically intractable loci .................................................................................... 77
  3.5 A. Generating dRPLT by CrIGCkid ...................................................................................... 79
  3.5 B. Generating dRPLTu by CrIGCkid .................................................................................... 80
  3.5 C. CrIGCkid does not occur in females, and infertility phenotypes lessen over time. 84
3.6 Characterization of dRPLT and dRPLTu docking lines .................................................... 84
  3.6 A. dRPLT is homozygous lethal, experiences proteasome dysfunction,
  and is likely a Rpt1 knockout ........................................................................................................ 84
  3.6 B. dRPLTu homozygotes are viable but display Minute-like fertility phenotypes .... 85
3.7 Gene drives in dRPLT and dRPLTu may be lethal .............................................................. 87
  3.7 A. Inserting B2GD into dRPLT or dRPLTu results in death
  of all transgenic individuals ........................................................................................................ 87
  3.7 B. Lethality of gene drives in docking sites may be due to
  aberrant Cas9 overexpression ...................................................................................................... 89
3.8 Developing drive-enabling germline-Cas9 promoters
independent of docking sites ........................................................................................................ 92
  3.8 A. Vasa2 transgenic promoter is too leaky for drive .......................................................... 92
  3.8 B. A Vasa protein fusion to fine-tune Cas9 expression ....................................................... 94
  3.8 C. Two versions of the ZPG promoter for gene drive Cas9 expression ........................... 95
DISCUSSION ......................................................................................................................... 98
METHODS .............................................................................................................................. 106
REFERENCES .......................................................................................................................... 122

CHAPTER 4: A transgenic tool to assess Anopheles mating competitiveness in the field .............. 127

ABSTRACT ................................................................................................................................. 128
INTRODUCTION .......................................................................................................................... 129
RESULTS ..................................................................................................................................... 131
4.1 Generation and characterization of the plugin-tdTomato transgenic line................................. 131
4.2 Plugin-tdTomato localizes to vesicles within the anterior compartment of the MAGs .................. 134
4.3 Plugin-tdTomato is sexually transferred to the female atrium in the mating plug ..................... 136
4.4 Testing PluTo males for in vivo assessment of successful mating ............................................ 136
4.5 Expression of the plugin-tdTomato fusion protein does not affect the mating competitiveness of PluTo males and does not impair female reproductive fitness .............................................................. 137

DISCUSSION ............................................................................................................................. 139

CONCLUSIONS .......................................................................................................................... 141

METHODS .................................................................................................................................. 141
REFERENCES ............................................................................................................................. 148

CHAPTER 5: Discussion .............................................................................................................. 152

5.1 Overview ............................................................................................................................. 153
5.2 CRISPR-mediated genetically sterile males for Sterile Insect Technique ................................. 153
5.3 Developing a transgenic tool to assess Anopheles mating competitiveness in the field ............ 156
5.4 Developing evolutionarily stable gene drives for Anopheles gambiae population replacement ...... 157
5.5 Concluding remarks ............................................................................................................. 161
5.6 REFERENCES ....................................................................................................................... 162

APPENDIX A: Plasmid and transgene maps ................................................................................. 166


APPENDIX C: Full Text: Engineering the control of mosquito-borne infectious diseases .................. 178

APPENDIX D: Full Text: Concerning RNA-guided gene drives for the alteration of wild populations... 188
LIST OF FIGURES AND TABLES

CHAPTER 1.

Figure 1.1 | Dynamics of population suppression and population replacement vector control strategies .................................................................................................................. 7
Figure 1.2 | A basic diagram of ZFN, TALEN, and CRISPR/Cas9 DNA cleavage................................................. 14
Figure 1.3 | Biased inheritance of Gene drives through mosquito populations and the genetic outcomes of cleavage and homing................................................................. 17

CHAPTER 2.

Figure 2.1 | Transgenic map and fluorescence properties of VasCas9 and gZPG transgenic lines........................................................................................................ 50
Figure 2.2 | The spermless male testicular phenotype........................................................................................................ 51
Figure 2.3 | A composite image of mutations observed in {+/gZPG;+/VasCas9} male testes.............. 52
Figure 2.4 | Genetically spermless males cause significant infertility in females............................................. 53
Figure 2.5 | ZPG mutant females have significantly ablated ovarian development and are defective in 20E synthesis ......................................................................................... 55
Figure 2.6 | Schematic for the Inducible Eunuch system.............................................................................................. 60

CHAPTER 3.

Figure 3.1 | An evolutionarily stable CRISPR gene drive design for population replacement .... 73
Figure 3.2 | Constructing gene drives into Rpl11-Rpt1.......................................................................................... 75
Table 3.1 | Summary and comparison of characteristics of the dRPLT and dRPLTu gene drive docking sites................................................................................................................. 76
Figure 3.3 | Putative mechanism for IGC to generate gene drive docking lines .................................................... 79
Table 3.2 | Summary of IGCdRPLTu<sup>pB</sup> insertion sites......................................................................................... 80
Figure 3.4 | PCR confirmation of dRPLTu knockin into Rpl11-Rpt1 following IGC.............................................. 81
Figure 3.6 | Characterization of CrIGCkid to create dRPLTu......................................................................................... 83
Figure 3.7 | Fitness characterization of dRPLT.............................................................................................................. 85
Figure 3.8 | dRPLT and dRPLTu individuals display many hallmark Minute phenotypes......................................... 86
Figure 3.9 | Gene drive and distal gene drive transgene maps developed in this work............................................ 88
Table 3.3 | Number embryos injected, transgenics recovered and survived to adulthood................................. 88
Figure 3.10 | Western blot analysis for Cas9 expression in drive lines........................................................................ 89
Figure 3.11 | No biased inheritance of mNosGD<sup>dRPLT</sup>.......................................................................................... 90
Figure 3.12 | B2dGD<sup>pB</sup> does not facilitating drive of dRPLTu in males...................................................................... 91
Figure 3.13 | Rescuing mNosGD<sup>dRPLT</sup> Cas9 function with Vasa-Cas9 results in death or infertility of all females and infertility in males.................................................................................. 94
Figure 3.14 | A Vasa fusion distal drive, VFdGD^ab is fertile but fails to facilitate drive of dRPLTu. 95
Figure 3.15 | A ZPG2 distal drive, ZPG2dGD^ab fails to facilitate drive of dRPLTu................................. 96
Figure 3.16 | A ZPG1 distal drive, ZPG1dGD^ab causes complete female infertility, and incomplete male infertility which decreases in subsequent egg-lays ................. 97
Figure 3.17 | Summary of Cas9 Goldilocks Conundrum................................................................. 105

CHAPTER 4

Figure 4.1 | PluTo transgene design and male phenotype................................................................. 132
Figure 4.2 | Endogenous Plugin and PluTo transgene expression....................................................... 133
Figure 4.3 | Endogenous Plugin localization in the MAGs ................................................................. 134
Figure 4.4 | Plugin-tdTomato localization in the MAGs and mating plug.............................................. 135
Figure 4.5 | PluTo transgenic males allow in vivo assessment of male mating success within the female.................................................................................................................. 137
Figure 4.6 | Reproductive fitness of females mated to PluTo males....................................................... 138
GENOTYPE NOMENCLATURE

To clarify the transgenes and genotypes outlined herein, here are presented the nomenclature rules used in this work. They are inspired by *Drosophila* nomenclature standards, but have been modified for the specific needs of the work undertaken here while attempting to maximize clarity and ease of readability.

**Italics are reserved for gene and locus names**, with capitalized letters denoting wild type loci, and lowercase letters denoting a disrupted or knockout gene product, occasionally specified with Δ. Sterile males are termed ‘ZPG mutants’, because as mosaics, they cannot be considered true knockouts.

**Transgene names vary**, but attempt to convey important information about the transgene’s function, purpose, and properties. The first time the transgene name is discussed, the abbreviation for its name is listed. All relevant transgene maps can be found in **APPENDIX A**, and for sequence specifics, the email (asmidler@gmail.com) can be reached for inquiries *ad infinitum*.

**Superscripts attempt to convey locus information for transgenes.** For example if a drive is inserted into the dRPLT docking line, it will be conveyed as a $d_{RPLT}$ superscript adjacent to the transgene name. Therefore the mNosGD line inserted in dRPLT is denoted mNosGD$^{dRPLT}$. Similarly if a transgene is piggyBac based, and the insertion(s) has been identified and purified as an iso-female family, it will be denoted simply as $pB^\#$. If an experiment was performed with a mixed population with a single clonal transgene inserted within many varied and unspecified sites in the genome in a mixed population, it will be denoted $pB^{mix}$.

**Mixed genotypes and phenotypes follow similar conventions to those used in Drosophila.** Genotypes are fully contained within brackets {}, and the genotype of each locus is denoted over a slash ‘/’. A wild type allele is conveyed with the standard ‘+’ symbol, and the genotypes of separate loci are separated by a semicolon ‘;’. Therefore the genotype of an individual heterozygous for each of the dRPLT and VasCas9 transgenes will be denoted {+/dRPLT ;+/VasCas9}. Similarly an individual transheterozygous for the mNosGD inserted into dRPLT, and unintegrated dRPLT will be denoted {mNosGD$^{dRPLT}$/dRPLT}. 

x
ACKNOWLEDGEMENTS

Thank you:

To all my scientific mentors through the years. To my childhood teachers who first instilled in me a reverence for genetics, to my genetics summer camp counselors, and to my undergraduate advisors, Dr. Jeff Dangl and Dr. Nuria S. Coll who believed in me enough to let me loose in a lab. The seemingly small doors you opened have become monumental to me. To Eric Marois, who first inspired my love of mosquito genetic engineering - I can’t imagine a life in which we did not meet.

To Flaminia. Who saw the potential behind that uncompromising awkwardness. To your unending patience and support despite our disagreements and my bullheadedness. For pushing me to be better, but giving me the space to follow my heart, and for letting me learn the hard way. For the opportunities you’ve provided, and for the lessons you’ve taught. I thank you from the bottom of my heart, I would never be the scientist I am today without you.

To Kevin Esvelt. For our inspiring conversations, and belief in the impossible. For showing me the immense motivation required to change the world, and for bringing me along as you do so. I am forever grateful for your guidance, inspiration, and generosity.

To George Church. For showing me how to think like an innovator, and how to embrace the improbable. For your phenomenal mentorship, and the examples of excellence you provide. For your truly generous and unwavering support, and for teaching me that I belong at the table.

To other members of my Dissertation Advisory Committee: Dr. Dyann Wirth, Dr. Matthias Marti, and Dr. Dan Neafsey. Thank you for keeping me focused on the important questions, and for providing your incredible feedback and guidance over the years. My project would not be where it is today without your
wisdom. To my Dissertation Defense Committee: Dr. Manoj Duraisingh, Dr. Jay Mitchell, Dr. Conor McMeniman, and Dr. Jeff Dvorin, thank you for your time and expertise as I defend my PhD.

Thank you to the funding that made this work possible, the Ruth L. Kirschtein NIH Fellowship.

Thank you to my fellow lab mates. You all have been amazing, and I’ve learned so much from you. Thank you for the support and comradery through the years. You each hold a special place in my heart, and it is impossible to properly thank you all. Especially to Rob, Emily, Manu, and Eryney for all your guidance, support and help through the years.

To my friends. To Mike and RJ, for showing me what academic dedication looked like. For teaching me how to write from the ground up, and for helping me discover my calling. To Gabe for being the best listener alive; for being inspiring, encouraging and pragmatic somehow all simultaneously, and to Erica, Vanessa and Eli, for being my Bean dip therapist, and a shoulder to cry on, and for keeping my teeth sharp.

To my family. Mom and Dad, I would never be the woman I am today without you. For teaching me how to be strong, smart, and independent, and for tolerating my crazy childhood years. For the unwavering love and laughs, and for always being there through it all.

To Dan, my first nerd partner in crime. From competing at doing Punnett Squares at the breakfast table, to constructing suits of chain-maille. You have kept me grounded, humbled, and inspired.

To Alex, my love. For being my rock [troll]. When the world around me is full of variables, you are my constant. For teaching me how to love life, and how to live it well. For showing me the joy that serenity can bring, and for teaching me the limits of my eye-rolling capabilities. For loving me as a garbage person, as well as a knockout, and for being the best father our dogter could ask for. You have no idea how much I look forward to our lives together.
Respective Contributions: This chapter was written by Andrea Smidler, with editorial input from Flaminia Catteruccia. Portions of this chapter include work from published Research and Review Articles co-written by Andrea Smidler in collaboration with colleagues. This chapter includes some images, phrases, and ideas from these articles with permission for use under Creative Commons Attribution 4.0. These articles are included in full in the Appendix for reference.

APPENDIX C


APPENDIX D

1.1 The burden of malaria and its current control

Vector-borne diseases have plagued humanity since the dawn of time, possibly killing half of all humans who have ever lived. The burden these diseases impose on mankind has left scars across our genomes as a testament to their prevalence, persistence, and pandemicity, and despite humanity’s best efforts, we have yet to eradicate even one. Some of the most destructive pathogens include the Dengue, Yellow Fever, Zika, and Chikungunya arboviruses, but the malaria-causing Plasmodium protozoan parasites are the most deadly. Spread by the bite of an anopheline mosquito, these parasites cause the death of approximately half a million people annually with another 200 million incapacitated for days—grinding lives and productivity to a halt, and affecting some countries’ GDP by as much as 6%.

Since the beginning of this century, funding for malaria research and control has increased almost 10-fold. Despite recent strides towards developing effective prophylactics and vaccines, the World Health Organization (WHO) predicts that “mosquito control is the only intervention that can reduce transmission from very high levels to close to zero.” Increased application of vector-control measures has contributed to a staggering reduction in mortality rates, with the use of long lasting insecticide-treated nets (LLINs) and indoor residual spraying (IRS) contributing to over 75% cases averted since the turn of the century.

However, these once-reliable control methods are becoming increasingly ineffective due to insecticide resistance mechanisms emerging in mosquito populations.

LLINs are fine-mesh nets hung above beds which protect people from mosquito bites when they are asleep, and when impregnated with insecticides they provide further protection to the community. This intervention targets vector populations which blood-feed indoors and at night, generally targeting a female once per gonotrophic (egg development) cycle, but neglect mosquitoes using different feeding strategies. Following a ramp-up of LLINs distribution, it is estimated that nearly half of the malaria at-risk population are now protected by bednets, but their widespread use has induced resistance to
pyrethroids in mosquitoes\textsuperscript{17-21}, to the point that has become difficult to find populations that are fully susceptible\textsuperscript{22}. Some reports even suggest that mosquitoes may be evolving novel feeding behaviors to circumvent these interventions\textsuperscript{23-25}, suggesting we may be reaching the limit of LLIN effectiveness.

Similar considerations are applicable to the use of IRS, which relies on insecticide application to interior house walls where female mosquitoes usually rest after taking a blood meal. Because this intervention targets the interiors of human dwellings, disease-transmitting populations which feed and rest outdoors are not affected by this strategy. Due to the reduced contact risks for humans compared to insecticides on LLINs, a number of insecticide classes are approved for IRS use, including organophosphates and carbamates. Even the highly toxic and persistent DDT can be utilized for malaria control purposes in some cases of extreme resistance to all other insecticides\textsuperscript{26}. Currently only an estimated 7\% of the African at-risk population lives in homes treated with IRS\textsuperscript{16}, however these numbers are increasing following a shift towards using non-pyrethroids\textsuperscript{17,21}, in response to increasing resistance amongst wild populations.

Indeed, the majority of anophelines are now resistant to pyrethroids\textsuperscript{27}, and resistance rates to carbamates and organophosphates are on the rise\textsuperscript{28}. Despite best practices recommending annual rotation of IRS insecticides to limit the emergence of resistance\textsuperscript{21}, alarmingly mosquito populations have emerged with resistance to all four classes of insecticides available for malaria control\textsuperscript{29,30} – with some resistance mechanisms causing resistance to multiple classes simultaneously\textsuperscript{29,31}. Since 2010, 68 countries have reported resistance to at least one class of insecticide with 57 of those countries reporting resistance to 2 or more\textsuperscript{14}. Although this suggests that the lethal effect of even newly-developed insecticides may be thwarted at an unexpectedly fast time-scale, both LLINs and IRS remain the core intervention strategies used to control malaria worldwide\textsuperscript{32} making the development of novel vector control technologies – especially those capable of targeting outdoor biting and resting populations – increasingly urgent.
1.2. A brief history of malaria and its eradication from Italy and United States - A precedent for mosquito elimination to control malaria

While deliberately eradicating an entire species seems an ecological nightmare, it is not without precedent. Malaria has so significantly changed the course of human history that it can be argued to have shaped the fate of empires, explaining the urgency with which many countries have sought to eliminate this disease through elimination of its mosquito vector. We will briefly examine malaria’s impact on human history, and the control efforts in two countries where the disease was successfully eliminated, in an attempt to glean insights on how similar campaigns may look like in the future, what the ecological impacts might be, and which motivations may drive them.

Malaria was endemic to Italy throughout most of recorded history, with first-century Roman scholars understanding enough about its epidemiology to encourage building of homes on high land away from swamps. In fact, “having forsaken its best farmland to malaria, ancient Rome could not feed itself, but [following expansion of the empire] the riches of conquest paid for grain, olives, fish sauce, and oil imported from North Africa. It paid for elaborate aqueducts, allowing wealthy Romans to move away from the most mosquito-ridden water’s edge.” Malaria posed such an immense burden that Romans would pray to the demon goddess of fever, Febris, in three dedicated temples around the city to attempt to beg for respite. Therefore it is no surprise that when the technology arose, the Italian people sought to eradicate the malaria mosquito at their first opportunity. Cases were treated with quinine, larvae were killed by application of petroleum derivatives and “Paris Green” (copper acetoarsenite), and breeding sites removed through land reclamation. These efforts resulted in significant disease reduction, with some regions falling to almost zero by 1939. However, only once DDT became available after the conclusion of WWII was eradication complete with virtually all cases eliminated by 1948.
At approximately the same time the United States began a similar eradication campaign\(^{38}\). Malaria had raged endemically in the United States since before it was a nation, with some epidemics occurring as far north as Sheffield Massachussets\(^{39}\). However most scientists agree that North America was free of malaria before colonization\(^{34}\), with one of – if not – the first cases being imported by a Jamestown settler infected with \textit{Plasmodium vivax}\(^{40}\). Later, the more deadly \textit{Plasmodium falciparum} was introduced during the slave trade\(^{41}\) and was spread endemically by native \textit{Anopheles quadrimaculatus}, \textit{Anopheles crucians}, and \textit{Anopheles pseudopunctipennis}\(^{42}\). It was such a significant part of American life that it may have even contributed to the formation of the nation, with the environmental historian J.R. McNeill once writing; “Those tiny amazons [mosquitoes] conducted covert biological warfare against the British army”\(^{43}\). When the British Army attempted to invade South Carolina in 1780 they were ultimately thwarted by malaria, succumbing to infection due to their lack of exposure to the parasite through childhood\(^{44}\). The disease burden in the region was so great that it led Johan David Shoepf to write in 1788; “Carolina was in the spring a paradise, in the summer a hell, and in the autumn a hospital”\(^{45}\).

The battle with malaria raged through the Civil War, and “mosquito soldiers”\(^{46}\) may have even prolonged it. Despite the Confederate army being greatly outnumbered, the malaria resistance its soldiers had acquired from repeated exposure gave them an advantage against the Union when fighting in Southern territories, potentially extending the war by years\(^{46}\). Therefore almost 80 years later when WWII ignited, the impact of malaria on any war effort was not lost on the American government\(^{47}\), so in 1942 the predecessor of the CDC, The Office of Malaria Control in War Areas, was founded\(^{47}\). Its primary objective was to direct malaria eradication efforts, which officially commenced in 1947\(^{47}\) but had technically begun years prior, seeking to reduce mosquito numbers by mass spraying of DTT as well as oiling swamps to drown aquatic-stage larvae\(^{47}\). DTT was highly effective against mosquitoes and helped suppress their populations to near extinction, but its peripheral ecological effects later became known following the publication of the book \textit{Silent Spring}\(^{48}\). Besides causing a significant ecological impact, DDT did not achieve
complete elimination of the target mosquito populations, which were able to rebound at the end of the intervention thereby restoring original local ecologies.

Mosquitoes are one of many small insect prey species, and are not known to be the sole pollinators of any plants, perhaps with the exception of one orchid species in North America\textsuperscript{49,50}. It is generally agreed by the scientific community that if they were eliminated, their ecological niche could be easily filled by other insects\textsuperscript{51}. However only recently have studies begun to better understand the mosquito’s role in its ecological habitat\textsuperscript{52}. While no resounding ecological consequences were observed in Italy or the United States after their near-elimination, it would be foolish to presume that all ecologies are identical. Therefore suppression of mosquito vectors – even if only temporary – should be carried out only with the outmost care and with constant ecological monitoring\textsuperscript{53}.

**1.3 Mosquito genetic control – An introduction**

Genetic control of vector-borne diseases through manipulation of the vector has long been considered a panacea for disease control, as genetic technologies have consistently proven effective in humanity’s fight against insects for the better part of the past century. Briefly, first proposed in the 1930s by E.F. Knipling\textsuperscript{54}, Sterile Insect Technique, a technique which suppresses wild populations following release of Sterile males (SIT, discussed further in Section 1.4A) enabled the eradication of a major agricultural pest, the screwworm \textit{Cochliomyia hominivorax}, from all of North and Central America\textsuperscript{55,56}. Following the development of genetic sexing strains\textsuperscript{57} (discussed further in Section 1.10B), eradication of the Mediterranean fruit fly from the United States and Mexico was achieved soon after\textsuperscript{58}. More recently these advances were followed by the first releases of transgenic insects to suppress Dengue-spreading \textit{Aedes} populations in Grand Cayman\textsuperscript{59} and Brazil\textsuperscript{60} (discussed in Section 1.4B). Now, with the advent of CRISPR/Cas9 officially heralding the “dawn of the gene-editing age”\textsuperscript{61}, breakthroughs in developing
genetic control strategies against important disease vectors have occurred at a break-neck pace, providing new lymph to the prospect of eradicating the first vector-borne disease.

Technologies aimed at controlling vector-borne diseases using genetic modification of the insect vectors can be broadly grouped into two general categories; those that seek to eradicate the vector population (Figure 1.1A), and those that seek to replace them with populations incapable of spreading disease (Figure 1.1B). Termed suppression and replacement technologies respectively, they are not mutually exclusive and can be achieved using a variety of methodologies. These range from mass release of sterile males which sterilize the females with whom they mate (Figure 1.1A), to drive systems which – through biased inheritance – spread factors resulting in population crashes, or which can replace pathogen-permissive populations with pathogen-refractory ones (Figure 1.1B). Each type of design has its own benefits and drawbacks ecologically, biologically, and logistically (Reviewed in1,65) with one of the major distinctions being whether their modifications are designed to disappear following release (self-limiting) or are designed to persist on a long-term basis (self-sustaining).

Figure 1.1 | Dynamics of [A] population suppression and [B] population replacement vector control strategies. [A] Population suppression techniques generally require iterative releases of sterilizing mosquitoes (green arrows) over a number of generations (X-axis), resulting in decreased population size. [B] Population replacement strategies require fewer releases (green arrow) which results in the original wild type population going to extinction (grey), and being replaced with the engineered strain (green) ideally with the engineered drive strain going to allelic fixation (Drive frequency). Reproduced from 1 with permission from Creative Commons 4.0. Also available in original text in Appendix C.
1.4 Population suppression systems

1.4 A. Sterile Insect Technique for mosquito population suppression.

Some of the first genetic insect control campaigns for insect population suppression used a technology termed Sterile Insect Technique (SIT). “Classic” SIT is the use of chemicals or radiation to sterilize males, which when released en masse, compete with wild males for females mates, and effectively sterilize the females following copulation. This system is most efficient in monandrous species in which females mate once per lifetime, and in which failure to transfer viable sperm to the female renders her permanently infertile. Effectively acting as a species-specific chemical-free insecticide by their ability to suppress specific populations, such vector control programs require massive numbers of males be released iteratively until population elimination is achieved. Importantly for mosquitoes in particular, male-exclusive releases are compulsory as females bite and spread disease.

SIT was first successfully implemented to eliminate a number of important agricultural pests, most notably the Screwworm and Medfly, however its successful implementation in mosquitoes has proven elusive. While chemo- or radiation-based sterilization of male Aedes albopictus, An. arabiensis, An. stephensi and An. gambiae can be achieved, it generally causes a reduction in male mating fitness. Because mosquito males must compete in swarms for rare mates, such fitness defects necessitate that more males, than initially predicted by simple models, need be released to achieve suppression making such systems impractical. However, a notable hurdle for genetic SIT systems exists; in order to produce the substantial numbers required, the desired sterility phenotypes must either be suppressed during mass rearing or induced for release. Towards this end, alternative technologies for mass male sterilization have been designed and implemented for GM mosquito control.
1.4 B. RIDL-based transgenic population suppression systems.

In the face of these challenges, the Classical SIT system was updated for more effective use in mosquitoes. In the absence of effective sterilization systems using classic methods, the company OXITEC developed a transgenic system in *Aedes aegypti*, which although distinct from SIT\(^74\), can achieve similar aims\(^75,76\). Termed Release of Insects carrying Dominant Lethals (RIDL™), a few permutations of this system were developed, all largely characterized by release of males carrying a tetracycline-controlled lethal transgene that kills their offspring. RIDL transgenic systems are suppressed by the addition of tetracycline during rearing, enabling mass quantities of mosquitoes to be produced, which are then activated following release resulting in offspring death. The absence of tetracycline in the wild larval environment permits aberrant accumulation of protein products from the transgene ultimately causing lethality\(^75,76\). Therefore in this system the males are fertile although their offspring are not viable, resulting in a population suppression effect similar to that of classic SIT\(^1\).

A similar system, fsRIDL™, causes lethality specifically in females rather than all individuals, by disrupting their ability to fly. By expressing the aforementioned lethal transgene specifically in the female’s flight muscles, the daughters of released males still contribute to larval suppression by competition, but never fly away to bite, mate, or spread disease\(^77\). First developed in *Ae. aegypti*, but eventually transferred to *Ae. albopictus*\(^78\) and *An. stephensi*\(^79\), this transgenic system has the added benefit that males are unaffected by the transgene – thus allowing them to mate and continue to sire offspring. This permits the population suppression effect to persist for a few generations following the initial release, making the initial investment more effective, until the fitness costs of the large transgenic cargo causes all transgenics to be outcompeted and die. In the end, both the RIDL and fsRIDL technologies rely on mass release of males whose offspring are ultimately inviable, and in which the transgene does not persist in the population long term, but does achieve transient population suppression.
With successful field trials suppressing *Ae. aegypti* populations by as much as 80-95%\(^{80}\), Oxitec recently announced its intent to develop similar technologies in the Central American malaria vector *An. albimanus* in conjunction with the Bill and Melinda Gates Foundation\(^{81}\). Despite having already developed the technology in the anopheline species *An. stephensi*\(^{79}\), they notably have never announced intention to develop their technology in any African malaria vectors – leaving a critical gap in the development of genetic sterilization technologies in these important species.

### 1.5 Population replacement systems

While eradication of mosquitoes may be alluring to some, it remains mostly unfeasible across many large geographic expanses of Africa, or where vector populations are particularly dense\(^1,82,83\). Since the creation of the first transgenic anopheline\(^84\), scientists have discovered a number of native genes or synthetic genetic cargoes that can render mosquitoes refractory to parasite transmission (reviewed in\(^{85,86}\), discussed further in Section 1.5E). But to affect transmission in the field, these traits would need to be spread into every wild mosquito. If we could develop a method for biasing inheritance of these genetic cargoes, they may be able to spread to most members of the population over many generations in spite of fitness costs. Expanding such technologies to other organisms beyond mosquitoes could have broader impacts, enabling us to address several other major world problems, including the rise of pesticide and herbicide resistance, and the agricultural and environmental damage wrought by invasive species\(^2\).

In nature, certain selfish genes ‘drive’ themselves through populations by increasing their inheritance probability\(^87\). Some well-characterized examples include endonuclease genes that copy themselves into homologous chromosomes\(^88\), segregation distorters that destroy competing chromosomes during meiotic germline development\(^89\), transposons that insert multiple copies of themselves throughout the genome\(^90\), and Medea elements that kill competing siblings who do not inherit them\(^91,92\). In this class we can also broadly place vertically-transmitted bacterial endosymbionts such as *Wolbachia*\(^93\) that bias their
inheritance by manipulating reproduction of their insect hosts. Such selfish genetic elements have been identified in species ranging from flies\textsuperscript{94,95}, yeast\textsuperscript{96}, and nematodes\textsuperscript{97}; to corn\textsuperscript{98}, beetles\textsuperscript{91}, and lemmings\textsuperscript{99} – with many hundreds likely lurking in species yet to be studied\textsuperscript{87}.

1.5 A. Endonuclease gene drives – Mode of action and history of development.

One such type of selfish gene spurred early scientific curiosity by the mode in which it spread, and has ultimately inspired a new field of genetic engineering. Naturally occurring homing endonuclease (HEG) gene drives are selfish genetic elements, which encode a large endonuclease within a specific locus, capable of cleaving the wild type homologous locus. When the cell repairs this break via the Homology Directed Repair (HDR) DNA-repair pathway\textsuperscript{88}, it is forced to use the HEG-encoding gene as the template for repair, inducing the cell to copy its sequence into the break. This copying process is termed ‘homing’, and it ultimately results in the heterozygous cell becoming homozygous. When this occurs in the germline of sexually reproducing species, the gene is inherited accordingly, resulting in super-Mendelian inheritance of the HEG in the progeny. Over many generations, this caused the HEG allele to increase in frequency, effectively ‘driving’ it through the population (Figure 1.1B, Figure 1.3A). Over time, the HEG drive – and any associated genetic cargo – spreads into all individuals in a population, enabling deliberate engineering of the entire wild populace\textsuperscript{83,87,100-102}. This basic copy-and-paste mode of action provides the basis for many modern synthetic gene drives, including the gene drives outlined in this dissertation.

In 2003, Austin Burt first suggested that naturally-occurring meganucleases, such as Onul, I-Ppol or I-SceI, could be modified to recognize new host sequences in the insect genome to enable manipulation of natural insect populations\textsuperscript{103}. Several early efforts focused on building such HEG-enabled gene drives in mosquitoes with the aim to block malaria transmission\textsuperscript{104-109}, with the first successful demonstration of the drive phenomenon in mosquitoes occurring in 2011\textsuperscript{107}. Concurrently, efforts were underway to redesign HEGs to target the ribosomal rDNA repeats on the \textit{Anopheles} X chromosome\textsuperscript{106}, whose
expression from the male Y-chromosome would create a Selfish-Y system capable of ‘shredding’ X-containing sperm during spermatogenesis, ensuring male-only progeny. This type of ‘genetic load’ drive would spread rapidly but the resulting all-male population would inevitably crash, the speed of which depending on the dynamics of gene flow. Among a variety of technical challenges, development was ultimately thwarted by difficulties in engineering the HEGs to target new host sequences, making implementation impractical if not impossible.

1.5 B. TALEN and ZFN-mediated gene drives.

At this time, newer designer nuclease arose enabling more straightforward targeting of genomic sequences by designer DNA-binding modules. These zinc finger (ZF), or transcription activator-like effector (TALE) DNA-binding proteins, are composed of multiple large DNA binding domains fused in sequence, with each domain characterized by slight amino acid differences enabling binding to distinct DNA bases. Assembling multiple DNA-binding domains into repetitive modules allows for binding up to 20 consecutive bases, and fusing a pair to the homodimerizing endonucleolytic domains of the type II endonuclease, Fok1, enables targeted DNA cleavage of unique genomic sites (Figure 1.2A,B). This technology was successfully used to generate the first knockout and knock-in mosquito mutants.

During the course of my Master’s Thesis, from 2011-2012, research I conducted at the University of Strasbourg, France, successfully developed the use of paired transgenic TALE endonucleases (TALENs) to make the first knockout anopheline (Full text in Appendix B). However, attempts to build a basic gene drive targeting a neutral genomic site using the same transgenic TALEN technology resulted in shuffling of the repetitive TALE DNA-binding domain sequences, rendering the transgene nonfunctional and precluding the possibility of drive (Smidler, Marois et al. unpublished). Similar TALEN- and ZFN-enabled gene drives published in Drosophila suffered from the same instability due to their repetitive nature.
Taken together, these studies revealed that highly repetitive gene drives would not likely be tolerated in biological systems, necessitating development of less repetitive endonucleases for gene drives.

1.5 C. CRISPR-mediated gene drives.

Following the discovery that it could be used as a powerful engineering tool\textsuperscript{128,129}, the Clustered Regularly Interspaced Short Palindromic Repeat/Cas9 system (CRISPR/Cas9) has completely revolutionized synthetic biology. Along with many other groups, we recognized its promise for use in gene drives\textsuperscript{2,53} following the first application in cell systems\textsuperscript{130-133}. Based on the Type II CRISPR acquired immune systems which enables bacteria to ‘remember’ viral infections and recognize and cleave the DNA of new invaders\textsuperscript{134,135}, CRISPR has been adopted to edit the genomes of a number of species (Reviewed in\textsuperscript{136}). The Cas9 endonuclease can be targeted to nearly any DNA sequence in the genome through the action of a synthetic small guide RNA (gRNA), with the only DNA sequence requirement being the bases ‘NGG’ (termed a PAM) at the 3’end. The gRNA is less than 100 bp in length, and is composed of the first 17-20 bp ‘spacer’ sequence which directs Cas9 to the ‘target’ sequence through direct Watson-Crick-Franklin basepairing, followed by a sequence which forms a stem-loop secondary structure recognized by Cas9\textsuperscript{128}. The Cas9 enzyme and small guide RNA work together as a two-part system to bind and cleave DNA, and are capable of targeting every 8 bp on average\textsuperscript{128} (Figure 1.2C). Moreover, because unlike TALENs and ZFNs targeting different sequences does not require any protein redesign – instead only requiring redesign of a small gRNA – this system is much more inherently flexible, enabling straightforward redesign and simultaneous targeting of multiple sequences with less sequence repetitiveness. Because CRISPR-mediated genome editing relies on the same copy-and-paste mechanism as prior endonuclease gene drives, we reasoned it could be capable of enabling powerful gene drives if properly engineered\textsuperscript{12,137}. 
1.5 D. Safety considerations concerning gene drive

Due to this ease of design the general flexibility of CRISPR to engineer innumerable species, CRISPR gene drives could pose a threat as a potential dual-use technology in some species. With this in mind we were motivated to discuss safety regulations for monitoring and development even prior to beginning laboratory experiments on the topic\textsuperscript{13}. However, in keeping with the molecular biological focus of this dissertation, this work will not be discussed further.
1.5 E. Anti-parasitic cargoes for population replacement.

The population-replacement gene drives outlined in Chapter 3 have the ultimate aim to spread anti-malarial genetic cargoes into wild populations. A number of potential mechanisms capable of rendering mosquitoes genetically refractory to malaria infection have been developed, including activation or augmentation of the mosquito host immunity\textsuperscript{138-146}, impairment of parasite development by disrupting factors necessary for parasite invasion of the midgut or salivary glands\textsuperscript{147,148}, and introduction of artificial constructs with parasite-killing properties\textsuperscript{149-151}.

Mosquitoes mount a multifaceted immune response to invading parasites and viruses (reviewed in \textsuperscript{138,139}), which can be artificially manipulated to reduce infection or transmission of the parasite. The NF-κB-like immune pathway in mosquitoes has proven to be a good target for upregulating the native immune response. As an example, overexpression of the NF-κB transcription factor \textit{Rel2} causes a significant decrease in parasite prevalence in engineered \textit{Ae. aegypti} and \textit{An. stephensi}\textsuperscript{140,141}, and upregulation of \textit{Rel2}'s downstream effector genes also causes a reduction in parasite load\textsuperscript{142,143}. While promising, no transgenic system relying solely on native immune activation has been successful at completely-ablating parasite development, necessitating the identification of additional anti-malarial cargoes for population replacement.

Parasite killing can also be achieved by transgenic expression of parasite-targeting single-chain antibodies borrowed from mammals\textsuperscript{152}, which when expressed in the midgut\textsuperscript{153}, salivary glands\textsuperscript{154}, or both tissues\textsuperscript{155}, can significantly reduce parasite loads. While originally developed in \textit{An. stephensi}, these combined transgenes may achieve similar inhibition of parasite development in \textit{An. gambiae} and therefore represent good cargo candidates for population replacement gene drives.
Disruption of host-pathogen interactions required for parasite invasion and replication also represents a promising strategy. For example, transgenic expression of the small synthetic peptide SM1 that binds to both midgut and salivary gland epithelia has been shown to block putative receptors required for invasion of the rodent malaria parasite *P. berghei*, resulting in reduced parasite loads\(^\text{156}\). However this strategy was ineffective against the human malaria parasite *P. falciparum* infections\(^\text{157}\). In another example, knockout of the enzyme *kynurenine 3-monooxygenase* (kmo) resulted in a decrease in Xanthuric Acid (XA) production in the midgut\(^\text{148}\) – a waste product of tryptophan metabolism, and a known *Plasmodium* exflagellation stimulator\(^\text{147}\). Such mutants displayed significant decreases in oocyst intensity, suggesting that if the phenotype can be recapitulated in infected *An. gambiae*, knockout during the course of drive spread by the addition of drive-encoded kmo-targeting gRNAs could be a viable candidate anti-malarial cargo. Despite these promising advances, no anti-malarial gene described to date achieves complete suppression of the deadliest malaria parasite, *P. falciparum*, necessitating further development for population replacement strategies.

### 1.6 The problem of drive-resistant alleles (DRAs)

Gene drives only copy and spread when the CRISPR-generated DNA breaks are repaired by the HDR pathway. HDR rates are known to vary across cell types\(^\text{133}\), developmental stages\(^\text{158,159}\), species\(^\text{160}\), and phase of the cell cycle\(^\text{161}\). Use of any one of the alternative repair pathways, such as the error-prone non-homologous end-joining (NHEJ) or microhomology-mediated end joining (MMEJ), will not only result in failure to copy the drive, but may also prevent future drive spread via the generation of mutations within the target sequences recognized by the gRNA. Once one such mutation arises, or the drive encounters a naturally occurring one, the target sequence can no longer be recognized by the gRNA and cannot be cleaved by Cas9. The inability to cleave and home renders the drive no more than a large burdensome transgene. If the fitness burden of these uncleavable alleles is negligible, mutations will increase in
frequency and will out compete the drive, forcing it to extinction. Such alleles thus earn the name ‘drive-resistant alleles’ (shortened to ‘DRAs’ for brevity) (Figure 1.3B). For clarity, and in keeping with nomenclature standards of the field\textsuperscript{62}, DRAs which maintain the wild type phenotype will be termed ‘r1DRAs’, while those that disrupt gene function will be termed ‘r2DRAs’ within this text.

Figure 1.3 | Biased inheritance of Gene drives through mosquito populations and the genetic outcomes of cleavage and homing. [A] During perfect drive, mosquitoes heterozygous for the gene drive (blue) always pass on the transgene to the progeny even when mating to wild type (grey). Over many generations, the biased inheritance of the drive results in the drive increasing in frequency, until all individuals in the population have it. [B] Individuals encoding the gene drive will initiate DNA cleavage of the corresponding wild type chromosome. This DNA break can repair by a number of endogenous cellular repair pathways, which can result in multiple distinct outcomes. The DNA break can repair back to the original wild type sequence, in which case cleavage can re-occur iteratively until one of the other two outcomes have occurred. The second outcome results in mutation at the site of the break, often the result of an End-Joining repair pathway. These mutations can inhibit subsequent cleavage and homing and are termed Drive Resistant alleles (DRAs). The final repair outcome is that of successful drive homing into the break, which occurs as a result of repair by the Homology Directed Repair (HDR) pathway. This homing has resulted in the duplication of the drive sequence from the original chromosome into the homolog, effectively making a heterozygous cell homozygous for the drive. If occurring in the germline, this causes biased inheritance of the drive. Reproduced from\textsuperscript{2} under Creative Commons 4.0. Also available in original text in Appendix D.
Modeling predicts that the key factor to drive success is its ability to suppress DRA formation, with homing efficiency, fitness burden, and introduction frequency only minimally contributing to drive success\textsuperscript{163}. The generation of innovative drive architectures that are evolutionarily stable\textsuperscript{164} and capable of overcoming this issue is therefore a priority for the successful implementation of gene drives, as discussed in one of our studies (Appendix D\textsuperscript{2}) and the topic of a Patent Application\textsuperscript{137}.

1.7 Learning from other CRISPR gene drives

In 2015 the first demonstrations of CRISPR-mediated gene drives in mosquitoes were published for population suppression and population replacement in \textit{An. gambiae} and \textit{An. stephensi} respectively\textsuperscript{62,165}. These gene drives were characterized by use of a single gRNA targeting a sequence which could tolerate some degree of mutation. While these papers demonstrated successful drive in many individual couplings, they both failed to demonstrate invasion of the drive into cage populations presumably due to inhibition by DRAs. The replacement drive in \textit{An. stephensi} addressed this in depth by demonstrating significant DRA formation (>77\%) in the progeny of females\textsuperscript{64}, likely due to leaky maternal deposition of Cas9 – expressed by the germline \textit{Vasa2} promoter\textsuperscript{166} – into developing oocytes at a stage in which homing cannot occur. The population suppression drive in \textit{An. gambiae} addressed this phenomenon in later publications by characterizing the generation of DRAs in the gRNA binding site, and demonstrated that they not only prevented drives from reaching fixation but also competed them to extinction\textsuperscript{167}. These examples underscore the importance of designing drives with architectures that i) prevent the emergence of DRAs within the population, and ii) can ‘delete’ DRAs were they to arise.

Similar drives developed concurrently in \textit{Drosophila} also demonstrated the emergence of drive-resistant mutant alleles, which could be slightly attenuated by using the less leaky \textit{nanos} germline promoter for Cas9 expression\textsuperscript{168}. Modified population suppression gene drive designs soon followed in \textit{An. gambiae} which also attenuated DRA formation of earlier drive designs\textsuperscript{62} by use of the more tightly regulated
germline promoters *nanos* and *zpg*, however achieving only slight decreases in the rate of DRA-based drive inhibition\textsuperscript{169}. These were some of the first demonstrations that fine-tuning Cas9 expression to specific cell types and stages could have a notable impact on not just homing efficiency, but also could contribute to minimizing DRA formation.

In our opinion, all drive designs relying on a single gRNA for cleavage should not be considered truly ‘evolutionarily stable’\textsuperscript{170,171}, as even those targeting the most conserved loci run the serious risk of developing DRAs, thwarting drive spread. The best way to overcome the issue of previously existing or recently evolved DRAs is to target multiple sites\textsuperscript{172}. Because mutations in target sites are evolutionarily favored only when they are beneficial in providing resistance to the gene drive, using many target sites would reduce the chances of mutations emerging at all of the sites, as long as cutting rates are high\textsuperscript{103}. In fact, numerous groups have modelled the rate of DRA formation and the number of gRNAs required to achieve fixation in spite of this with \~4 gRNAs predicted to be sufficient for most systems\textsuperscript{172}. However, drives relying on multiple gRNAs have specific design requirements to prevent cleavage of the drive-containing chromosome, which necessitates the target sequence be removed or changed during homing to prevent self-cleavage within the drive chromosome. A multi-gRNA design was recently demonstrated in *Drosophila* in which four gRNAs targeted a gene for female fertility to induce population suppression\textsuperscript{173}.

While this drive demonstrated super-Mendelian inheritance, it was highly ineffective at homing despite nearly 100% cleavage rate in the majority of individuals. It was also plagued by the creation of DRAs, many of which corresponded to aberrant sequence shuffling over repetitive regions of the drive, or by incomplete or aborted homing. This drive may have experienced instability due to its design. Two of the four gRNAs targeted internal sequences far from the homology arms, while the remaining two targeted regions adjacent to the homology boundary. With rates of HDR-mediated repair being highly dependent on the homology’s proximity to the cut site\textsuperscript{174}, we suspect that only the two outermost gRNAs generated cleavage events which enabled HDR repair. Further, the authors reported that one of the outermost
gRNAs was a disproportionately prolific cleaver while the other gRNA near the homology boundary was not, suggesting this drive may have struggled to complete HDR repair at this end. Taken together these suggest that gRNA localization proximal to the homology boundary may be critical to attenuate instability.

Another population suppression gene drive in An. gambiae attempted evolutionarily stability by targeting highly conserved DNA sequences intolerant of SNPs due to their position over a critical splice site. This drive was designed to interrupt splicing of the female-specific isoform of doublesex, a gene required for insect sex-determination. Either disruptive mutations or successful homing prevented female-specific splicing, resulting in an intersex phenotype and complete female sterility, while leaving males unaffected. This drive homed efficiently via a single gRNA causing extinction of cage populations in only 7-11 generations, without succumbing to competition from DRAs. However, the drive was released at what we consider an unrealistic allele frequency of 25%, thus not permitting the population sufficient time to evolve DRAs as would be expected at lower release frequencies. Furthermore, the site targeted by the single gRNA is polymorphic in wild populations, and while the authors demonstrated that such a polymorphism can be cleaved by their gRNA in vitro, the in vivo cutting efficiency over this polymorphism may not be sufficient to prevent resistance. Importantly the existence of a naturally occurring polymorphism demonstrates that this site can in fact tolerate some variation, suggesting that r1DRAs will eventually emerge. Therefore although this type of design architecture is promising, it is important to note it is not wholly ‘resistance proof’, and moreover it could be only utilized for population suppression campaigns, and not replacement, as it relies on knocking out an essential gene.

1.8 Fitness considerations for gene drives

Gene drives are large, nuclease-encoding, genetically burdensome transgenes whose fitness costs on the individual and population are a significant and critical factor to consider when designing potential drives for release. Two examples of these potential costs are provided below.
1.8 A. CRISPR off-target mutagenesis.

Despite its promise and flexibility, the CRISPR system has a well-documented tendency to cleave off-target sequences\textsuperscript{175}, sometimes targeting sites with as many as five basepair mismatches between the target and gRNA\textsuperscript{131,174,176-178}. The unintended DNA damage may be even greater than that reported so far, as commonly used methods to identify off-target mutations were recently shown to miss perturbations such as large chromosome rearrangements, deletions, and insertions\textsuperscript{179}. CRISPR paired-nickase strategies have been developed that have more stringent sequence requirements and can reduce off-target effects by 50-1,500-fold\textsuperscript{180}. However because CRISPR-based gene editing tools are still incipient, no one has yet tested this strategy for gene drive.

Fitness effects due to off-target mutagenesis are likely to become more pronounced over time as drives spread longer, farther, and faster. A gene drive expressing CRISPR for many generations would be expected to accumulate many off-target mutations over time, causing increasing fitness defects in the drive-containing genetic background until it homes into a new genetic background, potentially inhibiting drive spread over the long term, a side-effect ignored during modeling. Moreover, Cas9-expressing transgenic lines in many species have been hinted to experience fitness defects. For instance some Cas9-expressing Drosophila lines have notable fitness costs (personal communication Dr. Kevin Esvelt, MIT), and attempts to generate An. stephensi lines expressing Cas9 ubiquitously resulted in early larval lethality (unpublished, personal communication Dr. Anthony James, UC Irvine). Fitness costs in An. stephensi were contained only when Cas9 expression was confined to the germline, following our suggestion, resulting in the generation of the first successful population replacement gene drive in mosquitoes\textsuperscript{165}.

Interestingly, work from our lab suggests that Cas9-expression may not only affect mosquito host fitness, but also transmission of malaria parasite. Analysis of a germline Cas9-expressing transgenic line developed in the course of this work (further detailed in Section 2.1) revealed that females expressing the nuclease
in the absence of any gRNA supported significantly higher levels of *P. falciparum* infections. This effect was not mitigated by RNAi suppression of Cas9 or by outcrossing, suggesting Cas9 had caused a permanent alteration in the genome (unpublished, Dr. W. Robert Shaw). While the exact mechanism behind this phenotype is still unknown, these results are a stark warning that Cas9 activity in the absence of gRNAs may introduce unwanted mutations in the mosquito genome, as previously demonstrated by Sundaresan et al. *in vitro*[^181], with possible negative consequences for disease transmission.

1.8 B. Gene drive-imposed evolutionary selection pressures.

Not all gene drive designs exert the same selective pressures on their mosquito host. Theoretically a drive designed to efficiently avoid the emergence of DRAs should be self-optimizing through purifying selection[^182,183]. Under purifying selection, broken or suboptimal drives within a population will be counterselected, while those most efficient will spread prolifically. Gene drives are transgenes which are subject to at least the same rate of baseline random cellular mutagenesis as other genomic sequences. Therefore Cas9- and gRNA-encoding sequences are expected to eventually mutate, rendering gene drives simply non-driving transgenes. However, even if such mutations occur in some individuals, as long as others still harbor a functioning drive, then drives should continue to spread if DRAs are absent from the population[^184]. An entire field of study is dedicated to exploring this phenomenon in naturally occurring HEG selfish genetic elements. For these reasons, such HEGs have been described as “molecular parasites, which have to be regarded as distinct evolutionary units, and whose fate is intertwined, but separate, from the fate of the [] host organism”[^184]. Similar to what would be expected of well-designed population-replacement CRISPR gene drives, naturally occurring HEGs undergo cycles of self-optimization until fixation is achieved, followed by mutation, then counter-selection, near extinction, and then reinvasion[^184]. Such self-optimizing properties have enabled some HEGS to drive within their hosts for over 500 million years[^185], has allowed them to hop into distantly related species[^186-188], and has made them hop back-and-
forth between closely related organisms as evidenced by phylogeny discordance between many HEGs and their hosts\textsuperscript{189-193}. This interspecies transfer and persistence of HEGs underscores the natural ability of such selfish genetic elements to self-optimize, and the potential for a well-designed CRISPR gene drive to achieve the same.

Naturally occurring HEGs can however only achieve fixation if they impose minimal fitness costs on their host\textsuperscript{184}. It is therefore expected that any CRISPR gene drives with significant effects on host fitness may struggle to spread. For example, CRISPR gene drives which cause infertility or population suppression\textsuperscript{62,170} exert extremely high selection pressure on their mosquito hosts because survival of the species depends on emergence of resistance mechanisms. Similarly, population replacement gene drives seeking to spread infection-inhibiting factors may impose selective pressures depending on the phenotypes they cause. For example, while a cargo ablating the mosquito salivary glands might not be able to transmit pathogens\textsuperscript{194}, significant selection pressure to maintain critical organ function could thwart its spread. Therefore, less disruptive cargoes, such as those which simply enhance the mosquito’s ability to mount an immune response against the parasite (Section 1.5E), may be preferred to achieve robust population replacement.

Intriguingly, synthetic CRISPR-mediated population replacement gene drives could also further self-optimize beyond their original design. It is possible that such drives may, for example, select for polymorphisms in Cas9 regulatory regions to increase or fine-tune expression to ideal cell types and developmental stages, or may select for mutations in Cas9 itself to reduce off-target mutagenesis. Such drives may select for altered gRNA sequences to enable drive into polymorphic sequences, or even duplicate gRNAs to enable faster drive. After many generations a drive could therefore have a very different genetic make-up relative to its original design, posing the important question of possible associated risks.
It is important to note that even the most robust CRISPR-mediated population-replacement gene drives are likely to succumb to a variety of other inhibition mechanisms evolved by eukaryotes to combat selfish genetic elements. Selfish genetic elements have evolved innumerable times throughout history, and one type, transposons, account for nearly one-third to one-half of the *Drosophila* and human genomes respectively\(^{195}\). Many organisms have evolved robust systems to defend against transposable elements\(^{196}\), because remobilization results in genomic instability due to mutation at excision and insertion sites\(^{195}\). These selfish genetic elements all share the feature of generating long-noncoding RNAs with significant secondary structure, capable of being recognized by the cell as ‘non-self’\(^{197-199}\). Generally recognized by formation of hairpins in nascent transcripts, the genetic elements which encode them ultimately become the target of gene silencing\(^{200,201}\) by the PIWI pathway, first post-transcriptionally, and then followed by more permanent heterochromatinization of the encoding locus. But there is further evidence that the selfish-genetic element burden is so high in insects, that some species have evolved additional RNAi-like silencing\(^{202}\) pathways to deal with the burden imposed by transposons. Perhaps unsurprisingly then, attempts to create dsRNA-expressing transgenic systems in *Aedes* mosquitoes resulted in the loss of RNAi-silencing from the transgene after 17 successive generations while leaving the tightly-associated fluorescence reporter gene unaffected\(^ {203}\). And subsequent work by the same group to generate a completely new set of dsRNA-expressing transgenic lines experienced the same fate\(^ {204}\). These observations suggest that mosquitoes retain powerful cellular mechanisms to suppress selfish genetic elements through the recognition and selective silencing of transgenic stem-loop RNAs. How does this affect CRISPR-enabled gene drives? Perhaps the power of such gene drives – the ability to target multiple sites by gRNAs – could also be their downfall. Although slightly different structurally, the gRNAs encoded by CRISPR gene drives are distinctly similar to long noncoding RNAs (IncRNAs); small transposon-associated RNA species\(^ {205}\), with no protein-encoding counterparts in the genome, and with distinct stem-loop secondary structures. Similarly, gRNAs could be targeted for inhibition by similar cellular mechanisms.
used to silence the noncoding RNAs of transposons due to their similar structure and properties. Although the current focus of the field is to suppress DRAs to enable drive spread, it is possible that once this issue is overcome gene drive scientists may also face the hurdle of overcoming gRNA silencing.

1.9 Designing evolutionarily stable gene drives for population replacement

As mentioned before, for a gene drive to be evolutionarily stable, DRAs must be prevented and if they do arise a mechanism must be designed to delete them. This could be achieved using a drive architecture which is inserted within – but does not disrupt – the C-terminus of a haploinsufficient gene (also termed haplolethal), i.e. a gene in which loss-of-function in a single diploid allele results in cell lethality. During drive homing, most mutations would result in cell death, precluding the spread of DRAs and ensuring that correct drive homing is the only outcome which restores gene functionality and permits cell survival. In contrast, designs targeting exclusively haplосufficient genes would generate many r2DRAs, greatly inhibiting drive spread in \{r2DRA/drive\} heterozygotes by their persistence.

Haplolethal genes are rare within eukaryotic genomes, and nearly all function in ribosome protein synthesis\textsuperscript{206}, with significant conservation from yeast to humans\textsuperscript{207}. Functionally, their haploinsufficiency is due to the average cell’s requirement for an astonishing number of ribosome units, with individual mammalian cells requiring as many as 10 million ribosomes each\textsuperscript{208}. Despite being caused by different mutations in distinct subunits, all mutants share similar growth defects whose phenotypes are practically identical\textsuperscript{209}. Although haplolethal genes are difficult to study due to their inability to tolerate genetic perturbation, a number of screens in \textit{D. melanogaster} have identified ribosomal genes by disruption of their regulatory sequences\textsuperscript{209}. Such mutants are termed Minutes for their diminished size and growth defects; with nearly all characterized by prolonged development, female-infertility and low male fertility, small body size, abnormally short bristles on the body, and failure to eclose\textsuperscript{210}. Minute phenotypes fall on a dose-dependent spectrum with more intense phenotypes corresponding to greater loss of ribosome
protein\textsuperscript{209}, but significant phenotypes can be observed with as little as a ~10% decrease in mRNA levels\textsuperscript{211}. To date the only known \textit{Minute} demonstrated in mosquitoes has been an ablation of ovarian development following transient ribosome depletion\textsuperscript{212}, and most ribosomal genes in \textit{An. gambiae} have over 80% protein identity to \textit{D. melanogaster} despite being separated by >200 million years of evolution\textsuperscript{213}, supporting functional conservation of this critical cellular process. Therefore, while on the one hand gene drives designed within ribosomal genes would be extremely challenging to engineer, leaving no room for error, on the other hand success would provide perhaps the strongest selection possible against the emergence of DRAs.

1.10 Mosquito genetic control strategies need more tools

Field releases of GM mosquitoes have already been carried out in the Cayman Islands, Malaysia and Brazil since 2009\textsuperscript{60,80}, and releases of population suppressing gene drive mosquitoes are on the horizon. With GM mosquito releases in Africa becoming increasingly likely, the lack of effective tools to monitor and assess the effectiveness of field these trials has become apparent. Regardless of whether SIT-like trials (\textsection 1.4A) or gene drive releases (\textsection 1.5) are being considered, all strategies require the exclusive release of males because females bite and spread disease. The fitness and competitiveness of these males in a wild mating setting is critical to trial success and is a very important variable to monitor. This is particularly relevant in mosquito species that, such as the most important malaria vectors \textit{An. gambiae}, \textit{An. coluzzii} and \textit{An. funestus}, mate in large swarms where competition between males is fierce\textsuperscript{214}. However there are currently no straightforward methods to determine the mating fitness of released males without sacrificing the female. Currently available methods necessitate dissections of females shortly after copulation to assess the presence of sperm in the spermatheca and the mating plug in the atrium. The only non-invasive assessment of male mating success available at present is to blood feed mated females, allow them to oviposit, and screen their progeny for specific genetic traits to determine
paternity, thereby making analysis of mating events laborious. Novel methods to study the fitness and mating competitiveness of these released males are needed if GM field releases are to be successfully monitored. Additionally, genetic strategies for vector control rely on the release of male-only populations, because female mosquitoes bite and spread disease. This necessitates the generation of genetic sexing technologies to separate males from females prior to a release.

Mechanical sorting methods are possible in *Aedes* and *Culex* species due to sexual size dimorphism\textsuperscript{215}, with robots recently being developed to enable this on a large scale\textsuperscript{216-218}. However this technology is not readily transferable to anophelines due to close similarities between the sexes, necessitating development of genetic technologies to aid in the sorting process. Transgenic lines have been developed which differentially express fluorescent markers in the sexes\textsuperscript{110,219-221}, enabling fluorescence-based sorting to remove females. However these systems suffer from multiple problems including the fact that half of the population is discarded (females) and sorting must occur at the larval stage\textsuperscript{110,219} precluding the option of shipping pre-sorted embryos. Systems that selectively kill females are therefore more desirable. In an effort to make males incapable of siring daughters, transgenic lines have been developed which encode HEGs capable of selectively destroying X chromosome-containing sperm\textsuperscript{106} (discussed previously in Section 1.5A). While early versions of this system were plagued by infertility caused by sperm-mediated deposition of the HEG into zygotes destroying the maternal X chromosome\textsuperscript{106}, subsequent optimization has led to systems capable of generating ~95% male-only populations with minimal fertility costs\textsuperscript{109,222}. Although significantly improved, this system is still not sufficiently stringent for most control programs, necessitating development of sexing strategies achieving nearly 100% separation. While the work outlined in this dissertation does not attempt to develop genetic sex-separation technologies, it does set out to develop innovative transgenic systems for mosquito vector control, and in doing so, has kept a constant focus on ensuring compatibility and complementarity to many of the sex-separation technologies currently under development.
1.11 Summary of Aims

The work described in this dissertation aims at developing novel tools to aid genetic control programs in *An. gambiae* towards malaria elimination.

Over the course of this work I have aimed to develop both self-limiting population suppression, and self-sustaining population-replacement systems,— as well as a suite of supporting genetic tools to enable better field trials and fundamental studies into basic biology of the deadly primary African malaria vector, *Anopheles gambiae*. Later chapters outline my studies describing work aimed at developing CRISPR-based systems to genetically sterilize males for vector control programs (Chapter 2), generating synthetic selfish genetic elements for spreading disease-refractory cargoes (Chapter 3), and validating a transgenic system for monitoring effectiveness of field trials based on the release of genetically modified (GM) mosquitoes (Chapter 4).

In Chapter 2 we outline the development of CRISPR technology in *An. gambiae* to achieve ablation of the adult male germline, resulting in spermless males for use in SIT vector control programs. The genetic system we generate facilitates generation of large quantities of sterile males in mass, and results in 95% male infertility. We explore the causes of residual male fertility and assess that failure of CRISPR to mutate the target site as well as the induction of fertility-maintaining mutations both result in maintenance of male spermatogenesis in a minority of males. Following optimization, this novel genetic system could be adapted for use in SIT in *An. gambiae*.

In Chapter 3 we make progress toward the generation of CRISPR-enabled, evolutionarily stable gene drives for population replacement, and develop a suite of novel genetic tools to facilitate their creation and testing. We develop a novel method — based on the cellular phenomenon of Interlocus Gene Conversion — for gene knock-in of complex templates into genetically intractable loci to create gene drive
docking lines. We demonstrate successful survival of individuals carrying recoding of an endogenous haplolethal ribosome gene, and characterize them as Minute ribosome mutants. In attempts to create a functioning drive, we demonstrate that strict regulation of Cas9 expression is critically important for designs targeting haplolethal genes, and identify areas requiring significant optimization for successful drive. Our work also suggests that gRNA-expressing transgenes may be epigenetically silenced over time, leading to loss of function, a finding with potential impact for future gene drive campaigns. Combined, our results suggest that drive designs targeting haploinsufficient essential genes in the primordial germ line may be fundamentally inviable.

Finally, in Chapter 4 we develop a novel transgenic tool to enable the non-invasive assessment of male mating competitiveness in GM field trials. We characterize a Male-Accessory Gland-specific promoter, and harness it to create a transgenic line in which males transfer a fluorescent mating plug to females. We demonstrate that these males have mating competitiveness comparable to wild type individuals and show no additional fitness defects, proving that this tool could be included in GM mosquito trials to facilitate the identification of successful mating events and determine the mating competitiveness of released males.
1.12 References


33 Bruce-Chwatt, L. J. & De Zulueta, J. *The rise and fall of malaria in Europe: a historico-epidemiological study.* (Published for the Regional Office for Europe of the World Health ..., 1980).

34 Shah, S. *The fever: How malaria has ruled humankind for 500,000 years.* (Sarah Crichton Books, 2010).


46 Bell, A. M. *Mosquito soldiers: malaria, yellow fever, and the course of the American Civil War.* (LSU Press, 2010).


48 Carson, R. *Silent spring.* (Houghton Mifflin Harcourt, 2002).


Knipling, E. F. Sterile-Male Method of Population Control: Successful with some insects, the method may also be effective when applied to other noxious animals. *Science* 130, 902-904 (1959).


CHAPTER 2.

CRISPR-mediated germline knockouts for genetic sterilization of male *Anopheles gambiae*.

Andrea Smidler¹,², Doug Paton¹, George Church², Kevin Esvelt³, William R. Shaw¹,†, Flaminia Catteruccia¹,†

¹Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

³Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

† These authors contributed equally to this work

**Respective Contributions:** This chapter was written by Andrea Smidler with editorial input from Flaminia Catteruccia. ALS conceived of the study and performed all transgene design, cloning, transgenesis, mutant analysis, husbandry, data analysis, and drafted the manuscript. ALS and KE together performed transgene design. ALS and WRS executed the study. DP performed microscopy. GMC contributed to the study and provided CRISPR expertise. FC contributed to the design of the study, edited the manuscript, and provided entomological and reproductive expertise. Portions of this chapter were based off a published work performed in collaboration with other colleagues:

ABSTRACT
Mosquito control has consistently proven to be the most effective way to prevent the transmission of mosquito-borne diseases. In the fight against malaria, control of the *Anopheles* mosquito with the use of long-lasting insecticide treated bed nets (LLINs) and indoor residual spraying (IRS) has helped reduce the number of new cases by almost 40% since the turn of the century. However, LLIN- and IRS strategies are being rendered increasingly ineffective by the spread of insecticide resistance in mosquitoes, making development of novel vector control technologies important for continued control of this disease. Insect control programs targeting other species have relied on releasing sterilized males who cause infertility in the females with whom they mate. This technique, termed Sterile Insect technique (SIT) has, however, failed in *Anopheles* species, largely due to strong fitness defects in released males caused by the sterilization process. Here we outline a novel CRISPR-based genetic sterilization system for the selective ablation of male sperm cells in *Anopheles gambiae*. Specifically, robust mosaic mutagenesis of the ZGP (AGAP006241) gene in F1 individuals resulting from a cross of a germline-expressing Cas9 line and ZPG-targeting gRNAs, resulted in complete sterility in 95% of mutant males. Mating of zpg males to wild type females caused high levels of female infertility suggesting potential application for vector control strategies although analysis of fertile mutant zpg individuals showed fertility-maintaining mutations. With further optimization, this genetic system could be adopted as the basis for sterile male release campaigns to combat the spread of malaria.
INTRODUCTION

Strategies that target insect vectors are central to the control of vector-borne diseases. Among control strategies targeting mosquito fertility, Sterile Insect Technique (SIT), a strategy based on the mass release of irradiated males to sterilize females upon mating\(^2\), has been successfully implemented against a variety of burdensome agricultural insect pests\(^3\)–\(^5\) (previously discussed in Section 1.4A). Iterative mass release of sterilized males in the wild results in infertility in females, causing population suppression. However adapting this technology to mosquitoes has been challenging, mostly because the use of chemo- and irradiation-based sterilization reduces the mating competitiveness of males\(^6\)–\(^10\).

To circumvent the potential problems linked to sterility by irradiation or chemosterilization, strategies based on genetic manipulation of insect fertility are being developed. Lethal transgenic systems have been developed in the arbovirus vectors, \textit{Aedes} (discussed in Section 1.4B), which preserve male fertility but kill their offspring at later, post-embryonic developmental stages\(^11\),\(^12\). Among the most promising strategies are gene drives based on CRISPR/Cas9 genome editing technology\(^13\),\(^14\) which are capable of spreading infertility through natural mosquito populations by destroying endogenous genes required for fertility while positively biasing their own inheritance (previously discussed in Section 1.5C). The potential of gene drives to induce sterility in \textit{An. gambiae} populations has been recently demonstrated under laboratory conditions, through the biased inheritance of female-sterilizing and androgenizing factors\(^13\),\(^14\), however there are multiple issues related to the use of these genetic systems. Firstly, they are prone to rapid deactivation by the evolution of resistance mutations that render them not functional\(^14\). Secondly, the self-autonomous propagation properties of a successful gene drive makes it difficult if not impossible to control once released in the wild\(^15\), raising concerns for their spread across national boundaries\(^16\) and supporting the need for safer genetic sterilization systems for the control of \textit{Anopheles} mosquitoes. In this work we aimed to develop safe, controllable, and non-invasive CRISPR-based technologies for genetic control of wild \textit{An. gambiae} populations.
The gene Zero Population Growth (ZPG) has been shown to be required for germ cell development in Drosophila\textsuperscript{17,18} and mosquitoes\textsuperscript{19}. It is a germline-specific gap junction innexin which plays a crucial role in early germ cell differentiation and survival\textsuperscript{17}. ZPG is composed of four transmembrane domains which participate in forming pores on the surface of cells in the developing germline, through which close-range signaling occurs with the adjacent somatic support cells to promote homeostatic germ cell development\textsuperscript{17}. Silencing ZPG by transient RNAi injections in \textit{An. gambiae} resulted in sterile males with phenotypically atrophied testes. Importantly, these males initiated typical post-mating responses in females following copulation\textsuperscript{19}, making this an ideal gene target for genetic sterilization. These males were also shown to have similar mating competitiveness to wild type (wt) males\textsuperscript{19}. However, the embryo-microinjection procedure required to generate these males is laborious and is not conducive to large-scale studies. Therefore, here we set out to develop a transgenic CRISPR system to mutate ZPG and generate large numbers of sterile \textit{An. gambiae} males. Crossing germline Cas9 and ZPG-targeting gRNA strains resulted in synchronous RNA-guided dominant biallelic knockouts of ZPG in the developing germline, fully preventing sperm development in 95\% of males. Moreover, our system generated heritable gene knockouts in F2 progeny, further supporting the use of this type of system to generate stable knockout lines for reverse genetic studies\textsuperscript{20}. 
RESULTS

2.1 Generation of transgenic CRISPR-expression lines

To generate CRISPR spermless males we created two transgenic lines, one expressing Cas9 specifically in the germline and another expressing three gRNAs targeting ZPG (AGAP006241) as previously described\(^1\).

The Cas9 line (termed **VasCas9**) is characterized by a selectable neuronal 3xP3-DsRed marker, and expresses Cas9\(^2\) via the Vas2 promoter\(^2\) which enables robust expression in germ cells of both males and females (Figure 2.1A). (NB: This transgenic is used throughout the work described in this dissertation). A second transgenic line, termed **gZPG**, expresses of three gRNAs (a-c) by the Pol III U6 promoter (AGAP013557)\(^1\) designed to target two sequences (the first exon and the 3’UTR) in the endogenous ZPG locus, to generate large deletions and premature stop codons. This transgene is visually identifiable by a 3xP3-EYFP selectable marker which enables fluorescent identification in all individuals throughout the lifecycle, and additional Vasa2-EYFP marker for fluorescent visualization of the germline (Figure 2.1B).

![Figure 2.1](image-url) | Transgenic map and fluorescence properties of VasCas9 and gZPG transgenic lines, previously described in\(^1\). [A] The VasCas9 germline expressing transgenic. It is similarly designed to previous germline endonuclease expression systems developed for gene knockout in *An. gambiae*\(^2\). The canonical Vasa2 promoter\(^2\) (indigo) composed of 2,300 bp upstream of the Vasa gene (AGAP008578) expresses *S. pyogenes* Cas9 (salmon)\(^2\), followed by an Sva40 (tan) transcription terminator. It is cloned into the pDSAR transgenesis plasmid\(^2\), providing a neuronal 3xP3-DsRed selectable marker, and inserted into the neutral X1 transgenic docking site by phiC31 transgenesis\(^2\). The larval fluorescence pattern is shown at right. [B] The gZPG gRNA-expressing transgenic line. It uses the Pol III U6 promoter composed of 322 bp upstream of AGAP013557 (navy) to express a cassette of two tethered gRNAs, gRNAa and gRNAc, and a second U6 promoter expressed a third gRNA, gRNAb (blue). A germline fluorescent marker is included composed of the Vasa2 promoter (indigo) for EYFP expression (chartreuse), which can be easily visualized in the developing male germline of late age larvae (white arrows). A 3xP3-EYFP fluorescent selectable marker is included for neuronal transgenic identification (pale blue, and chartreuse) and the fluorescent pattern of transgenic larvae is shown at right. It is cloned into the pDSAY *An. gambiae* transgenesis plasmid\(^2\), and inserted into the neutral X13 docking site by phiC31 trasgenesis\(^2\).
2.2 ZPG mutants fail to develop normal testes

Crossing gZPG males to VasCas9 females yields F1 progeny with the \(+/gZPG;+/VasCas9\) genotype harboring a single copy of each transgene, identified by dual 3xP3-EYFP and 3xP3-DsRed fluorescence. These F1 individuals undergo significant mosaic mutagenesis which results in failure to develop normal testes, which can be observed in the pupal stage as the absence of fluorescence from the Vas2-EYFP reporter in \(+/gZPG;+/VasCas9\) males (Figure 2.2A). Dissecting these tissues from 2-day old adult males reveals atrophied testes with no visible sperm in contrast to wt controls (Figure 2.2B).

![Figure 2.2| The spermless male testicular phenotype observed by pupal fluorescence and following dissection. [A] Absence of developed testes indicated by absence of testis-derived Vas2-EYFP fluorescence in the ventral pupal tail. A comparison of the fluorescence pattern of individuals with the \(+/gZPG;+/VasCas9\) genotype (left) and control siblings not inheriting the VasCas9 transgene with the \(+/gZPG;+/+\) genotype (right). Individuals with both transgenes characterized by both the 3xP3-EYFP and 3xP3-DsRed neuronal selection markers (diamond-like stripe) fail to manifest any observable testicular-derived Vas2-EYFP expression (blue arrows). Control siblings inheriting just the gZPG transgene, evidenced by the absence of the 3xP3-DsRed selectable marker (dotted outline), demonstrate significant Vas2-EYFP testicular fluorescence from the cassette on gZPG (white arrows). [B] Following dissection of two-day old males, the absence of sperm in the testes (ii, blue arrows) can be observed in \(+/gZPG;+/VasCas9\) males compared to controls (i, white arrows). The male accessory glands are labeled as ‘MAGs’ for clarity. Panel (vi) is a visualization of the absence of sperm in the testes shown in (ii), and is a composite of the inset Differential Inference Contrast (DIC) (iii), DAPI stained (iv) and Vas2-EYFP marker (v).]
Sequencing the germline of {+/gZPG;+/VasCas9} individuals reveals a variety of CRISPR-induced mutations, mostly large deletions between the gRNA target sites (Figure 2.3) but also some insertions as well as small deletions. Notably, many mutations appear localized over gRNA-b, suggesting advantageous cleavage properties. Often we observed different mutations within the same individual, suggesting that active mosaic mutagenesis is occurring throughout the development of the individual mosquito.

**2.3 Male ZPG mutants are highly sterile**

The absence of visible sperm in {+/gZPG;+/VasCas9} males (Figure 2.2B) suggests they may be sterile, making them candidates for use in SIT programs. To test this, we released 26 {+/gZPG;+/VasCas9} males (3-day old) into a cage with 175 wt virgin females, and allowed them to mate for 2 nights. Females were then blood fed and allowed to lay eggs. Out of the 4,132 eggs laid only 126 larvae hatched (3% fertility), suggesting males were highly sterile. To determine whether every male had some degree of fertility or these larvae were sired by a few fully fertile males, we performed forced mating assays with individual spermless (n=26) and wt (n=17) males to wt females and assayed for fertility phenotypes. While the vast majority of females (96%, n=25/26) were completely sterile, some showed normal fertility levels and produced broods with the expected transgene ratio, demonstrating that some {+/gZPG;+/VasCas9} males maintain normal levels of fertility (Figure 2.4A). Additional control experiments confirmed that sterility is a product of both transgenes being inherited and expressed within a single male (Figure 2.4B).
Figure 2.4 | Genetically spermless males cause significant infertility in females following copulation independent of each transgene alone. [A] Percent fertility of individual females force-mated to {+/gZPG;+/VasCas9} (n = 25) or wild type (+/+) males (n = 17). Of the {+/gZPG;+/VasCas9} males assayed, 96% were completely sterile (n = 24/25) with a single escapee fertile male siring a brood with normal percent fertility (blue asterisk). The broods sired by wild type males were 87% fertile on average (SD ± 0.2), and was statistically different from the genetically sterile males (p < 0.0001, Mann Whitney). [B] Infertility is not caused by the individual gZPG or VasCas9 transgenes alone. Percent fertility of individual females force-mated to {+/gZPG;+/VasCas9} (n = 15), {VasCas9/VasCas9} (n = 12), or {gZPG/gZPG} (n = 8) was scored. Of the {+/gZPG;+/VasCas9} males assayed, 93% were completely sterile (n = 14/15) with a single escapee fertile male siring a brood with normal percent fertility (blue asterisk). The broods sired by {VasCas9/VasCas9} and {gZPG/gZPG} were 93% (SD ± 0.1) and 79% (SD ± 3.3) fertile respectively, which were both significantly different than the fertility observed in {+/gZPG;+/VasCas9} (p > 0.0001 and p > 0.05 respectively, Multi-Way AVOVA).

2.4 Characterization of fertile ZPG mutant males

To determine the cause of the residual fertility in otherwise genetically sterile males, we sequence-analyzed the residual testicular tissue or offspring of {+/gZPG;+VasCas9} males to determine the nature of the mutations present. This analysis revealed multiple possible causes for residual male fertility, including lack of mutations and mutations which did not cause frame shifts or stop codons. Some {+/gZPG;+VasCas9} males showed a dimorphic phenotype in their testes, with one spermless and one phenotypically normal testis, suggesting mutagenesis was occurring separately within each tissue. In one individual, sequencing each testis separately revealed that the atrophied testis harbored an 8bp deletion.
under gRNA-C, while the phenotypically normal testis had no identifiable CRISPR mutations. Sequencing the somatic tissues of this male revealed it was heterozygous for a naturally occurring 2bp Asp53Gly polymorphism within the gRNA-C binding site, suggesting that gRNA-C binding over this sequence may have been prevented in the fertile testis and had instead successfully occurred in the other tissue. At large, however, these findings suggest that some males remain fertile because of fundamental failure to mutate both copies of ZPG, and that using gRNAs targeting sites devoid of naturally occurring polymorphisms may enable complete male sterilization. Redesign of the system to target highly conserved regions within this gene could also minimize such escapees.

Interestingly, sequencing progeny from escapee {+/gZPG;+/VasCas9} fertile males and wild type females revealed larvae heterozygous for a 69 bp deletion between gRNA-A and gRNA-C roughly corresponding to the putative 4th transmembrane domain of ZPG17. Because this deletion would likely not affect localization of the other transmembrane domains nor would it affect the extracellular cysteines critical for innexin pore formation (Reviewed in26), we postulate that this ZPG mutant might permit sufficient protein function required for fertility, although further characterization would be required to confirm this. However, this does present further proof that germline CRISPR-expression systems can generate heritable mutations for stable mutant lines, with applications for enabling reverse genetic studies20.

2.5 Female ZPG mutants display severe ovarian atrophy

While not otherwise discussed in this work, ZPG mutant sisters display significant ovarian phenotypes and merit brief discussion. Female {+/gZPG; +/VasCas9} individuals undergo active mosaic mutagenesis during development resulting in ovaries with varied phenotypes (Figure 2.5). In females 48h hours post blood-feeding, significantly less yolk and lipid deposition is observed compared to control females, and often varies for each follicle (Figure 2.5A). On closer inspection, some ovaries are partially ablated (Figure 2.5B), while others experience complete ablation of one lobe and development of the other (Figure 2.5C,D).
Consistent the role of the ovaries in synthesis of the important insect hormone, 20-hydroxyecdysone (20E)\textsuperscript{27,28}, females with the {+/gZPG; +/VasCas9} genotype contain significantly less 20E than controls 28h following blood-feeding regardless of phenotype (Figure 2.5E), and when binned by phenotype, females with no ovarian development in both lobes have almost none\textsuperscript{1}. Taken together, these females provide a valuable biological tool for study of mosquito reproduction and in the role of ovarian tissue signaling on important physiological processes.

Figure 2.5 | ZPG mutant females have significantly ablated ovarian development and are defective in 20E synthesis. [A] Ovary pairs from seven mutant ZPG females adjacent to wild type ovaries 48h post-blood feeding. Mutant ZPG ovaries display varied phenotypes ranging in lipid and yolk deposition levels, with some lobes completely failing development. Select ovaries, B, C and D inset expanded for analysis. [B] Mutant ZPG ovaries display polymorphic phenotypes consistent with differential mutagenesis within different segments of ovaries. Here the bottom half of one lobe is completely ablated (blue arrow), with upper half of lobe developed. [C, D] Here the right lobe in each ovary pair is completely ablated undergoing no yolk or lipid deposition in contrast to the opposing lobe. [E] Females with the ovariless genotype display significantly decreased 20E levels 28h post-bloodfeeding compared to controls (p<0.01, Unpaired T-test, Mean with SEM)
DISCUSSION

Here we outline a system for generating genetically sterilized male Anopheles gambiae for biological study and as candidates for SIT release. We show that crossing transgenics expressing Cas9 in the germline, to transgenics expressing gRNAs targeting ZPG, produces a majority of sterile males in the F1 progeny. These sterile males have atrophied testes, no observable sperm, and harbor numerous CRISPR-generated mutant alleles that arise by active mosaic mutagenesis during development. We show that approximately 95% of these males completely sterilize the females they whom they mate. Coupled with previous research that suggests ZPG knockdown males have comparable mating fitness to wild type males, this suggests that – following optimization of the system – ZPG mutant males could be utilized in SIT programs.

We further show that ZPG fertile males can arise by multiple mechanisms including failure to mutate potentially due to inhibited gRNA binding in the presence of SNPs, and the acquisition of de novo CRISPR-generated fertility-maintaining mutations. Finally, we demonstrate that this system can be used to create heritable knockouts in a gene of interest, evidenced by heterozygous mutant offspring sired by Zpg mutant males, a fundamental tool for basic biological studies in this important vector species.

Anophelines mate in large swarms where competition between males is fierce. While previous work in the lab demonstrate that ZPG knockdown males without testes maintained similar fitness properties, the transgenes used here may effect males fitness in ways dsRNA knockdown did not, so future analysis of male mating competitiveness is critical. Therefore, this genetic system requires similar validation to confirm them as better alternatives to chemo- and irradiation-based sterilization procedures.

This system requires optimization before release, as the 5% of males escaping sterilization is an unacceptably large proportion for mass release programs. To increase sterility, gRNAs could be targeted to less polymorphic sites, and in patterns that prevent function-maintaining mutations. Further, gRNAs targeting additional germline-essential genes could be supplemented, reducing mutagenic escape by...
orders of magnitude with each additional gene targeted. Such genes could be any anopheline homologs to those required for the Drosophila male germline, including Tudor (AGAP008268), B2-tubulin (AGAP008622), or perhaps even Vasa itself (AGAP00857). However before this can be undertaken, understanding the properties required for optimal gRNA cleavage in anophelines is critical. Others have shown that gRNAs vary in their mutagenic potential, an observation also supported by our findings. This could be caused by multiple factors in gRNA design whose elucidation is beyond the scope of this work, but it underscores the need for further research into anopheline gRNA optimization.

A major hurdle for development of genetic sterilization systems is that maintenance of purely sterile, male-only, lines is impossible by nature of their inability to breed. Therefore, all such systems require some degree of inducibility to trigger female killing and to suppress sterility until immediately prior to release. RIDL systems in Aedes achieve this by induction of a lethal transgene following release, which is suppressed by addition of tetracycline during rearing. In our system, because Δzpg knockout males and females are infertile, inducibility is achieved by crossing together two different transgenic populations, which alone do not have fertility defects, making mass rearing more efficient. However, our system does not currently address the requirement for female removal. Female killing was demonstrated in similar systems developed contemporaneously in Drosophila, and following the recent elucidation of sex determination factors in anophelines, introducing these types of modifications to our ZPG mutant system could be implemented with relative ease.

Beyond potential vector control applications, this ZPG mutant system could be used to explore a variety of other biological questions of interest. Firstly, the role of sperm in initiating female-post mating responses is still largely unexplored. An. gambiae females display two major responses after copulation; the stimulation of oviposition following blood-feeding, and the induction of refractoriness to further mating; both initiated following sexual transfer of factors from the male to the female atrium during copulation. Although a previous study showed that sperm is not involved in triggering these female
responses\textsuperscript{19}, the use of transgenic spermless males may help identify more subtle effects linked to sperm transfer and storage. Indeed, in \textit{Drosophila} sperm is needed to extend the mating refractoriness period to up to a week due to male-transferred proteins bound to sperm tails which are released slowly over time\textsuperscript{35-37}. Application of this genetic sterilization system could therefore be used to study the effect of sperm on similar post-mating responses in female mosquitoes, opening up an intriguing avenue of research of some importance to the vector control community.

Our study shows that with correct target selection and precise temporal expression, CRISPR can generate mosaic biallelic knockout of entire tissues, in this case generating genetically spermless males. In this way our system is somewhat analogous to some ‘floxed’ murine models\textsuperscript{38}, revealing the potential for this type of system to be a similarly powerful tool to study basic insect biology. Its adaptation to knockout other organs could enable salivary gland-less mosquitoes\textsuperscript{39}, those lacking haemocytes\textsuperscript{40}, or male accessory glands, which could revolutionize the basic research in this important insect pest, and may lead to insights across a spectrum of topics of interest to entomologists.

Our work presents a novel tool for inducible genetic sterilization in the important \textit{An. gambiae} disease vector, making SIT programs to suppress this mosquito possible. This system facilitates efficient mass rearing, and robust sterilization phenotypes upon induction. While it requires significant optimization to maximize male sterility, and modification to kill females, it provides a valuable proof-of-principle that similar transgenic sterilization systems hold promise to enable genetic SIT in this deadly vector.

**FUTURE DIRECTIONS**

For successful field releases, near complete sterilization of males must be achieved. To accomplish this, we attempted development of a RIDL-inspired, germline-specific, tetracycline-inducible genetic rescue system. Though not otherwise discussed in this work, the transgenic, \textbf{TOZR (Tet-Off ZPG Rescue), was
developed to provide a suppressible copy of ZPG when in a Δzpg mutant background, and was composed of the ZPG promoter and mRNA separated by a Tet-Off switch\textsuperscript{41}. However we discovered TOZR did not respond to tetracycline supplementation in contrast to similar systems developed in other tissues\textsuperscript{42}, suggesting that the privileged germline tissues\textsuperscript{43,44} may be impermeable to tetracycline regulation, narrowing the potential suite of tools available to develop effective inducible genetic sterilization.

Therefore to meet the needs of modern SIT, we theoretically developed a transgenic crossing system to mass produce genetically sterilized – but not transgenic – male-only populations. This high-throughput system builds upon the ZPG mutant system described above, but does not rely on flawless mosaic gRNA mutagenesis, relies on proven RIDL technologies, kills all females, and generates wholly non-transgenic males with guaranteed sterility. We have termed it the **Inducible Eunuch system**, and it uses specific crosses between two lines to make the sterile male-only populations required for SIT (Figure 2.6). The two lines, \textbf{\textalpha}\textit{line} and \textbf{\textbeta}\textit{line}, both carry biallelic Δzpg, a transgenic rescue copy of ZPG, \textit{a 3xP3-DsRed} marker, and tet-suppressible RIDL, in a balancer site maintaining 50% transgene frequency in the stock population. The \textbf{\textalpha}\textit{line} specifically carries a \textit{3xP3-EYFP} marker and an additional RIDL construct on the Y chromosome, while the \textbf{\textbeta}\textit{line} carries the same transgene on the X. Both stock lines would be stable and have fertility maintained by the heterozygous dominant ZPG-rescue masking the underlying Δzpg phenotype. Fluorescent sorting larvae\textsuperscript{45} of the \textbf{\textalpha}\textit{line} for a specific subset of females, and the \textbf{\textbeta}\textit{line} for a subset of males would generate the parental genotypes and sexes needed to initiate the final sterile male cross. Crossing these males and females together in the absence of tetracycline would result in all daughters dying from the X-lined induction of RIDL, and death of all fertile genotypes by induction of the RIDL cassette associated with the ZPG-rescue, leaving only the Δzpg non-transgenic sons alive for release. The small nonsense Δzpg mutation carried by these males would make them indistinguishable from naturally occurring mutants, making possible the argument that they are a ‘milder’ GMO. This may provide an ideal system for SIT-like campaigns in mosquitoes and other insects in the future. Having developed a method
to generate ZPG knockouts, and having developed docking lines which can act as balancers (outlined in Chapter 3), construction of such a system may be possible in Anopheles gambiae in the near future.

Figure 2.6 | Schematic for the Inducible Eunuch system. This system allows for mass rear and crossing of two lines with endogenous ZPG knocked out, which results in killing of all transgenics and females, leaving only non-transgenic sterile males for release. Fluorescent sorting larvae of the α line with a single copy of 3xP3-DsRED and non-3xP3-EYFP by COPAS^45 would yield only ZPG-rescue heterozygous females, and sorting larvae from β line with the 3xP3-DsRED and 3xP3-EYFP pattern would yield only heterozygous males. This single high-throughput sorting step would generate the parental genotypes and sexes needed to initiate the final sterile male cross. Crossing these males and females together in the absence of tetracycline would result in all females dying by nature of RIDL construct induction from the X chromosome, and all those fertile individuals inheriting a ZPG-rescue cassette would die by nature of adjacent RIDL induction. This would leave the only surviving genotype as males, who have not inherited any transgenes, and who therefore manifest the Δzpg phenotype.
METHODS

Generation of transgenic mosquito lines

gRNA design: Design of gRNAs for these lines was previously reported in\(^1\). Briefly, the ZPG locus (AGAP006241) was PCR amplified and sequenced across multiple individuals within our An. gambiae lines to identify any SNPs present. Putative gRNA candidates were identified by in silico tools available through the Broad Institute (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and Zhang Laboratory at MIT (http://crispr.mit.edu\(^{46}\)). Three gRNA targets were chosen to maximize the probability of mutagenesis early in the coding sequence, with the additional aim of achieving large deletions. Two gRNA candidates were chosen, gRNAa and gRNAc, targeting the sequences (5’ GCGGCTCAGTGTGACGG 3’) and (5’ CCGATCGACTGCGTGATCGGATC 3’) within Exon 1 located 71 b and 150 bp from the stop codon respectively. They were further chosen for their localization over semi-unique restriction enzyme sites AleI and PvuI respectively to enable PCR-based identification of mutants, as previously described in\(^{22}\). gRNAb (5’ CCAAGTGTTTGCATTCCTGGCGG 3’) was designed to target the 3’UTR sequence to facilitate generation of large deletions. gRNAs under the control of the U6\(_{57}\) promoter concurrently developed for use by others\(^{13}\) (322 bp upstream of AGAP013557) were ordered as gBlocks (Integrated DNA technologies, Skokie, IL) in two cassettes. gRNAa and gRNAb were synthesized as a tethered pair connected by a 21 bp sequence ( 5’ TTCACTGTGCGCATTATAT 3’) predicted to not interfere with gRNA folding secondary structure (RNAfold)\(^{47}\). The gRNAC was synthesized as an isolated gBlock.

Plasmid Construction: As described previously\(^1\), plasmids were constructed using standard molecular biological techniques GoldenGate cloning\(^{48,49}\), into the An. gambiae transgenesis plasmids pDSAY and pDSAR\(^{25}\). VasCas9: To build VasCas9, the 2.3 kB Vasa2 promoter\(^{22}\) was PCR amplified from genomic DNA using primers (5’ CAGGTCTCAATCCCGATGTAGA CCGAG 3’) and (5’ CGGTCTCACATATTGGTTTCTTTTATTCCAGG 3’) and was cloned immediately upstream of SpCas9.
amplified from plasmid PX165 (Addgene #48137)\textsuperscript{21} using primers (5’ CAGGTCTCATATGGACTATAAGGACCACCGAGGAG 3’) and (5’ CAGGTCTCAAGCTTACTTTTTTTTCCTGGCC 3’). These fragments were Goldengate cloned into the multiple cloning site of the pDSAR vector which provides an Sv40 terminator for protein transcription termination, an attB site to facilitate phiC31 transgenesis into well established An. gambiae transgenesis docking lines containing an attP, and a 3xP3-DsRed fluorescence selectable marker \textsuperscript{25}.  

**gZPG:** To build gZPG, the two previously discussed gRNA-containing gBlocks were Golden gate cloned into the multiple cloning site of the pDSAY transgenesis plasmid\textsuperscript{25}. To facilitate in vivo validation of the presence or absence of a germline, a Vas2-EYFP fluorescence cassette was further cloned into the unique Ascl site on the pDSAY plasmid backbone by Golden gate ligation. For this cassette, the Vas2 promoter was PCR amplified using the primers (5’ CAGGTCTCACGCGATGTAGAACGCGAGCAAA 3’) and (5’ CAGGTCTCACATATTGTTTCCTTTCTTTATTCACCGG 3’) and EYFP was PCR amplified with ( 5’ CAGGTCTCAATGGTGAGCAAGGGCG 3’) and (5’ CAGGTCTCAAGCTTACTTTGTACAGCTCGTCCATGCC 3’). Complete plasmids were sequence verified by Macrogen Sequencing services (Rockville, MD, USA).

**Transgenesis:** Transgenesis procedures were carried out effectively as described in \textsuperscript{1,50,51}. The gZPG construct (350 ng/µl) was co-injected with a phiC31-integrase expressing helper plasmid (80 ng/µl) into the posterior end of >3h-old aligned X13 docking line\textsuperscript{25} An. gambiae embryos (n=1663), and the VasCas9 plasmid (350 ng/µl) was similarly injected into X1 docking line\textsuperscript{25} embryos (n=2585). Survivors were reared to adulthood and outcrossed in bulk to large cages of wild type G3 virgin adults (n>1000) of the opposite sex. New transformants were identified and isolated as newly hatched larvae in the subsequent F1 generation by fluorescence. First generation F1 transformants were outcrossed to wild type G3 to introduce genetic diversity before intercrossing in the subsequent F2 generation to establish homozygous lines. Homozygous lines were established by identification of homozygous individuals by fluorescence intensity and subsequent PCR verification.
**Generation of spermless \(+/gZPG;+/VasCas9\) males**

To generate spermless males in bulk, \(gZPG/gZPG\) males were crossed to \(VasCas9/VaCas9\) females. Females were chosen to give the VasCas9 transgene because Vas2 maternal deposition of Cas9 protein into the developing embryo causes earlier and more significant mutagenesis resulting in a higher penetrance of the spermless phenotype. Genotype of the resulting offspring were confirmed by dual \(3\times P3\)-EYFP and \(3\times P3\)-DsRed fluorescence, and males were sex separated during the pupal stage to guarantee virginity.

**Microscopy**

Imaging of transgenic larvae and ventral pupal tails was carried out under a Leica M80 fluorescence dissecting microscope following immobilization on ice and positioning by paintbrush. Imaging of microscopic testes structure was carried out under a Zeiss Inverted Observer Z1.

**Mutation Analysis**

Male \(+/gZPG;+/VasCas9\) mutant testes or surviving larvae were analysed for mutations by PCR and sequenced for mutations. DNA extraction was carried out using the Qiagen DNeasy Blood & Tissue Kit, and PCR was carried out using a variety of primers flanking the ZPG locus ranging in fragment size from 700 bp to 5 kB, including the forward primers \(5'\ CGTTTTCTTCACTCTCGGCACG 3'\), \(5'\ GCAGCTTCTGGTAGTCGATGTCG 3'\), and \(5'\ CCATTGTTGTTGGCTGAAAGC 3'\), and reverse primers \(5'\ GACCAGAGCCGAAAAAGATC 3'\), \(5'\ GAGGAACGCGGGTTTTTTTG 3'\), and \(5'\ GTGAAATGTTGGCCGGATC 3'\). PCR products were cloned into the CloneJet PCR Cloning Kit (ThermoFisher Scientific) to isolate PCR products corresponding to individual alleles, and plated on Ampicillin (100 µg/mL) LB media plates. Individual colonies were either picked, cultured in liquid media, plasmid miniprepped (SpinSmart Plasmid Miniprep DNA Purification kit, Denville Scientific) and sequenced using the universal pJET2.1F or pJET2.1R primers (Macrogen USA), or the entire agar plate was
sent for direct colony sequencing (Macrogen USA). Resulting sequencing reads were aligned to an annotated Snapgene 3.2.1 file of the ZPG gene sequence.

**Infertility mating assays**

**Bulk mating:** 30 Male {+/gZPG;+/VasCas9} were sexed as pupae and allowed to eclose into a 25cm x 25cm BugDorm cage (MegaView Science co, Taiwan). Four drowned as pupae, leaving 26 living males for the experiment. Female wild type pupae of the G3 line were sexed on the same day and allowed to eclose in a separate cage. The absence of contaminating males was confirmed the next morning, and 176 females were removed and mouth aspirated into the cage containing the {+/gZPG;+/VasCas9} males. They were allowed to mate *ad libitum* for 4 nights, and were bloodfed on day 5 until significant diuresis was observed. An oviposition site consisting of a Whatman® filter paper cone (90mm, Grade 2, Sigma-Aldrich) within a urinalysis cup containing 80 mL DI water was inserted into the cage on day 7. The oviposition cup was removed day 8, and larvae were counted and scored for transgene frequency day 9. Eggs were counted on day 11 and 12.

**Individual forced-mating assays:** 5 day-post eclosure virgin males of their respective phenotypes and blood-fed virgin wild type females of the G3 line were force-mated to guarantee paternity (method available at [https://www.beiresources.org/MR4Home.aspx](https://www.beiresources.org/MR4Home.aspx)). Male carcasses were saved for subsequent mutation analysis. Successful mating was validated by verification of the mating plug in the female atrium by autoflourescence visible through the female cuticle (previously demonstrated in 34) and females were isolated to oviposit within individual paper cups lined with filter paper and filled with 1 cm DI water. The number of eggs laid and larvae hatched were counted from each female’s brood, and larvae screened for transgene fluorescence to verify paternity. Escapee larvae sired by genetically sterilized {+/gZPG;+/VasCas9} fathers were collected for subsequent sequence analysis.
REFERENCES


Criscione, F., Qi, Y. & Tu, Z. GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in Anopheles stephensi. *eLife* **5**, e19281 (2016).


CHAPTER 3.
Advancement towards developing evolutionarily stable gene drives for *Anopheles gambiae* population replacement

Andrea L. Smidler\(^1,2\), Eryney A. Marrogi\(^2\), Enzo Mameli\(^1\), George M. Church\(^2\), Kevin Esvelt\(^3,\)\(^†\), Flaminia Catteruccia\(^1,\)\(^†\)

\(^1\)Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

\(^2\)Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

\(^3\)Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

\(^†\) Co-corresponding, please direct genetics inquiries to KME, and entomologic inquiries to FC

**Respective Contributions:** This chapter was written by Andrea Smidler, with editorial input from Flaminia Catteruccia. ALS and KME conceived of the study and designed drive architectures. ALS performed all transgene design, cloning, transgenesis, husbandry, data analysis, and drafted the manuscript. ALS and EAM executed the study and Western blots. EM assisted in embryonic microinjections. GMC contributed to the study, conceived the CrIGCkid phenomenon, and provided CRISPR expertise. FC contributed to the design of the study and provided entomological and reproductive expertise. Portions of this chapter include ideas from published Research Articles co-written by Andrea Smidler in collaboration with colleagues. These articles are included in full in the Appendix for reference.

**APPENDIX D**

ABSTRACT

Artificial selfish-genetic elements enabled by CRISPR/Cas9 promise to spread of disease-refractory genes into wild mosquito populations, providing a novel tool to combat malaria. All gene drives described to date have either already selected for resistance mutations, or are likely to do so in the near future, necessitating development of improved designs. While some drive designs seek to eliminate local populations, replacement drives instead disperse neutral cargoes through them, imposing fewer fitness costs and increasing the probability of success. In this work, we develop a suite of tools to construct, and attempt generation of, evolutionary stable gene drives for population replacement. We target the haplolethal and haplosufficient RpL11 and Rpt1 genes for insertion – but not disruption – with a drive. As these genes are vastly intolerant to missense and nonsense mutations, successful drive homing is the only survivable outcome for the cell, preventing and deleting resistance alleles in the process. We first make gene drive docking lines in this genetically intractable locus, recoding RpL11 and Rpt1 during knockin, making them resistant to self-cleavage and forcing HDR to delete resistance alleles. For this, we modify the HACK system developed in Drosophila for use as a novel knockin method in Anopheles gambiae using CRISPR DNA cleavage to stimulate interlocus gene conversion (IGC) from transgenic donors, and term the technology CrIGCkid. Using CrIGCkid we integrate multiple gene drive docking lines, and characterize the requirements for reproducibility, however these docking sites may cause aberrant lethality following gene drive integration, and their phenotypes may preclude drive in females. We further develop multiple germline promoters for Cas9 expression, and attempt ectopic drive of the docking sites with them, revealing the critically narrow expression profile required for drive-enabling Cas9 in this system; as lethality, infertility, or failure to drive was observed from all promoters tested. Our findings suggest that multi-gRNA drive designs may not stimulate HDR in the manner initially predicted, and that targeting haplolethal genes during spermatogenesis may be fundamentally incompatible with fertility as putative driving individuals are highly infertile and the only viable larvae recovered had escaped mutagenesis.
Finally, we observe that phenotypes reliant on robust gRNA expression lessen over time, consistent with silencing of transgenes expressing long noncoding RNAs observed by others, raising doubts about the long-term viability of CRISPR gene drives. Taken together, our results stress the numerous difficulties associated with the design and construction of evolutionarily stable gene drives, and demonstrates that constructing successful gene drives still has many hurdles to overcome.

INTRODUCTION

A large fraction of the research on *Anopheles* innate immunity aims to develop tools to make mosquitoes incapable of malaria transmission. Once transmission-blocking genetic factors have been identified, selfish genetic elements can be used to spread them through wild populations rendering entire species refractory to pathogen transmission *(Section 1.5)*. These gene drive systems would replace disease-susceptible populations with disease-refractory ones through the biased inheritance of an anti-malarial cargo, aiding vector control campaigns. The advent of CRISPR/Cas9 technology *(Section 1.5C)* has dramatically increased the chances of developing and releasing population replacement gene drives. As observed in recent studies, CRISPR/Cas9 gene drives however face numerous hurdles, as for instance they are plagued by nearly immediate evolution of genetic resistance to drive spread *(Section 1.6)*. These include *de-novo* mutations under the gRNA binding site which occur as the result of DNA break repair by End Joining (EJ) instead of the Homology Directed Repair (HDR) pathway required for successful drive homing. In the presence of these resistant mutations (drive-resistant alleles, DRAs), gene drives cannot home and are ultimately outcompeted to extinction in the population. While recent drive designs have attempted to circumvent the issue of DRAs by homing into and knocking-out a conserved splice site critical for female development *(Section 1.7)*, such designs are not suitable for population replacement strategies which require maintenance of essential gene function. And in light of modeling suggesting four
gRNAs are required for population replacement\textsuperscript{5}, designs using less should be considered susceptible to DRA-based inhibition and not evolutionarily stable\textsuperscript{6}.

To enable more effective CRISPR-mediated gene drives and delay the emergence of DRAs, we designed a gene drive system inserted within, but not disrupting the function of, a haplolethal and a haplosufficient essential gene. Because these genes do not tolerate mutations, successful drive copying would be the only viable outcome for the cell – and that any nonlethal DRAs (r1DRAs) created in the process would be deleted (Section 1.9).

RESULTS

3.1 Gene drive design

Here we develop evolutionarily stable drives for population replacement in An. gambiae based on designs previously proposed\textsuperscript{1} (Appendix D), and proven in yeast\textsuperscript{7}. Our evolutionarily stable drive design is inserted within – but does not disrupt the function of – a haplolethal gene (Figure 3.1). We delete a C-terminal segment and 3’UTR, as well as the intervening intergenic sequence, and replace it with our gene drive architecture. This includes reconstitution of protein coding sequence with a recoded DNA sequence\textsuperscript{8}, and surrogate replacement 3’UTRs from a paralog (Figure 3.1A). Replacing the endogenous coding sequence with a divergent recoded one maintains endogenous protein function\textsuperscript{9}, prevents cutting of the drive-containing chromosome, and prevents HDR within the region by forcing it to occur beyond the recoding boundary; enabling deletion of r1DRAs during homing (Figure 3.1C). Reintroduction of a paralogous surrogate 3’UTR maintains homeostatic protein regulation while minimizing homology to the homolog. The CRISPR-encoding gene drive components including germline Cas9, gRNAs, and any associated genetic cargo, would be immediately adjacent (Figure 3.1B), occupying the area corresponding to the intergenic sequence deleted from wild type. The gRNAs target the wild type haplolethal C-terminal sequence at multiple, non-overlapping, sites (Figure 3.1D) while being unable to cleave the drive-containing
chromosome due to recoding. If DNA breaks repair by any number of EJ mechanisms most mutations disrupt protein function and cause cell death (Figure 3.1F), prohibiting transmission of r2DRAs to the next generation. In this system, because most haplolethal genes cannot tolerate any missense or nonsense mutations, r1DRAs are limited to SNPs in wobble bases, necessitating simultaneous r1DRAs at all four gRNA targets for full resistance; a highly improbable event. So long as cleavage by one gRNA initiates HDR, successful homing – which must occur beyond the recoding (Figure 3.1C, magenta arrows) – deletes all r1DRAs that have arisen (Figure 3.1E). In effect, this design makes most DRAs lethal, and deletes the nonlethal during homing, making successful drive the only outcome permissive of cell survival. To our knowledge, such a design is the only homing-based architecture capable of evolutionarily stable population replacement proposed to date. Other replacement systems that increase evolutionary stability by targeting cellular EJ-repair mechanisms for suppression, or use single gRNAs, will still permit DRA emergence at some frequency.

Figure 3.1 | An evolutionarily stable CRISPR gene drive design for population replacement. At top are the two homologs, one drive-containing and one wild type. [A] DNA sequence recoding (red bars) replaces C-terminal sequence of an essential gene (mint). [B] Cas9 and multiple gRNAs are encoded on the drive (Cas9, salmon; gRNAs, red shades). [C] Boundaries of homology indicating where HDR can occur (also indicated left). [D] The haplolethal wild type target recognized by gRNAs and Cas9(salmon PacMan). [E] Outcome of successful drive homing into the homolog results in a homozygous cell. [F] Outcome of EJ repair results in lethal mutation.
3.2 Gene drive docking site locus identification

Our design necessitates insertion into the C-terminal end of a haplolethal gene for selection pressure against DRA formation. Further enhancement is achievable if both drive ends are inserted into haplolethal genes, necessitating identification of a locus with two such genes in a tail-to-tail orientation. Ideal haplolethal genes are those involved in ribosome protein synthesis (Section 1.9), however we failed to identify any such gene pairs in the annotated An. gambiae genome11. Expanding the search to include conserved haplosufficient genes, we identified the Rpl11-Rpt1 locus on chromosome 2L with the 60S ribosomal subunit (AGAP011173) facing the 26S proteasome subunit T1 (AGAP011174). Though less ideal than adjacent haplolethal genes, haplosufficient Rpt1 would still provide selection against DRAs.

3.3 Designing two docking sites in Rpl11-Rpt1 – benefits and drawbacks of each design

To develop evolutionarily stable gene drives within this locus, we pursued drive insertion as a two-step process (Figure 3.2). Step 1 creates a gene drive docking site by HDR-based knockin, replacing the endogenous 3’ coding sequences and UTRs with recoding, a fluorescent selectable marker, and two attP\textsubscript{s} for subsequent phiC31 drive transgenesis\textsuperscript{12} (Figure 3.2, Step 1), permitting study of recoding’s effect in isolation from a drive. Step 2 inserts the CRISPR gene drive into the docking site by phiC31 RCME transgenesis\textsuperscript{13,14}, exchanging the docking site fluorescent marker for the drive (Figure 3.2, Step 2).
A number of assumptions were made when developing this drive design. The first assumption, shared by many scientists developing gene drives, is that expression of CRISPR constructs is maintained across many generations (i.e. the behavior of a line at generation 1 (G1) should be similar to its behavior at G10). The second assumption is that the recoding and surrogate 3’UTRs can fully recapitulate endogenous gene function. The third assumption is that CRISPR-expressing gene drives inserted into the docking sites transgenically (Figure 3.2 Step 2) should not show position effects\textsuperscript{15}. The fourth assumption is that multi-gRNA designs can stimulate robust HDR (although this assumption has recently been questioned\textsuperscript{16}). The fifth assumption is that targeting ribosome and proteasome genes in the primordial germ line (the cell type targeted for drive) is compatible with fertility. And the final assumption is that, because successful drive homing is the only survivable outcome for the cell, CRISPR expression in any cell types incapable of HDR will cause cell death and may affect mosquito fitness accordingly. Through the course of this work, we demonstrate evidence that many of these assumptions may be incorrect.
3.3 A. Designing dRPLT – a gene drive docking site with Rpt1 knocked-out. We developed two gene drive docking lines in RpL11-Rpt1 following the basic design outlined above, with the properties of each summarized in Table 3.1. The first, dRPLT (docking site in RpL11-Rpt1), lacks a 3’UTR replacement for Rpt1, rendering heterozygotes functionally hemizygous for the proteasome subunit. Rpt1 hemizygocity makes the remaining copy haploinsufficient and thereby of DRA-intolerant, however also rendering dRPLT – and inserted drives – homozygotes inviable. Drive could still be tested by use of a single-sex-specific promoter for Cas9 expression, permitting heterozygous dominant trait fixation instead of homozygous allelic fixation, and would maintain the most stringent selective pressures against DRA-formation. However, homozygous lethality of this line would make it cumbersome to rear, difficult to manipulate, and potentially unfit for field release, motivating development of a second alternative docking line, termed dRPLTu, containing a complete 3’ UTR (dRPLT phenotype characterization follows in Section 3.5A).

Table 3.1 | Summary and comparison of characteristics of the dRPLT and dRPLTu gene drive docking sites.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>dRPLT</th>
<th>dRPLTu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recoding in Rpt1 and RpL11?</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Replacement 3’ UTR for RpL11 from RpL32?</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Replacement 3’ UTR for Rpt1 from RpL32?</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Fluorescent marker</td>
<td>3xP3-EYFP</td>
<td>3xP3-m2Turquoise</td>
</tr>
<tr>
<td>Rpt1 function and expected line phenotype</td>
<td>Rpt1 knockout (drp1)</td>
<td>Rpt1 not knocked-out</td>
</tr>
<tr>
<td></td>
<td>Homozygous lethal</td>
<td>Homozygous viable</td>
</tr>
<tr>
<td>Drive-resistant allele selection properties</td>
<td>Following Rpt1 knock-out in dRPLT, the remaining wild type allele of Rpt1 is rendered hemizygous. Therefore DRA’s are less tolerated in the wild type copy.</td>
<td>Rpt1 function maintained in dRPLTu. DRAs will be tolerated in wild type allele, and will be allowed to persist in population as heterozygotes inhibiting drive.</td>
</tr>
<tr>
<td>Drive properties following insertion into this site</td>
<td>GDs will be homozygous lethal. Population replacement GDs can achieve heterozygous dominant ‘trait fixation’ but not classic allele fixation.</td>
<td>GDs should be homozygous viable. GDs can achieve classic allele fixation in the population.</td>
</tr>
<tr>
<td>Drive design requirements for population replacement drives.</td>
<td>Single sex-specific germline promoter required for Cas9 expression.</td>
<td>No limits on drive design; both male, female and dual-sex germline promoters are acceptable for Cas9 expression.</td>
</tr>
<tr>
<td>Husbandry considerations</td>
<td>Line must be constantly selected to maintain as heterozygous population.</td>
<td>A true-breeding homozygous stable line should be establishable.</td>
</tr>
</tbody>
</table>
3.3 B. Designing dRPLTu – a gene drive docking site maintaining Rpt1 function. The second docking line, dRPLTu (docking site in RpL11-Rpt1 with UTR), is identical to dRPLT but exchanges a m2Turquoise fluorescent selectable marker (henceforth CFP) for EYFP, and adds a replacement 3’UTR from Rpt3. dRPLTu was designed to maintain wild type-like Rpt1 function, permitting establishment of homozygous lines. Drives in this site could spread to allelic fixation, but will permit the emergence of r2DRAs in Rpt1.

3.4 Knockin of dRPLT docking line into RpL11-Rpt1 by HDR

We first pursued knockin of the dRPLT docking line template into embryos by standard HDR-based knockin methods\textsuperscript{17,18}. The HDR-donor plasmid, HDRdRPLT\textsuperscript{donor}, contains design features outlined in Section 3.3A, and is mapped in Figure 3.2 and Appendix A. Injecting HDRdRPLT\textsuperscript{donor} and in vitro-synthetized gRNAs targeting RpL11 and Rpt1 into 2,732 VasCas9 embryos (Section 2.1) resulted in high embryonic lethality (8.9% survival compared to standard 15-33% survival with other transgenes), and failed to result in knockin progeny. Further PCR on carcasses of injected individuals revealed no evidence of dRPLT-integrated cells. These results suggest RpL11-Rpt1 is at least partially genetically intractable to delivery of DNA by this method, and knockin by direct embryo injection may not be achievable. Indeed An. gambiae transgenesis injections occur into the syncytial blastoderm such that most nuclei undergo mutagenesis while few knockin by HDR, killing the embryo, and necessitating development of single-cell knockin technologies. Therefore to knockin the docking lines we designed a new method based on the cellular phenomenon of interlocus gene conversion (IGC), inspired by a similar system developed in Drosophila\textsuperscript{19}, to achieve integration of a distal donor transgene into our target locus.

3.5 Developing transgenic Interlocus Gene Conversion (IGC) for knockin into genetically intractable loci

Inspired by a press release from Editas Medicine\textsuperscript{20} which reported a novel CRISPR-induced DNA repair mechanism of hemoglobin beta with hemoglobin delta, we decided to recapitulate this gene conversion phenomenon transgenically for knockin of HDRdRPLT\textsuperscript{donor}.
To harness interlocus gene conversion (IGC) for knockin in *An. gambiae*, we designed a transgenic crossing scheme to provide the donor template, gRNAs, and Cas9 transgenically in the adult germline with fluorescent markers orientated for identification of gene conversion events in the offspring (Putative mechanism, Figure 3.3). We term this system **CrlIGCkid** (**CR**ISPR-mediated **I**nterlocus **G**ene Conversion-derived **Knockin**; /ˈkrɪkɪd/). For this we modified the HDRdRPLT<sup>donor</sup> to contain the Rpt1 3'UTR surrogate for dRPLTu (termed **HDRdRPLTu<sup>donor</sup>**), and added to both donors 8 gRNAs targeting *RpL11-Rpt1*, a second fluorescent marker, *act*-DsRed, outside the homology arms of the donor cassette. These transgenes were subcloned into piggyBac transposons<sup>11</sup> for semi-random transgenesis throughout the genome, and were termed **IGCdRPLT<sup>p8</sup>** and **IGCdRPLTu<sup>p8</sup>** respectively (Figure 3.3, Figure 3.4 respectively, Appendix A).

We hypothesized that crossing **IGCdRPLT<sup>p8</sup>-** or **IGCdRPLTu<sup>p8</sup>-**-containing transgenics to VasCas9 may enable knockin of the donor template – including recoding and associated 3xP3-FP<sup>†</sup> marker – into *RpL11-Rpt1* in the adult germline by an IGC-like mechanism. Successful repair of *RpL11-Rpt1* by IGC would copy the 3xP3-FP marker into the locus – leaving the *act*-DsRed marker outside of the homology arms unintegrated – allowing for fluorescent screening of knockin larvae by a 3xP3-FP-positive, *act*-DsRed-negative pattern. These larvae would be the desired dRPLT and dRPLTu lines, and could be confirmed by PCR. However CrlIGCkid from donor transgenes on the same chromosome as *RpL11-Rpt1* (chromosome 3) would be difficult to screen, as knockin larvae would still be positive for *act*-DsRed from the linked donor. However, PCR could identify such ‘hidden’ docking site larvae, and would provide insight into the dynamics of CRISPR-induced IGC gene knockin. Transgenesis injections established four unique **IGCdRPLT<sup>p8</sup>** transgenic families and five donor **IGCdRPLTu<sup>p8</sup>** families (named **IGCdRPLTu<sup>p81</sup>**, **IGCdRPLTu<sup>p82</sup>** **IGCdRPLTu<sup>p83</sup>**, **IGCdRPLTu<sup>p84</sup>**, and **IGCdRPLTu<sup>p85</sup>**) to generate dRPLT and dRPLTu respectively via CrlIGCkid.

<sup>1</sup> The *act*-DsRed uses the Actin5c promoter for intense expression in the gut and the rest of body.
<sup>†</sup> Because dRPLT and dRPLTu use different fluorescent proteins for selection (*3xP3-EYFP, 3xP3-CFP*, respectively) but are similar in function, when referring to fluorescence of either line, it will be shortened to **3xP3-FP**.
Generating dRPLT by CrIGCkid. Crossing females from the four IGCdRPLT family families to VasCas9 males yielded F1 offspring with the {+/IGCdRPLT; +/VasCas9} genotype capable of CrIGCkid in the germline. Outcrossing these CrIGCkid male offspring to wild type females yielded six larvae from one family with 3xP3-EYFP unlinked from act-DsRed, suggesting IGC-mediated knockin of dRPLT had occurred. PCR confirmed knockin, and the dRPLT docking line was founded (Section 3.6A). We performed this replicate five generations (G5) after establishment of IGCdRPLT transgenics, however subsequent replicates at G11 failed to reproduce CrIGCkid, preventing further characterization of the CrIGCkid of dRPLT.
3.5 B. Generating dRPLTu by CrIGCkid. Before initiating dRPLTu knockin by CrIGCkid, we determined the IGCdRPLTu<sup>pB</sup> insertion sites of the five families (Table 3.2) to characterize the influence of donor locus on CrIGCkid efficiency.

<table>
<thead>
<tr>
<th>Line name</th>
<th>Insert number</th>
<th>Insert location</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGCdRPLTu&lt;sup&gt;pB&lt;sub&gt;1&lt;/sub&gt;&lt;/sup&gt;</td>
<td>3</td>
<td>2L.20C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3R.29B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown – likely chr.X</td>
</tr>
<tr>
<td>IGCdRPLTu&lt;sup&gt;pB&lt;sub&gt;2&lt;/sub&gt;&lt;/sup&gt;</td>
<td>1</td>
<td>2R.17C</td>
</tr>
<tr>
<td>IGCdRPLTu&lt;sup&gt;pB&lt;sub&gt;3&lt;/sub&gt;&lt;/sup&gt;</td>
<td>2</td>
<td>Unknown – likely chr.X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown – likely chr.X</td>
</tr>
<tr>
<td>IGCdRPLTu&lt;sup&gt;pB&lt;sub&gt;4&lt;/sub&gt;&lt;/sup&gt;</td>
<td>1</td>
<td>3L.38C</td>
</tr>
<tr>
<td>IGCdRPLTu&lt;sup&gt;pB&lt;sub&gt;5&lt;/sub&gt;&lt;/sup&gt;</td>
<td>1</td>
<td>3L.43D</td>
</tr>
</tbody>
</table>

Table 3.2 | Summary of IGCdRPLTu<sup>pB</sup> insertion sites

Summary of IGCdRPLTu<sup>pB</sup> insertion sites for five iso-female families used in CrIGCkid characterization assays. Line name shown at left, number of transgene insertions in middle column, and chromosome arm and band shown at right. If inverse PCR confirmed insertion sequence, but did not align to the annotated genome it is denoted as “Unknown”, and is likely inserted into repetitive regions of X chromosome.

For CrIGCkid of dRPLTu, we crossed G5 {+/IGCdRPLTu<sup>pB</sup>; +/VasCas9} males from each family to wt females and screened for knockin larvae. Two families, IGCdRPLTu<sup>pB<sub>4</sub></sup> and IGCdRPLTu<sup>pB<sub>5</sub></sup>, produced larvae suggestive of CrIGCkid by fluorescence. From these individuals we established the dRPLTu docking line (Section 3.6B) and validated dRPLTu knockin into Rpl11-Rpt1 by PCR. PCR confirming insertion of each end of dRPLTu into Rpl11-Rpt1 was performed with multiple primer sets, with one side shown below (Figure 3.4). These findings demonstrate that CrIGCkid is a reproducible process for knockin into otherwise genetically intractable genes, and can be used to generate docking sites in haplolethal genes – a critical step for making evolutionarily stable population replacement gene drives in many organisms.
In these experiments we observed that families of males undergoing CrIGCkid demonstrated decreased fertility (Figure 3.6A, Percent fertility) suggesting CrIGCkid may be genetically burdensome. The first three broods from ‘CrIGCkiding’ ([+/IGCdRPLTu^pB4; +/VasCas9]) males were wholly infertile followed by fluorescent identification of single CrIGCkid-positive larvae in each of two subsequent broods (Figure 3.6A, Percent fertility). PCR analysis on the final brood revealed that 77% (n = 27/35) of larvae co-inheriting IGcdRPLTu^pB also harbored a hidden dRPLTu site not visible due to fluorescence overlap (Figure 3.6A, Percent ‘hidden’ dRPLTu larvae). CrIGCkid of dRPLTu from IGCdRPLTu^pB5 could not be formally quantified in this replicate (Figure 3.6A, nX_dRPLTu), but was confirmed by PCR. Both lines carry IGCdRPLTu^pB
on chromosome 3L, 15.8 Mb and 9.7Mb away from Rpl11-Rpt1 respectively, suggesting CrIGCkid occurs in a proximity dependent manner in concordance with IGC observed in other organisms (Reviewed in\textsuperscript{2}). Observation of more ‘hidden’ than visible dRPLTu larvae suggests CrIGCkid occurs more frequently intrachromosomally than with the homolog. To further characterize CrIGCkid in these males, we performed further replicates at G12. In G12 no CrIGCkid was observed from IGCDRPLTu\textsuperscript{a84}-containing males, and none were identified by PCR (n=68) suggesting either CrIGCkid is a stochastic process occurring rarely, or that IGCDRPLTu\textsuperscript{a84} function had degraded over time. Meanwhile IGCDRPLTu\textsuperscript{a85}-containing males sired many dRPLTu larvae in each of three broods (Figure 3.6A, Percent visibly identifiable dRPLTu larvae). PCR for hidden dRPLTu revealed its presence in 23\% (n = 33/141) of IGCDRPLTu\textsuperscript{a85}-containing individuals. Estimating the total hidden dRPLTu larvae in the brood (537 larvae x 0.23 = 124), reveals a similar number to those observed directly by fluorescence (n=115) (Figure 3.6A, Percent projected ‘hidden’ dRPLTu larvae). The close linkage of the IGCDRPLU\textsuperscript{a85} donor to the Rpl11-Rpt1 target (9.7 Mb) therefore suggests it was capable of templating CrIGCkid interchromosomally, capable of CrIGCkid followed by homing of dRPLTu into the homolog, or is separated from Rpl11-Rpt1 by a recombination hotspot.

The absence of CrIGCkid from each brood suggests it may be stochastic occurring in a rare subset of male germlines. To explore this, at G16 we crossed cohorts of ten males undergoing CrIGCkid from the IGCDRPLTu\textsuperscript{a84} and IGCDRPLTu\textsuperscript{a85} families to 200 wt females. No visibly identifiable or hidden dRPLTu sites were identified among all larvae scored (n > 5,000) from five cohorts using IGCDRPLTu\textsuperscript{a84} and four cohorts using IGCDRPLTu\textsuperscript{a85}. This suggests that either CrIGCkid is sufficiently rare that it was unobserved, or – because these experiments occurred at IGCDRPLTu\textsuperscript{a8} G16 while VasCas9 function remained unchanged over many years – that the gRNA-expressing transgenes facilitating knockin may be losing function over time. Further characterization of CrIGCkid for gene knockin is beyond the scope of this work at this time.
Figure 3.6 | Characterization of CrIGCkid to create dRPLTu in [A] males and [B] females. Black and grey denote results from a replicate at G5 and blue denotes a second replicate at G12. **Percent fertility:** Each point is a single egg-lay (one brood) from a single cage of {+/IGCdRPLTu\textsuperscript{pB}; +/VasCas9} males [A] or females [B], outcrossed to reciprocal wt mates. Open points are broods without egg laid and filled grey points are broods with only infertile eggs (G5 only). Closed points are percent fertility if >1 larvae hatched. Number in parentheses is number eggs laid. Brood data was considered invalid and discarded if all larvae could not be scored living. **Percent visible dRPLTu larvae:** Percent larvae with the visible dRPLTu phenotype of {3xP3-CFP-positive; act-DsRed-negative} among total, pie charts summed across all broods. Total summed larvae hatched and scored as n=# below, number of broods summed in parentheses, ‘n=a’ within pie chart is number of visible dRPLTu larvae, ‘n/a’ is no data collected. \(n_{\text{X dRPLTu}}\) is total visible dRPLTu larvae identified in invalid broods with total broods in parentheses. ‘n/a’ denotes no data collected. **Percent projected dRPLTu larvae:** For replicates with visible dRPLTu larvae, and for which ‘hidden’ dRPLTu larvae were PCR-screened (G5 IG CdRPLTu\textsuperscript{pB4} and G12 IG CdRPLTu\textsuperscript{pB5} grey outlined or blue outlined pie chart respectively), percent calculated ‘hidden’ dRPLTu (77% and 23% respectively) was multiplied by the total number of IG CdRPLTu\textsuperscript{pB}-positive larvae observed (36 and 537 respectively) to approximate the number of ‘hidden’ dRPLTu larvae present in the total population (mauve, Projected {+/IGCdRPLTu\textsuperscript{pB}; +/-dRPLTu}).

### A. Characterizing CrIGCkid in males

<table>
<thead>
<tr>
<th>Chromosome positions</th>
<th>Percent fertility</th>
</tr>
</thead>
<tbody>
<tr>
<td>2L. 20 C</td>
<td>unk</td>
</tr>
<tr>
<td>3R. 29 B</td>
<td>unk</td>
</tr>
<tr>
<td>2R. 17 C</td>
<td>unk</td>
</tr>
<tr>
<td>3L. 38 C</td>
<td>unk</td>
</tr>
<tr>
<td>3L. 43 D</td>
<td>unk</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent visible dRPLTu larvae</th>
<th>G5</th>
<th>G12</th>
</tr>
</thead>
<tbody>
<tr>
<td>n= 10180 (2)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 2872 (2)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 2031 (2)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 174 (2)</td>
<td>n= 7226 (2)</td>
<td>n= 115</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 48 (3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent ‘hidden’ dRPLTu larvae</th>
<th>G5</th>
<th>G12</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 1773 (3)</td>
<td>n= 1287 (3)</td>
<td>n= 115</td>
</tr>
<tr>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 3773 (3)</td>
<td>n= 1287 (3)</td>
<td>n= 115</td>
</tr>
<tr>
<td>77% [n= 27/35 (1)]</td>
<td>23% [n= 33/141 (3)]</td>
<td></td>
</tr>
<tr>
<td>% [n= 0/68 (2)]</td>
<td>% [n= 0/62 (2)]</td>
<td></td>
</tr>
</tbody>
</table>

### B. Characterizing CrIGCkid in females

<table>
<thead>
<tr>
<th>Chromosome positions</th>
<th>Percent fertility</th>
</tr>
</thead>
<tbody>
<tr>
<td>2L. 20 C</td>
<td>unk</td>
</tr>
<tr>
<td>3R. 29 B</td>
<td>unk</td>
</tr>
<tr>
<td>2R. 17 C</td>
<td>unk</td>
</tr>
<tr>
<td>3L. 38 C</td>
<td>unk</td>
</tr>
<tr>
<td>3L. 43 D</td>
<td>unk</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent visible dRPLTu larvae</th>
<th>G5</th>
<th>G12</th>
</tr>
</thead>
<tbody>
<tr>
<td>n= 91 (3)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 1349 (3)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 1359 (3)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n= 403 (3)</td>
<td>n= 781 (3)</td>
<td>n= 0</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 48 (3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent ‘hidden’ dRPLTu larvae</th>
<th>G5</th>
<th>G12</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 75 (1)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 115</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 124</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>n, X_{dRPLTu} = 635</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

| Visible (+/++; +/dRPLTu)       | Wild type (+/+) |
|                               | Projected (+/IGCdRPLTu\textsuperscript{pB}; ++/+)
|                               | Projected (+/IGCdRPLTu\textsuperscript{pB}; +/+dRPLTu) |
3.5 C. CrIGCkid does not occur in females, and infertility phenotypes lessen over time. To determine if CrIGCkid could also occur in females, we analyzed the offspring of G5 \{+/IGCdRPLTu^pb;+/VasCas9\} females from the five families. We discovered most families were highly infertile, with the exception of line IGCdRPLTu^pb3, with IGCdRPLTu^pb4 females laying no eggs across all four broods (Figure 3.6B, Percent fertility). Among the larvae produced by other families, none displayed the fluorescence pattern indicative of CrIGCkid (Figure 3.6B, Percent visible dRPLTu larvae). Later replicates at G12 on IGCdRPLTu^pb4- and IGCdRPLTu^pb5-containing females confirmed no knockin larvae by fluorescence nor PCR (n = 0/45 and n = 0/62; IGCdRPLTu^pb4 and IGCdRPLTu^pb5 respectively), however overall fertility had increased (p<0.05 and ns respectively, Mann-Whitney) (Figure 3.6B, Percent fertility, blue points). Together these results suggest CrIGCkid does not occur in females, at least not in a manner which permits larval survival, and increasing fertility further points towards attenuated IGCdRPLTu^pb transgene function.

3.6 Characterization of dRPLT and dRPLTu docking lines

3.6 A. dRPLT is homozygous lethal, experiences proteasome dysfunction, and is likely a Rpt1 knockout. Omission of a 3’UTR during Rpt1 recoding should cause dRPLT recessive lethality due to its requirement for sub-cellular localization\(^2\). Indeed, intercrossing \{+/dRPLT\} yielded no homozygous offspring. To monitor lethality, we crossed \{+/dRPLT\} to a second transgenic within dRPLT, mNosGD\(_{dRPLT}\) (discussed in section 3.7B, maintains all properties of dRPLT), harboring a 3xP3-DsRed marker enabling identification of homozygotes by \{3xP3-EYFP/3xP3-DsRed\} fluorescence. The crosses yielded the expected frequencies of larval genotypes (Figure 3.7A | p > 0.05 Chi-squared test), however all \{dRPLT/mNosGD\(_{dRPLT}\)\} individuals died within 24 hours, phenocopying Drosophila \(\Delta rpt1\)\(^2\). Western blots confirmed proteasome dysfunction by significant poly-ubiquitin accumulation\(^2\)\(^3\)\(^5\)\(^6\) in homozygotes compared to controls (Figure 3.7B). Together these findings confirm knockout of Rpt1 and knockin of dRPLT.
Figure 3.7 | Fitness characterization of dRPLT. [A] Genotypes of one-day-old offspring from {+/dRPLT} x {+/mNosGD<sub>dRPLTu</sub>} cross, scored by 3xP3-EYFP or 3xP3-DsRed fluorescence. Nonsignificantly different from expected Mendelian ratios of 1:1:1:1 (Chi-squared test p > 0.05). [B] Western blot for proteasome dysfunction in wild type (+/+), heterozygous (+/dRPLT) or (+/ mNosGD<sub>dRPLTu</sub>), and homozygous (dRPLT/mNosGD<sub>dRPLTu</sub>) siblings. Supernatant and pellet shown, probing for poly-ubiquitin. 30 larvae per sample, with murine tissue culture positive control.

3.6 B. dRPLTu homozygotes are viable but display Minute-like fertility phenotypes. In contrast to dRPLT, dRPLTu homozygous larvae survive to adulthood (Figure 3.4). Western blot has not yet confirmed proteasome activity, but homozygous viability suggests the replacement 3’UTR from Rpt3 is sufficient to maintain most proteasome function (Table 3.1). However we observed that many homozygotes pupated later than siblings, and either failed to eclose from the pupal case (n = 7/34) or drowned (n= 21/34), with only 17% surviving (n = 6/34). These observations are consistent with Drosophila Minute phenotypes characterized by delayed pupation, pupal death, and failure to eclose.

To further explore Minute phenotypes in dRPLT and dRPLTu, we performed larval competition assays between docking site homozygotes, heterozygotes and wt siblings. In all experiments wt siblings pupated significantly earlier than docking site siblings when reared with together (Figure 3.8A | p < 0.001, and p < 0.001 , Wilcoxon rank), suggesting both lines experience the classic pupation delay observed in other Minutes. However, while fewer {+/dRPLT} larvae survived to pupation than expected (Figure 3.8A | p < 0.01, Chi-square test), dRPLTu larvae had normal survival rates, suggesting that heterozygous proteasome dysfunction – rather than ribosome dysfunction – makes dRPLT larvae more susceptible to larval competition.
Figure 3.8 | dRPLT and dRPLTu individuals display many hallmark Minute phenotypes. [A] dRPLT and dRPLTu individuals form pupae later than wild type siblings. Graphs show sum of three biological replicates, each with 50 wt or 50 {+/dRPLT} siblings, or 25 wt, 50 {+/dRPLTu}, and 25 {dPRLTu/dRPLTu} siblings. Date of pupae formation was scored each day by fluorescence until all pupated, and the percent pupae formed each day of total calculated and plotted. Mean and SD shown, significance calculated as inverse survival curve by Wilcoxon rank (p < 0.001, p < 0.001). Total number pupated by genotype: dRPLT differs significantly from expected 50/50 ratio (Chi-Squared, p < 0.005). dRPLTu does not differ significantly from 25/75 fluorescence ratio (Chi-squared p > 0.01). [B] Midgut (white arrow) and ovaries (grey arrows) of {dRPLTu/dRPLTu} 10-day old female mated with males ad libitum, 48 h after second blood feed. Blood-meal is fully digested but ovaries are undeveloped. [C] Closer examination of {dRPLTu/dRPLTu} 48 h after second blood-meal reveals slight but stunted follicular development. [D] Differences in body size of wild type female {+/+} adjacent to {dPRLTu/dPRPLTu} sister.

These lines showed other defects generally associated with Minutes. While attempting to establish a stable {dRPLTu/dRPLTu} line, we observed that homozygous females could mate and blood-feed, but never laid eggs after many (n>3) blood meals. Dissection 48 hours post-blood feed revealed small and undeveloped ovaries when they should be voluminous (Figure 3.8B, example normal 48h post-blood feed ovaries shown in Figure 2.5A). Further examination revealed that follicle growth was stunted (Figure
3.8C), displaying a string-of-pearls-like phenotype similar to those observed in *Drosophila Minutes*\(^{28}\) caused by insufficient protein synthesis during the ribosome-intense process of oogenesis\(^{28}\). While this phenotype requires further characterization, it is likely due to insufficient ribosome activity, and will prevent establishment of true-breeding \{dRPLTu/dRPLTu\} lines. Finally, a clear macroscopic difference in adult body size and length was observed (Figure 3.8D), further supporting the hypothesis that dRPLTu homozygotes are *Minutes*. While Western blots to determine if homozygotes experience decreased *RpL11* expression were inconclusive, collectively the phenotypes observed strongly point towards a ribosome deficiency due to some aspect of recoding or misregulation by the *RpL32* replacement 3’UTR.

3.7 Gene drives in dRPLT and dRPLTu may be lethal

3.7 A. Inserting B2GD into dRPLT or dRPLTu results in death of all transgenic individuals. With creation of the drive docking lines complete, we began transgenic insertion of gene drives. We constructed a spermatogenic specific gene drive plasmid using the β2-tubulin promoter\(^{29,30}\) for Cas9 expression and the eight gRNAs proven effective to facilitate CrIgCkid, with further 3xP3-DsRed fluorescence and attB sites for insertion onto the docking sites. The plasmid is termed B2GD (β2-enabled Gene drive), and the resulting transgenic is **B2GD\(^{dRPLT}\)** (β2-tubulin enabled Gene drive in the dRPLT site)(Figure 3.9).

Injections into >7,000 {+/dRPLT} mixed embryos gave four B2GD\(^{dRPLT}\) larvae, but all died in development. Such low transgenesis rates and death of all nascent larvae is rare, raising suspicions of a biological culprit. To determine if the homozygous viable dRPLTu site could tolerate transgenesis, we injected 1009 embryos with B2GD and recovered three **B2GD\(^{dRPLTu}\)** nascent transgenics, however followed again by death. Despite low numbers, failure of nascent transgenics to reach adulthood is rare in our hands (Table 3.3), suggesting that gene drive transgenesis into both docking sites is incompatible with life.
Figure 3.9 | Gene drive and distal gene drive transgene maps developed (Appendix A). Orientations of inserts are correct, size not to scale. Gene drives (GDs) are transgenes for which insertion into the docking lines was attempted (shown in dRPLT). Distal Gene drives (dGDs) are in random loci with piggyBac terminal repeats shown (forest green). Rpl11 (mint) and Rpt1 (navy) shown, recoding (red bars) and Rpl11 replacement 3’UTR (violet) consistent with prior images. Cas9 (salmon), recombinated phiC31 attR and attL sequences (black), and gRNAs are shown in blocks of four (red). Promoters for Cas9 begin with ‘p’ and color-coded by gene (mNos (rose), Vasa (indigo), β2 (lime), ZPG 2 kb (rust), ZPG 1 kb (copper)). LacZ multiple cloning site for anti-malarial cargoes in blue. 3xP3-DsRed selectable markers shown in pale blue and red-orange, and act-DsRed selectable markers shown in grey and red-orange. Vasa protein fusion is shown in pale indigo with the XTEN protein linker shown in aqua.

Table 3.3 | Number embryos injected, transgenics recovered and survived to adulthood

<table>
<thead>
<tr>
<th>Transgene name</th>
<th>Embryos injected</th>
<th>Transgenics recovered</th>
<th>Transgenics survived to adulthood</th>
</tr>
</thead>
<tbody>
<tr>
<td>PluTo</td>
<td>786</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Vas-Cas9</td>
<td>1908</td>
<td>35+</td>
<td>35+</td>
</tr>
<tr>
<td>TOZR</td>
<td>631</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Plugin-E22O</td>
<td>735</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>gZPG</td>
<td>1661</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>gChloeA</td>
<td>1567</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>mNosGDdRPLT</td>
<td>1783</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B2GDdRPLT</td>
<td>7014</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>B2GDdRPTu</td>
<td>1009</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Survival rates of select phiC31 transgenesis developed during the course of this work. Transgene names at left are reported for clarity but not all discussed here.
3.7 B. Lethality of gene drives in docking sites may be due to aberrant Cas9 overexpression. We hypothesize that Rpl11 or Rpt1 regulatory elements may be aberrantly affecting Cas9 expression when in the docking site. In our system, Cas9 expression in any tissue incapable of HDR is expected to be lethal, however direct assay of Cas9 levels on nascent transgenics is impossible given low transgenesis rates and early death of B2GD\textsuperscript{dRPLT} and B2GD\textsuperscript{dRPLTu} individuals, necessitating analysis by indirect methods.

To determine if dRPLT can tolerate insertion of drive-like transgenes independent of confounding Cas9 expression, we inserted a modified gene drive capable of gRNA expression but incapable of Cas9 translation. Enabled by a truncated (1,439 bp) Nanos germline promoter (AGA006098 this Nanos gene drive is termed mNosGD, (Figure 3.9, previously referenced in Section 3.6A) and expresses the same eight gRNAs used in B2GD (Section 3.7A) and for CrIGCkid (Section 3.5B). It was transgenically inserted into dRPLT, sequence verified to contain no polymorphisms, confirmed to display phenotypes consistent with transgenic gRNA expression (Figure 3.11), and has been maintained as heterozygotes for 44+ generations. Thus demonstrating dRPLT can tolerate insertion of gene drive-like gRNA-encoding transgenes incapable of Cas9 protein expression.

**Figure 3.10 | Western blot analysis for Cas9 expression in drive lines.** 7 5-10 day old male lower abdomens and female ovaries dissected and probed for Cas9 (160 kDA) or VasaFusionCas9 225 kDA. Controls: actin loading ctl. (42 kDA), Wild type G3 (black) negative ctl. and VasaCas9 (red) positive ctl. Drives: mNosGD\textsuperscript{dRPLT} (rose) gene drive at G40, and distal drives B2dGD\textsuperscript{B} (lime), ZPG2dGD\textsuperscript{B} (rust), and VFdGD\textsuperscript{B} (indigo), ZPG1dGD\textsuperscript{B} not shown.
At G6 we confirmed mNosGD\textsuperscript{dRPLT} females express germline Cas9 mRNA but no protein by Western (not shown), however Western blot at G40 either hints at possible expression or a blot artifact (Figure 3.10). While different versions of the Nanos promoter can enable drive\textsuperscript{31}, the absence of drive-enabling Cas9 expression in mNosGD\textsuperscript{dRPLT} was confirmed as no biased inheritance observed in G1 or G2 (Figure 3.11). Subsequent replicates at approximately G40 confirmed the continued absence of biased inheritance, suggesting the Western artifact not likely Cas9 expression in this line (Figure 3.10).

Figure 3.11 | No biased inheritance of mNosGD\textsuperscript{dRPLT}. Scatter plot of transgene frequency in progeny of adults tested for drive. Female and male mosquitoes heterozygous for the mNosGD\textsuperscript{dRPLT} allele, outcrossed to wild type, were scored for biased inheritance of mNosGD\textsuperscript{dRPLT} by fluorescence in the progeny. Male \{+/mNosGD\textsuperscript{dRPLT}\} is multiple broods from the single founder male. Female \{+/mNosGD\textsuperscript{dRPLT}\} were analyzed for drive in G2 and points single broods from isolated females. A replicate at approximately G40 is a bulk brood from 6 females (n = 593 larvae scored). The 50% expected Mendelian ratio is marked by a dotted line, with Mean and (± SEM) shown.

To determine if B2GD CRISPR-expressing cassettes are lethal independent of insertion in the docking site, we inserted β2-Cas9 and gRNA \textit{Rpl11-Rpt1}-targeting expression cassettes randomly throughout the genome by piggyBac transgenesis\textsuperscript{32}. If non-lethal, these transgenics should be easily established, and if capable of drive they should facilitate biased inheritance of dRPLT or dRPLTu when co-inherited in the same individual. Much like a daisy drive\textsuperscript{33}, we term these distal gene drives (dGD) in this work, with the one enabled by β2-Cas9 expression termed B2dGD\textsuperscript{pB} (Figure 3.9). Injecting B2dGD\textsuperscript{pB} into >700 embryos yielded a surplus of nascent transgenic larvae (>90 larvae), expressing polymorphic fluorescence.
phenotypes indicative of variable insertion sites, and Cas9 expression was verified by Western (Figure 3.10). This suggests that transgenic expression of B2GD CRISPR-components targeting Rpl11-Rpt1 are nonlethal when inserted within a variety of genomic loci, but are lethal when directly inserted into dRPLT or dRPLTu.

To test the ability of B2dGD to bias inheritance of dRPLTu, we scored {+/B2dGD\textsuperscript{pBmix};+/dRPLTu} males and females from bulk and single-family populations for drive, but none revealed biased inheritance (Figure 3.12). The β2 promoter has been previously demonstrated to enable spermatogenic endonuclease expression\textsuperscript{30}, but its expression begins in early spermatocytes\textsuperscript{34} possibly too late after meiosis I to stimulate interchromosomal homing, however IFA on our lines will follow to confirm this. Importantly, confirmation of Cas9 expression suggests that mosquito male germlines can tolerate some Rpl11-Rpt1-targeting CRISPR expression, likely in later stages of spermatogenesis.

**Figure 3.12 | B2dGD\textsuperscript{pB} does not facilitating drive of dRPLTu in males.** Female and male mosquitoes heterozygous for the B2dGD\textsuperscript{pB} and dRPLTu transgenes {+/B2dGD\textsuperscript{pB};+/dRPLTu} were outcrossed to wild type and analyzed for the biased inheritance of dRPLTu in the progeny. Scatter plot of the transgene frequency observed in the progeny with each filled dot representing the frequency of a single brood of either a single female founder (\{+/B2dGD\textsuperscript{pB};+/dRPLTu\}♀) or a cohort of brothers from an iso-female line (\{+/B2dGD\textsuperscript{pB};+/dRPLTu\}♂). Closed triangles show three broods from the same cage of iso-female brothers. Closed triangle with an asterisk denotes a single brood demonstrating statistically significant drive (60.2% allele frequency, \(p < 0.01\), O. vs E Two-tailed), but subsequent broods failed to recapitulate. Open diamonds show drive frequencies from broods sired by mixed population adult males (non-iso-female) {+/B2dGD\textsuperscript{pBmix};+/dRPLTu}. The 50% expected Mendelian ratio is marked by a dotted line, with Mean and (± SEM) shown.
Together these data are enigmatic; B2dGD\textsuperscript{18} transgenics expresses CRISPR and are viable from many genomic positions but do not enable drive of dRPLTu, while their functional clone, B2GD, kills all nascent transgenic individuals when inserted into dRPLT or dRPLTu, and further experiments confirm dRPLT can tolerate insertion of non-Cas9-translating drive-like transgenes. Taken together these suggests that inserting transgenes capable of \textit{any} Cas9 protein expression into the docking lines causes lethality when in the presence of ubiquitously expressed transgenic gRNAs, supporting the hypothesis that regulatory sequences controlling the locus may be able to cause aberrant Cas9 expression and associated lethality. With this in mind, we set out to identify and characterize promoter candidates capable of drive-enabling Cas9 expression, independent of the confounding effects associated with insertion into the docking lines, using the distal gene drive (dGD) system introduced earlier (Section 3.7B).

### 3.8 Developing drive-enabling germline-Cas9 promoters independent of docking sites.

Due to lethality associated with inserting Cas9-expressing transgenes into dRPLT and dRPLTu, we set out to characterize germline promoters for Cas9 expression from more neutral loci. These experiments could facilitate initial characterization of the DRA-inhibiting properties of these drive designs, and would provide a starting point for subsequent fine-tuning of Cas9 expression within dRPLT or dRPLTu.

#### 3.8 A. Vasa2 transgenic promoter is too leaky for drive.

The \textit{Vasa2} promoter has been used for Cas9 expression in successful gene drives in both \textit{An. gambiae}\textsuperscript{35} and \textit{An. stephensi}\textsuperscript{27}. However this blood-meal inducible transgenic promoter is leaky into adjacent tissues\textsuperscript{36} and causes maternal deposition into the developing embryo\textsuperscript{37}. While not an issue for other drive designs, Cas9 expression in any tissues incapable of HDR would cause significant fitness defects or death in our system, making \textit{Vasa2} a poor candidate for our drive designs. We verified that \textit{Vasa2-Cas9} embryo deposition is lethal by injecting a \textit{Vasa2-Cas9} gene drive construct (Figure 3.9, VasGD) into embryos, which recovered only 1.7% survivors (n=17/1010), the lowest percent survival of embryo microinjections to date in our hands. To verify Cas9 expression from
the VasGD transgenesis plasmid was inducing lethality, we co-injected a dsCas9-expressing RNAi plasmid to silence Cas9, which resulted in a significant survival increase (9.5 % survival, n = 105 survivors /1108 injected, p < 0.0001, Fisher’s exact, Two-sided). These data therefore confirm that a Vasa2-enabled gene drive would likely be embryonic lethal in our drive design.

We next attempted to rescue the non-translating Cas9 in mNosGD\textsuperscript{dRPLT} with the Vasa2-Cas9 expressed by the VasCas9 line by crossing these transgenic lines together. In the event of drive, individuals with both transgenes (\{+/mNosGD\textsuperscript{dRPLT}; +/Vas-Cas9\}) should show biased inheritance of mNosGD\textsuperscript{dRPLT}. In a first experiment using G20 of mNosGD\textsuperscript{dRPLT}, we discovered that driving females died within 24 hours post-bloodfeeding (n = 6), while control sisters survived and laid normal broods (Figure 3.13, n = 9 females, Chi-square test), suggesting Vasa2-Cas9 is leaky following blood feeding and causes lethal \textit{Rpl11-Rpt1} mutagenesis. Interestingly, subsequent replicates at approximately G40 produced viable females that however showed complete infertility (n=10) compared to control sisters which laid normal broods (n=12) (Figure 3.13). The lessening of the lethal phenotype over time suggests that transgenic expression weakens with subsequent generations, though it is still sufficient to cause complete infertility. We cannot test this hypothesis directly due to absence of RNA samples from earlier generations.

We then scored for biased inheritance of mNosGD\textsuperscript{dRPLT} from \{+/mNosGD\textsuperscript{dRPLT}; +/Vas-Cas9\} driving males and nondriving control brothers. From bulk cage outcrosses we discovered that control males sired larvae with expected transgene frequencies (n = 1321 total larvae, 49.0% transgene frequency) and normal fertility levels (89.9% fertility), while driving males sired no larvae despite 482 eggs laid (Figure 3.13). These results confirm that Vasa2-Cas9 drives are incompatible with our system design, suggesting more fine-tuned promoters need to be developed.
Figure 3.13 | Rescuing mNosGD<sup>DRPLT</sup> Cas9 function with Vasa-Cas9 results in death or infertility of all females and infertility in males. Female and male mosquitoes heterozygous for the mNosGD<sup>DRPLT</sup> and Vasa-Cas9 transgenes (+/mNosGD<sup>DRPLT</sup>;+/Vas-Cas9) genotype were outcrossed to wild type and analyzed for the biased inheritance of mNosGD<sup>DRPLT</sup> in the progeny compared to sibling controls (+/++;+/Vas-Cas9). Both transgenes are unlinked but share a common 3xP3-DsRed selectable marker, therefore drive would be indicated by >75% 3xP3-DsRed-positive larvae from {+/mNosGD<sup>DRPLT</sup>;+/VasCas9} parents, and >50% 3xP3-DsRed-positive larvae from {+/++;+/VasCas9} control siblings. Females were analyzed at approximately G20 and G40 of mNosGD.<br>

In a subsequent brood from this experiment driving ({+/mNosGD<sup>DRPLT</sup>;+/Vas-Cas9}) males were incompletely infertile, siring 19 larvae from n>600 infertile eggs. Interestingly, the larvae were exclusively nontransgenic, inheriting neither CRISPR-expressing transgene VasCas9 nor mNosGD<sup>DRPLT</sup>. These larvae were not the result of contamination, suggesting a rare event had occurred likely in the germline of a single {+/mNosGD<sup>DRPLT</sup>;+/Vas-Cas9} male. Together these suggest a Vasa2-Cas9 gene drive would cause significant male infertility, and rare viable sperm do not inherit either CRISPR-expressing transgenes.

3.8 B. A Vasa protein fusion to fine-tune Cas9 expression. To overcome this issue, we attempted to precisely express Cas9 only in the primordial germ cells using the Vasa2 promoter to express a fusion between the full-length Vasa protein and Cas9 by an Xten linker<sup>38</sup>, termed VFdGD<sup>38</sup> (Vasa Fusion distal
gene drive, Figure 3.9). The rational for this design was based on knowledge in other Diptera where Vasa localization to the PGC nuage is dependent on its physical interactions with other proteins such as Oscar^39. However IFAs could not verify if the Vasa-Cas9 fusion protein was correctly localized to the germ cells, despite Western confirming its expression. Driving \(+/VfGD^{pB}; +/dRPLTu\) males demonstrated no drive of dRPLTu across two broods (Figure 3.14, \(n = 149/294, 50.7\%\); \(n = 221/411\) 53.8%; both ns, Two-tailed Binomial) in concordance with no VasaFusionCas9 in male germlines (Figure 3.10). However females also did not display drive (Figure 3.14, \(n = 398/853, 46.7\%\), ns Two-tailed binomial) despite VasaFusionCas9 expression (Figure 3.10), and despite having normal levels of fertility. Therefore we set out to develop alternative germline promoters with lower baseline expression which may be capable of drive.

3.8 C. Two versions of the ZPG promoter for gene drive Cas9 expression. To develop alternate germline promoters for Cas9 drive expression, we developed distal gene drives using the 5’ and 3’ regulatory sequences of the Zero population growth (ZPG) Innexin2 gene (AGAP006241). A first transgenic construct using 2 kb of the 5’ promoter sequence and 1 kb of 3’UTR sequence amplified from the genome was termed ZPG2dGD^{pB} (for ZPG 2 kb promoter distal Gene drive). However this promoter failed to achieve

![Figure 3.14](image)
Cas9 expression in adult germlines (Figure 3.10), and did not facilitate dRPLTu drive in bulk female experiments nor from iso-female families (Figure 3.15, p>0.05, Chi-squared). Males from the families in which females demonstrated slight but insignificant drive were either infertile, or showed a bias against dRPLTu inheritance. Taken together, these suggest that the ZPG2 does not facilitate drive-enabling Cas9 expression.

Figure 3.15 | A ZPG2 distal drive, ZPG2dGDΔβ fails to facilitate drive of dRPLTu. Female and male mosquitoes heterozygous for the ZPG2dGDΔβ and dRPLTu transgenes {+/ ZPG2dGDΔβ;+/dRPLTu} were outcrossed to wild type and analyzed for the biased inheritance of dRPLTu in the progeny. Scatter plot of the dRPLTu transgene frequency observed in the progeny with each filled point representing the frequency of a single brood of either a single female founder ({{+/ZPG2dGDΔβ;+/dRPLTu} ♀} or a cohort of brothers from an iso-female line ({{+/ZPG2dGDΔβ;+/dRPLTu}♂}) and were chosen due to >55% dRPLTu frequency in female founders. Closed dots denote a single brood analyzed from a single family. Closed up triangles, down triangles, and squares specify multiple broods analyzed from three distinct families. Down triangle brood with 66.7% dRPLTu frequency only representative of 12 larvae, and biased inheritance not recapitulated in subsequent broods. Open triangles on 0 denote infertile broods from respective families. Open diamonds denote large bulk experiments from a mixed insertion population of {+/ZPG2dGDΔβmix;+/dRPLTu} with ZPG2dGDΔβ at varying genomic loci. The 50% expected Mendelian ratio is marked by a dotted line, with Mean and (± SEM) shown.

More recently, an alternative ZPG promoter was published capable of robust and tight Cas9 expression for gene drive. This ZPG promoter was ~1 kB long, while the 3’UTR used for termination was nearly identical to that used in ZPG2dGDΔβ. Therefore we designed a distal drive using these regulatory sequences ordered as clonal IDT gBlocks and developed a transgenic line termed ZPG1dGDΔβ (ZPG 1 kB promoter distal gene drive). In bulk outcrosses of 10 driving {+/ZPG2dGDΔβ;+/dRPLTu} females, all broods (n=4) laid no eggs despite having undergone oogenesis, with the exception of one female who laid an infertile egg batch. In light of similar observations in G40 female {+/mNosGDdRPLT;+/Vas-Cas9} drive rescue experiments (Figure 3.13), these results suggest that Rpl11-Rpt1-targeting, drive-enabling Cas9-
expression in the germline causes female infertility, consistent with *Minute-like* infertility phenotypes
associated with ribosome depletion\(^{27,28,40}\). In the absence of fine-tuning of Cas9 expression, drive designs
targeting haplolethal genes in the primordial germline may not be viable in females, especially given
significant female-infertility phenotypes known to occur in female *Minute* mutants.

To analyze distal drive properties of dRPLTu by ZPG1dGD\(^{\text{gb}}\) in males, we scored for dRPLTu biased
inheritance in the offspring of 35 driving \{+/ZPG2dGD\(^{\text{gbmix}}\), +/dRPLTu\} males. Brood 1 was highly infertile
(n=134 larvae, >1000 eggs laid), however among hatched larvae, only one inherited dRPLTu and three
inherited ZPG2dGD\(^{\text{gb}}\), suggesting a bias against inheritance of dRPLTu or the CRISPR-encoding parental
distal drive; ‘anti-drive’ (Figure 3.16). These findings phenocopy those observed in the offspring of a single
brood of \{+/mNosGD\(^{\text{dRPL}}\), +/Vas-Cas9\} males (discussed in Section 3.8A), in which males capable of drive
were largely infertile, but when fertile, they only sired larvae without CRISPR-encoding transgenes. Brood
2 was completely infertile despite >1000 eggs laid, while Brood 3 was similarly fertile to Brood 1, but was not characterized by negative biases in transgene frequency. The sequential nature of these brood phenotypes therefore possibly suggests occurrence of a time-dependent anti-drive or fertility effect. Sequence analysis on the fertile larvae sired in these broods revealed no evidence of DRA mutations under any of the eight gRNA binding sites, suggesting those larvae completely escaped CRISPR mutagenesis.

**DISCUSSION**

Our work makes important advancements towards the development of evolutionarily stable gene drives for population replacement in the important *An. gambiae* malaria vector, and suggests that developing such homing-based drives in haplolethal genes still has many significant hurdles to overcome.

*CrlGCKid is a novel knockin technology for insertion into difficult-to-manipulate loci in anophelines; it is robust, reproducible, proximity-dependent, and occurs exclusively in males.*

To make gene drive docking lines, we develop a novel method in the anopheline GM toolkit for reproducible CRISPR-mediated gene knockin, CrlGCKid, for insertion into haplolethal genes – a critical step for making evolutionarily stable gene drives for population replacement. CrlGCKid of two distinct donor templates, dRPLT and dRPLTu, over multiple replicates demonstrates that the technique is predictable and reproducible (Section 3.5B,C), and most likely occurs in a linked manner between the target and donor locus. Templating can occur from either side of the target site, with identical donors in distant loci templating CrlGCKid variably, consistent with IGC in other organisms\(^{19,22}\). Together these confirm the flexibility of this system, but provide general design guidelines when implementing CrlGCKid from other loci and in other organisms.
Within these experiments CrIGCkid is only observed in males, possibly due to the thousands of recombinatorial possibilities producible during spermatogenesis\(^1\). *Dipteran* male germlines can reestablish the stem cell niche from a surviving spermatogonial cell in the event of significant genetic perturbation (reviewed in \(^2\)), suggesting lethal CRISPR mutagenesis may bias for the desired knockin outcome which is resistant to further cleavage. The absence of CrIGCkid from females is likely due to the effects of CRISPR mutagenesis causing ribosome depletion which is critical to oogenesis\(^2\) explaining *Minute*-like infertility phenotypes (Figure 3.8C). Despite reproducibility, CrIGCkid was still a rare event, whose identification was facilitated by *Anopheles*’ colossal fecundity, suggesting it may not be reproducible in species with lower reproductive outputs. Similar systems have also been successful in *Drosophila*\(^1\) and are also characterized by high fecundity.

**Recoding haplolethal Rpl11 causes *Minute*-like phenotypes.**

To construct our gene dive design, endogenous *Rpl11* required sequence recoding. Swapping the codons in this way does not alter amino acid sequence and should maintain protein function, however tRNAs vary in cellular abundance\(^3\), necessitating codon optimization for optimal translation efficiency\(^4\). Cells require as many as 10 million ribosomes per cell\(^5\), making flawless translation of these proteins critically important. To our knowledge, Volohonski *et al.*\(^1\) presents the only study on anopheline codon frequency, but analyzes codons from a single haplolethal gene among a few other genes analyzed, suggesting it is not a comprehensive sampling. With the absence of an alternative, the gene recoding undertaken here was performed based on their analysis, and in the absence of formal studies on anopheline ribosomal 3’UTR regulation, a replacement for *Rpl11* was chosen from *Rpl32* based on shared functionality and stochasticity, and minimal homology. To our knowledge, while creation of the docking lines presents the first reported haplolethal gene recoding in higher eukaryotes, both however experience pupation delay, and dRPLTu homozygotes display small body size, eclosion failure, and female infertility – the classic symptoms of ribosome deficiency (Figure 3.8). Despite the absence of direct protein evidence, these
phenotypes convincingly suggest dRPLT and dRPLTu are anopheline Minutes, and underscore the need for research into anopheline codon optimization and ribosome 3'UTR regulation before successful development of evolutionarily stable drives can be achieved within ribosome genes.

**gRNA-expressing transgenes may be silencing over time.**

Females undergoing CrIGCkid became less infertile with increasing age of the gRNA-expressing donor line (Figure 3.6B), and CrIGCkid itself was never observed beyond G12. Similarly, as the mNosGD<sup>dRPLT</sup> line aged, driving {+/VasCas9; +/mNosGD<sup>dRPLT</sup>} females – which initially died 24H after blood feeding – began surviving (albeit infertile), consistent with a decrease in the function of transgenically expressed Cas9 or gRNAs (Figure 3.13). Because all experiments were carried out over four years using the same VasCas9 line, which has never otherwise varied in function, this suggests the function of the gRNA-expressing transgene was dampened over time. These findings are reminiscent of those observed in dsRNA-expressing *Aedes* transgenics, where the dsRNA-mediated anti-pathogen phenotype was lost over 17 generations<sup>46</sup>, therefore we postulate that the stem-loop like structure of gRNAs and dsRNAs are sufficiently alike they may be triggering similar cellular silencing mechanisms, perhaps originally evolved to silence transposon selfish-genetic elements<sup>47-50</sup>. The PIWI pathway is hypothesized to have evolved for the purpose of silencing active transposons in the germline<sup>47,51</sup>. It silences transposons in the germline<sup>52</sup> by recognition of ‘foreign’ long noncoding RNAs, followed by post-transcriptional silencing and subsequent heterochromatinization of the genetic element, possibly taking many generations. In flies activation of one such transposon, the P-element, is sufficiently genetically burdensome that it causes complete infertility in young flies, but following PIWI activation the P-element is silenced and fertility restored in a single generation as an individual fly ages<sup>53</sup> reminiscent of our findings in Section 3.8C. PIWI may be silencing non-coding RNA transcribing transgenes in mosquitoes, and its disruption may be required to maintain transgene function over generations. While still in need of more robust study, these clues raise serious concerns about the ability of gene drives to spread in the long term, even after
overcoming the hurdle of DRAs, underscoring the necessity for further research into the naturally evolved mechanisms for transposon silencing in many mosquito vector species in which gene drives are being developed.

The Goldilocks Cas9 conundrum – aberrant Cas9 expression within dRPLT and dRPLTu

All CRISPR gene drives are the same in two respects; if there is too little Cas9 in the germ line then drive cannot occur, and if Cas9 expression is localized correctly in the primordial germline, then drive can occur. However our drive is unique in that overabundance of Cas9 in cell types incapable of HDR results in cell death by Rpl11 mutagenesis.

During attempts to insert B2GD into the docking sites, all nascent transgenics died, raising suspicion of a biological culprit. Because position effect causes expression variation from an identical transgenes in different loci, we suspected that the high transcriptional activity of the adjacent Rpl11-Rpt1 genes may be affecting expression of Cas9 cassettes in the docking sites. However direct examination on nascent transgenic larvae was impossible due to transgenesis inefficiency and early death, necessitating indirect testing. To confirm docking sites could tolerate transgene insertion without lethality, we inserted the drive-like mNosGD transgene lacking Cas9 translation into dRPLT and maintained it for >40 generations (Section 3.7B). To determine if the B2GD transgene itself was lethal, we inserted it at multiple other sites in the genome without fitness defect. While we later discovered that β2-Cas9 distal gene drives could not bias dRPLTu inheritance likely due to late Cas9 expression, their insertion into docking sites caused lethality, suggesting augmented lethal expression in this position. These suggest that the position effects influencing dRPLT and dRPLTu may cause aberrant Cas9 expression from integrated transgenes into intolerant tissues, emphasizing the need for developing mechanisms to fine tune CRISPR expression within these sites. While this could be achieved following introduction of insulating sequences such as gypsy insulators, they were beyond the scope of this work at this time.
The Goldilocks Cas9 Conundrum — many failed promoters; the killers, the sterilizers, and the idlers

Inability to insert drives onto the docking sites necessitated assay of the system’s homing capabilities ectopically by CRISPR expression from other loci. In this manner, consistent with the Vasa2 promoter’s known drive-enabling properties\textsuperscript{17,18,37} we confirmed that Vasa-Cas9 expressed gene drives in our design would cause embryonic lethality, female infertility, and high levels of male fertility (Figure 3.13, Section 3.8A), and attempts to fine-tune expression through the use of fusion proteins proved unsuccessful (Section 3.8B). Similarly, we confirmed the published drive-enabling ZPG1 promoter, known to be even more tightly regulated than Vas2, also causes complete female sterility, and high levels of infertility in males while not demonstrating biased inheritance (Figure 3.16). To identify a Cas9 promoter with dampened and less leaky expression we developed many distal gene drive lines which failed to show any drive phenotype (B2dGD\textsuperscript{pB}, ZPG2dGD\textsuperscript{pB}, VFdGD\textsuperscript{pB}, Summarized in Figure 3.17), suggesting that identifying the perfect promoter for Cas9 expression may be challenging.

The Goldilocks Cas9 Conundrum — the anti-drive

Female infertility caused by ribosome depletion suggests that the highest probability of observing drive is through males. However male infertility has also been observed in some strong \textit{Drosophila Minute} lines\textsuperscript{27}, suggesting they too may incur sterility from \textit{Rpl11} mutagenesis. In experiments carried out in driving males co-inheriting mNosGD\textsuperscript{dRPLT} and VasCas9, most broods were infertile (Figure 3.13), as were most broods attempting distal drive of dRPLTu by ZPG1dGD\textsuperscript{pB}. These data suggest that drive-enabling Cas9 expression in males also expressing \textit{Rpl11}-targeting gRNAs, results in infertility suggesting that ribosomal targeting and mutagenesis of the male primordial germline may be fundamentally incompatible with fertility without significant fine-tuning.

Taken together these data bring into focus the critically precise Cas9 expression profile required to permit drive in our system. On one end of the spectrum, we observed that Cas9 expression in later stages of
spermatogenesis permits fertility, while being too late to enable drive (B2dGD\textsuperscript{dRPL} Figure 3.12), and on the other end of the spectrum, Cas9 ‘overexpression’ causes complete infertility or death (Figure 3.13, Figure 3.16). This emphasizes the precision with which Cas9 must be tuned along a narrow spectrum – one that ranges from non-driving fertility, to driving fertility, to complete infertility. In our system, if mosquitoes express known drive-enabling Cas9, and are fertile, then we should observe drive. If we fail to observe drive in these conditions, then we can glean valuable insights about our drive designs and the assumptions we made during design. We have observed a few such events, but the results were unexpected. In these experiments, not only was drive not observed, but the opposite occurred – a bias against the inheritance of CRISPR-encoding transgenes. In broods from males with mNanosGD\textsuperscript{dRPL} rescued by VasCas9, we observed a single brood with completely nontransgenic larvae (when at least half should have inherited mNanosGD\textsuperscript{dRPL}), and in a similar brood from ZPG1dGD\textsuperscript{dRPLTu we observed a strong bias against inheritance of the CRISPR-encoding distal drive in early broods. In the escaping fertile larvae from these broods, there was no evidence of any DRAs at any of the eight gRNA binding sites, suggesting these larvae had escaped mutagenesis (as is known to occur in other drives\textsuperscript{2,18}). Interestingly, as the mosquito aged, the ability of larvae to tolerate inheritance of the CRISPR-expressing transgene increased, a pattern consistent with possible PIWI silencing of the transgene as the sire aged. These findings suggest that Cas9 expression was sufficiently fine-tuned to permit fertility in the presence of driving Cas9, but that drive did not occur. This implies that targeting \textit{Rpl11} with CRISPR in the primordial germline may be fundamentally incompatible with fertility, as only those sperm which escaped mutagenesis sired viable larvae. Importantly, because successful drive homing should enable survival of the germ cells, yet this is not observed, this suggests that repair of breaks by HDR was not the predominant pathway for this drive design. Taken in conjunction with the findings of Oberhofer et al\textsuperscript{16}, this raises the possibility that multi-gRNA drive designs may not stimulate HDR as originally hypothesized, and that end-joining mechanisms may precede or supersede HDR in these designs. While redesigning the drives to
utilize a Cas9 paired nickase strategy\textsuperscript{55} to stimulate HDR may be a possibility, it was beyond the scope of this work. Another strategy to stimulate HDR by Cas9 via fusion to geminin, which aims to localize Cas9 to post-S-phase stages of the cell cycle\textsuperscript{56} is not recommended, as drive must happen between homologs, not sister chromatids, making homing during pre-S-phase stages preferable. In all cases, these gene drives must be redesigned to stimulate robust HDR in the presence DNA-breaks from multiple gRNAs before success can be achieved.

To date, the only proposed gene drive designs designed for evolutionarily stable population replacement are those which rely on targeting haplolethal genes with multiple gRNAs to provide strong selection pressures against the evolution of DRAs\textsuperscript{1}. Therefore this work raises the possibility that such homing based gene drives for population replacement may never come to fruition if i) Cas9 cannot be tuned perfectly to permit targeting of ribosome genes in the germline while maintaining fertility and drive (or if a tolerant non-ribosomal haplolethal gene cannot be identified), and ii) if multi-gRNA designs cannot be adapted to stimulate robust homing.
Figure 3.17 | Summary Goldilocks Cas9 Conundrum. Summary of experiments attempting to demonstrate drive under differing Cas9-expression profiles.
METHODS

gRNA Design For CrIGCkid and gene drives

To design gRNAs against Rpl11 and Rpt1 genes we verified the sequence of these genes in our laboratory stock lines by PCR and sequencing of 12 individuals. For each gene we identified two gRNA pairs targeting the 3’ region of the coding sequence (for a total of 8 gRNAs) in an paired-nickase orientation using the gRNA design tools outlined in Chapter 2. All 8 gRNAs used in CrIGCkid, gene drive, and distal gene drive experiments were clonal. The gRNA pairs were designed as tandem pairs linked by a 15-25bp RNA linker predicted to minimize aberrant secondary structures following analysis in RANfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The gRNAs were ordered as gBlocks with a U6 promoter for expression (outlined in Chapter 2). The pair gRNA\textsubscript{n} and gRNA\textsubscript{o} targeted the exon 3 Rpl11 sequences [5’ AAACCGTAGATACCGATCGTCGG 3’] and [5’ TTGGACTTCTACGTTGTGCTTGG 3’] separated by a 24 bp random linker [5’ TCAGATAAAGGACACCTGT3’]. The pair gRNA\textsubscript{l} and gRNA\textsubscript{m} target the exon 4 Rpl11 sequences [5’ TCGCGTCCTCCTTGGTCAGACGG 3’] and [5’ TGGTTCCAGCAGAAGTACGACGG 3’] separated by a 28 bp random sequence [5’ TCACTCTGACCAAGTACGACGG 3’]. The pair gRNA\textsubscript{s} and gRNA\textsubscript{r} target the exon 6 sequences [5’ TCACCTTGACCGCCTCGAGG 3’] and [5’ CCAAAGTTGATCGACGG 3’] separated by a random [5’ GAGGGGATCAAGGATCGACGG 3’] linker. The pair gRNA\textsubscript{p} and gRNA\textsubscript{q} target the Rpt1 exon 6 sequences [5’ TGTCGGTCGGACGGGACATTCG 3’] and [5’ GAATTGATATGTGCCTCTCAGG 3’] separated by a 24 bp [ 5’ ACTCTGTCTGCTCTGATCG 3’] linker.

Recoding Rpl11 and Rpt1 genes

To recode the endogenous Rpl11 and Rpt1 genes the boundaries for recoding were determined to be immediately upstream of the ‘outermost’ gRNA pairs designed. Recoding was performed manually using
the codon optimization table outlined in Volohonski et al., with as many codons as possible being replaced with the next-most common codon, and with no codons of >10% frequency being used.

Recoding of Rpl11 commences 620 bp from the start codon, pre-splices (removes) the 4th intron, with 55 bp of the remaining 178 bp (31%) being recoded. Recoding was performed to leave the original PAM sequences of gRNA binding sites intact, to enable subsequent reversal drives to use the recoded sequences as secondary gRNA binding sites. From outermost to innermost, the four gRNAs targeting Rpl11 have 65%, 65%, 75% and 75% identity to the recoded version of the gene respectively. To maintain ribosome function, a 209 bp replacement 3’UTR from Rpl32 (AGAP002122) was added immediately following the stop codon. Rpl32 is has a similar regulation profile and similar mRNA expression levels to Rpl11 (https://www.vectorbase.org), while having no predictable homology with the endogenous Rpl11 3’UTR sequence.

Recoding of Rpt1 commences 1366 bp from the start codon and alters 91 bp of the remaining 253 bp (36% sequence recoding). Recoding was performed to leave the original PAM sequences of gRNA binding sites intact, to enable subsequent reversal drives to use the recoded sequences as secondary gRNA binding sites. The four gRNAs targeting Rpt1 have 65%, 60%, 65% and 50% identity to the recoded version of the gene respectively. A 3’UTR was omitted in dRPLT, however in dRPLTu a 198 bp replacement 3’UTR from Rpt3 (AGAP003008) was added immediately following the stop codon to maintain wild type RPT1 subcellular localization.

**Plasmid Construction**

All plasmids and cloning was carried out through standard molecular biological methods including standard molecular cloning, GoldenGate cloning, or Gibson cloning. Synthetic sequences were ordered as IDT gBlocks, and naturally occurring sequences were PCR amplified from genomic DNA samples. For clarity here, and due to the variety of cloning methods used, sequences added to the 5’
terminus of primers intended for restriction enzyme digest or Gibson cloning are omitted and only the sequence annealing to the intended DNA sequence for amplification is presented.

**HDRdRPLT donor**: This gene drive docking site homology donor was designed for knockin of dRPLT into *Rpl11-Rpt1* in the genome by HDR. The plasmid was constructed within the pDSAY transgenesis backbone which provides a Kanamycin resistance cassette, and a 3xP3-EYFP cassette with Sv40 terminator\textsuperscript{12}. First assembled were the homology arms, recoding and 3’UTRs. The homology arm encompassing *Rpt1* was PCR amplified from *Anopheles gambiae* genomic DNA using primers [5’ GTTGTC ACCAAAGTCTGGTGGCG 3’] and [5’ CCCTCCAGATCCGGCCAGCCGAAC 3’]. This was cloned into the AflII site on PD5AY. Next in sequence and in frame with the *Rpt1* homology fragment was cloned the codon recoding (ordered as a gBlock) with the sequence of [5’ CCGtACcCAtATtTTtAAaATcCAtGCgaGaagcATGagcGTgGAaCGtGAtATcCGGTTcGAatTGcTaGCgaGgCTcTGtCc CcAAtagcACgGGaGCTGaAAATtCgaagtGTcTGCA CGGAaGCaGgATGTTcGCTATaCGGGCTaGgaGAAaATcGcCc ACCGaAAaGAtTTTCTtGAAaGCcGTgAAATtAAatcgTAtGCgAAaTTTtcGCGACcCccGGTAcATGACcTA tAATTAG 3’]. Within this sequence, recoded DNA bases are denoted in lowercase and original wild type base pairs in capitals. In this donor plasmid, no additional 3 UTR was added after the stop codon provided in this recoding. This was followed by a unique sequence not otherwise discussed in this work. This sequence included deliberately designed and optimized gRNA binding sites for second generation gene drives capable of targeting and homing through the recoded sequences in this first generation gene drive, to increase drive effectiveness, and included a further 100 random base pairs to enable synthesis. This sequence was ordered as a gBlock with [5’ CCCAAGACCACATAAATCGAGACAGAAATCCCGCCATAAATGCCAATGGTGTTTCTGATGCGATAACTCAA GTAGATGGCCTACTCCCAAGACATTTTCGTGTTCTAAGAAGCCTCGGAGCATCCAAGACTACGACCTGGAAGGTAC GTATCCATCATATTTTTTGAAACCAGCATAAAGCTGGCATTCCCTTCCGTAAAGCGG 3’] . Following this unique sequence phiC31 attP sequences were ordered as clonal gBlocks [ 5’
GTACTGACGGACACACCGAAGCCCCGGCGGCAACCCTCAGCGGATGCCCCGGGGCTTCACGTTTTCCCAGGTCAG
AAGCGGTTTTTCGGGAATGTCGGCCAAACTGGGTAACCTTTGAGGTTCTCTCAATGGTGGGCGTAGGGTTGCGCAC
ATGACACAAGGGTTGGACTGGAGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAGCGCGA
3’] to be compatible with attB sequences on Anopheles gambiae transgenesis plasmids. Two attP sites were cloned immediately upstream and downstream of the 3xP3-EYFP-Sv40 fluorescence cassette on pDSAY to enable later RCME with gene drive transgenes onto the docking site, and the LacZ MCS, and attB sequences were removed from pDSAY in the process. Into the Ascl site on pDSAY was then cloned the RpL11 homology and recoding sequence in the native orientation to enable compatible HDR knockin with the endogenous RpL11-Rpt1 locus. The RpL11 sequence was PCR amplified from genomic DNA using the primers [5’ TGTATTGGAAAGTTACGTTACCGT 3’] and [5’ ATCGTACTTGATGCCCAGATC 3’]. This was cloned upstream of the RpL11 recoded DNA sequence ordered as a gBlock with the final intron pre-spliced [ 5’ CCGACCATTGCTGACGCTGTTACCGGCGGTGCTTACGACCGTCAGTCGCGCGAGCGCGA 3’]. Again with recoded DNA bases denoted in lowercase and original wild type basepairs in capitals. This is followed by a surrogate replacement 3’UTR from Rpl32 amplified from genomic using primers [5’ GCCCTGGTGCAAGAGAG 3’] and [5’ AAATTTGCCGTTTAAGCGTAAT 3’]. This is followed by an additional Sv40 terminator sequence in an attempt to guarantee termination. This is followed by an additional unique sequence to provide additional gRNA binding sites for next generation drives homing through the first generation drives designed here. It was ordered as a gBlock [ 5’ AACAAACGCGAGATACCAGGAGTCTGCACCGGgATaTaCAGtTCcACgctCAGCTACGCGGGAGGgATtTTaAAaATaTGg GtAcCGGGGTGCCACCTGTACGGCTCGTTGAAGCCGATTAGTACAATAGATTTATTCAACCCCAAGGTCTACACTC CCGGCTACTTGATATGTCGCGTCCCggGGgGTcGCcgaAAtTTcGTtTTCAATCGGACTTcTTaTTcAcGCT TCcAGG 3’]. This sequence immediately leads into the attP site previously cloned after the 3xP3-EYFP-Sv40 cassette discussed earlier. For a graphical plasmid map refer to Appendix A.
**IGCdRPLT^pB:** This transgenic homology donor plasmid was designed to integrate the dRPLT donor sequence into the genome and to facilitate subsequent interlocus gene conversion-mediated knockin of the dRPLT docking site into endogenous *RpL11-Rpt1*. To build this plasmid, the entire donor cassette was PCR amplified (encompassing both homology arms, recoding and fluorescence markers) from HDRdRPLT^donor using the primers [5’ GTTGCAACCAAGTCTGTTGGCG 3’] and [5’ TGGTGAAAGTACGTATCGTG 3’]. This was cloned into the pXL-BACII-LoxP-3xP3DsRed-LoxP (Addgene plasmid 26852) multiple cloning site to provide piggyBac terminal repeats for transgenesis. Outside of the homology arms the 8 gRNAs were cloned in four gBlocks with gRNAp gRNAq, gRNAr, and gRNAs upstream of the *RpL11* homology arm and gRNAI, gRNAm, gRNAo and gRNAo upstream of the *Rpt1* homology arm. Outside of the homology arms, the 3xP3 promoter originally present on the pXL-BACII-LoxP-3xP3DsRed-LoxP plasmid was swapped for an Actin5c promoter amplified from *Drosophila melanogaster* genomic DNA with primers [5’ CGCATGTGCTTGTGTGTGAG 3’] and [5’ TGTAAGCTGCAATGGAAAGAATGC 3’]. On the backbone of this plasmid was cloned a Vasa-PiggyBac transposase expression cassette composed of the Vas2 promoter amplified with [5’ CAGGTCTCACATGCGATGTAGAACGCGAGCAAA 3’] and [5’ ATTGTTTCCTTTTTATTCACCGG 3’] expressing piggyBac transposase amplified with [5’ CAGGTCTCAATGGGCTCTAGCCTGGAC 3’] and [5’ CAGGTCTCACATGTCAGAAAACGCTCTGGCAGCATG 3’].

**IGCdRPLTu^pB:** This transgenic homology donor plasmid was designed to integrate the dRPLTu donor sequence into the genome and to facilitate subsequent interlocus gene conversion-mediated knockin of the dRPLTu docking site into endogenous *RpL11-Rpt1*. This plasmid was built from the IGCdRPLT^pB donor plasmid. For this the 3xP3-EYFP selectable marker cassette on IGCdRPLT^pB was removed and replaced the a m2Tursuoise fluorescence sequence amplified from pDSAT^12 using the primers [5’ ATGGTGAGCAAGGGCGAG 3’] and [5’ TCATCCGGACTTGTACACG 3’]. Also onto this plasmid, a surrogate replacement 3’UTR for *Rpt1* was added from *Rpt3* which was amplified from genomic DNA by [5’
ACACAGCAGAAGCAG 3’] and [5’ CCATTGCTTAACTTCCG 3’] and was cloned immediately following the stop codon within Rpt1 recoding.

**B2GD:** This transgenesis plasmid is a gene drive designed to insert within the docking sites by phiC31 transgenesis, carries the 8 gRNAs used throughout this work and is enabled by the β2-tubulin promoter for Cas9 expression. To build this plasmid, first the pDSAR transgenesis backbone plasmid was modified to remove the original attB site, and two new attB sites with the sequence of [5’ tcgaCGATGTAGGTCACaGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACT CCACCTCACATCTGGCTCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCA CCGGGACGCCCTGCCCCAAACCCTGGCCGGCGCGGTGCTACGGTACACGGGACTGCGACGCGGCGA CGGTCGGATACCGGGCCCGTCACCGGGGTTCTCGACGTCACCGGGCGGCA 3’] were cloned into the NheI and BsrGI restriction enzyme sites on the pDSAR backbone. During the insertion of one attB’s an AvrII restriction enzyme site was added, into which a Vasa2-PhiC31 integrase-expressing cassette was inserted. The entire integrase-expression cassette was PCR amplified from integrase helper plasmid with primers [5’ CGATGTAGAACGCGAGCAAATTC 3’] and [5’ CTAGCAATCGGTGCAAGCTTTCAC 3’]. Next the gNRAl, gRNAm, gRNAo – containing gBlocks were cloned into the AflII enzyme site with 3’ ends oriented towards the LacZ MCS adding an SbfI site followed by reconstituting the AflII sequence on the 3’ end. The gRNAp, gRNAq, gRNAr, gRNAs- containing gBlocks were then cloned into the Ascl site oriented with 3’ ends facing the LacZ MCS to reduce the probability of plasmid deletion between the gRNA blocks instead forcing shuffling events to result in an inversion and not a deletion. The B2Cas9 expression cassette was next cloned within the SbfI and AflII sites in two different versions. A version of the β2 promoter corresponding to the one outlined in composed of 2,234 bp upstream of the start codon were used for expression. It was amplified from genomic DNA by [5’ GAAGACGTTCGGAAACGAGC 3’] and [5’ TTTTGAAACTGTGAAAGCGCG 3’], cloned upstream of Cas9 amplified from VasCas9 using primers [5’ ATCCCGATGTAGAAGCAGGCGAGC 3’] and [5’ TTACTTTTTTTTTGGCTCGG 3’], followed by a 1,500 bp β2
3’UTR amplified from genomic using primers [5’ AACTAAAGCTAAATTGAACACCC 3’] and [5’ TCCGACATTAACATCTACGTGC 3’]. This plasmid was used for transgenesis injections into dRPLT. A second version of the plasmid using a shorter β2 promoter (500 bp) was used for injections into dRPLTu based on sequences characterized in 62,63, shown to be identical in function to 29 but significantly shorter in length. This promoter sequence was amplified with primers [5’ AGCGTTCTATAATTGATATAGTTTTG 3’] and [5’ AAGCTTGATATCTTTCGAAACTGTG 3’], cloned upstream of Cas9 amplified as before, followed by a 503 bp β2 tub 3’ UTR amplified with [5’ TCCGGATGAAAGCTTAACTAAATTGAACACCC 3’] and [5’ CGATTAAAGGACCATTCC 3’].

mNosGD: This transgenesis plasmid is a gene drive designed to insert within the docking sites by phiC31 transgenesis, carries the 8 gRNAs used throughout this work and is enabled by a truncated Nanos (AGAP006098) promoter for Cas9 mRNA but not protein expression. This gene drive was cloned in the same manner as B2GD above, with the exception of enabling Cas9 expression by a Nanos promoter and 3’UTR in lieu of β2-tubulin’s. The promoter is composed of the 1,439 bp immediately upstream of the start codon and was amplified from genomic DNA by [5’ AATCGATTAACGGCAAC 3’] and [5’ CTTGCTTTCTAGAACAAGG 3’]. This was cloned upstream of eSpCas964, in an attempt to improve on-target specificity, after demonstrating its ability to mutate target sites in An. gambiae embryos following in-vitro microinjection experiments. These experiments are not otherwise discussed in this work to maintain brevity and clarity, and this is the only implementation of eSpCas9 in this work. This eSpCas9 cassette was amplified with the primers [5’ ATGGACTATAAGGACCAGCAG 3’] and [5’ CTACTTTTTCTGCTGCGGC 3’]. Immediately following, 1,052 bp corresponding to the Nanos 3’UTR was PCR amplified with [5’ TAGGACAGAGTCTGTTTCG 3’] and [5’ GTCTTCACTCGAGAAGGTAAATG 3’].

VasGD: This transgenesis plasmid is a gene drive designed to insert within the docking sites by phiC31 transgenesis, carries the 8 gRNAs used throughout this work and is enabled by the Vas2 promoter for Cas9 expression. This gene drive was cloned in the same manner as B2GD above, with the exception of enabling
Cas9 expression by a Vas2 promoter and Sv40 terminator. The endogenous Vasa 3’UTR was decided to not be used as I had previously demonstrated that the Sv40 terminator in combination with this promoter was sufficient to enable robust germline endonuclease expression\(^\text{65}\), and that the combination also robustly enabled germline Cas9 expression specifically\(^\text{66}\). Therefore the Vasa2-Cas9-Sv40 cassette was amplified as one large fragment from the VasCas9 transgene using primers [ 5’ ATCCCGATGTAGAACCGAGC 3’] and [5’ GATACATTGATGAGTGGACCAACCC 3’] cloned into the drive plasmid.

**B2dGD**\(^\text{68}\): This transgenesis plasmid is a distal gene drive designed to insert randomly throughout the genome by piggyBac transgenesis, carries the 8 gRNAs used throughout this work, and is enabled by the short β2-tubulin promoter\(^\text{62}\) and 3’UTR for Cas9 expression. This transgenic also sought to disprove fundamental B2GD lethality due to CRISPR expression alone. To construct this plasmid, first all the donor sequences between the homology arms on IGCDRPLT\(^\text{68}\) were removed, leaving the piggyBac terminal repeats, Vasa-transposase cassette, 8 gRNAs, and act-DsRed fluorescent marker. Into the gap caused by removal of the donor sequence, the B2-Cas9 expression cassette was inserted. This cassette was PCR amplified from the ‘shorter’ B2GD described above using he primers [ 5’ AGCGTTCTAAATTGATATAGTTTTG 3’] and [5’ CGATTTAAGGACCGATTCC 3’]

**VFdGD**\(^\text{68}\): This transgenesis plasmid is a distal gene drive designed to insert randomly throughout the genome by piggyBac transgenesis, carries the 8 gRNAs used throughout this work, and is enabled by the VasaFusion protein system for Cas9 expression, and was constructed in the same manner as B2dGD\(^\text{68}\). The Vasa2 promoter was first PCR amplified with primers [ 5’ ATCCCGATGTAGAACCGAGC 3’] and [5’ ATTGTTCCTTCTTTATTCCGGG 3’] directly from VasGD. Next was cloned in a synthesized Vasa-encoding pre-spliced DNA sequence prefaced by a 3xFLAG tag followed by a Sv40 NLS ordered as a gBlock with the sequence [ 5’ ATGGACTATAAGGACCACGGAGACTACAAGGATCATGATATTGCCATCAAAGACGATGACGATAAGATGGCC 3’]
TCGAGCCGGAGGAAGAGTGGGAG 3’]. This was followed by an XTEN linker demonstrated previously to enable linkage of Cas9 to other large proteins67,68 with the sequence [5’ TCGGGCAGCGAGCGCCGCCACCTCGGAGTGGGCCACCCCCCGAGTGG 3’]. This is followed by linkage to Cas9 amplified as previously described, followed by a 991 bp Vasa 3’UTR amplified from genomic by [5’ CTTGGGGTGGGGTGTATATGTG 3’] and [5’ AGAAAATGTGGGCCATTAACAGC 3’].

ZPG2dGDpB: This transgenesis plasmid is a distal gene drive designed to insert randomly throughout the genome by piggyBac transgenesis, carries the 8 gRNAs used throughout this work, and is enabled by an ~2 kB ZPG promoter sequence, and is constructed in the same manner as B2dGDpB. This promoter was previously developed in the lab for use in a sterilization rescue transgenic not otherwise discussed in this work, termed TOZR. This promoter sequence was effectively used to express the Tet-off transcriptional activator, and had been demonstrated to successfully express Tet-Off mRNA, but was never able to be validated by preliminary Western blots (not discussed). Given the lack of an alternative, verification of mRNA was considered sufficient evidence to warrant use of this germline promoter for construction of a distal gene drive. The 1,998 bp preceding the ZPG start codon were PCR amplified from genomic DNA by primers [5’ GGTGTCCAGCTTTGAATATTC 3’] and [5’ CTCGATGCTGTTGTTGGTGG 3’]. This promoter was cloned immediately upstream of the Cas9 previously discussed, followed by a 1,016 bp 3’ UTR sequence amplified by [5’ AGGACGGCGAGAAGTAATCATATG 3’] and [5’ GGGAATTCTAACCGTTAATCGATTC 3’].

ZPG1dGDpB: This transgenesis plasmid is a distal gene drive designed to insert randomly throughout the genome by piggyBac transgenesis, carries the 8 gRNAs used throughout this work, and is enabled by an ~1 kB ZPG promoter sequence previously proven to enable robust drive-enabling Cas9 expression while being more fine-tuned than Vasa2 4,31. It is constructed in the same manner as B2dGDpB. The promoter and 3’UTR sequences were ordered as clonal gBlocks to those previously published. The 1,074 bp promoter gBlock was ordered with the sequence [5’ CAGCGCTGCGGCTTGGGGACAGCTCCGGCTGGCTGTTTCTCTCCTCCTCCTCGTCG 3’].
ACCAGTTTCAGTTTGGCTGAGCGGTAAGCCTGCTGTTTCCTGCTGATCATCGGGACCATTTTGATGGGCCATCCGACC
ACCACCACCATCACCCACCGGGCTCCATTCTTAGGGGAGATACACATCCTCCGGGGCAGGACATGTGCTGCT
GCCAAGGGTGCCATTCGTTTTTGGCTGAAAGCAAAAGCAATAGTGTGTTTTCTGCTGACACGATAATT
CGTTTCTTGCCGCTAGACACAAAAACACTGCTACTGGGAGGAATATTGTCAGCCTTAACTACTACCCTCA
AAGTTATGCTCACTCGTTGGTATAATGGACCTACCGAGCCGGGTTCACACTACAAAAGCAAGATTATAGCAGA
CACAGCGAAAAACTAGTAATTTTCTATCGAAAGCGCCAAGTTAGTGGTTTGCCTGACAGAAGATTATAGC
TGCTGCGGGATAAAACCCGAGACGACCCATGTCGCCCTCGCTGCTACCCGGGATAATCGAGACG
ATGGAGTGGGTATAATTATATATCCTACATCGTAACTGTCACAACACGTGCTGAAAGCAAAAGCAATAGTA
TTATTTTGAAGAACAGCACAAGATTTTGGCTGACAAATATATTATTGTTTGGTTGTTAATT
TTACATTGTAAGAAACAGCACAAGATTTTGGCTGACAAATATATTATTGTTTGGTTGTTAATT
CTCGCTGATTATCCCTCTCTGCTCTCTCTGCTACTGCTGCTGCCTCGTCGCCCTCCGGGATAATCGAGAGTA
GCCATTTTAAAATTGACTACACACCCGAGACGACCGTGGTGAGCTCTTTCAAGTTCTTCTGGACCAAAAGAG
AGAGAATACCGCCCGGACAGTGCCCGGAGTGATCGACATAGAAAATCGCCCATCATGTGCCACTGAGGCGAAC
CGGCGTAAGCTGTTGCAGTTTCCCAGTACAATCCCGCATATAACAAACAGCCCAACCAAAATACAC
GCATCGAG 3'], and was cloned immediately upstream of Cas9 as previously described. This is followed by a 1,037 bp 3’UTR ordered as a gBlock with the sequence [ 5’
CAGCGCTGGCGGTATTGCAACTTGACATTC
TGCTGCGGGATAAACCGCGACGGGCTACCATGGCGCACCTGTCAGATGGCTGTCAAATTTGGCCCGGTTTGCGAT
ATGGAGTGGGTGAAATTATATCCCACTCGCTGATCGTGAAAATAGACACCTGAAAACAATAATTGTTGTGTTAATT
Embryonic microinjection and transgenesis

Procedures carried out essentially as described in\textsuperscript{12,14,69} with 350ng/µl of transgenesis plasmid injected and 80 ng/µl of helper plasmid injected if by PhiC31 transgenesis. Transgenesis into dRPLT or dRPLTLu were into the mixed offspring of {+/dRPLTx} intercrossing adults due to the homozygous lethality and infertility of the lines respectively as previously discussed. Due to inability to isolate heterozygous and homozygous docking site embryos from wild type siblings, all genotypes were indiscriminately aligned and injected, leading to reduced transgenesis efficiencies as wild type embryos would not be capable of transgenesis, and homozygotes would be unable to begat transgenic offspring. Following injections into the docking sites, individuals were outcrossed to the wild type G3 laboratory stock line in bulk, and nascent transgenic F1 offspring were identified by fluorescence. PiggyBac transgenesis based injections to establish the IGC\textsubscript{dRPLT}\textsuperscript{pB} and IGC\textsubscript{dRPLTu}\textsuperscript{pB} donor lines were carried out directly within wild type G3 embryos. Injections to establish the piggyBac distal gene drives were either carried out within G3 embryos, and survivors immediately outcrossed to {+/dRPLTu} to assay for distal drive immediately in the F1’s, or the reciprocal was carried out with injections occurring directly in mixed {+/dRPLTu} embryos followed by F0 survivor outcross to G3 depending on the availability of each line. Under no circumstances were injections carried out in {+/dRPLTu} embryos followed by F0 survivor outcross to {+/dRPLTu} in order to guarantee all newly established distal drive transgenics could only be heterozygous for the dRPLTu site,
guaranteeing not false positives for drive. Precise injection and crossing parameters are available upon request.

**PCR validation of dRPLT and dRPLTu insertion:**

PCR validation of docking site insertion occurred using many combinations of primers. Some amplicons were sequenced to verify the product. Primer combinations used consist of those that specifically bind within the newly introduced sequence – often recoding, a replacement UTR, an attB or a fluorescent marker – and amplify over the homology arms and into a sequence within the genome. Not all primers worked given the polymorphic nature of the reference genome to our lines, and the polymorphic nature of the 5’ regulatory sequences targeted. However all primers listed here worked to produce a valid PCR product of the predicted size in at least reaction, and not all possible combinations were tested. Primers used specific to the polymorphic genomic sequences beyond the *RpL11* homology arm include [5’ GTGGAGCGTTTTTCCACTTTGC 3’], [5’ CGGGACGTTTTATGAATGTATAACC 3’], [5’ GTAATTTGCTATGCAATTGCTTG 3’], [5’ CGGGGATCCGGTTTTAATG 3’], [5’ GCTACACAGACAAAAAGTGCATC 3’], [5’ CAATTGCTGTCGACTTGAAGGC 3’] and [5’ GCGATACCTTGCTATTGAGTGCG 3’]. PCRs specific to the integrated sequence used include those binding to the *RpL11* recoding include [5’ GGCCCCAGACCCACATAAAAATCG 3’], [5’ GAGGGCCATAATGCCAATGGTC 3’], [5’ GCGCCACATTATATCTGGACG 3’], and [5’ GCGCCACATTATATCTGGACG 3’]. Primers specific to the *Rpt1* recoding used include [5’ GATATCACGTTCCACGCTCATG 3’], [5’ GTATAGCGAACATCCCTGCTTG 3’], [5’ CAAGACGTCTCCGAATTCCCTTC 3’], and [5’ GATATCACGTTCCACGCTCATG 3’]. A Primer specific to the 3’UTR replacement for RpL11 from RpL32 includes [5’ CTCTCCTAAAACGGATCTTCTTTCTTC 3’]. Primers specific to the EYFP cassette within dRPLT include [5’ CTGAACCTTGTGGCCTTTACGTCG 3’], and a primer specific to m2 Turquoise is [5’ GTCGTCCTTGAAGAGATGGTG 3’]. Primers specific to the attP or adjacent sequences include [5’ GGGTTCGAAATCGATAAGCTTG 3’], [5’ CTTAAGCTTATCGATACGCGTGACG 3’], and [5’ CATTATTAGGCCCTGGATTCTGTGG 3’]. Primers specific to the *Rpt1* replacement 3’UTR from *Rpt3* include
[5’ CAAGACGTCTCCGAATTCCTTC 3’]. Primers specific to the genomic sequences beyond the Rpt1 homology arm includes [5’ CACACCTTTACATCGCTCGC 3’], [5’ GGCCTTCTCTGACCGATTCT 3’], and [5’ GACTGTGCAAATTTTGTACATTGG 3’].

**Larval Fluorescence Scoring**

Egg broods were laid into an oviposition cup composed of a Whatman® filter paper cone (90mm, Grade 2, Sigma-Aldrich) within a urinalysis cup containing 80 mL DI water. The small amount of water in the bottom of the cone (10 mL) collected and concentrated larvae for further analysis. Larvae were scored the afternoon of hatching, approximately 2 hours after first hatchlings emerged, and late hatchling larvae were scored the next morning. Larvae were removed from the cone of water in the base of the dish by a 1 mL pipette with the tip clipped to have an opening circumference between 1-3mm. The larvae were then further concentrated onto a sieve with a fine mesh to not injure hatchlings. Larvae were then pipetted onto Teflon-printed diagnostic slides, and immobilized on ice or on a Corning® CoolBox™ XT Cooling workstation. Larvae were then analyzed under a Leica M80 fluorescence dissecting scope for fluorescence. Notably CrIGCkid larvae analysis occurred under a wideband GFP filter enabling simultaneous visualization of the presence for intense gut-derived act-DsRED and/or 3xP3-FP fluorescence. Larvae of desire fluorescence phenotypes were then isolated from siblings by serial dilution with a clipped tip 200µL pipette until isolated, and then reared apropriately.

**Western blots:**

Western blots were carried out essentially as described in 70. Tissue samples, quantity, age, and dissection status are provided in respective figure legends. They were dissected and prepared in Np-40 cell lysis buffer system (ThermoFisher Scientific) supplemented with 1mM PMSF in DMSO and 40µ Protease Inhibitor . Western blots to detect Cas9 were either carried out with an α-FLAG tag antibody previously described70(an epitope included on all Cas9 constructs) when blots performed before G20, or directly
with an α-Cas9 antibody compatible with all Cas9’s used in this work from Cell Signaling Technologies ® (Cas9 XP ®, Rabbit mAB # 19526, 1:1000 dilution) when blotted after G20.

Westerns targeting poly-ubiquitin aggregates were carried out with 30, 1-day old, larvae of each genotype. Early replicates attempted larval treatment with the known proteasome inhibitor, Mg132, by overnight incubation of larvae in a 10mM solution in Di water. Larvae survived the night, and failed to demonstrate polyubiquitin aggregates in subsequent western. Therefore a murine tissue culture sample with known proteasome defects was used as a positive control composed of 1µg/µL C2C12 cell line treated with Bortezomib. Inspired by 25, Polyubiquitin aggregates were finally visualized following release from aggregates via treatment of the membrane with 6M GuHCl, 20mM Tris pH 7.5, 1mM PMSF and 5mM β-ME for 30 min at 4°, before blocking and antibody incubation with 1:500 solution of the Fk2 Mono- and polyubiquitinated antibody (Enzo Life Sciences, BML-PW150-0025).

Western blots targeting RpL11 were undertaken using four different primary antibodies, all of which failed to identify the anopheles homolog of RpL11: RpL11 polyclonal (PA5-27468, ThermoFisher Scientific), RpL11 polyclonal (GTX101651, GeneTex), RpL11 polyclonal (aa128-178, LS-C290580, LifeSpan BioSciences), and RpL11 Polyclonal (LS-C190025, LifeSpan BioSciences).

**Larval competition assays:**

For larval competition trays they were set up similarly. Hatchling larvae from intercrossing heterozygous {+/dRPLT} or {+/dRPLTu} parents. For dPRLT, larvae were screened 20H post hatching, and three cohorts were isolated, each containing exactly 50 wt and 50 dRPLT heterozygous larvae, with great care taken to exclude homozygotes by fluorescence and size differentiation. All three cohorts were reared according to standard husbandry protocols outlined in the MR4 manual (https://www.beiresources.org/Portals/2/VectorResources/2016%20Methods%20in%20Anopheles%20Research%20full%20manual.pdf). Larvae were reared in adjacent trays at 28° in a climate-controlled
insectary. All trays were given as similar quantities of larval food as possible. Trays were monitored each day for development of pupae. On the day of first pupae observed for that tray, it was denoted “day one of pupation”, the pupa was removed, scored for fluorescence and sex. Each day pupae were removed and scored similarly until no living pupae or larvae were present. Larval competition trays assaying dRPLTu development were carried out in an identical manner with the exception of the starting quantities of each genotype. In this case, to more closely mimic the normal distribution of larval genotype frequencies, 25 wt, 50 heterozygous, and 25 homozygous larvae were seeded into each tray at commencement of experiment. Pupae for this experiment were scored for simply the presence or absence of fluorescence, preventing quantification of dRPLTu heterozygotes vs homozygotes at this time, however daily samples were collected to enable later analysis of the differences between these two genotypes.
REFERENCES


CHAPTER 4.

A transgenic tool to assess *Anopheles* mating competitiveness in the field

Andrea L. Smidler¹, Sean N. Scott¹, Enzo Mameli¹, W. Robert Shaw¹† and Flaminia Catteruccia¹†

1 Harvard T.H. Chan School of Public Health, Department of Immunology and Infectious Diseases, Boston MA, USA

† These authors contributed equally to this work

Respective Contributions: This chapter is based off a published work performed in collaboration with colleagues. ALS, WRS, and FC conceived of the research study. ALS performed all transgene design, cloning and embryonic microinjections, and initial transgene characterization. WRS conceived and coordinated experimental design and oversaw sample preparation and collection and analyzed data. SNS performed sample preparation and performed the IFA, Western blot, and RNA timecourse experiments. EM performed electron microscopy experiments. ALS and SNS shared in line husbandry, and EM, WRS and SNS performed mating captures. FC contributed to the design of the study. ALS, WRS and FC wrote the manuscript.

ABSTRACT

Malaria parasites, transmitted by the bite of an anopheline mosquito, pose an immense public health burden on many tropical and subtropical regions. The most important malaria vectors in sub-Saharan Africa are mosquitoes of the Anopheles gambiae complex including A. gambiae sensu stricto. Given the increasing rates of insecticide resistance in these mosquitoes, alternative control strategies based on the release of genetically modified males are being evaluated to stop transmission by these disease vectors. These strategies rely on the mating competitiveness of release males, however currently there is no method to determine male mating success without sacrificing the female. Here we exploit the male An. gambiae reproductive characteristic of transferring seminal secretions, made in the male accessory glands (MAGs), into the female atrium during mating to generate a transgenic system capable of monitoring copula pairs. We develop and validate the use of a MAG-specific promoter to fluorescently label the mating plug and visualize the occurrence of insemination in vivo. We used the promoter region of the major mating plug protein, Plugin, to control the expression of a Plugin-tdTomato (PluTo) fusion protein, hypothesizing that this fusion protein could be incorporated into the plug for sexual transfer to the female. A. gambiae PluTo transgenic males showed strong red fluorescence specifically in the MAGs and with a pattern closely matching endogenous Plugin expression. Moreover, the fusion protein was integrated into the mating plug and transferred to the female atrium during mating where it could be visualized microscopically in vivo without sacrificing the female. PluTo males were equally as competitive at mating as wild type males, and females mated to these males did not show any reduction in reproductive fitness. The validation of the first MAG-specific promoter in transgenic A. gambiae facilitates the live detection of successful insemination hours after copulations has occurred. This provides a valuable tool for the assessment of male mating competitiveness not only in laboratory experiments but also in semi-field and field studies aimed at testing the feasibility of releasing genetically modified mosquitoes for disease control.
INTRODUCTION

Targeting fertility for insect control has been successful in a variety of species, and a number of strategies aimed at inducing sterility in field populations are currently being developed in mosquitoes. Sterile Insect Technique (SIT), a strategy based on the mass release of irradiated males to sterilize females upon mating \(^7\), has been successfully implemented against populations of the screwworm Cochliomyia hominivorax and the Mediterranean fruit fly Ceratitis capitata to reduce the economic burden of these important agricultural pests \(^8\)-\(^10\). Adapting SIT and other similar sterilizing technologies (such as chemosterilization) to *Anopheles* vectors of human malaria has been challenging, and requires optimization to maximize competitiveness of colonized, mass-reared and sterilized males \(^11\)-\(^15\). To circumvent the potential problems linked to sterility by irradiation or chemosterilants, strategies based on genetic manipulation of insect fertility are being considered. Among the most promising strategies, gene drives are in development for *Anopheles* which could spread infertility through natural mosquito populations by positively biasing their own inheritance. The potential of gene drives to induce sterility in field populations has been recently demonstrated in cages of *A. gambiae*\(^16\), where gene drive transgenes targeting female fertility genes spread for a few generations\(^17\). Regardless of their mode of action, all genetic control strategies require the release of male mosquitoes that have high mating competitiveness and can successfully mate with field females, making studies into male reproductive fitness a critical requisite for successful implementation.

*A. gambiae* are largely monandrous, which means they mate only once in their lifetime. During copulation seminal secretions produced by the male accessory glands (MAGs) are transferred to the female atrium in the form of a coagulated mating plug \(^18\). The plug is composed of seminal secretions including proteins \(^19\) and the steroid hormone 20-hydroxyecdysone (20E) \(^20\), and is digested by the female over a period of 24 h following copulation \(^21\). The loss of female receptivity to further mating is at least partially due to the
transfer and function of the steroid hormone 20E\textsuperscript{22-24}, which also triggers oviposition in blood fed females and affects other important aspects of the female post-mating physiology, including egg development and fertility\textsuperscript{22,23,25}. One of the most abundant mating plug proteins is Plugin, a MAG-specific glutamine-rich structural protein that becomes incorporated into the plug upon cross-linking to itself and other seminal proteins by the action of the plug-forming transglutaminase AgTG3\textsuperscript{21,26}. Transcriptional silencing of AgTG3 partially prevents plug formation and transfer, which results in severe sperm storage defects in the female, causing infertility\textsuperscript{21}. Other less characterized mating plug factors that may play a role in female post-mating physiology include short accessory gland proteins (Acps), a number of proteases and peptidases\textsuperscript{27-29}, serine protease inhibitors which play a role in mammalian fertility\textsuperscript{30}, and cysteine-rich secretory folding proteins (CRISPs) that are involved in gamete interactions\textsuperscript{31}. While many of these proteins are characterized in other organisms, their role in male mating fitness has not yet been elucidated in A. gambiae, making development of novel MAG-specific tools important for addressing these outstanding questions. Studies on the evolutionary trajectory of mating plug formation have shown that transfer of a mating plug is a derived trait not limited to A. gambiae but widely conserved across anopheline species from Africa and South East Asia, while being absent in males of Central American species like A. albimanus\textsuperscript{24}.

The presence of a mating plug within the female atrium following mating is considered a de facto marker of successful copulation given its relevance for sperm storage\textsuperscript{21,22,32}. To facilitate measuring male mating success in vivo, we developed a novel transgenic line, Plugin-tdTomato (PluTo), for use as a marker for the occurrence of insemination in females. PluTo males express a Plugin-tdTomato fusion protein specifically in their MAGs via the Plugin promoter, without negatively affecting endogenous Plugin levels. We show that Plugin-tdTomato is incorporated into the mating plug and is transferred to the female during mating, where it is detectable by microscopic examination after copulation. Importantly, mating experiments in competition with wild type males suggest that male mating competitiveness is not affected by expression
of this transgene. Moreover, transfer of this fusion protein does not affect the reproductive fitness of females mated to PluTo males. These data demonstrate that this transgenic construct could be useful for determining the mating competitiveness of *A. gambiae* strains, as well as of other anopheline species that transfer a mating plug, in semi-field and field conditions.

RESULTS

4.1 Generation and characterization of the plugin-tdTomato transgenic line.

Plugin is specifically and abundantly expressed within the MAGs\(^{21,33}\). During mating, this seminal protein is transferred to females as part of the mating plug, a gelatinous structure that is composed of multiple other proteins as well as the steroid hormone 20E\(^{21}\). To create a transgenic line exhibiting robust MAG-specific expression, we cloned a region comprising 2,688 bp to drive expression of *tdTomato*\(^{34}\) upstream of the *Plugin* start codon, likely encompassing the *Plugin* promoter. We then used this region fused to a *Plugin* cDNA construct\(^{21}\) to achieve incorporation into the mating plug for transfer to females. The two protein-coding regions of the fusion construct were connected by a 6-serine residue linker (reviewed in \(^{35}\)) (Figure 4.1A). The transgenic cassette was cloned into the pDSAY transgenesis plasmid\(^{36}\) and was injected into *A. gambiae* embryos from the X1 docking line which contains an attP site for φC31-mediated integration site on chromosome 2L\(^{36}\), in conjunction with an integrase-expressing helper plasmid\(^{36}\). A single transgenic line was generated and named *Plugin-Tomato* (PluTo). Transgene insertion into the X1 docking site was confirmed by PCR.
Figure 4.1 | PluTo transgene design and male phenotype. [A] The PluTo transgene is composed of 2688 bp of the 5’ regulatory region immediately upstream of the start codon of AGAP009368 followed by Plugin cDNA fused to a 6-serine linker followed by tdTomato., with 3xP3-EYFP for selection. Each coding sequence has an SV40 terminator to prevent transcriptional read-through. Following insertion, the flanking recombination sites attL and attR are generated in the genome [B] Adult PluTo transgenic males display strong tdTomato fluorescence in their MAGs (lower panels). The fluorescence is clearly and easily identifiable under a fluorescent dissecting microscope through the adult male cuticle (upper panels). Scale bar – 100 µm.

This transgenic line displayed strong tdTomato fluorescence in the MAGs, which was visible through the male cuticle beginning at the early adult stage (Figure 4.1B). No discernable tdTomato fluorescence was observed in other male or female tissues. In agreement with our microscopic analysis, qRT-PCR demonstrated that the Plugin promoter was capable of robust Plugin-tdTomato expression in the MAGs comparable to endogenous Plugin expression, with negligible transcripts observed in the male rest-of-body and female whole-body tissues (Figure 4.2A). Moreover, both endogenous Plugin and Plugin-tdTomato fusion proteins were observed exclusively in the MAGs at ~80kDA and ~150kDA respectively, using antibodies targeting Plugin and tdTomato (Figure 4.2B, bands 2 and 4 respectively)\(^2\). Several tdTomato-specific bands were observed not recognized by α-Plugin, suggesting of cleavage of the fusion protein or use of an alternative start codon after the antibody binding site (Figure 4.2B, band 3).
Figure 4.2 | Endogenous Plugin and PluTo transgene expression. [A] Quantitative qRT-PCR of endogenous Plugin and Plugin-tdTomato expression levels in male and female PluTo and G3 pupae and adult tissues over time. An age of zero days refers to uneclosed pupae (not dissected) and ages 1 and 2 are minimal ages in days post-eclosion (MAGs were dissected out of males). In MAGs, transgenic Plugin-tdTomato was induced similarly to endogenous Plugin levels, and age significantly affected expression. Figure 4.2 (Continued) In male rest of body tissues, transgenic Plugin-tdTomato was expressed at negligible, although higher, levels at ages 1 and 2 days. In female whole body tissues, transgenic Plugin-tdTomato was also expressed at negligible levels. Endogenous Plugin levels could not be reliably detected in wild type females aged 1 or 2 days. Analysis of variance followed by post-hoc testing within ages showed the following significant differences: MAGs (left panel): ANOVA (Age F(2,15) = 16.1, p = 0.0002, Genotype-Primer combination F(2,15) = 2.7, p = 0.099, Interaction F(4,15) = 1.92, p = 0.160). Tukey’s post-hoc test: Age 1 – PluTo Plugin-tdTomato vs G3 Plugin (mean difference ± SEM = 43.3 ± 16.63), adj. p value = 0.0359. Male Rest of body (center panel) ANOVA (Genotype-Primer combination F(2,11) = 12.3, p = 0.0015, Age F(1,11) = 1.99, p = 0.186, Interaction F(2,11) = 0.53, p = 0.601). Tukey’s post-hoc tests: Age 1 – PluTo Plugin-tdTomato vs PluTo Plugin (mean difference ± SEM = 0.021 ± 0.005), adj. p value = 0.0067; Age 2 – PluTo Plugin-tdTomato vs PluTo Plugin (mean difference ± SEM = 0.015 ± 0.005), adj. p value = 0.0406; Age 2 – PluTo Plugin-tdTomato vs G3 Plugin (mean difference ± SEM = 0.015 ± 0.005), adj. p value = 0.037. The downward error bar for Age 1 G3 Plugin cannot be plotted on a logarithmic axis as it extends below 0. Female Rest of body (right panel): ANOVA (Genotype-Primer combination F(2,16) = 4.88, p = 0.022, Age F(2,16) = 0.638, p = 0.541, Interaction F(4,16) = 0.6, p = 0.668). Tukey’s post-hoc tests: no pairwise comparisons were detected with adj. p value < 0.05. [B] Western blot of endogenous Plugin and Plugin-tdTomato in adult male and female tissues with α-Plugin in green, and α-tdTomato in red. Endogenous Plugin is observed at similar levels in the MAGs of both wild type G3 and PluTo transgenics at ~80 kDa (band 2). The α-Plugin antibody also recognizes an additional band at 27 kDa (band 1), which may be non-specific or cleaved protein. Full length transgenic Plugin-tdTomato is observed at ~150 kDa recognized by both α-Plugin antibody and α-tdTomato (band 4). An additional Plugin-tdTomato species at 120 kDa is recognized exclusively by the α-tdTomato antibody (band 3), which may represent cleaved protein, or result from an alternative start codon excluding the α-Plugin antibody binding site. Multiple high molecular weight bands (bands 6) are observed as dimers and multimers of Plugin and Plugin-tdTomato species in wild type G3 and PluTo males, respectively. Detectable Plugin and Plugin-tdTomato protein are not observed in male rest-of-body nor female whole-body tissues. The α-tdTomato antibody recognizes unspecific low molecular weight moieties in these tissues (bands 5).
4.2 Plugin-tdTomato localizes to vesicles within the anterior compartment of the MAGs

To determine the expression of Plugin-tdTomato in more detail and compare it to endogenous Plugin, we first characterized the localization of Plugin in the MAGs of wild type mosquitoes by cryo-immune electron microscopy (IEM) using an α-Plugin antibody. In Anopheles, the MAGs are composed of a thin epithelial sheath encasing two secretory compartments, a large anterior compartment and a smaller posterior compartment, each containing a single layer of holocrine secretory cells that produce MAG substances of different electron densities. This global organization was observed in our IEM analysis (Figure 4.3A), in which we detected multiple Plugin-positive electron-dense vesicles in cells of the anterior compartment only (Figure 4.3B–E).

Figure 4.3 | Endogenous Plugin localization in the MAGs. [A] Transmission electron micrograph of the boundary between anterior (ant) and posterior (post) MAG compartments in wild type males. Red box indicates area magnified in panel B, white box indicates area magnified in panel D. Inset image provided for orientation within MAGs and approximate area magnified. Scale bar 5 µm. [B] The anterior compartment has peripherally located nuclei (nuc) and contains electron-dense vesicles that fill the lumen (lum). The central aedeagus (aed) is marked for orientation. White box indicates area magnified in C. Scale bar 2 µm. [C] Plugin is localized within the electron-dense secretory vesicles by gold-labelled α-Plugin (white arrowheads). Scale bar 0.5 µm. [D] Vesicles within the anterior (ant) and posterior (post) compartments are different in size and electron density, suggesting different contents. White box indicates area magnified in E. Scale bar 2 µm. [E] Plugin is localized only within the more electron-dense secretory vesicles of the anterior compartment by gold-labelled α-Plugin (white arrowheads). Scale bar 0.5 µm.
We then performed live fluorescence microscopy and immunofluorescence analysis (IFA) using the α-Plugin antibody on PluTo MAGs. Plugin-tdTomato was localized to vesicles within the anterior MAG compartment (Figure 4.4A, B), similar to Plugin in the cryo-EM. Vesicles remained fluorescent as they moved into the interior of the anterior compartment, away from the epithelium. In the IFA, Plugin was localized to channels formed by a muscle network on the outside of the MAGs (Figure 4.4A, B), as described previously. It was not detected in vesicles in the interior of the glands, except when MAG epithelium integrity was disrupted (Figure 4.4B), suggesting incomplete penetration of the Plugin antibody, and/or the absence of regulatory elements affecting the precise localization pattern of Plugin-tdTomato within the anterior compartment. No Plugin-tdTomato or Plugin staining was detected in the posterior compartment or in the aedeagus.

Figure 4.4 | Plugin-tdTomato localization in the MAGs and mating plug. In all panels, Plugin-tdTomato fluorescence is shown in red, Plugin in green, and DNA stained with DAPI in blue. [A] Plugin-tdTomato-filled vesicles occurred throughout the secretory cells of the anterior MAG compartment (ant) but was not detected in the posterior compartment (post) or aedeagus (aed). Stacked Z-projection of the top 3.5 µm. The box in the merge panel represents the region shown in B. Scale bar – 100 µm. [B] Close-up of Plugin and Plugin-tdTomato where MAG epithelial integrity has been disrupted manually. Some Plugin-tdTomato vesicles are co-stained with Plugin (arrowheads). Stacked Z-projection of the top 3.5 µm. Scale bar – 20 µm. [C] The female reproductive tract, consisting of the ovaries (Ov), atrium (At), and the sperm storage organ spermatheca (Sp) immediately after mating. The red fluorescent mating plug is visible within the atrium. Scale bar – 100 µm. [D] A single Z-slice of the mating plug within the female atrium immediately post-mating. Plugin-tdTomato is visible throughout the mating plug, except the tip. Scale bar – 50 µm.
4.3 Plugin-tdTomato is sexually transferred to the female atrium in the mating plug

To determine if Plugin-tdTomato was incorporated into the plug and transferred to females, we performed forced-mating assays \(^{22,38}\) with PluTo males and wild type females, immediately followed by fluorescence imaging. We observed significant tdTomato fluorescence highlighting the characteristic structure of the mating plug within the female atrium (Figure 4.4C). Dissection of the female reproductive tract immediately after mating identified a strong fluorescent signal specifically along the length of the mating plug, except the proximal tip, demonstrating that the fusion protein had been incorporated in the plug and transferred during copulation (Figure 4.4D).

4.4 Testing PluTo males for in vivo assessment of successful mating

We next performed a time course analysis to determine whether sexual transfer of a mating plug containing the Plugin-tdTomato fusion could be reliably detected without sacrificing the female. Females mated to PluTo males were analyzed for mating plug fluorescence through the female cuticle at different time points after mating. As previously determined \(^{39}\), the plug shows naturally-occurring autofluorescence under a filter for green fluorescent protein, but this signal, likely derived from plug lipids, is only reliably detected for 2 hours post mating (Figure 4.5). However, in females mated to PluTo males, strong, sharp fluorescence was clearly visible under a red filter for at least 8 hours after mating, demonstrating that this fusion protein can be used to reliably determine the occurrence of successful insemination in a non-invasive manner for a significantly longer span of time (Figure 4.5). We did not systematically test later time points as the plug becomes digested and difficult to detect reliably by fluorescence.
4.5 Expression of the plugin-tdTomato fusion protein does not affect the mating competitiveness of PluTo males and does not impair female reproductive fitness

Given our ability to detect a fluorescent mating plug in mated female for at least 8 hours after copulation, we reasoned that the Plugin-tdTomato fusion protein could be used to assay male mating competitiveness, an important aspect determining the potential success of genetic control programs based on male releases. We therefore assessed whether expression of the transgene affected the mating fitness of PluTo males by performing mating capture assays where 100 PluTo and 100 wild type G3 males were released in cages with 100 wild type virgin females, and mating couples were captured while in copula. Fluorescence analysis of 73 couples collected across 4 replicate cages showed no difference in the mating competitiveness of the two groups of males (p = 0.3492, two-tailed binomial test), with 44%
females inseminated by PluTo individuals. Analysis of the remaining females that were not caught in copula but were left in the cage with males confirmed this result, with 33 of 69 mated females (48%) showing the PluTo transgene in sperm stored in their spermathecae ($p = 0.810$, two-tailed binomial test). Moreover, in separate experiments, females mated to PluTo males had comparable reproductive fitness as females mated to wild type G3 males, measured as the number of eggs developed after blood feeding, and the rates of oviposition and infertility of the broods (Figure 4.6A–C). Taken together, these results show that PluTo males induce normal post-mating behavior in females with which they mate, suggesting that Plugin-tdTomato may be a suitable construct for evaluating fitness in mosquito lines for semi-field and field releases.

Figure 4.6 | Reproductive fitness of females mated to PluTo males. [A–C] Wild type females mated to either wild type or PluTo males were scored for eggs developed [A], egg laying behavior [B], and the rate of infertility within broods [C]. In A and C, each dot indicates the brood of an individual female. No significant differences were observed between females in the two groups (A: Student’s t-test on square root-transformed data, $p = 0.38$; B: Fisher’s exact test, $p = 0.189$; C: Mann Whitney U test $p = 0.187$).
DISCUSSION

The control of the *Anopheles* mosquito via the use of LLINs or IRS is key to malaria control strategies and has contributed considerably to the reduction in number of cases and deaths witnessed since the beginning of the century. However, the effectiveness of our best vector control tools is jeopardized by the rapid insurgence and spread of insecticide resistance in *Anopheles* populations, incentivizing the generation of novel methods of genetic control. These methods rely on the release of males that have been modified to express genetic traits that either induce sterility in their female mates, or confer resistance to *Plasmodium* parasites that cause malaria. Some past attempts to release chemical- and radiation- sterilized anophelines reportedly reduced male mating competitiveness, which limits the ability of males to sterilize females or to transmit desired genetic traits. While release of sterile males works effectively in other insects, the intricate mating biology of many anophelines including the major Afrotropical vectors *A. gambiae, A. coluzzii, A. arabiensis* and *A. funestus* will make developing vector control programs relying on male release difficult in these species. These species indeed are characterized by copulations in mating swarms, which are formed every night at dusk by large groups of males. Given the highly skewed sex ratio within these swarms, competition between males for the many fewer females is fierce and a number of factors ranging from the desired genetic manipulations induced in the laboratory to the colonization process and mass rearing, can negatively impact male mating success. Determining the mating competitiveness of males prior to a release is therefore essential for the success of future genetic control efforts.

Currently there are no straightforward methods to determine the mating fitness of released males without sacrificing the female. Available methods necessitate dissections of females shortly after copulation to assess the presence of sperm in the spermatheca and the mating plug in the atrium. While the mating plug is slightly autofluorescent, we show here that this signal can only be reliably detected for
a short period of time after copulation. To achieve non-invasive determination of male mating success in competition assays, the only option at present is to blood feed mated females, allow them to oviposit, and screen their progeny for specific genetic traits of the relevant male to determine progeny paternity, while keeping the males alive and in isolation if needed. Our study provides an easy, non-invasive and effective method to assess male mating success in a short period of time and without the need to kill the female. As an example, mating plug-labeled males could be released in a swarm (whether natural or in semi-field cages), and after capturing mating couples, females could be fluorescently analyzed for the occurrence of insemination. Although this method relies on microscopes equipped with fluorescence, which are not likely to be available in remote field settings, the possibility to detect plugs for 8 hours after copulation would allow identification of transgenic matings back in the laboratory. Moreover, although in this study we only tested live females after mating, it is possible that fluorescence will also persist in dead females, such as those captured in light traps, as the fluorescent signal of transgenic lines can generally be observed for a few hours after death.

The Plugin upstream regulatory region used in this study was sufficient to achieve MAG-specific expression of our transgene. Such a promoter will facilitate the study of seminal secretions produced by the male glands, for instance via the generation of transgenic RNA interference (RNAi) lines stably expressing double-stranded RNAs (dsRNAs) against targets of interest. Expressing dsRNA transgenes will help to circumvent the limited silencing efficiency achieved by transient injections of dsRNA molecules targeting genes expressed in the MAGs 21,22, allowing detailed functional analyses.

Finally, by fusing a fluorescent protein to the coding region of Plugin, we ensured incorporation of the marker into the mating plug and its transfer to the female. This system therefore allows for efficient transfer of desired factors from the male to the female, and their slow release once in the female atrium. Interestingly, this property could be utilized to deliver female-specific sterilizing compounds or other factors that may reduce the reproductive fitness of mated females. Although such factors are not readily
available, future optimization of the system may afford specific activation of sterilants or toxins only after delivery to the female reproductive tract.

CONCLUSIONS

In conclusion, this study generates a valuable tool for assessing the mating competitiveness of males in semi-field and field studies as well as for laboratory experiments aimed at determining the function of MAG seminal secretions. This tool may prove invaluable when testing the feasibility of releasing genetically modified mosquitoes for the control of malaria-transmitting Anopheles populations. Importantly, given the observation that males from the most important Afrotropical anopheline species transfer mating plugs during copulation 24,44, this system will be readily transferable to other species like A. arabiensis and A. funestus, key vectors for Plasmodium falciparum transmission in sub-Saharan Africa.

METHODS

Plasmid construction

To clone the PluTo transgene, we separately cloned the promoter, Plugin coding sequence (AGAP009368), and tdTomato with 6S linker into the pDSAY transgenesis vector backbone 36 by standard Golden Gate cloning methods 45. We amplified 2,688 bp of the region upstream of the Plugin start codon from genomic DNA isolated from wild type A. gambiae (G3 strain) by PCR using GGpP FWD [5’ CAGGTCTCAATCCTTGTAGGGCTTGTTGACGGG 3’] and GGpP REV [5’ CAGGTCTCATCATGTCTACGGTTGAATCAGTGATACAAGCAAA 3’] primers to provide a Golden Gate compatible overhang for the pDSAY destination vector. We amplified the Plugin coding sequence from wild type G3 male abdominal cDNA extracts by PCR using GGcP FWD [5’
CCTCTCAATGAAGGCTTTGGTAGCTCTGCTCTG 3’] and GGcP REV [5’ CAGGTCTCACCTTCGACGACCAGCACA3’] to remove the stop codon, and fuse seamlessly to the Plugin promoter fragment. tdTomato was PCR amplified from a plasmid provided by the Tsien Lab 34 using GGtdT FWD [5’ CAGGTCTCATCCAGCTCCTCCTCCATGGTGAGCAAGGGCGAGG 3’] and GGtdT REV [5’ CAGGTCTCAAAGCTTACTTGTACAGCTCGTCCATGCC3’] primers to add a 6-serine linker and provide a Golden Gate compatible overhang to facilitate cloning into pDSAY. These fragments were cloned into the pDSAY φC31 transgenesis vector backbone 36, which provides an SV40 terminator at the 3’ end of Plugin-tdTomato as well as a selectable marker cassette (3xP3-EYFP) to detect transgenic individuals (Figure 1A).

**Transgenesis and PCR confirmation**

Embryo microinjections were performed essentially as described 46,47 with additional co-injection of a Vasa-φC31 integrase helper plasmid 36 (350 ng/µl and 80 ng/µl of transgenesis and helper plasmids, respectively). A total of 996 X1 docking line embryos were injected, with 172 embryos surviving to hatching. Following outcrossing of the injected P0 individuals to wild type G3, a single F1 transgenic PluTo male was isolated and outcrossed to G3 females. F2 transgenic heterozygotes were intercrossed to yield F3 homozygotes. Homozygotes were identified by strong YFP fluorescence intensity and isolated. Transgene insertion into X1 was confirmed by PCR on three distinct PluTo males and three unintegrated X1 docking line males using the NEB Q5® High-Fidelity 2 X Master Mix kit. Whole body mosquitoes were collected and DNA prepared using the Qiagen DNeasy Blood & Tissue Kit. Transgene insertion was confirmed by PCR with EYFP REV [5’GTCGTCCTTGAAGAAGATGGTG 3’] and X1 FWD [5’ AGGGAAGATTGGAATCCATC 3’]. Control PCRs for the empty and unintegrated docking site and S7 controls were performed with X1 FWD [5’ AGGGAAGATTGGAATCCATC 3’], X1 REFV [5’ ACTGCAACCCATTCACTG 3’], and S7 FWD [5’ GGCATCATCATCTACGTGC 3’], S7 REV [5’ GTAGCTGCTGAAAATTCGG 3’], respectively.
Mosquito maintenance

*A. gambiae* mosquitoes (wild type G3 and PluTo transgenic strains) were reared in cages at 28 °C and 70 % humidity on a 12 h/12 h cycle. Adult mosquitoes were fed a 10 % (wt/vol) glucose solution *ad libitum* and were given a human blood meal once a week for the purposes of line maintenance. For colony cages, adults were kept in mixed sex cages for up to two weeks post eclosion and one-week post egg-lay. For experiments, males and females were separated as pupae to ensure the virgin status of females and kept in separate cages as experimental conditions demanded. Mating couples were collected either *in copula*, or through forced mating assays (protocol available on https://www.beiresources.org/Publications/MethodsinAnophelesResearch.aspx) and mating was confirmed via mating plug autofluorescence or quantitative PCR (below).

Tissue sample collection for qRT-PCR

Male and female G3 and PluTo mosquitoes were sexed and separated into four cages. From each of these cages, sample groups were sacrificed at 24 h intervals from pupation (d0) until two days post eclosion (d2). To test whether Plugin or Plugin-tdTomato expression is constrained to the male accessory glands (MAGs), MAG samples (n = 10) were collected from eclosed males and compared to the rest-of-body (ROB) (n = 5) while female eclosed mosquitoes were collected as whole body samples (n = 5). ROB samples were collected in 200 μL of RNAlater® (Sigma-Aldrich), while MAG samples were transferred to 15 μL of RNAlater using a needle. All samples were stored at –20 °C until specimen collection was completed. Three replicates were performed.

RNA extraction and cDNA synthesis

ROB samples were defrosted on ice, RNAlater was removed and replaced with 200 μL TRI Reagent® (Applied Biosystems), while 185 μL TRI reagent was added to MAG samples to dilute out the RNA-later and avoid tissue loss. RNA was extracted according to the manufacturer’s instructions. RNA was quantified
by using a NanoDrop Spectrophotometer (Thermo Scientific) and DNase treated using Turbo DNase (Thermo-Fisher). cDNA synthesis was performed as in \(^{48}\). Approximately 4 µg of RNA was used in 100 µL cDNA reactions. Reactions were diluted to 200 µL with nuclease-free water for storage at –20 °C.

**Quantitative RT-PCR**

Samples for quantitative RT-PCR were diluted tenfold and quantified in triplicate using standard curves. PCRs were run in Fast SYBR Green Master Mix (Thermo-Fisher) on a Step One Plus thermocycler (Applied Biosystems). Endogenous Plugin cDNA was amplified using 9368_FWD [5’ TGATTCAACCGTAGACATGAAGG 3’] and 9368_REV [5’ CCACCATAACAACGGACACGAC 3’] primers. Transgenic Plugin-tdTomato was amplified using qPluginTOM_F [5’ ATCTCAACAGGAGCCCAATG 3’] and qPluginTOM_R [5’ CCCTTGCTCACCATGGAG 3’]. Quantities were normalized against the ribosomal protein RpL19 using previously described primers \(^{48}\).

**Western blotting**

Male accessory glands (n = 10) and ROB or whole body (n = 5) samples from 7-day-old PluTo and G3 age-matched males and females were collected by dissection on ice. Samples were homogenized on ice in protein extraction buffer (25 mM Tris HCl pH 7.4, 150 mM NaCl, 0.1 % SDS, 1 % Triton X-100, 10 mM EDTA pH 8, 1 X protease inhibitor, 1 % phosphatase inhibitor) and quantified using a Bradford assay. Protein samples were run at 150 V and 250 mA for 1.5 h on an XCell SureLock vertical system in a 4–12 % Bis-Tris NuPAGE gel (Thermo-Fisher). Protein was then transferred to PVDF using iBLOT2 transfer system (Thermo-Fisher). After transfer, membrane was washed twice in 1 X PBS-T (0.1 % TWEEN in 1 X PBS) and blocked in an automated shaker for 1 h at room temperature in Odyssey® Blocking Buffer (PBS) (Li-cor). Membrane was then stained with goat α-tdTomato (Sicgen) and rabbit α-Plugin (GenScript Corp; \(^{21}\)) (1:1000 concentrations for both antibodies) on an automated shaker at 4 °C overnight. Membrane was washed with PBS-T, and stained with secondary antibodies (donkey α-rabbit 800 and donkey α-goat 680 (Li-cor),
both at 1:10,000) in blocking buffer with 0.01 % SDS for 1 h at room temperature. Membrane was washed thoroughly before imaging with the Odyssey® CLx imager. Images were analyzed in StudioImageLite (Li-cor).

**Cryo-immune Electron Microscopy**

Virgin 4-day-old male reproductive tracts (including the last three abdominal segments) were fixed overnight at 4 °C (4 % paraformaldehyde, 0.1 % glutaraldehyde in 0.1 M sodium phosphate buffer at pH 7.4). Samples were washed with 1 X PBS, and infiltrated with 2.3 M sucrose in 1 X PBS, 0.2 M glycine for 15 min. Tissues were mounted on a pin and frozen in liquid nitrogen prior to sectioning. Samples were sectioned by cryo-Ultra Microtome (Reichert-Jung, Ultracut)(-120 °C) to obtain thick sections (0.5 μm) for visual screening with toluidine blue, or ultrathin sections (70–80 nm) for target tissue analysis. Ultrathin sections were transferred to carbon-coated copper grids, blocked in 1 X PBS, 1 % BSA for 10 min and stained 30 min with primary antibody at RT (1:30 dilutions of rabbit α-Plugin). The grids were washed in 1 X PBS for 15 min and then labelled with Protein-A gold (Sigma-Aldrich, 15 nm diameter) diluted in blocking solution for 20 min, following a final wash in distilled water for 15 min. Contrasting of the labelled grids was carried out on ice in 0.3 % uranyl acetate in 2 % methyl cellulose for 10 min. All micrographs were captured using a JEOL 1200EX 80 kV electron microscope and recorded with an AMT 2k CCD camera. Three biological replicates were performed, and a representative selection are shown in the text. Samples were prepared by the Harvard Medical School EM core facility.

**Fluorescence imaging**

MAGs were collected from 5–7 day-old virgin PluTo males and fixed in 4 % paraformaldehyde solution for 1 h, then rinsed 1 X PBS. Tissues were soaked in 80 % ethanol at 4 °C for 3 min to remove lipids, rinsed 3 times in 1 X PBS, bleached in 3 % hydrogen peroxide for 5 min to quench endogenous autofluorescence, and rinsed in 1 X PBS 3 times. Samples were blocked and permeabilized overnight in blocking solution (1
% BSA, 0.3 % Tween-20 in 1 X PBS), then stained with rabbit α-Plugin (1:200) for approximately 12 h. Tissues were washed 5 times in 1 X PBS followed by a 30 min incubation in blocking solution. Tissues were then stained for 1 h at room temperature with secondary antibody (Alexafluor 647 Donkey anti-Rabbit (1:1000) in blocking solution, rinsed in blocking solution, then stained with DAPI (5 µg/ml) in blocking solution for 10 min. Tissues were washed at room temperature 5 times with blocking solution, once with PBS-T, and twice with 1 X PBS. Tissues were then mounted in Vectashield® Mounting Medium (Vector Laboratories) and imaged using an upright SPE confocal microscope (Harvard NeuroDiscovery Center Leica SPE). Live unfixed samples were dissected on ice in 1 X PBS and imaged immediately on a Leica dissecting microscope with fluorescence.

Mating plug digestion and reproductive fitness assays

Four-day-old virgin G3 females were forced mated to four-day-old PluTo or G3 males and screened for mating plugs. Females without a visible mating plug were excluded from the study. Females force mated to PluTo males were then further divided into 4 groups of 10 individuals and each group imaged under fluorescent lighting at t = 0.5, 1, 1.5, 2, 4, 6, and 8 h post mating. After imaging, females were blood fed, placed in isolated oviposition cups and given 4 d to lay eggs, after which total numbers of eggs were tallied and infertility was scored.

Mating Competitiveness Assay

Reproductive behavioral effects on male mating efficacy resulting from our transgene were evaluated using a mating competitiveness assay where 100 G3 and 100 PluTo males (4–5-day-old) were allowed to compete and mate freely for 100 G3 females (4-day-old) in an open cage environment. Couples were caught in copula for 1 h as previously described 48, and male genotype determined by fluorescence of the PluTo transgene. Uncaptured females were left in the cage for 24 h, after which females were collected and individually processed for DNA by brief homogenization in extraction buffer (0.12 % Tris-Cl, 0.037 %
EDTA, 0.29 % NaCl). Mating status and male genotype were determined using a modified qPCR protocol using both Y-specific primers YQPCR_FWD [5’ GGATCTGGCCAAGAGGAGTA 3’], YQPCR_REV [5’ CCCAACCAAGGTACTCTAAG3’], and tdTomato primers Q485 [5’ TGGAGTTCAAGACCCTAC 3’], Q486 [5’ GTGTCCACCGTTAGG3’].

**Statistical analysis**

Data were analyzed using GraphPad Prism 6.0, with statistical tests used indicated in the figure legends. Quantitative RT-PCR: since expression of Plugin-tdTomato was not detectable in the G3 background, precluding a full multi-factorial analysis (genotype x gene x time), we analyzed Plugin and Plugin-tdTomato expression levels using two-way analysis of variance incorporating age (0, 1, and 2 d) and genotype-primer combination (G3-Plugin, PluTo-Plugin, PluTo-Plugin-tdTomato) as factors, followed by Tukey’s multiple comparison test, with adjusted p-values for multiple testing.
REFERENCES


CHAPTER 5.

Discussion

Respective Contributions: This chapter was written by Andrea Smidler, with editorial input from Flaminia Catteruccia.
5.1 Overview
Malaria still kills nearly half a million people annually\(^1\), and as the rate of reduction of malaria cases has plateaued in recent years, novel vector control technologies targeting the anopheline vector are urgently needed to push towards eradication. The control of the mosquito vector has proven the best way to reduce transmission of this disease\(^2\). However, resistance to all classes of insecticides used in malaria control programs has emerged in *Anopheles* populations across Africa and Asia\(^3\). With the advent of CRISPR/Cas9 technology, genetic-based vector control strategies capable of controlling or eliminating the *Anopheles* mosquito vector have become a realistic possibility. Throughout this work, we have aimed to develop a variety of novel CRISPR-based tools for the control of wild populations of *Anopheles gambiae*, as well as genetic tools to aid in monitoring GM mosquito field releases.

5.2 CRISPR-mediated genetically sterile males for Sterile Insect Technique

In Chapter 2 we outline the development of a novel transgenic system for ablation of male reproductive tissues, which induces complete sterility in 95% of males. These males sterilize the females with which they mate, making them candidates for use in Sterile Insect Technique-like programs. The crossing system we develop demonstrates the power of expressing germline-constitutive Cas9 and ubiquitous gRNAs to efficiently mutate target genes and achieve gene knockouts\(^4\). Mosaic mutagenesis occurring throughout development is sufficiently strong to ablate the germ cells of both males and females. The power of the system is partially explained through its bias for the mutated outcome; every cleavage event which repairs correctly becomes a target for cleavage again, until mutagenesis occurs. The incomplete penetrance of the phenotype (in 5% of males) can be partially explained by the presence of polymorphisms inhibiting correct gRNA targeting, and the occurrence of fertility-maintaining mutations, in addition to variable gRNA cleavage efficacy. Although understanding this variability is beyond the scope of this work, this
finding underscores the need for further investigation of the factors influencing gRNA efficacy in anophelines, a line of research not only critical for the sterilization system outlined here, but also for the development of gene drives⁵.

Although highly significant, 95% sterilization is not sufficiently stringent for SIT strategies. To increase sterilization frequency, disruptive mutagenesis must be increased, and fertility-maintaining mutations minimized. To accomplish this, additional gRNAs could be incorporated to minimize the chance of fertility-maintaining mutations, perhaps by targeting or spanning splice junctions, or by targeting sites prone to disruptive MMEJ repair outcomes. To minimize gRNA-binding inhibition due to polymorphisms, elimination of SNPs during establishment of stock transgenics could be undertaken, however this becomes prohibitively complex if transgenic lines are consistently refreshed with wild individuals to maintain them as field-like. A way to circumvent this issue could be targeting multiple fertility genes simultaneously, thereby decreasing the probability of fertile escapees with each additional gene cleaved.

It is important to note that even once sterilization is optimized, sex-separation of spermless males prior to release remains an important hurdle. While fluorescence-based sex separation methods relying on testis-specific⁶, Y-linked⁷, or male splicing-specific fluorescence⁸,⁹ provide an important tool (Section 1.10), they require sorting at larval stages which, while not prohibitive, is decidedly less effective than direct female killing. CRISPR-based female-killing transgenic systems have been developed by causing shredding of the X-chromosome during spermatogenesis¹⁰, but incorporating this system into the ZPG mutants would be extremely challenging. Another option could be turning females into males by targeting sex-determining genes such as doublesex (DSX)¹¹. However sex determination is a complex biological cascade, and despite its crucial function, it varies to a surprising degree between species. In mammals ‘maleness’ is carried on the Y chromosome¹², in Drosophila it is determined by the number of X chromosomes¹³, and in Aedes it is independent of sex chromosomes entirely¹⁴. In An. stephensi and An. gambiae the Y-linked male determining factors Guy1 and Yob were recently discovered¹⁵,¹⁶, each playing a role in male-specific
DSX splicing and sex-chromosome dosage compensation. Although not fully penetrant, transgenic expression of these factors in their respective species causes significant female killing\textsuperscript{15,17}, making similar constructs promising candidates for developing male-only transgenic systems.

Besides preventing sperm development, the ZPG knock-out also ablated egg development in females, a finding that has facilitated the study of the function of these tissues in the development of Plasmodium parasites\textsuperscript{18}. Egg-less females helped demonstrate that vector control strategies affecting ovarian development can expedite parasite development and increase disease transmission\textsuperscript{19}, providing a warning against use of some population suppression drive strategies\textsuperscript{20}. The egg-less phenotype of ZPG mutant females however prevented the isolation of true-breeding knockout lines, forcing us to intercross Cas9 and gRNA transgenics each time ZPG mutant were required. A consequence of this limitation is that released sterile males would be, by necessity, transgenic. The fact that our system uses CRISPR to induce sterility may damage the chances of public acceptance of this technology\textsuperscript{21} unlike SIT systems based on male irradiation. In contrast to gene drives, the controllability of SIT systems makes them more easily accepted in conservative settings; however shared use of CRISPR may present a point of public confusion. This issue may be exacerbated by the fact that ZPG mutant escapees would sire CRISPR-expressing offspring capable of persisting in wild populations for a few generations, presenting a public perception hurdle not encountered by other SIT systems. Despite the limited ecological relevance of such escapees, the issue of residual fertility may thwart practical implementation of CRISPR-based SIT systems, in addition to the fact they are transgenic, stressing the importance of optimization before releases can be undertaken.

In Chapter 2 (Future Directions) we describe a next generation SIT system which could perfect the ZPG mutants for release. Termed the Inducible Eunuch system, this transgenic crossing scheme uses no CRISPR, enables mass rearing, and causes complete male sterilization and female killing, all while reducing escape rates to nearly zero due to its reliance on RIDL\textsuperscript{22} rather than on mosaic mutagenesis. Furthermore,
Inducible Eunuchs would be completely non-transgenic, harboring only a small sterility-causing indel, which may provide a tremendous advantage in terms of public perception. This advantage was demonstrated recently following community outreach campaigns to gain approval for release of anti-Lyme disease GM mice on the island of Nantucket, where the local community largely voted in favor of the exclusive use of mouse-derived DNA. In light of a recent Florida referendum failing to approve release of GM moquitoes, the importance of public opinion should not be discounted. Designing non-transgenic systems may be key to gaining public acceptance and accomplishing successful release of disease-combatting mosquitoes in the future, making development of systems like the Inducible Eunuchs alluring for field applications.

Following the creation of multiple An. gambiae population suppression gene drives, development of genetic SIT technologies in the species may seem lackluster, however their distinct properties make them an important tool still warranting development. Although both SIT and sterilizing gene drives will select for pre- and peri-copulatory resistance mechanisms in wild populations to circumvent eradication, only drives, by nature of their genetic parasitism, will select for genetic resistance mechanisms. Much how rotation of insecticides is recommended to reduce the emergence of resistance mechanisms, so too may rotating between SIT and localized drive suppression systems to limit the impact of genetic drive resistance prove best practice.

5.3 Developing a transgenic tool to assess Anopheles mating competitiveness in the field.

Importantly, as field releases of GM Aedes are underway, and releases of population suppressing gene drive mosquitoes are on the distant horizon, we urgently need better tools to monitor the mating competitiveness of released GM males. Our PluTo system developed in Chapter 4 achieves this goal in a simple yet effective way, allowing visualization of successful copulation long after mating via the sexual
transfer of a fluorescent mating plug. In contrast to strategies relying on visualization of fluorescent sperm transfer to the female\textsuperscript{6}, this method can be implemented without the need of dissection and can be used for mating competitiveness of spermless male systems. While a relatively small body of work and a small contribution to science, this system may prove a valuable asset for effective future GM field trials.

5.4 Developing evolutionarily stable gene drives for \textit{Anopheles gambiae} population replacement.

In \textbf{Chapter 3} our work presents important findings that inform the development of evolutionarily stable gene drives for population replacement in \textit{An. gambiae}, and suggests that developing homing-based drives in haplolethal ribosome genes may be biologically impossible without first overcoming significant hurdles. To build gene drives that maintain the strongest possible selection against DRAs, we designed them to home into – but not disrupt – a haplolethal and haplosufficient gene pair, using the selection pressure on these genes to guarantee homing and prevent DRAs. However, we found that manipulating these haplolethal genes leaves very small room for error, and consistently obstructed drive construction.

To develop the system, we first inserted a docking site into the \textit{Rpl11-Rpt1} target, followed by subsequent drive insertion using standard transgenesis techniques (\textbf{Section 3.3}). Both of these steps were fraught with difficulties preventing successful drive development. During attempts to establish gene drive docking lines by HDR, we observed considerable embryonic lethality preventing line establishment (\textbf{Section 3.4}), a hurdle that motivated development of CrIGCkid to facilitate knockin from transgenic donors by interlocus gene conversion. While CrIGCkid was initially successful at generating two distinct docking lines, it failed in later experiments, providing the first evidence of possible gRNA silencing, and preventing in depth characterization of this phenomenon (\textbf{Section 3.5}). Following their establishment, both docking lines displayed unexpected \textit{Minute} phenotypes (\textbf{Figure 3.8}) most likely caused by suboptimal recoding or
misregulation by a heterologous 3' UTR, urging further research into anopheline codon usage and UTR regulation to overcome infertility phenotypes, and enable future drive. Inserting gene drives into the docking sites proved further problematic. While multiple drive constructs could be inserted into the docking lines, those capable of Cas9 expression died prematurely (Section 3.7), suggesting either these sites have minimal tolerance for Cas9 expression, or inserted Cas9-expressing transgenes are aberrantly regulated compared to other loci, and in the presence of ubiquitous gRNAs causes lethal mutagenesis in intolerant tissues. This is not unexpected given the role position effect plays in influencing transgenic variability in mosquitoes, and the docking site’s intimate association with the highly transcriptionally-active Rpl11 and Rpt1 genes. While achieving tighter CRISPR regulation within these sites may be feasible through use of Cas9 nickases, gypsy insulators, or tissue-specific gRNAs, these experiments would have considerably expanded the scope and timeline of the current work, making further optimization impossible at this time.

In an attempt to characterize putative drive properties of the system in the absence of a functional self-autonomous drive, we tested the ability of the ‘empty’ docking site to be homed by trans-expressed CRISPR from four distal drives, and attempted rescue of the nondriving drive (mNosGD\textsuperscript{drPLT}) with ectopic VasaCas9. While many distal drives failed to express Cas9 and failed to drive, in experiments with Cas9 expressed by known drive-enabling promoters, all females either died following blood feeding (Figure 3.13 ~G20), or were infertile (Figure 3.13 ~G40, Figure 3.16). These phenotypes are consistent with oogenic defects associated with ribosome depletion, and were lessened over time, suggesting drive designs targeting ribosome genes may struggle to maintain female fertility, and providing further evidence of gRNA silencing. Similar experiments in driving males revealed nearly complete sterilization in early broods, ‘anti-drive’ transgene ratios in intermediate broods, and low fertility with normal transgenic ratios in late broods. Taken together with the findings that all surviving larvae were the product of sperm that had escaped mutagenesis, this suggests that targeting the ribosome by CRISPR in the primordial germline...
may be fundamentally incompatible with fertility, and that transgenic expression of CRISPR may dampen over time as a ‘driving’ male ages, consistent with transgene silencing by PIWI\textsuperscript{35-38}. Moreover, the absence of homing in these males indicates that multi-gRNA drive designs may promote End-joining repair over HDR as previously observed\textsuperscript{39}, or mutagenesis may precede or supersede HDR in the PGCs, implying that targeting the ribosome for drive may be fundamentally incompatible with maintaining male fertility, underscoring the critical need for Cas9 fine-tuning in these designs. Unless all of these hurdles can be overcome, these findings suggest that our gene drive design – and perhaps all ribosome-targeting replacement drive systems – may be biologically inviable, suggesting the development of homing-based evolutionarily stable gene drives for population replacement may take years of further improvements.

One troubling hurdle facing the future of gene drive design is the lessening over time of phenotypes reliant on robust gRNA expression, suggesting inactivation, we believe through mechanisms that may have evolved to silence transposon selfish-genetic elements\textsuperscript{35,36,40-44}. CRISPR gene drives may prove to be equally burdensome genetic parasites as transposons and retroviruses\textsuperscript{45-47}, and may succumb to deactivation accordingly. In fact while near extinction of suppression drives was thought to be caused by DRAs\textsuperscript{48}, failure to home due to gRNA silencing cannot be ruled out. If this is the case, even once DRA-inhibition is achieved, overcoming gRNA silencing may be the next big impediment facing drive scientists. However, it is noteworthy that we have never observed complete loss of gRNA-dependent phenotypes. The ZPG crosses still make germ cell mutants 42 months later, and mNosGD\textsuperscript{drpt} females supplemented with VasCas9 are still wholly infertile after 40 generations (when they died 24h post-blood feeding at G20), suggesting gRNA activity is not lost, only dampened – at least over the time scales observed within this work. While significant silencing could happen in a generation\textsuperscript{36}, complete silencing may never occur, allowing homing to continue for years – after all transposons still jump despite eons of counter-evolution\textsuperscript{49,50}. More research is required on the subject, not just for mosquito gene drives, but also for any
eukaryotes expressing transgenic gRNAs\textsuperscript{51}, and may prove a significant hiccup for numerous CRISPR-based transgenic applications.

If building multi-gRNA replacement drives in ribosome genes proves impossible, then the scientific community may have to relax its expectations for the ideal drive. If a suitable non-ribosomal haplolethal gene target can be identified, and multi-gRNA designs can be constructed capable of stimulating robust HDR, then such replacement drives may still be possible. The alternative is construction of a drive within haplosufficient targets, but inhibition by heterozygous r2DRAs will greatly slow drive spread, likely preventing population replacement. Integral drive designs have been proposed for evolutionarily stable population replacement, but their reliance on single gRNAs makes the evolution of r1DRAs almost certainly inevitable\textsuperscript{52}. On the other hand, population suppression gene drives impose such immense pressure on the mosquito host to evolve mechanisms of resistance that they too will likely encounter setbacks, even when targeting the most conserved sites\textsuperscript{11}. In one form or another, all drives face a battle against evolution, and every drive design conceived by (wo)man will eventually be deactivated.

There is however still hope for development of our original drive designs – using evolution as our guide. Similarly to naturally occurring selfish genetic elements which have persisted for billions of years\textsuperscript{50}, synthetic self-autonomous drive systems should also be self-optimizing once they overcome the hurdle of DRAs (Section 1.8B). So even though current designs do not drive, we may be able to harness the power of evolution to guide us towards building one. By beginning with a line with little to no baseline drive, and nearly undetectable Cas9, but with all other necessary drive characteristics, maintaining these proto-drives in the lab for many generations could select for mutants with perfected drive properties. Depending on how close the proto-drive is to functioning, and the feracity\textsuperscript{*} of the species in which it’s tested, such experiments may be possible within practical timescales, and in fact have already begun. By maintaining

---

\textsuperscript{*} The state of being feracious. Producing in abundance; fertile, fruitful. (Archaic) Last defined in Webster’s Revised Unabridged Dictionary, G. & C. Merriam, 1913
the nearly self-autonomous mNosGD<sub>dRPLT</sub> for over 40 generations, we have inadvertently begun to select for evolution of de novo drive. While no explicit biased inheritance of the transgene has been observed (Section 3.7B), artifacts visualized by Western hint that subtle or modified Cas9 expression may be occurring in some individuals at ~G40 (Figure 3.10) when it was absent at ~G20. And while early generations of this transgenic line were plagued by constant near-extinction events due to insertion within dRPLT, recent generations have not experienced this, no longer requiring heterozygote enrichment for maintenance. While recent generations of this line have not yet been re-sequenced, maintenance of gRNA function was confirmed recently by experimental phenotypes (Figure 3.13), suggesting that the transgene has either lost the properties which made it once unfit, or gained a de novo ability to prevent extinction (drive?) by yet unknown mechanisms. This line merits further investigation as it may reveal the secrets necessary for development of evolutionarily stable population replacement gene drives.

5.5 CONCLUDING REMARKS

From gene drives to sterile male systems, and from knockouts to knockins, and developing monitoring tools in between, this dissertation provides an extensive body of work, considerably expanding the genetic toolkit in Anopheles gambiae mosquitoes. Our work demonstrates novel principles for genetic sterilization, and presents significant advancements towards the development of evolutionarily stable gene drives for population replacement, revealing insights critical for all CRISPR-based gene drives in the process, and it is our hope that it may one day contribute to the eradication of one of humanity’s most devastating diseases.
5.6 REFERENCES


APPENDIX A

Plasmids and transgene Maps
Spermless males

VasCos9

gZPG

Inserting docking sites into Rpl11-Rpt1

Rpl11-Rpt1
Wt endogenous

HDRdRPL7
plasmid

IGCdRPL7

dRPL7 in genome

dRPL7 in genome

Gene drives (GDS)

mNosGD
in dRPL7

VasGD
in dRPL7

B2GD
iven dRPL7

Distal Gene drives (dGDS)

B2dGD

ZPG2dGD

ZPG1dGD

VFdGD

167
APPENDIX B

Targeted Mutagenesis in the Malaria Mosquito Using TALE Nucleases
Targeted Mutagenesis in the MalariaMosquito Using TALE Nuclease 

Andrea L. Smidler1,2,a, Olivier Terenzi1,2, Julien Soichot1,2, Elena A. Levashina1,2b, Eric Marois1,2*

1 Institut National de la Santé et de la Recherche Médicale U963, Strasbourg, France, 2 Centre National de la Recherche Scientifique UPR9022, Strasbourg, France

Abstract

Anopheles gambiae, the main mosquito vector of human malaria, is a challenging organism to manipulate genetically. As a consequence, reverse genetics studies in this disease vector have been largely limited to RNA interference experiments. Here, we report the targeted disruption of the immunity gene TEP1 using transgenic expression of Transcription-Activator Like Effector Nuclease (TALENs), and the isolation of several TEP1 mutant A. gambiae lines. These mutations inhibited protein production and rendered TEP1 mutants hypersusceptible to Plasmodium berghei. The TALEN technology opens up new avenues for genetic analysis in this disease vector and may offer novel biotechnology-based approaches for malaria control.

Introduction

Malaria is caused by Plasmodium parasites transmitted to their human hosts by the bite of anopheline mosquitoes. Malaria has been charged with causing more human deaths than any other disease in human history and continues to kill about 660,000 annually [1]. A reduction in the malaria death toll has been achieved thanks to vector control using insecticides and insecticide-impregnated bednets, better health care and progress in medical treatment, but this success is currently mitigated by the spread of resistance both of mosquitoes to insecticides and of Plasmodium to antimalarial drugs. Sequencing the genome of the main malaria vector, Anopheles gambiae [2], enabled the identification of hundreds of genes involved in the vector's capacity to transmit Plasmodium. Altering the mosquito genome in a way that abates Plasmodium transmission, through transgenesis or other sophisticated genetic engineering tools, can offer new perspectives in the fight against malaria. On one hand, experimentally altering mosquito genes of interest will advance our fundamental understanding of the biological interactions between mosquito and parasite, and may help target vulnerable points in the parasite cycle. On the other hand, the prospect of releasing engineered male mosquitoes to propagate malaria resistance genes through wild susceptible vector populations has been receiving increasing attention [3,4]. Novel methods to disrupt or alter target genes of interest in the malaria mosquito would promote rapid progress towards these goals. Furthermore, targeted genetic modifications that do not require the permanent introduction of transposons in the genome are particularly desirable, as they eliminate the potential risk of subsequent unplanned transposon mobilization by natural sources of transposition factors.

Recently, a novel class of DNA-binding protein domain derived from the Xanthomonas Transcription Activator Like Effector (TALE) proteins [5,6] has been successfully harnessed to custom-design sequence-specific endonucleases [7,8]. These TALE nucleases (TALENs), in which the TAL DNA-binding domain is fused to the FokI endonuclease domain, are easy to engineer (e.g., [9,10]), highly predictable in their sequence specificity, and highly mutagenic [11] making them an attractive alternative to Zinc Finger Nucleases that have less predictable binding specificities and require in vitro optimization [12]. Mutations arise by imprecise repair of the

---


Editor: George Dimopoulos, Johns Hopkins University, Bloomberg School of Public Health, United States of America

Received June 24, 2013; Accepted August 7, 2013; Published August 15, 2013

Copyright: © 2013 Smidler et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by an Agence Nationale de la Recherche (ANR) young researcher grant to E.M. (project GEMM), by the European Commission FP7 projects INFRAVEC (grant agreement no. 228421) and EVIMalar (grant agreement no. 242095), by the Institut National de la Santé et de la Recherche Médicale and by the Centre National de la Recherche Scientifique. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

*a Current address: Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, United States of America

*b Current address: Department of Vector Biology, Max Planck Institute for Infection Biology, Berlin, Germany
TALEN-generated double-stranded breaks by the non-homologous end joining (NHEJ) repair pathway. In a number of animal species, injection of mRNA encoding TALENs have readily allowed researchers to generate mutants in their genes of interest [13–19]. Very recently, mutagenesis of an eye pigment gene was achieved in Aedes mosquitoes using this method [20]. In rice, disease resistant mutants have been produced by transgenic expression of the TALENs, whose respective transgenes were eliminated by subsequent genetic crosses once the desired mutation had been fixed [21]. We hypothesized that a similar approach would be applicable to insect vectors and set out to use TALENs to target the TEP1 gene, a key component of the mosquito immune system.

**Results**

Since the embryo microinjection procedure is technically challenging in Anopheles gambiae (as judged by poor survival and relatively low success rate of transgenesis in this species), we expected that mutant recovery after direct injection of TALEN-encoding plasmids or mRNA might be difficult. For this reason and to enable controlled mutagenesis experiments, we preferred transgenic expression of the TALENs in the mosquito germ cells. To obtain the proof-of-principle for gene targeting in A. gambiae via transgenic TALENs, we selected the well-characterized immune gene TEP1 as a target. TEP1, a protein similar to vertebrate complement factor C3, binds Plasmodium parasites as they invade the mosquito intestine and kills them in a manner probably dependent on its thioester site located in the C-terminus [22–25]. TEP1 mutants will be instrumental in further dissecting the antiparasitic complement-like system in mosquitoes.

TALENs function in pairs, each member of which binds a chosen 12 to 24-nucleotide sequence. The two selected target sequences are separated by 14-16 nucleotides, the optimal distance for the two FokI domains of the TALENs to properly dimerize and create a double-stranded break. Cleavage of the bound DNA molecule occurs near the center of the sequence separating the two target sites. We designed a single pair of TALENs to target a site within the TEP1 gene centered on an NcoI restriction site (5’-CCATGG-3’), offering the possibility to easily screen individual mosquitoes for mutations that destroyed the NcoI site (Figure 1A). Mutations in this region are expected to strongly affect TEP1 protein function: frame-shifts resulting in premature stop codons would remove the C-terminal third of the protein including its thioester domain, while amino-acid deletions or insertions would alter the length of the alpha-helix connecting the CUB domain to the thioester domain [26], presumably resulting in destabilization of the protein’s structure.

A distinct transgenic mosquito line was generated for each TALEN of the pair, the expression of which was driven by the Aedes Vasa promoter, active in the mosquito germline [27]. Therefore, F1 mosquitoes arising from a cross between the two lines will express both TALENs simultaneously and are expected to produce F2 gametes carrying mutations in the TEP1 gene. To screen individual larvae within the F2 progeny, we PCR-amplified a TEP1 fragment spanning the target site and subjected the amplification product to a restriction digest with NcoI (Figure 1B). Of 310 screened F2 larvae, 16 (5.16%) carried a heterozygous mutation at the target locus, as evidenced by the appearance of a PCR product that NcoI was unable to cleave. Thus, at least 2.58% of TEP1 copies were mutated after exposure to one TALEN dose (i.e., one generation). This figure is a conservative estimate of mutation frequency, as we subsequently observed some mutations that left the NcoI recognition sequence intact. In order to obtain a mosquito population containing a higher frequency of mutations, we self-crossed successive generations of mosquitoes expressing both TALENs of the pair. This was facilitated by automated COPAS selection of larvae [28] that had inherited one copy of each of the left and right TALEN genes, which are respectively associated with a red and yellow fluorescent marker expressed in the nervous system [29]. At least 1000 double-TALEN larvae were COPAS-selected and cultured for each generation. In the 7th generation (i.e., exposure to 6 TALEN mutagenic doses), 51% (49 out of 96) of the examined individual mosquitoes carried a heterozygous mutation in TEP1. This indicated that the frequency of mutations increased faster than predicted if mutations accumulated linearly from one generation to the next. This observation is consistent with a model where mosquitoes already carrying a heterozygous mutation employ homologous recombination to repair new TALEN-induced breaks in the wild-type chromosome, thereby effectively copying the existing mutation onto the newly damaged chromosome. This suggests that in addition to NHEJ, TALEN-caused breaks can also be repaired by homologous recombination. Therefore, the observed number of NHEJ mutations is an underestimation of the true rate of TALEN activity.

We wondered if a single TALEN of the pair is capable of causing mutations in TEP1. To investigate this, we sampled mosquito larvae from lines carrying a single TALEN maintained at high population levels for 8 (right TALEN) or 10 (left TALEN) generations, and again sampled larvae after about 16 generations. We expected that such a high number of generations would have allowed rare “monotalenic” mutational events to accumulate in the population. Out of 96 individual larvae tested by PCR for each sample, none carried a mutation in the TEP1 NcoI site. The observed absence of mutations among 576 haploid genomes exposed to single TALEN activity for 7, 9 or 15 generations suggests that single TALENs never or rarely induce mutations at their target site. To strengthen this point, we purified DNA from 2600 pooled larvae whose genomes had been exposed to one TALEN of the pair for 7 or 9 generations, and PCR-amplified the target region. PCR products appeared to be fully cleaved by NcoI, pointing to the absence of TALEN-induced mutations. To increase the chance of detection of a minor fraction of mutated products, we purified the region of the gel in which uncleaved PCR products may exist, cloned them into a plasmid, and examined E. coli transformants containing single copies of the PCR products. Again, all cloned fragments were cleaved by NcoI. Although we note that deep sequencing of amplicons would provide a more sensitive assay to detect rare mutations, this result further suggests that single TALENs rarely or never generate mutations.
Figure 1. TALEN mutagenesis of the TEP1 gene. A: Fragment from the TEP1 gene showing the target site of the TALEN pair. Nucleotides bound by each TALEN are underlined, TALEN repeats are color-coded to show repeat/nucleotide specificity. The NcoI restriction site centrally located at the TALEN cleavage site is highlighted. Inset: scheme of the entire TEP1 protein showing the location of TALEN-induced mutations (SP: signal peptide; CUB: CUB domain interrupted by the TED: thioester domain; the star indicates the position of the thioester site). B: PCR assay to identify TEP1 mutant mosquitoes. A PCR product spanning the TALEN target site is generated from individual mosquitoes (small larva or a leg from a living adult) and incubated with NcoI. Full cleavage (w) denotes a wild-type individual. Partial cleavage (h) denotes a heterozygous TEP1 mutant. Absence of cleavage (H) corresponds to a homozygous TEP1 mutant. C: TALEN-induced mutations in the TEP1 gene. Left and right TALEN target nucleotide sequences are shown in green and blue respectively, with the 15bp spacer sequence between the TALENs in black. The NcoI restriction site is highlighted in orange. Deletions are designated by a red dash or by Δ+number of missing bases. Insertions are shown in lowercase red letters. Uppercase red letters correspond to natural polymorphisms between multiple TEP1 alleles.

doi: 10.1371/journal.pone.0074511.g001
The apparent absence of single TALEN background activity was likely facilitated by our design scheme in which we used obligate heterodimeric FokI domains in TALEN construction [30,31].

The TALEN-generated TEP1 mutations were very similar in nature to mutations obtained in other organisms (Figure 1C) and consisted mainly of small deletions and insertions (indels) that are the hallmark of imprecise NHEJ-mediated DNA repair. Some mutations deleted a number of nucleotides in multiples of three, resulting in the deletion of one to a few amino acids from the TEP1 protein. Other mutations introduced a frame-shift in the TEP1 coding sequence, resulting in the loss of the entire C-terminal half of the protein, which contains features crucial for TEP1 function including the thioester domain. Using a PCR selection procedure from single legs taken from live classes of mutations: line M mosquito gut. These results confirm the pivotal role of TEP1 in females ingested dramatically reduced numbers of 1500, each experiment also yielded mutant midguts bearing no C-terminus (Figure 2A). Immunoblotting analysis failed to detect TEP1 in the Mut and M lines, while it revealed strongly reduced protein levels in M and M lines (Figure 2B, top). Although small amounts of TEP1 protein could be detected in whole mosquito extracts of the M, M and M lines, these proteins did not undergo cleavage and were impaired in their secretion to the hemolymph, as no TEP1 signal was observed in immunoblotting of hemolymph samples (Figure 2B, bottom). Therefore, the alpha helix connecting the CUB and thioester domains of TEP1 [26,32] seems to be very sensitive to insertion or deletion of single amino acids that destabilize the structure and prevent proper protein synthesis and secretion.

We next assessed the phenotype of these mutant lines in comparison to the parental lines by infection assays using Plasmodium berghei-infected mice (Figure 3). Similar to the RNA interference knockdown phenotype of TEP1 [22], all mutations in the homozygous state resulted in a dramatic increase in the number of developing parasites within the mosquito gut. These results confirm the pivotal role of TEP1 in antiparasitic responses. While it cannot be fully concluded that RNAi-mediated TEP1 knockdown may also affect genes sharing some nucleotide identity with the TEP1 sequence, such as other closely-related genes in the TEP family [33], mutations in TEP1 should leave the expression of other genes unaffected.

Interestingly, while midguts from TEP1 mutant mosquitoes often displayed impressive oocyst numbers that could exceed 1500, each experiment also yielded mutant midguts bearing no or only a few oocysts. Upon blood feeding on a single infected mouse, only well-gorged mosquito females were selected for subsequent dissection. It is therefore unlikely that some females ingested dramatically reduced numbers of P. berghei gametocytes. Thus, it is apparent that TEP1-independent mechanisms are at work to limit Plasmodium infection in a subset of mosquitoes.

To examine the effect of TEP1 dosage in the process of parasite killing, we compared oocyst infection levels between TEP1 heterozygous mutant, homozygous mutant and control mosquitoes (Figure 3d). The heterozygous mutant had an intermediate susceptibility phenotype, which was closer to, but significantly different from, the control. This suggests that the efficiency of parasite killing depends on TEP1 protein levels. The antiparasitic role of TEP1 was discovered and characterized using RNA interference assays, in which synthetically produced double-stranded RNA homologous to a fragment of native TEP1 is injected at a high concentration in the body of adult mosquitoes. This technique was generalized for the functional characterization of hundreds of mosquito genes [34]. However, injection per se was reported to impinge on Plasmodium development by the potential induction of the wounding response [35]. Therefore the mutant TEP1 lines developed here will be useful in studies that must exclude the confounding effect of the injection procedure itself. To examine how the classical TEP1 RNAi phenotype compares with the mutant phenotype, we compared parasite loads in M to dsTEP1-injected mosquitoes of the parental line (Figure 3e).

The mutant and RNAi phenotypes were very similar, validating a posteriori the high efficiency of RNAi knockdown.

Discussion

Here, we obtained proof of principle that TALENs can be used for targeted mutagenesis in the genome of the malaria mosquito, which is notably difficult to manipulate genetically. Recently, Aryan et al. [20] reported disruption of the eye pigmentation kmo gene in the dengue vector mosquito Aedes aegypti by injection of TALEN-encoding plasmid DNA. In the same species, successful mutagenesis of GFP and of the odorant receptor co-receptor (orco) was achieved by injecting mRNA encoding Zinc Finger Nucleases [36]. In both cases and in other animal systems for which TALEN mutagenesis has been reported, mutants were obtained directly upon injection of DNA or RNA into embryos. In contrast, we employed transgenically-expressed TALENs for this purpose. Besides the assurance of expressing the pair of TALENs in all germ cells, this offered the possibility to increase the frequency of mutant alleles in successive TALEN-expressing generations of mosquitoes. We disrupted the antiparasitic gene TEP1 as a first target; future work will make use of the obtained hypersusceptible mutant lines to further dissect the role of this important anti-malarial factor in parasite killing. Beyond the research field of mosquito immunity, this study paves the way for numerous other applications such as obtaining mutations in Anopheles genes considered to be essential for Plasmodium parasite development [37–39]. Disrupting these genes, or altering specific domains on the encoded proteins, may render homozygous mutant mosquitoes unable to support parasite development. Such parasite-refractory mutant mosquitoes could be used in anti-malaria intervention schemes, including vector population replacement in endemic regions. Unless a gene-drive strategy is specifically designed to spread desired mutations, this could be achieved by the repeated release of mass-produced male mosquitoes. Where knockout mutants in
A given target gene might compromise the fitness of the mosquitoes, TALEN mutagenesis offers the possibility of a gene therapy to cure mosquitoes of malaria by selecting mutations that prevent a protein’s interactions with parasite factors while preserving its other vital functions. TALEN mutagenesis could also be employed to knock out male fertility genes for use in the Sterile Insect Technique to reduce vector populations [40]. Of note, the TALEN transgenes used to obtain a desired mutation can subsequently be discarded by selection, thereby rendering the obtained homozygous mutant mosquitoes transgene-free. This could facilitate mosquito
Figure 3. **TEP1 mutant mosquitoes are hypersusceptible to *P. berghei***. Mosquito females from five different homozygous mutant mosquito lines and from control parental lines were offered a blood meal on a *P. berghei*-infected mouse. Seven days after infection, the midgut was dissected and the number of oocysts developing in each midgut was evaluated. The statistical significance of differences in mean parasite numbers was measured with a Mann-Whitney test (mutant versus control) and with a Kruskall-Wallis test followed by Dunn’s post-test (to compare all groups in [d]). (a) M\(^\Delta T\) mosquitoes are compared to the two parental, non-mutant TALEN lines. (b) 4 different mutant lines are compared to the parental mosquito line that initially served to produce TALEN transgenic lines. This control line was verified to show the same level of susceptibility to *P. berghei* as the two TALEN daughter lines (not shown). In this experiment, three mutant lines showed significantly elevated parasite numbers compared to the control while the M\(^\Delta c2\) line did not, presumably due to a different physiological condition of this mosquito culture. In two independent experiments (c, d), we used mosquitoes of different genotypes marked with distinct fluorescence markers and cultured the larvae together in the same water to eliminate potential confounding factors due to rearing conditions. On the day of dissection, genotypes were separated on the basis of fluorescence. The same M\(^\Delta c2\) line shows significantly elevated parasite numbers. (d) Heterozygous M\(^\Delta c2\) mosquitoes are compared to control and homozygous M\(^\Delta c2\) mosquitoes: the susceptibility phenotype of the heterozygote is intermediate. (e) Control mosquitoes of the parental line, mosquitoes of the parental line injected with TEP1 double-stranded RNA, and homozygous TEP1 mutant mosquitoes of the M\(^\Delta T\) line are compared. TEP1 mutant and dsRNA-injected mosquitoes show comparable susceptibility to *P. berghei*.

doi: 10.1371/journal.pone.0074511.g003
release interventions that adhere to local regulations regarding genetically modified organisms.

Materials and Methods

Ethics statement

Our experimental protocols were approved by Comité de Qualification Institutionnel (CQI), the ethics evaluation committee of INSERM (IRB00003888, FW00005831). Mosquitoes were reared and blood-fed on anesthetized mice in compliance with French and European laws on animal house procedures (agreement E67-482-2 of the Direction of Veterinary services of the French Ministry of Agriculture).

Assembling the TALENs

TALENs targeting the sites shown in Figure 1 were constructed by Golden Gate Cloning according to [9]. For TALEN assembly, we prepared two transgenesis-compatible destination vectors (annotated sequences provided in Text S1) encoding a Venus yellow fluorescent and a DsRed fluorescent transgenesis reporter gene, respectively. These vectors also contain a phage \( \phi C31 \) attB site for genomic integration into transgenic lines harboring attP sites. The first module used in Golden Gate assembly provided the Vasa promoter characterized in [27]. The last module closed the TALEN assembly with either a FokI-DD or a FokI-RR domain [30], designed for obligate heterodimerization of the two TALENs and codon-optimized for \( A. \) gambiae. The annotated sequence of the three plasmids is provided in Text S1. The TALEN C and N-terminal domains flanking the repeat region were identical to those used in [7].

\( A. \) gambiae lines and mosquito transgenesis

The TALEN-encoding vectors were inserted by \( \phi C31 \) integrase-mediated transgenesis [41] into the genome of \( A. \) gambiae lines X1 and X13, which are derived from laboratory strain G3. These two lines carry a PiggyBac transgene on chromosome II, containing an \( \text{attP} \) docking site. Individual transgenic larvae carrying the inserted left or right TALEN genes at the X1 or X13 \( \text{attP} \) site were identified by their red or yellow fluorescence, respectively. Resulting adults were crossed to their non-fluorescent parental line. In the F2 generation, fluorescent homozygous larvae were COPAS-selected [28] to establish stable TALEN-expressing lines.

Identification of TEP1 Mutants

In putative mutants (mosquitoes arising from parents expressing both TALENs), we PCR amplified a region of \( \text{TEP1} \) spanning the TALEN target site using Phusion or Phire Polymerases (Thermo, Fisher) and primer 5'-TCAACCTGGACACTCAACAGAAGGCCGA-3' in combination with either 5'-GCATATCTTTGTGCCACACTTT-3' or 5'-GCCACCGTAACCAATTTCACA3'. The PCR product was digested with \( \text{Ncol} \) (Fermentas), the recognition site of which is centrally located in the sequence cut by the TALENs. PCR products corresponding to mutant \( \text{TEP1} \) alleles were not digested by \( \text{Ncol} \). These PCR products were sequenced directly in the case of homozygous individuals, or cloned (CloneJET PCR Cloning Kit, Fermentas) and subsequently sequenced.

Mutagenesis and mutant line recovery

For mutagenesis experiments, the two TALEN lines were crossed. From the F2 generation onwards, 1000 mosquito larvae that inherited a single copy of each TALEN were COPAS-purified to initiate the next generation. To isolate \( \text{TEP1} \) mutants, mosquitoes from this population were out-crossed to the parental line. The progeny carried a single TALEN and was therefore no longer subjected to TALEN mutagenesis. Among this progeny, we screened single adult mosquitoes by PCR on one leg using the Phire direct animal tissue PCR kit (Thermo, Fisher). The identified heterozygous mutants were individually crossed to the parental line. In the F2 progeny, we identified homozygous mutants by leg PCR and pooled them to start distinct homozygous mutant families.

RNAi and infection assays

RNAi silencing of \( \text{TEP1} \) by double-stranded RNA injection into the thorax of adult mosquitoes and mosquito infections on mice carrying \( \text{Plasmodium berghei} \) GFP-con 259c12 were performed as described [22]. To semi-automatically quantify the number of oocysts in photographs of dissected mosquito midguts, we used the "watershed segmentation" and "analyze particles" plugins of the ImageJ software after digital subtraction of the image background and smoothing of the signal. Oocyst counts obtained by this method are consistent with those obtained by manual counting of oocysts.

Immunoblotting

Whole-body mosquito extracts were obtained by grinding one adult female mosquito in 80 µl of protein sample buffer. The sample was denatured for 3 min at 95°C and centrifuged. 8 µl were loaded on 8% SDS-polyacrylamide gels. Hemolymph samples were prepared as described [42], 10 µl of the samples were loaded. Immunoblotting was performed using standard procedures [43] with rabbit polyclonal antibodies raised against the prophenoloxidase \( \text{PPO} \) [44] or against the C-terminal half of \( \text{TEP1} \) [45].

Supporting Information

Text S1. The nucleotide sequences of plasmids and building blocks used to construct the TALEN-expressing transgenesis vectors are provided. Building blocks not listed here are published in references 5,7.

(DOCX)

Acknowledgements

We thank the Boch and Bonas Laboratories (University of Halle, Germany) for sharing the TAL construction kit.
Author Contributions
Conceived and designed the experiments: AS OT JS EM. Analyzed the data: AS EM EAL. Wrote the manuscript: EM AS EAL.

References


APPENDIX C

Engineering the control of mosquito-borne infectious diseases
Engineering the control of mosquito-borne infectious diseases

Paolo Gabrieli, Andrea Smidler and Flaminia Catteruccia*

Abstract
Recent advances in genetic engineering are bringing new promise for controlling mosquito populations that transmit deadly pathogens. Here we discuss past and current efforts to engineer mosquito strains that are refractory to disease transmission or are suitable for suppressing wild disease-transmitting populations.

Introduction
Mosquitoes transmit a variety of infectious agents that are a scourge on humanity. Malaria, dengue fever, yellow fever, and other mosquito-borne infectious diseases infect millions of people and account for hundreds of thousands of deaths each year, posing a huge burden for public health and on the economic growth of countries where these diseases are endemic [1]. Given the lack of effective vaccines against many mosquito-borne pathogens, national programs are heavily reliant on the use of insecticides to control mosquito populations in order to stop disease transmission [2]. Unfortunately, the alarming pace of emergence of insecticide resistance in mosquitoes [3] is threatening chemical-based campaigns and is forcing scientists to develop alternative strategies to combat vector-borne diseases. Moreover, insecticide-treated bed nets and indoor residual sprays principally target mosquitoes that feed indoors at night and that rest inside houses, thereby neglecting those species that prefer to bite and rest outdoors or at earlier hours of the day, and inducing some degree of insecticide-avoidance behavior (behavioral resistance) in indoor-biting individuals [4-6].

Recent major advances in the field of genetic engineering are providing an unprecedented opportunity to conceive and create designer mosquito strains in order to control natural vector populations. From the generation of the first transgenic mosquitoes [7-10] to the creation of the first gene knock-outs [11-13], the discovery of genetic tools has revolutionized our ability to functionally study and edit the mosquito genome. In the fight against infectious diseases, vector populations can be modified using these tools in two principal ways: 1) they can be made refractory to disease transmission by the introduction of genes with anti-pathogenic properties; 2) they can be rendered sterile or modified in such ways that the population size will crash below the threshold necessary to support disease transmission (Figure 1) [14]. Both strategies have strengths and limitations that are inherent to their design and properties.

Genetic engineering technologies include those that allow heterologous gene expression and those that modify endogenous genes or entire portions of the mosquito genome. Here we review the genetic tools that are currently in use and those that promise to become available in the near future, with particular focus on those techniques that are capable of reprogramming the genomes of field populations. We also discuss current field trials in which genetically modified mosquitoes are being released, and will mention ecological hurdles and potential environmental and regulatory issues stemming from the release of genetically modified insects into the wild.

First generation of anti-pathogenic strains
The expression of exogenous genes - through the transposon-mediated integration of transgenes - was the first genomic technology to be developed in mosquitoes, and gave birth to the modern field of mosquito genome engineering [7-10]. In this initial system, different exogenous ‘effector’ genetic elements are cloned between the transposon terminal repeats (usually using the PiggyBac transposon [10]) to form a synthetic element that, in the presence of the integrating enzyme transposase, inserts into the mosquito genome at quasi-random loci (Figure 2a). In order to identify successful transformants, synthetic transposons are generally designed to carry a fluorescent reporter construct, such as the green fluorescent protein...
fluorescence and facilitating high-throughput sorting during larval development allowing easy detection of the neuronal 3xP3 promoter [19], which is expressed by automated live sorters [20]. Moreover, this system can incorporate cargoes with anti-pathogenic properties to render mosquitoes refractory to disease transmission.

Both Anopheles and Aedes mosquito species, the vectors of malaria and dengue, respectively, have been modified to reduce their vectorial capacity. To stop the development of Plasmodium parasites, the causative agents of malaria, scientists have developed transgenic Anopheles stephensi lines that express single chain variable fragment antibodies (scFvs) [21-23] or synthetic antimalarial factors [24,25] (Figure 2b). Transgenic lines that express ScFvs against the ookinete proteins Chitinase 1 and Pfs25 [38,39] or the predominant surface protein of the sporozoites, circumsporozoite protein [40,41], show reduced ookinete crossing of midgut walls or sporozoite invasion of the salivary glands, respectively. Similarly, An. stephensi strains have been generated that secrete the synthetic dodecapeptide SM1 (an acronym for salivary gland- and midgut-binding peptide 1) into the midgut lumen during blood feeding. SM1 binding to the epithelium - probably through a mosquito midgut receptor - prevents ookinetes from invading the midgut in the rodent malaria Plasmodium berghei model, thereby reducing both the prevalence and the intensity of infection [24]. Additionally, the incorporation of bee venom phospholipase A2 into transgenic An. stephensi inhibits ookinete invasion of the midgut by modifying epithelial membranes [25]. Anopheles gambiae, the principal vector of malaria in sub-Saharan Africa, has been engineered to ectopically express the endogenous antimicrobial peptide cecropin A [26] and the synthetic peptide Vida3 [27], a hybrid peptide based on natural antimicrobial peptide sequences that have strong activity against Plasmodium sporozoic forms [28].

Different laboratories have also developed Anopheles strains modified in key endogenous cellular pathways that regulate parasite development, namely the insulin-growth factor signaling (ISS) and the immune deficiency (IMD) pathways. In An. stephensi, overexpression of Akt, a critical regulator of ISS, elicits mitochondrial dysfunction that enhances parasite killing in the midgut, even if at some cost to mosquito survival [42,43]. To overcome fitness costs, an inhibitor of ISS, the phosphatase and tensin homolog (PTEN), was instead overexpressed [44]. PTEN inhibits phosphorylation of the ISS protein FOXO, and its expression blocks Plasmodium development by enhancing the integrity of the midgut barrier, although this causes an increase in the female lifespan with possible negative consequences for disease transmission [44]. In another study, An. stephensi mosquitoes were engineered to express the active form of the IMD-regulated NF-kB transcription factor Rel2-S. Rel2-S activates the expression of several antimicrobial and anti-Plasmodium peptides, and when overexpressed in the midgut and in the fat body, it strongly inhibits parasite development [45].
Engineering pathogen resistance has not been limited to anophelines. Dengue virus infections in *Aedes aegypti* mosquitoes have been attenuated by exploiting the natural antiviral RNA interference pathway. An inverted-repeat RNA capable of forming double-stranded RNAs that target the pre-membrane protein coding region of the DENV-2 serotype was expressed in the midgut [29] or in the salivary glands [30]. This modification reduced viral titers by more than five-fold compared to those in control mosquitoes. It should be noted, however, that multiple dengue serotypes (as well as multiple human malaria parasites) exist, complicating population replacement efforts aimed at spreading pathogen-refractory genes into wild populations.

**First generation of sterile strains for population suppression**

Early transposon-based technology has been also used to generate mosquito strains aimed at the suppression or elimination of vector populations through the release of sterile males (the sterile insect technique (SIT)) [15]. The alternative sister strategy is the release of insects carrying a dominant lethal (RIDL) modification [16]. SIT is based on the release of large numbers of sterile males,
usually sterilized with high doses of irradiation or chemical sterilants, that upon mating with field females produce no fertile progeny causing suppression or elimination of local populations (Figure 1a) [15]. The sterilization process usually induces severe fitness costs in the male, such that larger numbers of males than those initially predicted by simple models need to be released to achieve the desired level of suppression [46]. Genetic engineering can not only enable high-throughput sorting of male-only populations based on sex-specific fluorescent markers [47,48], but can also enable the design of strains in which specific sterility-inducing transgenes or genetic mutations have been introduced without causing the fitness costs associated with irradiation [49,50]. The most successful RIDL example is provided by the Ae. aegypti strain OXS13A [16], which carries an inducible dominant genetic system that kills late larval stages. This system is composed of a gene encoding the tetracycline transactivator (tTA) protein under the control of the tetracycline-responsive element (tRE). Binding of tetracycline to tTA prevents tTA from activating transcription; when tetracycline is removed, tTA instead binds to tRE, thereby inducing its own expression via a positive feedback loop. The accumulation of tTA is toxic to cells and ultimately leads to organismal death (Figure 2c). This repressible system allows the generation of males that are fertile in the laboratory but that, once released, sire unviable progeny upon mating with field females. These RIDL strains are already being released in different geographical locations as part of field trials.

A different approach, initially developed in Ae. aegypti and now transferred to Aedes albopictus and An. stephensi, is based on a bimodal system that severely impairs the functionality of the female flight muscles, disrupting the female's ability to fly (6RIDL) [51-53]. The first module consists of tTA under the control of the female-specific Actin-4 transcriptional regulatory elements, which drive gene expression in the indirect flight muscles of female pupae. The second module comprises a lethal gene (Nipp1Dm or michelob_x in Ae. aegypti, VP16 in Ae. albopictus and Nipp1Dm in An. stephensi) under the control of tRE. In the absence of tetracycline, expression of the lethal gene specifically in the female flight muscles causes cell death and inability to fly. As males are unaffected by the transgene, their release will generate flightless female progeny that are unable to mate, bite, and transmit disease, eventually leading to population suppression [51].

**Second generation transgenesis provides increased flexibility**

New genome-editing tools now allow scientists to modify endogenous genes with increasing flexibility and ease, and are being utilized in the laboratory with promising results to reduce the vectorial capacity of mosquito vectors (Figure 2d). The flexibility of these tools resides in the use of protein precursors that can be designed to bind sequences of interest within the mosquito genome [11-13]. Repetitive zinc finger (ZF) and transcription activator-like effector (TALE) modules have been successfully fused to the endonucleolytic domains of a type II endonuclease, normally FokI, to generate knock-out and knock-in mutants [11-13,34] (Figure 2e,f). These modified nucleases cause site-specific double-stranded DNA breaks that can be repaired by the non-homologous end-joining (NHEJ) pathway, an error-prone repair pathway that often results in small indels. As a basic proof-of-principle, this technology has been used to generate eye-color mutants (Figure 2e) [11], but it can also help elucidate pathways that are important for vector competence. For example, TALE nucleases (TALENs) have been used in An. gambiae to generate null mutants of the thioester-containing protein 1 (TEP1) gene, a complement-like factor that opsonizes Plasmodium parasites in the midgut and mediates their killing. Mutant strains are, therefore, hyper-susceptible to Plasmodium infection [13], and although not directly employable for malaria control, they allow detailed genetic analyses of anti-Plasmodium immune pathways. Similarly, the zinc-finger nuclease (ZFN)-mediated knock-out of the odorant receptor co-receptor (ORCO) in Ae. aegypti has enabled the analysis of pathways involved in host-seeking behavior for blood feeding [12], opening up new avenues for the development of mosquito repellents and attractants. In another study, the CO2 response of Ae. aegypti mosquitoes was analyzed in mutants that have a defect in the AeagGr3 gene, which encodes a subunit of the heteromeric CO2 receptor, contributing to our understanding of mosquito attraction to humans [34]. This mutant, the first knock-in to be reported in mosquitoes, was generated by the disruptive insertion of a fluorescent reporter gene into the AeagGr3 locus. Such knock-in technology could also be used to facilitate in-frame insertions of protein tags into genes of interest, further enabling the study of complex pathways in mosquitoes (Figure 2f).

Homing endonucleases (HEGs) have also been successfully used to manipulate the mosquito genome [32,54,55]. HEGs are double-stranded DNases targeting large (12 to 40 bp) asymmetric recognition sites that occur extremely rarely in genomes [56]. An. gambiae strains have been generated that express I-Ppol, a HEG that recognizes and cuts a site in a multi-copy rDNA gene, which in this species is located exclusively on the X chromosome [35,57]. When I-Ppol is expressed specifically during spermatogenesis, it cleaves these multiple target sequences causing shredding of the paternal X chromosomes in sperm cells [35,57]. This feature was originally meant to generate male-only populations by preventing fathers from transmitting the X chromosome to embryos; but I-Ppol expression
in sperm cells induces complete embryonic lethality, probably as a consequence of the shredding of the maternal X chromosome upon unintended transfer of the enzyme to the embryo [57]. These strains induce a high level of infertility in large cage trials, as discussed below [58]. An improved version of these strains, which carries a less thermostable version of I-Ppol with reduced in vivo half life, has been generated that is instead active only in the testes, causing the specific shredding of the paternal X chromosome in sperm without directly affecting the embryo [35] (Figure 2g). The resulting sex-distorter strains produce >95% male offspring and are able to suppress wild-type mosquito populations in laboratory cages [35].

**Gene drives for population replacement**

For the implementation of population replacement strategies aimed at curbing mosquito-borne diseases, the anti-pathogen constructs described above need to be driven genetically through natural populations so that the disease refractory traits will spread (Figure 2h). A number of artificial gene-drive systems capable of forcing their own spread in a non-Mendelian manner are being developed that could be used for this purpose. In the model organism *Drosophila melanogaster*, the first gene-drive mechanism was developed on the basis of a toxin-antidote system [59]. This synthetic system, named *Medea* after the mythological figure of the woman who killed her own children to take revenge on her husband's betrayal, is based on expression in the zygote of a toxic gene, such as a microRNA against a maternal mRNA essential for embryonic development [59,60]. Transgenic females carry an ‘antidote’, that is, an allele of the gene that is insensitive to the toxin, allowing transgenic progeny to survive and spread the transgene. Although *Medea* has yet to be adapted to disease vectors, HEG-based technologies have been suggested and tested as gene drives in mosquitoes [36,61]. In this system, the drive encodes DNA-cutting machinery that cleaves a wild-type target locus from a transgene located at the homologous locus. Repair of the DNA break by homologous recombination causes the transgene to copy into the cleaved locus, causing a hemizygous cell to become homozygous for the transgene (Figure 2h). If this mechanism occurs in the germline, the transgene can spread through the population, potentially carrying an anti-pathogenic construct with it. Proof-of-principle use of HEGs to facilitate gene-drive mechanisms in *An. gambiae* was based on the I-SceI enzyme, which targeted its own recognition sequence that had been artificially introduced into a GFP reporter gene [36]. Homing of the HEG into its target sequence, previously integrated into the mosquito genome, would therefore generate GFP null mutants. Small cage experiments indicated that I-SceI could rapidly invade the receptive target strain, providing the first evidence of the gene-drive capabilities of HEGs in mosquitoes [36].

The range of applications enabled by HEGs and other nuclease-based technologies (ZFNs and TALENs) has some limitations, especially in terms of specificity, flexibility and stability. For example, ZFNs do not always have the desired sequence specificity when assembled into arrays, which limits the number of loci that can be targeted [62]. HEGs have been shown to cleave non-target sites (for a review see [63]), and laborious in vitro studies are necessary to generate new enzymes that have the required sequence specificity [64]. Furthermore, as these systems cut a single genomic sequence at a time, new transgenic strains must be created for each target sequence. A new genome-engineering tool, CRISPR/Cas9 (for clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9), has the potential to overcome these limitations and stimulate the generation of effective gene drives for vector control. Discovered as the molecular machinery of a bacterial acquired immune defense system [65], CRISPR/Cas9 was soon co-opted to engineer the genomes of a wide variety of organisms with high flexibility and efficiency [33]. Cas9 is an endonucleolytic protein that can recognize and cleave specific genomic sequences with the help of a small artificial guide RNA (gRNA). When the gRNA and Cas9 form a complex, they catalyze DNA cleavage upon recognition of the target site by the gRNA. The reliance on easily designed gRNAs for the recognition of target sequences results in a significant increase in the number of genomic loci that can be cleaved when compared to other systems, as RNA-guided engineering does not require modification of the Cas9 protein itself. Moreover, a number of loci can be targeted simultaneously by providing multiple gRNAs, thereby reducing the possible emergence of resistance to cleavage [37]. Although research demonstrating the use of CRISPR/Cas9 in mosquitoes has yet to be published, it is likely that this technology will soon enable the development of innovative and evolutionarily stable gene drives for the control of vector-borne diseases. Nevertheless, further research is needed to demonstrate the improved performance of this system over already existing technology, including minimizing off-target cleavage events and the possibility to revert the effects of the introduced gene architectures [37].

**Current field trials utilizing genetically modified mosquitoes to fight disease**

Intensive research is ongoing to generate improved engineered strains that are suitable for vector-control programs, but the first generation of genetically modified mosquitoes is already being released in the field. Since 2009, the UK-based biotech company Oxitec has been
pushing the boundaries of genetic control by operating the first releases of transgenic Ae. aegypti RIDL strains to suppress wild populations [66-69]. Their aim is to test the efficacy of these strains as a tool against dengue, a viral disease for which no vaccine or effective drugs are available. Repeated releases of the RIDL strain OX513A achieved a sizable reduction of wild populations, bringing new promise for disease control. The first program was operated on Grand Cayman Island, a British Territory in the Caribbean [66]. An average of 465 males/hectare (ha)/week were released across 10 hectares over a 4-week period, representing about 16% of the male population in the field. A total of 9.6% of fluorescent larvae were detected from eggs collected in ovitraps three weeks after the release, demonstrating that RIDL males could mate with wild females and sire progeny, despite their reduced field competitiveness. A subsequent program, using 3,500 males/ha/week, was carried out over a 23-week period and achieved 80% suppression of the wild population in a 16-ha area [67]. To accomplish this task, 3.3 million engineered males were reared and released, stressing the need to optimize mass-rearing protocols [69]. OX513A was also released in a forested area in Pahang, Malaysia, and transgenic males were shown to live as long as their wild-type brothers from the same laboratory strain, even if their dispersal ability was reduced [68]. Releases of OX513A are currently being performed in Brazil [69], where additional trials are planned and the mosquito production factory is being expanded. Large outdoor field cages have also been employed to test the potential use of the flightless Ae. aegypti fsRIDL strain [51,70]. This strain did not, however, achieve complete suppression of target populations, suggesting that it may not be suitable for large-scale releases [70]. Reduced mating competitiveness of transgenic males probably contributed to test failure but other explanations, including the different genetic backgrounds of released individuals and wild populations, have also been proposed [70].

In the case of malaria vectors, large caged laboratory trials have been established to test the mating competitiveness of sterile An. gambiae males carrying the HEG I-PpoI. When released at 5- to 10-fold coverage in large cages, I-PpoI males induced high levels of infertility, leading to the suppression of caged populations in 4 to 5 weeks, despite showing reduced mating competitiveness [58]. Males carrying a less thermostable version of I-PpoI, which causes sex distortion rather than male infertility, also achieved elimination of caged populations within six generations when released at a 3x ratio [35]. Before the field release of these strains is contemplated, their competitive performance and sterilizing activity will need to be tested in semi-field settings, such as those provided by large outdoor enclosures, where mosquitoes are exposed to normal environmental conditions and must produce appropriate swarming and mating behavior [71].

**Ecological hurdles and environmental and regulatory considerations**

The implementation of genetically modified mosquitoes in vector control programs is challenged by a number of ecological, environmental and regulatory issues (summarized in Figure 3). Two crucial behavioral components of the released males are dispersal ability, which affects the possibility of targeting populations in impenetrable regions [68], and mating competitiveness, especially for species with complex sexual behaviors [72]. Indeed, the mating fitness of released males has proven to be an important limiting factor in previous campaigns aimed at reducing the size of Anopheles populations (for a comprehensive discussion of these issues see [73] and references therein). Generally, anopheline species mate in

---

**Figure 3 Challenges for the field release of transgenic mosquitoes.** This scheme summarizes the ecological, behavioral and regulatory issues faced by disease control programs based on the release of genetically modified mosquitoes. Ecological requirements are shown in green, behavioral requirements in orange, while regulatory issues are presented in blue. Light-grey sections highlight operational tools that may be used to comply with the requirements. Behavioral requirements include key fitness parameters such as the dispersal ability and mating competitiveness of released males, and can be tested in large laboratory cage trials and then in semi-field settings to select the mosquito strains with the greatest probability of success. Ecological hurdles comprise heterogeneity in the genetics, behavior and natural habitats of vector species (biodiversity), and possible unintended side-effects on non-target species or on the ecosystem. Monitoring of these effects must be constantly in progress in the release phase. The risks, safety and specificity of the engineered strains need to be evaluated by appropriate regulatory agencies, and early public engagement is a priority.
elaborate swarms that are highly demanding energetically, and in which males are subject to strong competition to find a mate [74]. Reduction of competitiveness can be caused by a number of factors including but not limited to mass rearing, inbreeding, transposon expression and insertion sites in the genome [75-77]. The latter issue can now be partially overcome by utilizing ‘docking’ strains that are selected on the basis of limited fitness costs, using the PhiC31 integration system [78].

Other ecological features, including the biodiversity of native vector species, will also determine the success of a release campaign (Figure 3). Malaria transmission is supported by over 30 major primary vectors [79], many of which are morphologically indistinguishable [80]. These often sympatric species exhibit distinct behaviors in terms of mating, blood feeding and resting, and inhabit diverse ecological niches, making their control extremely arduous [81]. Such complexity represents a significant hurdle to the implementation of genetic engineering for malaria control; elimination of this disease solely by transgenic means would require the simultaneous release of all malaria-transmitting species in any given area, a highly arduous task. By contrast, dengue virus transmission is supported by over 30 major primary vectors [79], many of which are morphologically indistinguishable [80].

Increased interest in gene-drive applications with the undisputable benefits of a world free of vector-borne pathogens, while ensuring that possible unanticipated ecological and environmental consequences are eliminated.

**Abbreviations**

Cas9: CRISPR-associated protein 9; CRISPR: Clustered regularly interspaced short palindromic repeats; fRRL: Female-specific RIDL; GFP: Green fluorescent protein; gRNA: Guide RNA; HEG: Homing endonuclease; IMD: Immune deficiency pathway; IIS: Insulin-growth factor signaling; NHEJ: Non-homologous end-joining; ORCO: Odonate receptor co-receptor; PTEN: Phosphatase and tensin homolog; RIDL: Release of insects carrying a dominant lethal; scFv: Single chain variable fragment antibody; ST: Sterile insect technique; SM1: Salivary gland- and midgut-binding peptide 1; TALE: Transcription activator-like effector; TALEN: Transcription activator-like effector nuclease; TEP1: Thioester-containing protein 1; tRE: Tetracycline-responsive element; tTA: Tetracycline transactivator; ZF: Zinc finger; ZFN: Zinc finger nucleases.

**Competing interests**

The authors declare that they have no competing interests.

**Acknowledgements**

Work in our laboratories is supported by the European Research Council FP7 ERC Starting Grant (project ‘Anorep’ (grant ID: 260897)), the Wellcome Trust (grant ID: 093553), and an NIH grant (grant ID: NIH 1R01AI104956-01A1) awarded to FC.

**Author details**

1. University of Pavia, Pavia 27100, Italy. 2. Department of Immunology and Infectious Diseases, Harvard School of Public Health, Avenue Louis Pasteur, Boston, MA 02115, USA. 3. Department of Genetics, Harvard Medical School, Avenue Louis Pasteur, Boston, MA 02115, USA. 4. Department of Microbiology, Perugia University, Perugia 06100, Italy.

Reference:


APPENDIX D

Concerning RNA-guide gene drives for the alteration of wild populations
Concerning RNA-guided gene drives for the alteration of wild populations

Abstract Gene drives may be capable of addressing ecological problems by altering entire populations of wild organisms, but their use has remained largely theoretical due to technical constraints. Here we consider the potential for RNA-guided gene drives based on the CRISPR nuclease Cas9 to serve as a general method for spreading altered traits through wild populations over many generations. We detail likely capabilities, discuss limitations, and provide novel precautionary strategies to control the spread of gene drives and reverse genomic changes. The ability to edit populations of sexual species would offer substantial benefits to humanity and the environment. For example, RNA-guided gene drives could potentially prevent the spread of disease, support agriculture by reversing pesticide and herbicide resistance in insects and weeds, and control damaging invasive species. However, the possibility of unwanted ecological effects and near-certainty of spread across political borders demand careful assessment of each potential application. We call for thoughtful, inclusive, and well-informed public discussions to explore the responsible use of this currently theoretical technology.

DOI: 10.7554/eLife.03401.001

KEVIN M ESVELT*, ANDREA L SMIDLER, FLAMINIA CATTERUCCIA* AND GEORGE M CHurch*

Introduction

Despite numerous advances, the field of molecular biology has often struggled to address key biological problems affecting public health and the environment. Until recently, editing the genomes of even model organisms has been difficult. Moreover, altered traits typically reduce evolutionary fitness and are consequently eliminated by natural selection. This restriction has profoundly limited our ability to alter ecosystems through molecular biology.

If we could develop a general method of ensuring that engineered traits would instead be favored by natural selection, then those traits could spread to most members of wild populations over many generations. This capability would allow us to address several major world problems, including the spread of insect-borne diseases, the rise of pesticide and herbicide resistance, and the agricultural and environmental damage wrought by invasive species.

Scientists have long known of naturally occurring selfish genetic elements that can increase the odds that they will be inherited. This advantage allows them to spread through populations even if they reduce the fitness of individual organisms. Many researchers have suggested that these elements might serve as the basis for ‘gene drives’ capable of spreading engineered traits through wild populations (Craig et al., 1960; Wood et al., 1977; Sinkins and Gould, 2006; Burt and Trivers, 2009; Alphey, 2014). Austin Burt was the first to propose gene drives based on site-specific ‘homing’ endonuclease genes over a decade ago (Burt, 2003). These genes bias inheritance by cutting the homologous chromosome, inducing the cell to copy them when it repairs the break. Several efforts have focused
on the possibility of using gene drives targeting mosquitoes to block malaria transmission (Scott et al., 2002; Windbichler et al., 2007, 2008, 2011; Li et al., 2013a; Galizi et al., 2014). However, development has been hindered by the difficulty of engineering homing endonucleases to cut new target sequences (Chan et al., 2013a; Thyme et al., 2013; Takeuchi et al., 2014). Attempts to build gene drives with more easily retargeted zinc-finger nucleases and TALENs suffered from instability due to the repetitive nature of the genes encoding them (Simoni et al., 2014).

The recent discovery and development of the RNA-guided Cas9 nuclease has dramatically enhanced our ability to engineer the genomes of diverse species. Originally isolated from ‘CRISPR’ acquired immune systems in bacteria, Cas9 is a non-repetitive enzyme that can be directed to cut almost any DNA sequence by simply expressing a ‘guide RNA’ containing that same sequence. In little more than a year following the first demonstrations in human cells, it has enabled gene insertion, deletion, and replacement in many different species (Bassett et al., 2013; Cho et al., 2013; Cong et al., 2013; DiCarlo et al., 2013; Friedland et al., 2013; Gratz et al., 2013; Hu et al., 2013; Hwang et al., 2013; Jiang et al., 2013a, 2013b; Jinek et al., 2013; Li et al., 2013b; Mali et al., 2013c; Tan et al., 2013; Upadhyay et al., 2013; Wang et al., 2013b).

Building RNA-guided gene drives based on the Cas9 nuclease is a logical way to overcome the targeting and stability problems hindering gene drive development. Less obvious is the extent to which the unique properties of Cas9 are well-suited to overcoming other molecular and evolutionary challenges inherent to the construction of safe and functional gene drives.

We submit that Cas9 is highly likely to enable scientists to construct efficient RNA-guided gene drives not only in mosquitoes, but in many other species. In addition to altering populations of insects to prevent them from spreading disease (Curtis, 1968), this advance would represent an entirely new approach to ecological engineering with many potential applications relevant to human health, agriculture, biodiversity, and ecological science.

The first technical descriptions of endonuclease gene drives were provided by Austin Burt in his landmark proposal to engineer wild populations more than a decade ago (Burt, 2003). Any of the rapidly expanding number of laboratories with expertise in Cas9-mediated genome engineering could attempt to build a gene drive by substituting Cas9 for the homing endonucleases described in his proposal. Indeed, the well-recognized potential for gene drives to combat vector-borne diseases such as malaria and dengue virtually ensures that this strategy will eventually be attempted in mosquitoes.

While considerable scholarship has been devoted to the question of how gene drives might be safely utilized in mosquitoes (Scott et al., 2002; Touré et al., 2004; Benedict et al., 2008; Marshall, 2009; UNEP, 2010; Reeves et al., 2012; David et al., 2013; Alphey, 2014), few if any studies have examined the potential ecological effects of gene drives in other species. After all, constructing a drive to spread a particular genomic alteration in a given species was simply not feasible with earlier genome editing methods. Disconcertingly, several published gene drive architectures could lead to extinction or other hazardous consequences if applied to sensitive species, demonstrating an urgent need for improved methods of controlling these elements. After consulting with experts in many fields as well as concerned environmental organizations, we are confident that the responsible development of RNA-guided gene drive technology is best served by full transparency and early engagement with the public.

Here we provide brief overviews of gene drives and Cas9-mediated genome engineering, detail the mechanistic reasons that RNA-guided gene drives are likely to be effective in many species, and outline probable capabilities and limitations. We further propose novel gene drive architectures that may substantially improve our control over gene drives and their effects, discuss possible applications, and suggest guidelines for the safe development and evaluation of this promising but as yet unrealized technology. A discussion of risk governance and regulation intended specifically for policymakers is published separately (Oye et al., 2014).

**Natural gene drives**

In nature, certain genes ‘drive’ themselves through populations by increasing the odds that they will be inherited (Burt and Trivers, 2009). Examples include endonuclease genes that copy themselves into chromosomes lacking them (Burt and Koufopanou, 2004), segregation distorters that destroy competing chromosomes during meiosis (Lytte, 1991), transposons that insert copies of themselves elsewhere in the genome (Charlesworth and Langley, 1989), Medea elements that eliminate competing siblings who do not inherit them (Beeman et al., 1992; Chen...
Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

Craig et al., 1960; rally occurring ‘meiotic’ or ‘gametic’ drives that bias extinction due to the accumulated load of recessive alleles cause the population to crash or even become impaired heterozygotes when rare, and eventually drives would spread rapidly through minimally copies are lost (Craig et al., 1960; Lyttle, 1977; Burt, 2003; Schliekelman et al., 2005; Deredec et al., 2008, 2011; North et al., 2013; Burt, 2014).

Whether a standard gene drive will spread through a target population depends on molecular factors such as homing efficiency, fitness cost, and evolutionary stability (Marshall and Hay, 2012); only the rate of spread is determined by the mating dynamics, generation time, and other characteristics of the target population. In contrast, models suggest that the deleterious and complex effects of genetic load and sex-biasing suppression drives render them more sensitive to population-specific ecological variables such as density-dependent selection (Burt, 2003; Schliekelman et al., 2005; Huang et al., 2007; Deredec et al., 2008; Marshall, 2009; Yahara et al., 2009; Deredec et al., 2011; Alphey and Bonsall, 2014).

No engineered endonuclease gene drive capable of spreading through a wild population has yet been published. However, the Crisanti and Russell laboratories have constructed gene drives that can only spread through laboratory mosquito (Windbichler et al., 2011) and fruit fly (Chan et al., 2011; Simoni et al., 2014) populations that have been engineered to contain the endonuclease cut site. The Burt and Crisanti laboratories are attempting to build a male-biasing suppression drive using an endonuclease that serendipitously cuts a conserved sequence repeated hundreds of times in the X chromosome of the mosquito Anopheles gambiae (Windbichler et al., 2007, 2008; Galizi et al., 2014). If successful, their work promises to substantially reduce the population of this important malaria vector.

All engineered gene drives based on homing endonucleases cut the natural recognition site of the relevant enzyme. Despite early hopes, it has proven difficult to engineer homing endonucleases to cleave new target sequences. Numerous laboratories have sought to accomplish this goal for well over a decade with only a few recent successes (Chan et al., 2013a; Thyme et al., 2013; Takeuchi et al., 2014). More recently, a team
constructed new versions of the fruit fly gene drive using modular zinc-finger nucleases or TALENs in place of the homing endonuclease (Simoni et al., 2014), both of which can be engineered to cut new target sequences. While initially successful at cutting and homing, both declined in effectiveness over time due to the evolutionary instability of the modular repeats inherent to those proteins.

These early attempts demonstrate that it is possible to build synthetic gene drives, but also emphasize the importance of cutting any desired gene and remaining stable during copying. The recent discovery of the RNA-guided Cas9 nuclease represents a possible solution.

**RNA-guided genome editing via the Cas9 nuclease**

One straightforward method of genome editing relies on the same mechanism employed by endonuclease gene drives: cut the target gene and supply an edited version for the cell to use as a template when it fixes the damage. Most eukaryotic genome engineering over the past decade was accomplished using zinc-finger nucleases (Urnov et al., 2005) or TALENs (Christian et al., 2010), both of which are modular proteins that can be redesigned or evolved to target new sequences, albeit only by specialist laboratories (Esvelt and Wang, 2013). Genome editing was democratized by the discovery and adaptation of Cas9, an enzyme that can be programmed to cut target DNA sequences specified by a guiding RNA molecule (Deltcheva et al., 2011; Jinek et al., 2012; Cho et al., 2013; Cong et al., 2013; Mali et al., 2013a; Jinek et al., 2013; Mali et al., 2013c).

Cas9 is a component of Type II CRISPR acquired immune systems in bacteria, which allow cells to ‘remember’ the sequences of previously encountered viral genomes and protect themselves by recognizing and cutting those sequences if encountered again. They accomplish this by incorporating DNA fragments into a memory element, transcribing it to produce RNAs with the same sequence, and directing Cas9 to cut any matching DNA sequences (Deltcheva et al., 2011). The only restriction is that Cas9 will only cut target ‘protospacer’ sequences that are flanked by a protospacer-adjacent motif (PAM) at the 3’ end. The most commonly used Cas9 ortholog has a PAM with only two required bases (NGG) and therefore can cut protospacers found approximately every 8 base pairs (Jinek et al., 2012).

Remarkably, it is possible to direct Cas9 to cut a specific protospacer in the genome using only a single guide RNA (sgRNA) less than 100 base pairs in length (Jinek et al., 2012). This guide RNA must begin with a 17-20 base pair ‘spacer’ sequence identical to the targeted protospacer sequence in the genome (Fu et al., 2014).
Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

The process of editing a target gene involves choosing protospacers within the gene, building one or more guide RNAs with matching spacers, and delivering Cas9, guide RNAs, and an edited repair template lacking those protospacers into the cell (Figure 3).

Cas9 is efficient enough to cut and edit multiple genes in a single experiment (Li et al., 2013c; Wang et al., 2013a). The enzyme is active in a wide variety of organisms and is also quite specific, cutting only protospacers that are nearly identical to the spacer sequence of the guide RNA (Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013). Moreover, methods that allow Cas9 to bind but not cut enable the expression of target genes to be regulated by selectively recruiting regulatory proteins attached to Cas9 or the guide RNA (Gilbert et al., 2013; Mali et al., 2013a). All of these applications were developed within the last two years.

Because RNA-guided genome editing relies on exactly the same copying mechanism as endonuclease gene drives, it is reasonable to ask whether it might be possible to build gene drives based on Cas9. In principle, RNA-guided gene drives might be capable of spreading almost any genomic alteration that can be generated using Cas9 through sexually reproducing populations.

**Will RNA-guided gene drives enable us to edit the genomes of wild populations?**

Although we cannot be certain until we try, current evidence suggests that RNA-guided gene drives will function in some and possibly most sexually reproducing species. Learning how to insert a drive into the germline and optimize its function in each new species will likely require months to years depending on generation length, with subsequent drives in the same species taking less time. Because inserting the drive into the germline with Cas9 involves the same molecular copying process as the drive itself will utilize, successful insertion may produce a working if not particularly efficient RNA-guided gene drive. But if population-level engineering is to become a reality, all molecular factors relevant to homing – cutting, specificity, copying, and evolutionary robustness – must be considered. Below, we provide a detailed technical analysis of the extent to which Cas9 can address each of these challenges. Capabilities, limitations, control strategies, and possible applications are discussed in subsequent sections.

**Cutting**

The first requirement for every endonuclease gene drive is to cut the target sequence. Incomplete cutting was a problem for the homing endonuclease drive constructed in transgenic mosquitoes (72% cutting) and also for the homing endonuclease, zinc-finger nuclease, and TALEN drives in fruit flies (37%, 86%, and 70% cutting) (Windbichler et al., 2011; Chan et al., 2013b; Simoni et al., 2014). The simplest way to increase cutting is to target multiple adjacent sequences. However, this is impractical for homing endonucleases and quite difficult for zinc-finger nucleases and TALENs, as each additional sequence requires a new nuclease protein to be engineered or evolved and then co-expressed.

In contrast, the RNA-guided Cas9 nuclease can be readily directed to cleave additional sequences by expressing additional guide RNAs...
Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

Esvelt et al. eLife 2014;3:e03401. DOI: 10.7554/eLife.03401

Feature article

out of six tested guide RNAs (Kondo and Ueda, 2013). The two least effective guide RNAs individually cut at rates exceeding 12% and 56%, but exhibited cutting rates above 91% when combined. Using more than two guide RNAs should further enhance cutting. The notable success of Cas9-based genome engineering in many different species, including studies that targeted every gene in the genome (Shalem et al., 2014; Wang et al., 2014), demonstrates that most sequences can be efficiently targeted independent of species and cell type. Thus, RNA-guided gene drives should be capable of efficiently cutting any given gene.

Specificity

Because cutting other sites in the genome may seriously compromise the fitness of the organism, the second requirement is to avoid cutting non-targeted sequences.

While several studies have reported that Cas9 is prone to cutting off-target sequences that are closely related to the target (Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013), more recent developments and strategies designed to improve specificity (Mali et al., 2013a; Fu et al., 2014; Guilinger et al., 2014; Tsai et al., 2014) have demonstrated that the off-target rate can be reduced to nearly undetectable levels (Figure 4). Notably, Cas9 does not appear to represent a noticeable fitness burden when expressed at a moderate level in fruit flies with or without guide RNAs (Kondo and Ueda, 2013). Organisms with larger genomes may require more careful target site selection due to the increased number of potential off-target sequences present.

Copying

The third and most challenging requirement involves ensuring that the cut sequence is repaired using homologous recombination (HR) to copy the drive rather than the competing non-homologous end-joining (NHEJ) pathway (Figure 4). HR rates are known to vary across cell types (Mali et al., 2013c), developmental stages (Fiorenza et al., 2001; Preston et al., 2006), species (Chan et al., 2013b), and the phase of the cell cycle (Saleh-Gohari and Helleday, 2004). For example, the endonuclease gene drive in mosquitoes was correctly copied following ~97% of cuts (Windbichler et al., 2011), while a similar drive in fruit flies was initially copied only 2% of the time (Chan et al., 2011) and never rose above 78% even with extensive combinatorial optimization of promoter and 3′ untranslated region. This difference is presumably due to the use of different guide RNAs, which can affect the efficiency of HR.

(Figure 4) The sequences of these additional guide RNAs can be altered so as to avoid creating unstable repeats within the drive cassette (Nishimasu et al., 2014; Simoni et al., 2014). Including more guides has been demonstrated to improve upon already high rates of cutting. For example, fruit flies expressing both Cas9 and guide RNAs in their germline exhibited target cutting rates exceeding 85–99% in males for four
Figure 4. Technical advantages of RNA-guided gene drives. Clockwise from lower left: The targeting flexibility of Cas9 permits the exclusive selection of target sequences with few potential off-targets in the genome. Targeting multiple sites increases the cutting frequency and hinders the evolution of drive-resistant alleles, which must accumulate mutations at all of the sites. The Cas9 nuclease is can be quite specific in the sequences that it targets; fruit flies do not exhibit notable fertility or fitness defects resulting from off-target cutting when both Cas9 nuclease and guide RNAs are expressed in the germline (Kondo and Ueda, 2013). Choosing target sites with few or no close relatives in the genome, using truncated guide RNAs (Fu et al., 2014), employing paired Cas9 nickases (Mali et al., 2013a) instead of nucleases, or utilizing Cas9-FokI fusion proteins (Guilinger et al., 2014; Tsai et al., 2014) can further increase specificity. Several of these strategies can reduce the off-target mutation rate to borderline undetectable levels (Fu et al., 2014; Guilinger et al., 2014; Tsai et al., 2014). The frequency at which the drive is correctly copied might be increased by using Cas9 as a transcriptional regulator to activate HR genes and repress NHEJ genes (Gilbert et al., 2013; Mali et al., 2013a) (Figure 4—figure supplement 1). By choosing target sites within an essential gene, any non-homologous end-joining event that deletes all of the target sites will cause lethality rather than creating a drive-resistant allele, further increasing the evolutionary robustness of the RNA-guided gene drive. Other options include using distinct promoters and guide RNAs to avoid repetitiveness and increase stability (Figure 4—figure supplement 2) or employing newly characterized, engineered, or evolved Cas9 variants with improved properties (Esvelt et al., 2011; Mali et al., 2013b). These optimization strategies have also been summarized in tabular form with additional details (Figure 4—figure supplement 3). DOI: 10.7554/eLife.03401.005

The following figure supplements are available for figure 4:

**Figure supplement 1.** Enhancing drive copying by regulating endogenous genes. DOI: 10.7554/eLife.03401.006

**Figure supplement 2.** Repetitiveness and evolutionary stability of multiple guide RNAs. DOI: 10.7554/eLife.03401.007

**Figure supplement 3.** Table of known technological advances that might be adapted to optimize gene drive efficiency. DOI: 10.7554/eLife.03401.008
Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

due to a lower rate of HR in fruit fly spermatocytes relative to mosquitoes (Chan et al., 2013b). Ideally, drives should be activated only in germline cells at developmental stages with a high rate of HR, but this may be challenging in many species.

Copying efficiencies may also depend on whether the cut produces 5′-overhangs, 3′-overhangs, or blunt ends (Kuhar et al., 2014). Because Cas9 nickases can generate either overhang type while Cas9 nucleases produce blunt ends, the enzyme can be adapted to the needs of the cell type and organism.

The ability to regulate gene expression with Cas9 might be used to temporarily increase the rate of homologous recombination while the drive is active (Figure 4). For instance, the Cas9 nuclease involved in cutting might simultaneously repress (Gilbert et al., 2013) genes involved in NHEJ and therefore increase the frequency of HR (Bozas et al., 2009) if supplied with a shortened guide RNA that directs it to bind and block transcription but not cut (Bikard et al., 2013; Sternberg et al., 2014). Alternatively, an orthogonal nuclease-null Cas9 protein (Esvelt et al., 2013) encoded within the drive cassette could repress NHEJ genes and activate HR genes before activating the Cas9 nuclease. Lastly, Cas9 might be used directly recruit key HR-directing proteins to the cut sites, potentially biasing repair towards that pathway.

Evolutionary stability

Even a perfectly efficient endonuclease gene drive is vulnerable to the evolution of drive resistance in the population. Whenever a cut is repaired using the NHEJ pathway, the result is typically a drive-resistant allele with insertions or deletions in the target sequence that prevent it from being cut by the endonuclease. Natural sequence polymorphisms in the population could also prevent cutting. These alleles will typically increase in abundance and eventually eliminate the drive because most drives – like most transgenes – are likely to slightly reduce the fitness of the organism. A second path to gene drive resistance would involve the target organism evolving a method of specifically inhibiting the drive endonuclease.

The best defense against previously existing or recently evolved drive-resistant alleles is to target multiple sites. Because mutations in target sites are evolutionarily favored only when they survive confrontation with the gene drive, using many target sites can render it statistically improbable for any one allele to survive long enough to accumulate mutations at all of the sites so long as cutting rates are high (Burt, 2003). However, very large populations – such as those of some insects – might require unfeasibly large numbers of guide RNAs to prevent resistance. In these cases it may be necessary to release several successive gene drives, each targeting multiple sites, to overcome resistant alleles as they emerge. From an evolutionary perspective, the ability to preclude resistance by targeting multiple sites is the single greatest advantage of RNA-guided gene drives.

We propose to extend this strategy by preferentially targeting multiple sites within the 3′ ends of genes important for fitness such that any repair event that deletes all of the target sites creates a deleterious allele that cannot compete with the spread of the drive (Figure 4). Whenever the drive is copied, the cut gene is replaced with a recoded version flanked by the other components of the drive. Recent work has demonstrated that most genes can be substantially recoded with little effect on organism fitness (Lajoie et al., 2013; the 3′ untranslated region might be replaced with an equivalent sequence from a related gene. Because there would be no homology between the recoded cut site and the drive components, the drive cassette would always be copied as a unit.

Relative to drive-resistant alleles, inhibitors of Cas9 are less likely to arise given the historical absence of RNA-guided nucleases from eukaryotes. Any inhibitors that do evolve would presumably target a particular Cas9 protein or guide RNA used in an earlier drive and could be evaded by building future drives using a Cas9 ortholog with a different guide RNA (Esvelt et al., 2013; Fonfara et al., 2013). Alternatively and least likely, organisms might evolve higher RNase activity to preferentially degrade all guide RNAs; this may be difficult to accomplish without harming overall fitness.

A final evolutionary concern relates to the stability of the gene drive cassette itself. The zinc-finger nuclease and TALEN-based gene drives in fruit flies suffered from recombination between repetitive sequences: only 75% and 40% of each respective drive was sufficiently intact after one copying event to catalyze a second round of copying. Because RNA-guided gene drives will not include such highly repetitive elements, they are likely to be more stable (Figure 4).

Development time

RNA-guided genome editing is advancing at a historically unprecedented pace. Because it is
now much easier to make transgenic organisms and therefore candidate gene drives, the design-build-test cycle for gene drives will often be limited only by the generation time of the organism in the laboratory. Moreover, many advances from genome engineering can be directly applied to RNA-guided gene drives. For example, all of the methods of increasing Cas9 specificity described above were developed for RNA-guided genome editing in the past 2 years. Future methods of increasing the rate of HR relative to NHEJ would be useful for both technologies. These factors suggest that scientists will enjoy an increasing number of tools well-suited to rapidly building and testing gene drives in addition to those we describe above.

None of this is to gloss over the many practical difficulties that are likely to arise when constructing a particular gene drive in a given species. Early success is as unlikely as ever when engineering complex biological systems. But if half a dozen or even a dozen design-build-test cycles are sufficient to produce moderately efficient gene drives, many molecular biology laboratories around the world will soon be capable of engineering wild populations.

**Gene drive limitations**

Given their potentially widespread availability, it will be essential to develop a comprehensive understanding of the fundamental limitations of genetic drive systems.

First and most important, gene drives require many generations to spread through populations. Once transgenic organisms bearing the gene drive are constructed in the laboratory, they must be released into the wild to mate with wild-type individuals in order to begin the process of spreading the drive through the wild population. The total time required to spread to all members depends on the number of drive-carrying individuals that are released, the generation time of the organism, the efficiency of homing, the impact of the drive on individual fitness, and the dynamics of mating and gene flow in the population, but in general it will take several dozen generations (Burt, 2003; Huang et al., 2007; Dereced et al., 2008; Marshall, 2009; Yahara et al., 2009; Dereced et al., 2011). Thus, drives will spread very quickly in fast-reproducing species but only slowly in long-lived organisms.

Second, gene drives cannot affect species that exclusively practice asexual reproduction through clonal division or self-fertilization. This category includes all viruses and bacteria as well as most unicellular organisms. Highly efficient standard drives might be able to slowly spread through populations that employ a mix of sexual and asexual reproduction, such as certain plants, but drives intended to suppress the population would presumably force target organisms to reproduce asexually in order to avoid suppression.

Third, drive-mediated genome alterations are not permanent on an evolutionary timescale. While gene drives can spread traits through populations even if they are costly to each individual organism, harmful traits will eventually be out-competed by more fit alleles after the drive has gone to fixation. Highly deleterious traits may be eliminated even more quickly, with non-functional versions appearing in large numbers even before the drive and its cargo can spread to all members of the population. Even when the trait is perfectly linked to the drive mechanism, the selection pressure favoring the continued function of Cas9 and the guide RNAs will relax once the drive reaches fixation. Maintaining deleterious traits within a population indefinitely is likely to require scheduled releases of new RNA-guided gene drives to periodically overwrite the broken versions in the environment.

Fourth, our current knowledge of the risk management (Scott et al., 2002; Touré et al., 2004; UNEP, 2010; McGraw and O’Neill, 2013;
Alphey, 2014) and containment (Benedict et al., 2008; Marshall, 2009) issues associated with gene drives is largely due to the efforts of researchers focused on mosquito-borne illnesses. Frameworks for evaluating ecological consequences are similarly focused on mosquitoes (David et al., 2013) and the few other organisms for which alternative genetic biocontrol methods have been considered (Dana et al., 2014). While these examples provide an invaluable starting point for investigations of RNA-guided gene drives targeting other organisms, studies examining the particular drive, population, and associated ecosystem in question will be needed.

**Safeguards and control strategies**

Given the potential for gene drives to alter entire wild populations and therefore ecosystems, the development of this technology must include robust safeguards and methods of control (Oye et al., 2014). Whereas existing gene drive proposals focus on adding genes (Ito et al., 2002), disrupting existing genes (Burt, 2003), or suppressing populations, RNA-guided gene drives will also be capable of replacing existing sequences with altered versions that have been recoded to remove the sites targeted by the drive (Figure 3). We hypothesize that the unique ability of RNA-guided gene drives to target any gene may allow them to control the effects of other gene drives or transgenes.

**Reversibility**

RNA-guided gene drives could reverse genome alterations that have already spread through populations. Suppose a given gene drive causes unexpected side-effects or is released without public consent. A ‘reversal’ drive released later could overwrite one or all genomic changes spread by the first drive (Figure 5A). The new sequence spread by the reversal drive must also be recoded relative to the original to keep the first drive from cutting it, but any amino acid changes introduced by the first drive could be undone. If necessary, a third drive could restore the exact wild-type sequence, leaving only the guide RNAs and the gene encoding Cas9 as signatures of past editing (Figure 5B).

The ability to update or reverse genomic alterations at the speed of a drive, not just a drive-resistant allele, represents an extremely important safety feature. Reversal drives could also remove conventionally inserted transgenes that entered wild populations by cross-breeding or natural mutations that spread in response to human-induced selective pressures. However, it is important to note that even if a reversal drive were to reach all members of the population, any ecological changes caused in the interim would not necessarily be reversed.

**Immunization**

RNA-guided gene drives could be used to block the spread of other gene drives. For example, an ‘im Immunizing’ drive could prevent a specific unwanted drive from being copied by recoding sequences targeted by the unwanted drive (Figure 5A). This could be done preemptively or reactively and would spread on a timescale comparable to that of the unwanted drive. A combined ‘immunizing reversal’ drive might spread through both wild-type individuals and those affected by an earlier gene drive, converting both types to a recoded version that could not be invaded by the unwanted drive (Figure 5B). This may represent the fastest method of neutralizing an already-released drive. As with a standard reversal drive, any ecological changes caused in the interim would not necessarily be reversed.

**Precisely targeting subpopulations**

RNA-guided gene drives might be confined to a single genetically distinct target species.
or even a subpopulation by targeting unique genes or sequence polymorphisms. Because these ‘precision drives’ will only cut the unique sequence, they will not be able to spread through non-target populations as long as that sequence is sufficiently distinct. We estimate that either the PAM or at least five base pairs of the spacer must differ within each target site in order to prevent the guide RNAs in the drive from evolving to recognize the equivalent non-target sequence (Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013a; Pattanayak et al., 2013).

Populations that are not genetically distinct but experience only intermittent gene flow, such as those on islands, might be given a unique sequence permitting them to be specifically targeted by precision drives later on. For example, releasing Drive A into the island population...
would recode a target gene, but exhibit no other effect (Figure 5A). Releasing Drive B, a precision drive which would exclusively spread through Drive A but not the wild-type allele, would similarly replace Drive A with a unique sequence. So long as Drive A does not escape the island before being replaced, the unique sequence in the island population would allow it to be targeted with future precision drives that could not spread through mainland populations (Figure 5C).

**Limiting population suppression**

Population suppression may be one of the most powerful applications of gene drives. The previously described genetic load and sex-biasing drives (Burt, 2003) could potentially lead to extinction (Deredec et al., 2008, 2011). While this outcome may be necessary to achieve compelling goals such as the eradication of malaria, other situations may call for more refined methods. Here we outline a handful of alternative architectures that would provide greater control over the extent of population suppression.

Chemical approaches to population control might utilize ‘sensitizing drives’ to render target organisms vulnerable to a particular molecule using one of three strategies (Figure 6). First, a sensitizing drive might reverse known mutations that confer resistance to existing pesticides or herbicides. Second, it might carry a prodrug-converting enzyme (Schellmann et al., 2010) that would render a prodrug molecule toxic to organisms that express it. Third, it could swap a conserved gene for a version that is strongly inhibited by a particular small molecule. Because sensitizing drives would have no effect in the absence of their associated molecule – and in some cases vice versa – they could grant very fine control over the geography and extent of population suppression with minimal ecological risk.

Temporal approaches to controlling populations would deliberately limit the lifetime of a suppression drive by rendering its effects evolutionarily unstable (Figure 6). For example, a male-determining or female-specific sterility gene carried by a standard drive on an autosome would suppress the target population, but the effect would be short-lived because any drive that acquired a loss-of-function mutation in the cargo gene would be strongly favored by natural selection due to its ability to produce fertile female offspring. Notably, turning existing female-specific sterility lines (Fu et al., 2010; Labbe et al., 2012; Alphey, 2014) into unstable drives may increase their effectiveness. Periodically releasing organisms carrying new unstable drives that are capable of replacing earlier broken versions could extend the suppression effect.

Genetic approaches to population control might initiate suppression only when two distinct ‘interacting drives’ encounter one another (Figure 6). For example, a cross between standard drives A and B might produce sterile females and fertile males that pass on the ‘sterile-daughter’ trait when crossed with females of any type. Scattering A- and B-carrying individuals throughout an existing population would produce many tiny pockets dominated by either A or B and very few organisms in between due to the infertility of AB females. Because each drive would spread from a small number of initially released individuals scattered over a wide area, this strategy may be capable of large-scale population suppression, but its effectiveness and resolution will depend on the average distance between released A and B individuals. Further suppressing the residual A and B populations could be accomplished by releasing only members of the opposite drive type. Modeling studies will be needed to determine whether this possibility is feasible for different species. Interestingly, the use of this drive type would effectively induce speciation in the affected population.

Finally, immunizing drives might protect specific subpopulations from the effects of full-scale suppression drives released elsewhere (Figure 6). Assuming some degree of gene flow, the immunized population will eventually replace the suppressed population, though this might be delayed if crossing the two drives generates a sterile-daughter effect as described above. Due to the comparatively uncontrolled spread of both drive types through the wild-type population, this method would only be suited to large geographic areas or subpopulations with limited gene flow. For example, immunization might be used to protect the native population of a species while suppressing or eradicating populations on other continents.

**Applications of RNA-guided gene drives**

RNA-guided gene drives have the potential to merge the fields of genomic and ecological engineering. They may enable us to address numerous problems in global health, agriculture, sustainability, ecological science, and many other areas (Figure 7). Of these opportunities, perhaps the most compelling involve curtailing the spread...
### Suppression Drive Type

| Chemical | Drive Confers Sensitivity to a Small Molecule |
| "Sensitizing Drive" | ![Diagram](image1.png) |
| Temporal | Suppression Drive is Unstable |
| "Unstable Drive" | ![Diagram](image2.png) |
| Genetic | Cross with Neighbour Initiates Suppression |
| "Interacting Drives" | ![Diagram](image3.png) |
| Genetic | Immunizing Drive Blocks Suppression Drive |
| "Immunizing Drive" | ![Diagram](image4.png) |

### Spatial Resolution

- Field
- City
- State
- Island/Continent

#### Figure 6

Controlling population suppression. Previously proposed genetic load and meiotic suppression drives spread without limit and may incur a substantial risk of extinction. Alternative gene drive types might be used to grant finer control over the extent of suppression. ‘Sensitization drives’ would be harmless save for conferring vulnerability to a particular chemical, which could then be used as a population-specific pesticide. Evolutionarily ‘unstable drives’ would place a limit on the average number of drive copying events and thus the extent of population suppression. ‘Interacting drives’ would initiate suppression only upon encountering a specific genetic signature in the population, in this case a different gene drive. The combination would create a sterile-daughter effect (Figure 6—figure supplement 1) capable of continuing suppression for several generations. Finally, an immunizing drive could protect a subpopulation from a full genetic load or male-biasing suppression drive employed elsewhere. Interacting drive and immunizing drive approaches would be effective on very large populations spread across substantial geographic areas (Figure 6—figure supplement 2) while suffering from correspondingly reduced geographic resolution and greater ecological risk (Figure 6—figure supplement 3). Resolutions are approximations only and will vary with the specific drive utilized in each class.

DOI: 10.7554/eLife.03401.010

The following figure supplements are available for figure 6:

- Figure supplement 1. Sample interacting drives that produce a sterile-daughter effect. DOI: 10.7554/eLife.03401.011

*Figure 6. Continued on next page*
of vector-borne infectious diseases, controlling agricultural pests, and reducing populations of environmentally and economically destructive invasive species.

**Eradicating insect-borne diseases**

The human toll inflicted by infectious diseases spread by insects is staggering. Malaria alone kills over 650,000 people each year, most of them children, while afflicting 200 million more with debilitating fevers (WHO, 2013). Dengue, yellow fever, chikungunya, trypanosomiasis, leishmaniasis, Chagas, and Lyme disease are also spread by insects. These afflictions could potentially be controlled or even eradicated by altering vector species to block transmission. Several laboratories have identified candidate gene disruptions or transgenes that interfere with the transmission of malaria (Ito et al., 2002; Dong et al., 2011; Isaacs et al., 2012) and other well-studied diseases (Franz et al., 2006). Depending on their effectiveness, these alterations may or may not allow the disease to be eradicated before the pathogen evolves resistance. Alternatively, the relevant vector species might be suppressed or eliminated using RNA-guided gene drives, then potentially reintroduced from sheltered laboratory or island populations once disease eradication is complete. In the case of malaria, gene drive strategies may represent particularly effective solutions to the emerging problem of mosquito vectors with an evolved preference to bite and rest outdoors, traits that render them resistant to current control strategies based on indoor insecticide spraying and bednets.

**Agricultural safety and sustainability**

The evolution of resistance to pesticides and herbicides is a major problem for agriculture. It has been assumed that resistant populations will remain resistant unless the relevant alleles impose a substantial fitness cost in the absence of pesticide or herbicide. We propose that RNA-guided sensitizing drives might replace resistant alleles with their ancestral equivalents to restore vulnerability. For example, sensitizing drives could potentially reverse the mutations allowing the western corn rootworm to resist Bt toxins (Gassmann et al., 2014) or horseweed and pigweed to resist the herbicide glyphosate (Gaines et al., 2010; Ge, 2010), which is currently essential to more sustainable no-till agriculture. Because these three organisms undergo one generation per year, comparatively large numbers of drive-bearing individuals must be released to quickly exert an effect, but fewer than are already released to control pests using the sterile-insect technique (Gould and Schliekelman, 2004; Dyck et al., 2005). Releases would need to occur in local areas not treated with pesticide or herbicide, which would quickly become reservoirs of sensitizing drives that could spread into adjacent
Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

Esvelt et al. eLife 2014;3:e03401. DOI: 10.7554/eLife.03401

Figure 7. Potential applications of RNA-guided gene drives. Clockwise from left. Disease vectors such as malarial mosquitoes might be engineered to resist pathogen acquisition or eliminated with a suppression drive. Wild populations that serve as reservoirs for human viruses could be immunized using Cas9, RNAi machinery, or elite controller antibodies carried by a gene drive. Reversal and immunization drives could help ensure that all transgenes are safe and controlled. Drives might quickly spread protective genes through threatened or soon-to-be-threatened species such as amphibians facing the expansion of chytrid fungus (Rosenblum et al., 2010). Invasive species might be locally controlled or eradicated without directly affecting others. Sensitizing drives could improve the sustainability and safety of pesticides and herbicides. Gene drives could test ecological hypotheses concerning gene flow, sex ratios, speciation, and evolution. Technical requirements for these applications vary with the drive type required (Figure 7—figure supplement 1).
DOI: 10.7554/eLife.03401.014

The following figure supplement is available for figure 7:
Figure supplement 1. Technical limitations of different gene drive architectures with implications for various applications.
DOI: 10.7554/eLife.03401.015

A second form of sensitizing drive could potentially render pest populations vulnerable to molecules that never previously affected them. For example, a gene important to fitness might be replaced with a version from another species or laboratory isolate whose function is sensitive to a particular compound. In principle, this approach could eventually lead to the development and use of safer and more species-specific pesticides and herbicides.

Controlling invasive species

One of the most environmentally damaging consequences of global economic activity is the introduction of invasive species, which often cause ecological disruption or even the extinction of native species. Isolated ecosystems such as those on small islands are especially vulnerable. RNA-guided suppression drives might be used to promote biodiversity by controlling or even eradicating invasive species from islands or possibly entire continents. The economics of invasive species control are also compelling: the top ten invasives in the United States cause an estimated $42 billion in damages every year (Pimentel et al., 2005). Black and brown rats alone cause $19 billion in damages and may be responsible for more extinctions than any other nonhuman species.

Deploying RNA-guided suppression drives against invasive species will incur two primary risks related to undesired spread. First, rare mating events may allow the drive to affect closely related species. Using precision drives to target sequences unique to the invasive species could mitigate or eliminate this problem. Second, the suppression drive might spread from the invasive population back into the native habitat, perhaps even through intentional human action.

Native populations might be protected using an immunizing drive, but doing so would risk transferring immunity back into invasive populations. Instead, we might grant the invasive population a unique sequence with a standard drive (Figure 5C), verify that these changes have not spread to the native population, and only then release a suppression drive targeting the recoded sequences while holding an immunizing drive in reserve. Another approach might utilize a sensitizing drive to render all populations newly
vulnerable to a specific compound, which could then be used as a pesticide for the local control of invasive populations. All of these possibilities will require modeling and experimentation to establish safety and feasibility before use.

Most importantly, all decisions involving the use of suppression drives must involve extensive deliberations including but not limited to ecologists and citizens of potentially affected communities.

Development and release precautions

Because any consequences of releasing RNA-guided gene drives into the environment would be shared by the local if not global community, research involving gene drives capable of spreading through wild-type populations should occur only after a careful and fully transparent review process. However, basic research into gene drives and methods of controlling their effects can proceed without risking this type of spread so long as appropriate ecological or molecular containment strategies are employed (Figure 8).

A great deal of information on probable ecological outcomes can be obtained without testing or even building replication-competent gene drives. For example, early studies might examine possible ecological effects by performing contained field trials with organisms that have been engineered to contain the desired change but do not possess a functional drive to spread it. Because they do involve transgenic organisms, these experiments are not completely without risk, but such transgenes are unlikely to spread in the absence of a drive.

We recommend that all laboratories seeking to build standard gene drives capable of spreading through wild populations simultaneously create reversal drives able to restore the original phenotype. Similarly, suppression drives should be constructed in tandem with a corresponding immunizing drive. These precautions would allow the effects of an accidental release to be swiftly if partially counteracted. The prevalence of gene drives in the environment could in principle be monitored by targeted amplification or metagenomic sequencing of environmental samples. Further investigation of possible monitoring strategies will be needed.

Transparency, public discussion, and evaluation

Technologies with the potential to significantly influence the lives of the general public demand societal review and consent. As self-propagating alterations of wild populations, RNA-guided gene drives will be capable of influencing entire ecosystems for good or for ill. As such, it is imperative that all research in this nascent field operate under conditions of full transparency, including independent scientific assessments of probable impacts and thoughtful, informed, and fully inclusive public discussions.

The decision of whether or not to utilize a gene drive for a given purpose should be based entirely on the probable benefits and risks of...
that specific drive. That is, each drive should be judged solely by its potential outcomes, such as its ability to accomplish the intended aims, its probable effects on other species, the risk of spreading into closely related species by rare mating events, and impacts on ecosystems and human societies. As scientists developing these technologies, it will be our responsibility to make all empirical data and predictive models freely available to the public in a transparent and understandable format. Above all else, we must openly share our level of confidence in these assessments as we determine how best to proceed.

Discussion

The potentially widespread implications of RNA-guided gene drives demand a thoughtful and collected response. Numerous practical difficulties must be overcome before gene drives will be in a position to address any of the suggested applications. Many of our proposals and predictions are likely to fall short simply because biological systems are complex and difficult to engineer. Even so, the current rate of scientific advancement related to Cas9 and the many outcomes accessible using the simplest of gene drives suggest that molecular biologists will soon be able to edit the genomes of wild populations, reverse or update those changes in response to field observations, and perhaps even engage in targeted population suppression.

What criteria might we use to evaluate an RNA-guided gene drive intended to address a given problem? There are compelling arguments in favor of eliminating insect-borne human diseases, developing and supporting more sustainable agricultural models, and controlling environmentally damaging invasive species. At the same time, there are valid concerns regarding our ability to accurately predict the ecological and human consequences of these interventions. By bringing these possibilities before the scientific community and the public prior to their realization in the laboratory, we hope to initiate transparent, inclusive, and well-informed discussions concerning the responsible evaluation and application of these nascent technologies (Oye et al., 2014).

Acknowledgements

We thank the participants of the MIT/Woodrow Wilson Center workshop ‘Creating a Research Agenda for the Ecological Implications of Synthetic Biology’ for sharing their thoughts concerning the potential impacts of this technology on various ecosystems (Drinkwater et al., 2014). We are grateful to S Lightfoot, K Oye, JP Collins, and M Wilkinson for critical reading of the manuscript, and to F Gould, T Wu, J Braff, and members of the Church and Oye laboratories for helpful discussions. KME was supported by a Wyss Technology Development Fellowship.

Funding

<table>
<thead>
<tr>
<th>Funder</th>
<th>Grant reference number</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wyss Institute for Biologically Inspired Engineering</td>
<td>Technology Development Fellowship</td>
<td>Kevin M Esvelt</td>
</tr>
<tr>
<td>Wyss Institute for Biologically Inspired Engineering</td>
<td>Synthetic Biology Platform funds</td>
<td>Kevin M Esvelt, Andrea L Smidler</td>
</tr>
</tbody>
</table>

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

KME, Conceived of the study, performed the analysis, designed new drive types, drafted and revised the article; ALS, Designed new drive types, drafted and revised the article; FC, GMC, Revised the article

Kevin M Esvelt Synthetic Biology Platform, Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, United States

Andrea L Smidler Synthetic Biology Platform, Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, United States; Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, United States

Flaminia Catteruccia Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, United States; Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università degli Studi di Perugia, Terni, Italy

George M Church Synthetic Biology Platform, Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, United States

Competing interests: KME: Filed for a patent concerning RNA-guided gene drives. ALS: Filed for a patent concerning RNA-guided gene drives. GMC: Filed for a patent concerning RNA-guided gene drives. The other author declares that no competing interests exist.

Received 18 May 2014
Accepted 09 July 2014
Published 17 July 2014
References


Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations


Emerging technology | Concerning RNA-guided gene drives for the alteration of wild populations

290


