IMPERIAL COLLEGE LONDON

DOCTORAL THESIS

Engineering a Minimal Gene Drive System for Integral Replacement in *Drosophila melanogaster*

Author:
Alexander NASH

Supervisor:
Dr. Nikolai WINDBICHLER

Dr. Giorgio GILESTRO

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Windbichler Group
Department of Life Sciences

May 19, 2020

Declaration of Authorship

I, Alexander Nash, declare that the work contained in this thesis is entirely of my own doing, and that where I have used the work of others it has been duly acknowledged and cited.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

IMPERIAL COLLEGE LONDON

Abstract

Faculty of Natural Sciences
Department of Life Sciences

Doctor of Philosophy

Engineering a Minimal Gene Drive System for Integral Replacement in *Drosophila melanogaster*

by Alexander NASH

Gene drives represent a powerful tool for the control of vector-borne diseases. By suppressing or replacing vector populations, laboratory studies have highlighted the potential for this group of tools to make a powerful impact on the burden of zoonotic disease. Current genetic drive systems have a number of limitations, namely their complexity, susceptibility to genetic resistance, and a high regulatory threshold. Here we suggest a novel design paradigm for the creation of replacement gene drives, which we have termed 'Integral Replacement'. By splitting drive constructs, and integrating components into endogenous loci, we have aimed to engineer a minimal drive system, with low fitness cost, higher resilience to resistance alleles, and with greater flexibility for field testing. In so doing we have generated a model that illustrates increases in efficacy versus existing drive systems, and expanded on work performed on intronic gRNA cassettes. We subsequently were able to build prototype Integral Gene Drive (IGD) components, and demonstrate their efficacy using the model organism *Drosophila melanogaster*, providing evidence for an initial proof-of-principle for this novel design paradigm.

Acknowledgements

Undertaking the work in this thesis is undoubtedly the greatest academic challenge I've faced, and it would not have been possible without a whole cast of supporters. So many people have contributed so much, and I'll do my best to thank them now. To Nikolai - thank you for time, patience, and advice. It's helped to keep me grounded, and focus on the bigger picture. Your allusions to Marx and Dune will be missed. Thanks to Giorgio for lending your advice, and supporting the path I chose, even though it diverged from our original plan. Thanks to the BBSRC for providing the means to carry out this work - providing equitable support for aspiring scientists remains an essential duty, long may it continue.

Thank you to everyone in the Windbichler and Christophides groups, for your constant support and friendship. Barbara, Angela, Astrid, Dave, and Paolo - my fly buddies. Thank you for teaching me so much about insect genetics and for always listening to my ideas with an open mind. To George Jr, Manjunatha, Paolo, Sofia, Tib, and Jan - your insight was invaluable, and much of this work has benefited greatly from your input. I owe much to George K. Christophides and Tony Southall, my review panel. Your insight and experience have helped me keep things together and on track. I must thank Tony in particular for being a constant sounding board for my ideas, and for all his direction on the finer points of fly pushing. The modelling conducted in this thesis owes much to Andrea K. Beaghton. Without a doubt, the work we conducted together was some of the most enriching for me. Thank you for taking the time to share your knowledge and for helping me understand the nuances and complexities of what we hope to achieve.

I would be remiss if I didn't acknowledge the tidal wave of support and encouragement I also received from the Crisanti group. Each and every one of you has helped in part to make this experience an enjoyable one. To Roberto, thank you for guidance and knowledge - you always had an answer when I was in doubt. Maria, your limitless determination and unabashed positivity were a regular source of strength for me. Drew and Kryos - the dream team. Thank you for always making time for my endless questions. Nace, thank you for always letting me rant about statistics and for setting me straight. Will, your extensive insight into amplicon sequencing helped to allay much of my confusion. Chryssa, your advice on flow cytometry was invaluable. Matt and Chris, thank you for stopping me from wrecking the insectary and for having a solution no matter the problem.

Along the way, I had the pleasure of working with a number of students. Tong, Giulia, and Oliver - your enthusiasm, dedication, and friendship kept me going when experiments failed and biology decided to ignore my plans. You all gave a great deal and I hope you enjoyed your time as much as I did. To my dear friends, Matt and Marta, thank you for keeping me sane. You always put a smile on my face. To Kadri Oras and Alla Madich at the University of Cambridge - the egg whisperers. Thank you for all the knowledge you lent, and for the embryos you injected. I'd also like to take the opportunity to thank all my friends and colleagues at Synthace,

for the opportunity to intern, and ultimately to join as a permanent team member. Vishal, Lily, Lucia, and Markus, thank you for the chances you have given me, and for all the support you have provided.

To my parents and my brother, thank you for the unending well of encouragement, resolve, and happiness - now and always.

None of this would have been possible without Katherine. No matter the scale of the challenge, you always helped me push forward. Your courage and brilliance have always been my greatest source of inspiration. This work is dedicated to you.

Personal Bibliography

Works published during PhD candidature: Nash, Alexander et al. (2019). "Integral gene drives for population replacement." In: Biology open 8.1, bio037762. ISSN: 2046-6390. DOI: 10.1242/bio.037762.

Contents

Declaration of Authorship i

Abstract ii

Acknowledgements iii

1 Introduction 1

1.1 The Challenge of Vector-Borne Disease . . . 1
1.1.1 The Vector . . . 2
1.1.2 The Parasite . . . 4
Blood-Stages of the *Plasmodium* Parasite . . . 4
Mosquito Stages of the *Plasmodium* Parasite . . . 5
1.2 Conventional Vector Control Strategies for Malaria . . . 7
Biological Controls . . . 8
Habitat Management . . . 9
Insecticides and Resistance . . . 9
1.3 Genetic Control Strategies . . . 10
1.3.1 Sterile Insect Technique . . . 11
1.3.2 The Advent of Insect Transgenesis . . . 12
1.3.3 RIDL Systems . . . 14
1.4 Introducing Gene Drive Systems . . . 16
1.4.1 Meiotic Drive Systems . . . 16
1.4.2 Underdominance Gene Drive Systems . . . 18
1.4.3 Using *Wolbachia* - A Naturally Occurring Vector Control System . . . 19
1.4.4 Killer-Rescue Systems . . . 22
1.4.5 Homing Gene Drive Systems . . . 24
1.5 Introducing CRISPR-Based Drives . . . 27
1.5.1 The Fundamentals of CRISPR-Mediated Immunity . . . 27
Limitations of CRISPR Editing . . . 29
1.5.2 Implementing CRISPR-mediated Homing Drives . . . 30
Biosafety and Ethical Considerations . . . 32
1.5.3 Population Modification Approaches . . . 35
Transmssion Blocking and Anti-microbial Peptides . . . 36
Antibodies . . . 37
Anti-Malarial Immune Regulators . . . 37
1.5.4 Current Shortcomings in Gene Drive Designs . . . 39

1.5.5 Introducing Integral Gene Drives . . . 40
1.5.6 Aims and Hypothesis . . . 43

2 General Materials and Methods 44
2.1 Chemicals . . . 44
2.2 Bacterial Transformation . . . 44
2.3 Alkaline Lysis Preparations of Plasmids . . . 45
2.3.1 Miniprep of Plasmid DNA . . . 45
2.3.2 Midiprep of Plasmid DNA . . . 45
2.4 Quantifying Plasmid and gDNA Concentrations . . . 45
2.5 Restriction Digestions . . . 45
2.6 Gel Purification of DNA . . . 46
2.7 Ligation Reactions . . . 46
2.8 Colony PCR . . . 46
2.9 Sanger Sequencing of Constructs . . . 47
2.10 Statistics . . . 47

3 Results - Modelling Integral Gene Drive for Population Replacement 48
3.1 Acknowledgements . . . 48
3.2 Chapter Aims . . . 48
3.3 Introduction . . . 48
3.4 Materials and Methods . . . 49
3.5 Results . . . 53
3.5.1 Designing IGD Drive Components . . . 53
3.5.2 Designing IGD Effector Components . . . 55
3.5.3 Practical Development of IGD Components . . . 58
3.5.4 Modeling IGD Dynamics . . . 58
Comparison of IGD and a Classical Replacement Drive . . . 61
3.5.5 Investigating a Three-Locus IGD Implementation . . . 68
3.5.6 Field Deployment and Evaluation of IGDs . . . 72
3.6 Discussion . . . 77

4 Results - Engineering Intronic gRNAs 80
4.1 Chapter Aims . . . 80
4.2 Introduction . . . 80
4.2.1 Intronic Splicing in *Drosophila* . . . 82
4.3 Materials and Methods . . . 84
4.3.1 Construction of Experimental Plasmids . . . 84
Cloning gRNA Spacers into pCFD3-dU6:3-gRNA Expression Vector . . . 85
4.3.2 Construction of Control Constructs . . . 88
4.3.3 Compensation Control Constructs . . . 90
4.3.4 Additional Controls . . . 91
4.3.5 Cultivating *Drosophila* S2 Cell Culture . . . 91

Maintenance of Cells . . . 91
Passaging Cells . . . 92
Cryopreservation of Cells . . . 92
4.3.6 Thawing Cells . . . 92
Mycoplasma Contamination Testing . . . 92
4.4 Transient Transfection of *Drosophila* S2 Cell Culture . . . 93
4.5 Flow Cytometry Conditions . . . 94
4.6 Results . . . 94
4.6.1 Constructing a Novel igRNA Test Platform . . . 94
4.6.2 Flow Cytometry Compensation . . . 97
4.6.3 Establishing the Feasibility of igRNA Expression . . . 98
4.6.4 Investigating Inactive igRNA Variants . . . 100
4.6.5 Demonstrating Inactivity of dCas9-AD . . . 102
4.7 Discussion . . . 103

5 Results - Implementing Integral Gene Drive in *D.melanogaster* 106
5.1 Acknowledgements . . . 106
5.2 Chapter Aims . . . 106
5.3 Introduction . . . 106
5.4 Materials and Methods . . . 109
5.4.1 *In vitro* Screening of Putative gRNA Targets . . . 109
5.4.2 Cloning of Experimental Constructs . . . 110
Cloning gRNA Spacers into pCFD3-dU6:3gRNA Expression Vector . . . 110
Assembling IGD Drive Components . . . 110
5.4.3 Line Generation and Biosecurity . . . 114
5.4.4 Cultivating *Drosophila* lines . . . 115
5.4.5 Crossing and Genotyping of *Drosophila* lines . . . 116
5.4.6 Cre:Lox Recombination of *Drosophila* lines . . . 117
5.4.7 *Drosophila* Development Studies . . . 118
5.4.8 Oviposition and Hatching Rate Assay . . . 118
5.4.9 *Drosophila* Homing Assays . . . 119
5.4.10 Genotyping Markerless *Rcd-1r*^{d} Progeny . . . 119
5.4.11 Amplicon Sequencing . . . 120
5.5 Model Generation and Parameters . . . 121
5.6 Results . . . 122
5.6.1 Engineering IGD Drive Components . . . 122
5.6.2 gRNA Design and Selection . . . 123
5.6.3 Generating Transgenic Lines . . . 125
5.6.4 Cre-Lox-Mediated Removal of a Reporter Cassette . . . 133
5.6.5 Development Assays . . . 137
Oviposition and Hatching Rate Assays . . . 139
5.6.6 Homing Assays . . . 140

Genotyping of a Reporterless Transgenic Line . . . 141
5.6.7 Investigating Indel Formation . . . 146
5.6.8 Integrating Sex-specific Parameters into the IGD Model . . . 147
5.7 Discussion . . . 148
Feasibility of Generating Transgenic Lines . . . 149
5.7.1 Assessing Fitness Costs to Transgenic Lines . . . 149
5.7.2 Investigating Indel Formation Due to Drive Activity . . . 151
5.7.3 Analysis of Homing Efficacy and *Rcd-1r* Expression . . . 151
Analysis of a Markerless IGD Drive Component . . . 155
5.7.4 Sex-Specific IGD Modelling . . . 155
5.7.5 Limitations of the IGD Paradigm . . . 156

6 Discussion and Future Perspectives **158**
6.1 Conclusions . . . 158
6.2 Future Work . . . 160
6.2.1 Investigating Expression of *Rcd-1r* . . . 160
6.2.2 Setup of a Population Trial . . . 161
6.2.3 Development of Suitable Effector Components . . . 161
6.2.4 Improving Gene Drive Performance . . . 162
6.2.5 Ethics of Gene Drive Technologies . . . 163

A Supplementary Information **166**
A.0.1 Additional Sequencing Primers . . . 166
A.0.2 IGD Sex-Specific Model Parameters . . . 168
A.0.3 Transgenic Lines and Target Loci . . . 170
A.0.4 Additional Amplicon Sequencing Information . . . 172

Bibliography **175**

List of Figures

1.1 Global Prevalence of Malaria . . . 2
1.2 Overview of the Mosquito Life Cycle . . . 3
1.3 Overview of the Parasite Life Cycle . . . 6
1.4 Population Replacement vs. Population Suppression . . . 11
1.5 RIDL System Topologies . . . 15
1.6 X-Shredding Mediated by an Autosomally Encoded Endonuclease . . . 17
1.7 One and Two-Locus Underdominance Drives . . . 19
1.8 Unidirectional *Wolbachia* Cytoplasmic Incompatibility . . . 20
1.9 Bidirectional *Wolbachia* Cytoplasmic Incompatibility . . . 21
1.10 Synthetic *Medea* Drive Systems . . . 23
1.11 Homing Endonuclease Gene Homing Activity . . . 25
1.12 ZFNs and TALENs Overview . . . 27
1.13 Integral Gene Drive Overview . . . 42

3.1 Intein and 2A Sequence Mediated Splicing . . . 57
3.2 Comparison of 'Classical' and Integral Gene Drive Dynamics . . . 63
3.3 Systematic Scan of IGD Model Parameters . . . 65
3.4 Drive Dynamics in a Background with Pre-Existing Resistance . . . 67
3.5 Allele Frequency Dynamics of a Two-Effector IGD Implementation - Pre-existing Resistance . . . 69
3.6 Allele Frequency Dynamics of a Two-Effector IGD Implementation - Secondary Release . . . 71
3.7 Comparison of Theoretical Regulatory Pipelines for Conventional and IGD Drive Strains . . . 73
3.8 Suggested pathway for the safe deployment of gene drive mosquitoes in a stepwise manner . . . 74
3.9 Considering Allele Frequency Dynamics of Limited Release Strategies . . . 76

4.1 Overview of the *Drosophila melanogaster* Life Cycle . . . 81
4.2 *D.melanogaster* splice site consensus sequences . . . 83
4.3 Reporter Construct Schematic . . . 84
4.4 pCFD3-dU6:3-gRNA gRNA Spacer Cloning Site . . . 85
4.5 igRNA Variant Schematic . . . 88
4.6 Cell Culture dCas9 Control Constructs . . . 89
4.7 Cell Culture Activation Control Constructs . . . 90

4.8 Compensation Control Construct - eGFP . . . 90
4.9 dCas9 Knockout Constructs . . . 91
4.10 Overview of Experimental Reporter Constructs . . . 97
4.11 Flow Cytometry Compensation Gating . . . 98
4.13 Overview of igRNA Platform Workflow . . . 99
4.12 Activation Experiments - *ftz* Intron . . . 100
4.14 Activation Experiments - Minimal Intron . . . 100
4.15 Positive Activation Experiment - *ftz* and Minimal Intron . . . 102
4.16 dCas9 Activity Assay . . . 103

5.1 Gene Drive Construct MK1 . . . 113
5.2 Gene Drive Construct MK1.1 . . . 113
5.3 Gene Drive Construct MK2 . . . 114
5.4 Overview of Gene Drive Constructs . . . 123
5.5 *In Vitro* sgRNA Screen . . . 124
5.6 Genotyping - *Rcd-1r*^{d,CFP} Founders . . . 127
5.7 Genotyping - *bam*^{d,CFP} Founders and Homozygote Pools . . . 128
5.8 Crossing Scheme - Overview . . . 129
5.9 Detailed Crossing Scheme - *Rcd-1r* and *bam* Transgenic Lines . . . 130
5.10 Genotyping - *bam*^{d,CFP}/TM6B Balanced Lines . . . 131
5.11 Genotyping - *Rcd-1r*^{d,CFP} Homozygote Drive Lines . . . 132
5.12 Genotyping - *Rcd-1r*^{d}/Cre Heterozygote Lines . . . 134
5.13 Genotyping - *Rcd-1r*^{d} Homozygote Lines . . . 136
5.14 Development Assays - Transgenic Lines . . . 138
5.15 Fertility and Hatching Rate Analysis . . . 140
5.16 Transmission Rate of Drive Alleles . . . 141
5.17 Experimental Pipeline for Determining Transmission Dynamics of a Markerless Line . . . 143
5.18 Limitations in Genotyping Markerless Individuals . . . 145
5.19 Amplicon Sequencing Analysis of *Rcd-1r*^{d,CFP} Offspring . . . 147
5.20 Sex-Specific Modelling Predictions . . . 148
5.21 Sex Determination in *D.melanogaster* . . . 153

A.1 Gene Drive Targets . . . 171
A.2 Additional Amplicon Sequencing Analysis . . . 172
A.3 Amplicon Sequencing - Mutation Distributions by Replicate . . . 173

List of Tables

2.1 Ligation Reaction Components . . . 46
2.2 Colony PCR Components . . . 46

3.1 Baseline Parameter Values for the IGD Model . . . 59
3.2 Baseline Parameter Values for the 'Classical' Model . . . 61

4.1 Intronic gRNA Platform - gRNA Spacers . . . 85
4.2 Guide RNA Ligation Reaction Reagents . . . 85
4.3 Guide RNA Vector Assembly Conditions . . . 86
4.4 PCR Primers Used for igRNA Insertion into the *ftz* intron . . . 86
4.5 Reporter Construct - igRNA Cassette Colony PCR Primers (*ftz* Intron) . . . 87
4.6 Reporter Construct - Minimal Intron Colony PCR Primers . . . 87
4.7 igRNA Cassette Assembly Primers - *ftz* Intron . . . 88
4.8 Activation Control - gRNA PCR Primers . . . 90
4.9 dCas9 Knockout Constructs - Primers . . . 91
4.10 Mycoplasma Screening PCR Primers . . . 93
4.11 Mycoplasma Screening PCR Setup . . . 93

5.1 T7 primers required for assembly of sgRNA cassettes . . . 109
5.2 PCR conditions for amplification of target amplicons from gDNA . . . 109
5.3 Amplicon primers for the *in vitro* gRNA activity assay, with corresponding fragment sizes following successful cleavage . . . 110
5.4 Single-stranded DNA oligos used to construct sgRNA sequences for insertion into pCFD3-dU6:3gRNA . . . 110
5.5 Primers used for cloning of fragments required for construction of IGD drive components . . . 112
5.6 *bam* and *Rcd-1r* Genotyping Primers . . . 117
5.7 Primers used to confirm successful Cre:Lox mediated excision of the 3xP3:CFP:SV40 reporter cassette . . . 117
5.8 Primers used for amplification of the *Rcd-1r* gRNA target site, for amplicon sequencing . . . 121
5.9 PCR reaction conditions for high-fidelity amplification of the *Rcd-1r* locus for amplicon sequencing . . . 121
5.10 Details of initial microinjections, with corresponding targets and recovery rates . 125
5.11 Improved microinjections with corresponding targets and recovery rates . . . 125

A.1 IGD Drive Component Sanger Sequencing Primers 166
A.2 Additional IGD Drive Component Sanger Sequencing Primers 167
A.3 Sequences for *ftz* and Minimal Intron 168
A.4 Baseline Parameter Values for the Sex-Specific IGD Model - Male Parameters . . 168
A.5 Baseline Parameter Values for the Sex-Specific IGD Model - Female Parameters . 169
A.6 Externally Sourced *Drosophila* Lines 170

List of Abbreviations

ACT **A**rtemisin **C**ombination **T**herapy

AD **A**ctivation **D**omain

AMP **A**nti-**m**icrobial **P**eptides

Cas9 **C**RISPR-**as**ociated protein **9**

CATCHA **Ca**s9 **T**riggered **C**hain **A**blation

CFP **C**yan **F**luorscent **P**rotein

CHACR **C**onstruct **H**itchhiking on the **A**utocatalytic **C**hain **R**eaction

CI **C**ytoplasmic **I**ncompatibility

CP **C**arboxy**p**eptidase

CRISPR **C**lustered **R**egularly **I**nterspersed **P**alindromic **R**epeats

dCas9 dead-Cas9

DNA **D**eoxyribo**n**ucleic **A**cid

DSB **D**ouble **S**tranded **B**reak

DVS **D**ominant **V**ector **S**pecies

EDTA **E**thylene**d**iamine**t**etraacetic **A**cid

eGFP **E**nhanced **G**reen **F**luorscent **P**rotein

ERACR **E**lements for **R**eversing **A**utocatalytic **C**hain **R**eaction

fsRIDL **f**emale-**s**pecific **R**elease of **I**nsects carrying **D**ominant **L**ethals

ftz *fushi **t**arazu*

gDNA **g**enomic **D**eoxyribo**n**ucleic **A**cid

gRNA **g**uide **R**ibonucleic **A**cid

HDR **H**omology **D**irected **R**epair

HEG **H**oming **E**ndonuclease **G**ene

HSV **H**erpes **S**implex **V**irus

IGD **I**ntegral **G**ene **D**rive

igRNA **i**ntronic **g**uide **R**ibo**n**ucleic **A**cid

IRES **I**nternal **R**ibosomal **E**ntry **S**ite

IRS **I**ndoor **R**esidual **S**praying

ITR **I**nverted **T**erminal **R**epeat

IVM **I**ntegrated **V**ector **M**anagement

IME **I**ntron-**m**ediated **E**nhancement

LB **L**uria-**B**ertani

LLINs **L**ong **L**asting **I**nsecticidal **N**et

MCS **M**ultiple **C**loning **S**ite

MEDEA **M**aternal **E**ffect **D**ominant **E**mbryonic **A**rrest

MESF **M**olecules of **E**quivalent **S**oluble **F**luorophore

MMEJ **M**icrohomology-**m**ediated **E**nd **J**oining

NASEM **N**ational **A**cademy of **S**cience, **E**ngineering, and **M**edicine

NGS **N**ext **G**eneration **S**equencing

NHEJ **N**on-**h**omologous **E**nd **J**oining

NLS **N**uclear **L**ocalisation **S**ignal

PAM **P**rotospacer **A**djacent **M**otif

PBO **P**iperonyl **B**ut**o**xide

PCR **P**olymerase **C**hain **R**eaction

PTEN **P**hosphatase and **Te**nsin **H**omolog

RIDL **R**elease of **I**nsects carrying **D**ominant **L**ethals

RMCE **R**ecombination **M**ediated **C**assette **E**xchange

RNA **R**ibo**n**ucleic **A**cid

sgRNA **s**ingle **g**uide **RNA**

SIT **S**terile **I**nsect **T**echnique

SOC **S**uper **O**ptimal broth with **C**atabolite repression

SSB **S**ingle **S**tranded **B**reak

TALE **T**ranscription **A**ctivator-**l**ike **E**ffectors

TALENs **T**ranscription **A**ctivator-**l**ike **E**ffector **N**ucleases

UTR **U**n**t**ranslated **R**egion

WHO **W**orld **H**ealth **O**rganisation

ZFN **Z**inc **F**inger **N**uclease

Chapter 1

Introduction

1.1 The Challenge of Vector-Borne Disease

Vector-borne disease has historically presented one of the greatest environmental challenges to human survival (Gubler, 1998). Indeed, between the 17th to early 20th centuries, insect-borne diseases were responsible for more fatalities than all other causes combined. Recent estimates indicate that vector-borne disease accounts for 17% of all infectious disease experienced globally, accounting for approximately 700,000 deaths in 2017 (World Health Organization, 2017a). The most significant of these diseases is malaria, which is caused by vector-host transmission of the *Plasmodium* parasite, and facilitated by females of the *Anopheles* genus.

In 2017, the World Health Organization estimated 219 million cases of the disease, resulting in 435,000 deaths (World Health Organization, 2018). It further estimated that despite record levels of funding being provided to tackle the global prevalence of malaria ($3.1Bn in 2017), there has not recently been a significant decrease in the estimated number of cases. Most susceptible are children under the age of 5, who comprise 61% of fatalities (World Health Organization, 2018).

Mosquitoes of the genus *Anopheles* are comprised of 465 identified species (Sinka et al., 2012), of which approximately 70 have the capacity to act as vectors for the causative agent of malaria, the *Plasmodium* parasite (Gilles and Warrell, 1996). Meta-analyses of established literature have since identified a subset of 41 dominant vector species (DVS), responsible for the majority of malarial transmission to humans, and which pose the greatest threat to public health (Hay et al., 2010). Sub-Saharan Africa remains the epicentre for malarial transmission globally (Fig.1.1) with the entire continent estimated to account for 92% of all malarial incidents and 93% of mortalities (World Health Organization, 2018). The DVS in this region have been identified as *An.gambiae*, *An.funestus*, and *An.arabiensis* (Sinka et al., 2012).

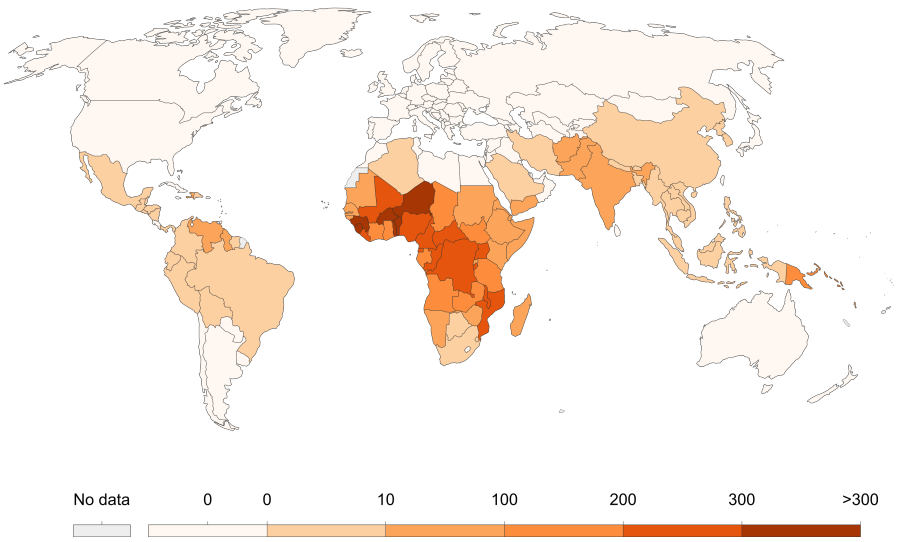


FIGURE 1.1: **Global Prevalence of Malaria (2017)**. The above heat map indicates the number of malaria cases reported, per 1000 people, in a given country (Global Burden of Disease Collaborative Network, 2018)

1.1.1 The Vector

Anopheline mosquitoes undergo four main stages of development (Fig.1.2), that are ubiquitous to all mosquito species. First, there is an aquatic phase, during which individuals mature from eggs to larvae, and then to pupae. This phase lasts from 7-14 days depending on species and ambient conditions. Pupae subsequently metamorphose into adults, leaving their aquatic environs. There are a number of parameters affecting habitat selection when a female chooses a site for oviposition. These parameters include the salt content, opacity, size, and permanence of the prospective body of water (Impoinvil et al., 2008). Exposure to sunlight, predation, and human interference are also considered (Walker and Lynch, 2007).

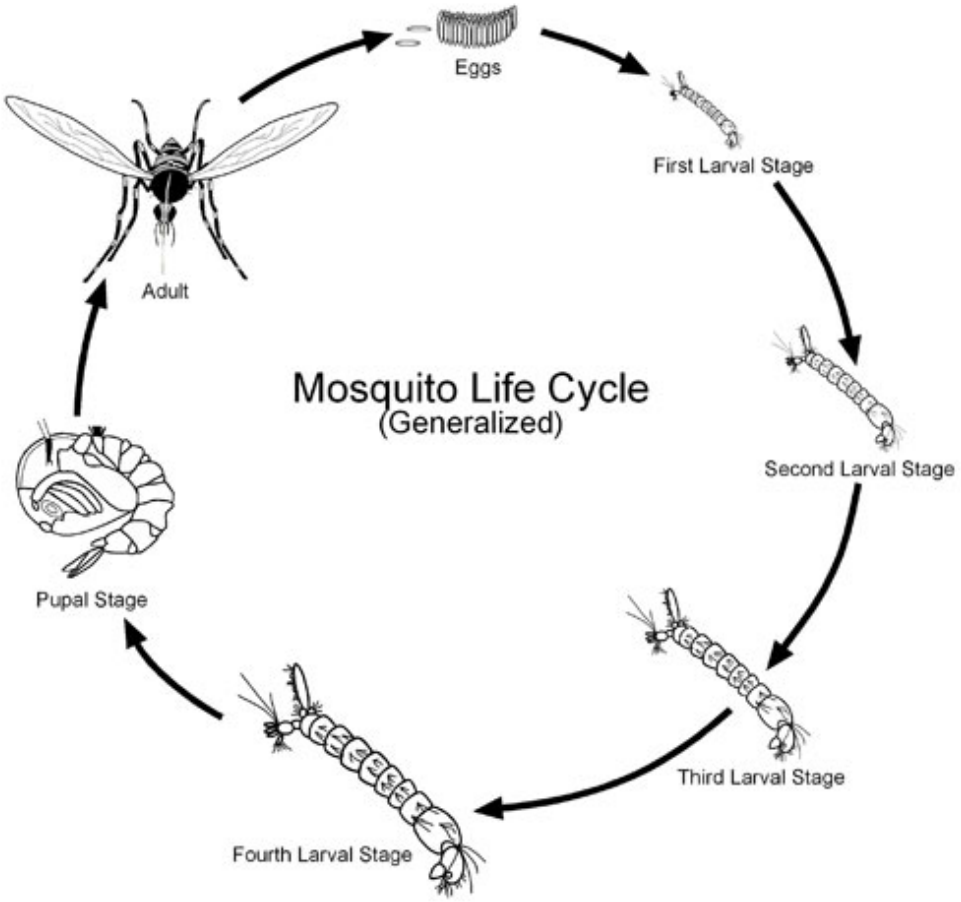


FIGURE 1.2: **Overview of the Mosquito Life Cycle**. Mosquitoes undergo main four stages of development, split into an aquatic phase (egg to pupae), and a terrestrial stage (adult), Charlesworth, 2008

Following oviposition, eggs hatch within 2-3 days, and nascent larvae emerge. They subsequently develop through four instar phases, frequently coming to the surface to respire through spiracles. Fourth instar larvae mature into pupae, which do not feed, but rapidly develop into adults that emerge from the cephalothorax. Adult mating is crepuscular and requires males to form large swarms. Females entering the swarm are identified based on the frequency of their wing beat during flight (Belton and Costello, 1979).

Males have a limited life-span, dying within a week of emerging from their pupal stage (Centers for Disease Control and Prevention, 2019). During this time, they primarily feed on plant nectar for sustenance. By contrast, females tend to live for up to a fortnight. Following mating, they store sperm received from males in the spermatheca, which acts as a reservoir allowing myriad of eggs to be fertilised. Anopheline females are anautogenous and require a protein-rich blood meal to facilitate maturation of eggs. Critically, Anopheline mosquitoes have a preference for human blood feeding (Takken and Verhulst, 2013). They secrete a number of anti-coagulant and anti-hemostatic factors from their salivary glands in order to facilitate blood feeding (Stark and James, 1996). It is at this critical point in the vector life cycle that a female *Anopheles* mosquito may become infected with the *Plasmodium* parasite, or may transmit it to a

new host. Females will continue to blood-feed and mate until their demise, further propagating the disease.

1.1.2 The Parasite

The five *Plasmodium* species known to be the causative agents of malaria in humans are *P. vivax*, *P. ovale*, *P. malariae*, *P. knowelsi*, and *P.falciparum*. In 2017, *P.falciparum* accounted for 99.7% of estimated malarial incidents in the WHO's designated Africa region (World Health Organisation, 2017), and was the major causative agent of malaria in the WHO designated South-East Asia, Eastern Mediterranean, and Western Pacfici regions. *P.vivax* was predominant in the Americas, accounting for 71.4% of estimated cases. These heteroxenous members of the Apicomplexan phylum have a complex life cycle, in both the female *Anopheles* mosquito, and in their human host. Given that the vast majority of malaria cases occur in Sub-Saharan Africa, this work will focus on discussions pertaining to *P.falciparum*.

Blood-Stages of the *Plasmodium* Parasite

When an infected female blood feeds on a host, sporozoites may enter the host bloodstream using gliding motility - a process by which the *Plasmodium* uses an actin-myosin motor to achieve locomotion through the host (Matuschewski and Schüler, 2008). As part of the symptomless pre-erythrocytic phase, parasites rapidly traverse to the liver, where upon they begin to mature into schizonts over the course of 2-10 days (Cowman et al., 2016). Upon achieving maturation, up to 40,000 merozoites are released per hepatocyte in a vesicle termed a merosome, beginning the erythrocytic cycle (Sturm et al., 2006). Merozoites proceed to invade erythrocytes and replicate (schizogony), producing 16-32 daughter merozoites. During this process, parent parasites differentiate into trophozoites and schizonts. Egress of replicated merozoites causes explosive rupture of the erythrocyte, release of said parasites, as well as a variety of toxins and parasite products that trigger the innate immune system (Autino et al., 2012). It is this mechanism that culminates in the characteristic symptoms of malaria, including a cyclical fever as waves of merozoites continue the erythrocytic cycle, continuously triggering a host-immune response.

As schizogony proceeds, a sub-population of parasites will enter the sexual development stage. This is often cited as a critical phase for medical interventions, as it is at this stage that the parasite will be able to infect Anopheline mosquitoes. Interruption here represents a means by which human-vector transmission may be halted, and the entire parasite life cycle is terminated. The exact moment at which gametocytogenesis occurs is as yet unclear (Josling and Llinás, 2015), although it is thought to be at a point during the erythrocytic cycle before schizogony, as all daughter merozoites within a schizont differentiate into gametocytes or merozoites. A number of environmental parameters such as drug exposure, or high levels of parasitemia have been linked to gametocytogenesis *in vitro*. Given that such conditions might result in host death or parasite elimination, from an evolutionary perspective it is logical that such events should trigger the onset of parasite sexual stages. Gametocyctogenesis takes between 10-12

days, whereupon male gametocytes (microgametocytes) and female gametocytes (macrogametocytes) are available in the blood for uptake by an *Anopheles* mosquito.

Mosquito Stages of the *Plasmodium* Parasite

As previously mentioned, the sexual stages of the parasite life cycle represent a key target for drug and vaccine development (Fig.1.3). At the point at which gametocytes enter the mosquito vector, there may be just tens of parasites. Having been ingested, the gametocytes are stimulated to mature into gametes (gametogenesis) in the lumen of the mosquito midgut. (Aly, Vaughan, and Kappe, 2009). These gametes in turn are able to fuse, fertilising the female, and causing the formation of a zygote. Following meiosis, the zygote develops into an ookinete, which possesses secretory organelles that facilitate its passage into the peritrophic matrix of the mosquito gut and through the midgut epithelium. Cell traversal triggers the ookinete to enter an immobile phase and to transform into an oocyst in the basal lamina.

The transformation from ookinete to sessile oocyst lasts 10-12 days. The parasite proceeds to replicate, producing many thousands of sporozoites. The sheer number of sporozoites was thought to be the cause of their subsequent egress (Sinden, 1974). However, a more recent knockout study in *P.berghei* has instead indicated that this occurs once a cysteine protease has been expressed that permits perforation of the walls of the oocyst (Aly and Matuschewski, 2005). Sporozoites are then transported through the hemolymph throughout the mosquito. Some will eventually pass the basal lamina of the salivary glands, which are recognised by the parasite using tissue-specific ligands that bind host-receptors on the tissue (Sinden and Matuschewski, 2005). The sporozoites adhere to the tissue and invade, entering the salivary gland. Upon blood-feeding, these infectious sporozoites enter the bite-site of a human host and enter the blood stream, beginning the cycle once more.

A range of medical interventions has been developed to combat infection with the *Plasmodium* parasite, using both anti-malarial drugs and vaccines. There are four main classes of drugs currently available (Travassos and Laufer, 2019) -

1. Quinoline derivatives (e.g. chloroquine, meloquine, and primaquine)
2. Antifolates such as sulfadoxine and pyrimethamine
3. Antimicrobials (tetracyclines, macrolides, lincosamides)
4. Artemesinin derivatives and combined therapies (e.g. artemether and arteether)

Increasing resistance to these drugs has resulted in increased use of artemisinin combination therapies (ACTs). Combination therapies employ artemisinins with a short half-life together with longer lasting drugs, aiming to rapidly clear parasitemia and prevent development of artemisinin-resitant parasites. Despite this, there have already been reports of desensitisation and resistance to ACTs in South-East Asia (Dondorp et al., 2009; Ashley et al., 2014). Resistance is associated with mutations of the kelch protein gene, which are reported to be implicated in cytoprotective responses to environmental stresses (Ariey et al., 2014). Due

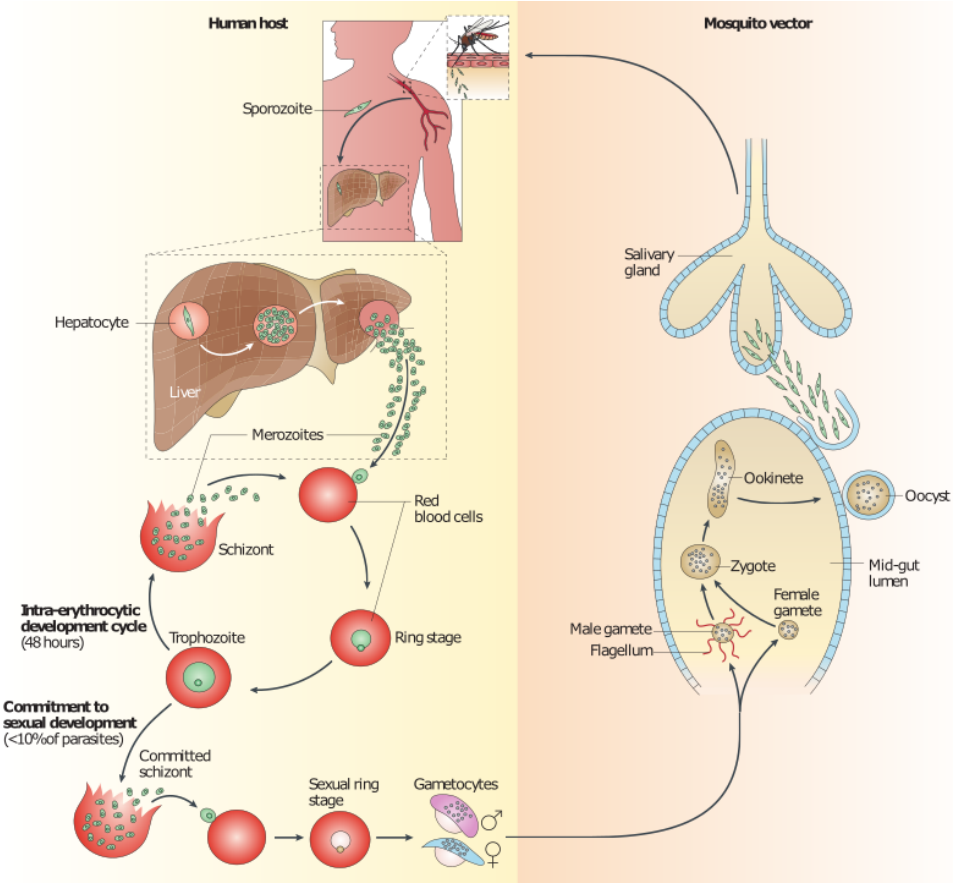


FIGURE 1.3: **Overview of the Parasite Life Cycle**. Initially, the parasite enters the bloodstream as sporozoites following biting from an infected female mosquito. Sporozoites migrate to the liver, where they invade hepatocytes and mature. Merozoites are released into the bloodstream where they invade erythrocytes and undergo multiple rounds of asexual replication. A small subpopulation commits to sexual development forming gametocytes. These gametocytes are ingested by female mosquitoes underaking blood-feeding. Gametocytes mature forming zygotes and ookinetes that are able to invade the midgut epithelium. Resultant oocysts form sporozoites that migrate to the salivary glands, ready to recommence the entire cycle. Josling and Llinás, 2015.

to the critical role artemisinins play in combatting malaria, the WHO has strongly mandated against the use of artemisinin monotherapies in an attempt to slow the development of resistant *Plasmodium* strains. Increasing resistance to all classes of antimalarial drug underscore the urgent need to continue to develop novel therapeutics. That being said, there are currently few compounds in development and the pharmaceutical pipeline is traditionally costly and time consuming, suggesting that any immediate developments are unlikely.

In the context of generating a vaccine against *P.falciparum*, the most developed candidate is RTS,S/AS01. The vaccine utilises PfCSP, the primary circumsporozoite protein, reducing infection caused by pre-erythrocytic sporozoites. To date, it remains the only vaccine to have successfully generated protection in a phase III clinical trial amongst young children (World Health Organisation, 2019). It is important to note that protection was only partial, age-dependent, and diminished following vaccination (RTSS Clinical Trials Partnership, 2015). As of April 2019, the RTS,S vaccine was approved by the WHO for deployment in a pilot programme in Malawi, Ghana and Kenya, where the vaccine was made available for children under the age of 2 years. The vaccine demonstrated a 36% reduction in severe malaria in the trial cohort, offering a degree of protection, but also showing that there is considerable work yet to be done. These shortcomings in the availability of effective therapeutics and vaccines underline the need for an alternative solution to the problem of zoonotic disease, and malaria in particular.

1.2 Conventional Vector Control Strategies for Malaria

Whilst a number of medical interventions currently exist, which aim to target the human-stage of the *Plasmodium* parasite, eradication historically has been most successful when focusing on control of malarial vectors (World Health Organization, 2017b). The work in this thesis focuses on developing a novel control tool for malarial vectors with the possibility of expanding to other agricultural and zoonotic pests. In 1998, the Roll Back Malaria Partnership was launched to eradicate malaria globally by 2010. Whilst this campaign and other independent interventions have resulted in a reduction of 37% in the global malaria incidence (Gari and Lindtjørn, 2018), there has been a decline in the pace of change.

Existing progress has largely been attributed to the expansion of key vector control strategies including indoor residual spraying (IRS) and long-lasting insecticidal nets (LLINs). Furthermore, the parasite has been demonstrated to evolve resistance to drug interventions, and many of those affected have poor access to healthcare (Guyant et al., 2015). Despite recent advances in vaccine and drug technologies, vector control strategies remain a potent means to limit the global burden of malaria. There are a number of traditional vector control strategies that have been deployed to limit disease transmission by Anophline mosquitoes and other zoonotic mosquito species -

Biological Controls

Biocontrol strategies tend to have a lower environmental impact than chemical pesticides and can be self-sustaining. The most prominent biological agents that have been deployed against Anopheline mosquitoes are natural predators. Predatory fish belonging to the genus *Gambusia* and *Poecilia* have been successfully utilised in mosquito control strategies in over 60 counties (Benelli, Jeffries, and Walker, 2016) and have proven particularly successful in targeting the aquatic stage of the mosquito life-cycle (Raghavendra et al., 2011). Other fish, such as *Tilapia* and *Aphanius*, have also been used. Some studies have sounded concern over the introduction of new predators into existing foodchains, speculating that such changes may have unforseen consequences on the broader eco-system (Rupp, 1996; Kats and Ferrer, 2003).

A number of members of the copepod subclass have been demonstrated to be efficient predators of mosquito larval stages too. Mass release of members of the genus *Mesocyclops* in Vietnam as part of a community program was shown to successfully eliminate the zoonotic vector*Aedes aegypti* from two communes within two years (Nam et al., 2005). The use of copepods as a means for vector control is well-suited to targeting mosquito larvae in urban habitats due to the low cost and relative ease with which they can be reared (Chitra et al., 2013). However, as with predatory fish their introduction is not without consequence, and many mosquito species tend towards habitats unsuitable for copepods (Hales and Van Panhuis, 2005). Mosquitoes of the genus *Toxorhyncites* have been implicated as a potential biocontrol agent for malarial mosquitoes. These mosquitoes do not blood feed in adulthood, but predate the larvae of other mosquito species and other aquatic organisms whilst in their larval state. They pose no threat to human health (Steffan and Evenhuis, 1981), but given their unrestricted diet may pose a significant risk to other aquatic organisms.

Other biocontrol approaches tend towards the use of microbial agents. *Bacillus thuringiensis israelensis* (*Bti*) and *Bacillus sphaeriucs* (*Bs*) are gram-positive, spore-forming bacteria. Their spore coatings contain toxins capable of acting in the midgut of mosquito larvae. Historically, *Bti* has been used against a range of mosquito disease vectors, including *An. albimanus*, *An. sinesis*, *An. stephensi*, *An. gambiae*, *An. malactus*, *An. maculipennis*, and *An. sundacius* (Raghavendra et al., 2011). Such biocontrol agents have a number of advantages when compared to chemical controls - they are effective at low concentration, are typically non-harmful to other fauna and flora, are cheap to produce in sufficient quantities, and have a low risk of producing resistance (though increasing tolerance has been noted in some field populations) (Tetreau et al., 2013).

Biological controls are not without limitations; in particular the introduction of novel species into an ecosystem can have severely detrimental effects (Benelli, Jeffries, and Walker, 2016). Moreover, biological controls tend to be slow-acting and lack the dramatic activity of insecticides. Any consequences of a biological control strategy may not become observable for a considerable time, by which point reversal may be costly or impractical. Typically biological

controls will be habitat specific and lack the broader applicability offered by modern genetic control methodologies.

Habitat Management

Integrated vector management (IVM) has focused on the synergistic use of multiple countermeasures to disrupt the spread of malaria. One component of such strategies is the reduction of the source of the Anopheline mosquito by systematic targeting of larval habitats. A number of land management techniques have been employed such as the targeted destruction of aquatic environments adjacent to rivers and lakes, and proper irrigation of wetlands and fields (Mutero et al., 2000; Yasuoka et al., 2006). Unfortunately, these approaches tend to be broadly destructive from an environmental standpoint and pressure mosquito populations to select more isolated breeding grounds (Roux and Robert, 2019). Isolation of mosquito habitats has an additional knock-on effect, limiting access to predators, who would otherwise act as a natural control of mosquito populations. If they require aquatic environments to survive, land management may also negatively affect their numbers. This decrease in predation can consequently permit an increase in mosquito numbers (Chase and Shulman, 2009). It is important to remark that in areas where malaria is endemic with large mosquito populations, habitat destruction does not sustainably reduce malarial transmission and should always be part of an IVM approach, whilst a more permanent solution is sought (Shaukat, Breman, and McKenzie, 2010).

Insecticides and Resistance

A critical component of integrated approaches to the control of mosquitoes and prevention of malaria transmission lies in the use of long-lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS). These relatively simple tools are estimated to have prevented 517 million cases of malaria between 2000 and 2015 (Bhatt et al., 2015). LLINs are considered to be primarily responsible for these results, though modelling has suggested it is insufficient when used alone (Griffin et al., 2010). There are six classes of insecticides which have been recommended for uses against mosquitoes (World Health Organization: Global Malaria Programme, 2013) - carbamates, pyrethroids, pyrroles, organochlorines, organophosphates, and phenyl pyrazoles. Until 2018, the only class of insecticides approved for use in LLINs was the pyrethroids (Choi, Pryce, and Garner, 2019). However, the emergence of pyrethroid-resistant mosquito populations has threatened the protection offered by LLINs with the potential to undermine a fundamental cornerstone of malaria prevention. To this end, novel compounds are being employed in LLINs and IRS; Piperonyl butoxide (PBO) is now being incorporated into LLINs as an enhancer, inhibiting oxidases critical to resistance in mosquitoes. The introduction of PBO was demonstrated to significantly increase the efficacy of LLINs. (Protopopoff et al., 2018).

IRS application is affected by a number of factors including: efficacy against mosquito populations, half-life of the insecticide, site and timing of the application of IRS, and the quality of

said application (Sherrard-Smith et al., 2018). Currently, both LLINs and IRS are inhibited in their efficacy by changes in mosquito behaviour resulting in novel forms of resistance. Studies have demonstrated that the the introduction of such vector control methods has resulted in critical behavioural modifications to the biting time and indoor/outdoor feeding preference in the Anopheline mosquito (Moiroux et al., 2012; Thomsen et al., 2017). Furthermore, such adaptations have severe implications for the efficacy of these types of vector control programme. Taken with the emergence of insecticide resistance they may pose a considerable risk to the long-term prospects of LLINs and IRS.

As with LLINs, chemical compounds approved for use in IRS have primarily been pyrethroids. Recently, approval has been achieved for the use of organophosphate and carbamate derivatives. Organophosphates and carbamates are effective against a broad range of invertebrates. They function by disrupting the activity of colinestererases responsible for regulation of the acetylcholine in the nervous system, resulting in paralysis and death (Fukuto, 1990). These classes of insecticides have the additional limitation of being toxic to many untargeted invertebrates and vertebrates. Resistance has been demonstrated in Anopheline mosquitoes, underscoring the need for continued biological surveillance if insecticides are to remain effective vector control tools. The limitations to the efficacy of chemical interventions underscores the need to develop alternative control methods. One such approach is the use of genetic control strategies.

1.3 Genetic Control Strategies

Significant research has been conducted to overcome challenges posited by more conventional vector control methodologies. This has given rise to a family of vector control strategies aimed at suppressing or modifying target populations at the genetic level. Genetic control is hereafter defined as "...dissemination, by mating or inheritance, of factors that reduce pest damage." (Alphey, 2014). In this way, these strategies are determined by the fact that a modified version of the vector host itself acts as the control agent, propagating useful heritable traits.

Genetic control strategies are designed with one of two outcomes in mind - population suppression or replacement (Fig.1.4). Suppression strategies aim to lower the total number of potential vector hosts, lowering the risk of vector-human disease transmission. By contrast, population replacement aims to replace the existing wild-type population with a modified population, carrying a desirable trait, which can be propagated between generations. Typically, expression of the designed trait will induce a fitness cost in a given individual, which may lower the efficacy of such strategies. On the other hand, population replacement strategies lower the probability of another pest occupying a vacated ecological niche, created by a suppression strategy (Alphey et al., 2010).

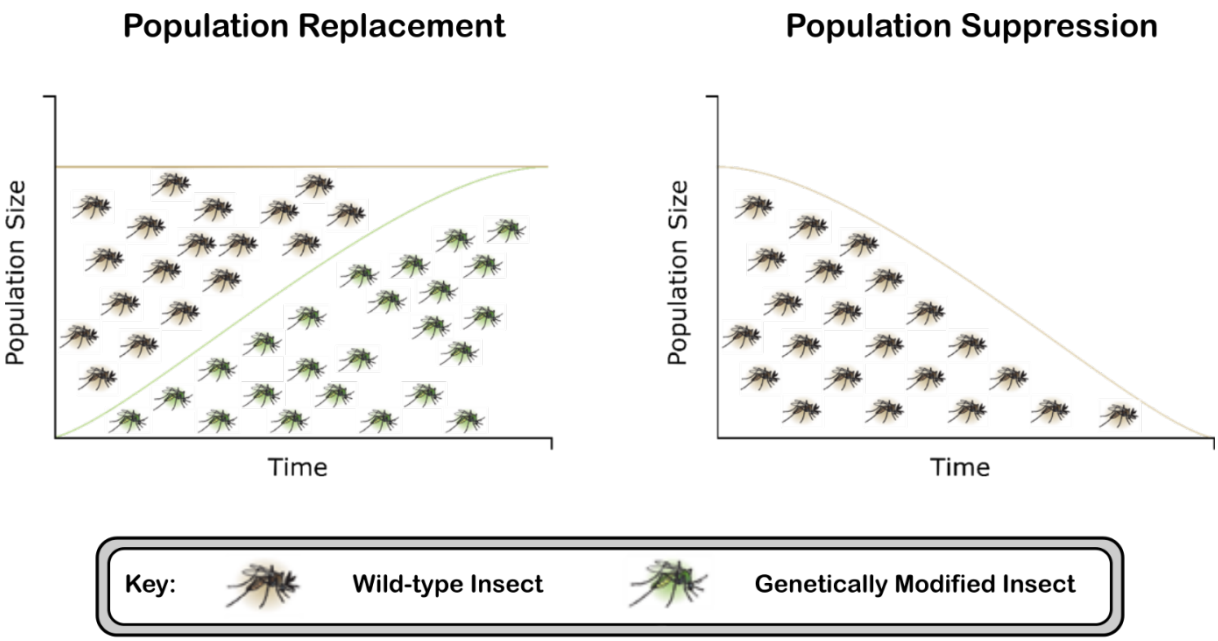


FIGURE 1.4: **Population Replacement vs. Population Suppression**. Population replacement strategies aim to replace an existing field population with modified individuals. These modified individuals carry a beneficial trait, i.e. rendering mosquitoes refractory to parasite infection. By contrast, population suppression strategies aim to spread a deleterious trait through a target population by introducing modified individuals. By modulating fertility or propagating toxins, they aim to entirely eliminate the population.

1.3.1 Sterile Insect Technique

The first genetic control strategy to have emerged is Sterile Insect Technique (SIT), proposed in 1940 by Serebrovsky (Serebrovsky, 1969, *Originally published in Russian in 1940*) and later developed by E.F.Knipling (Knipling, 1959), using release of sterilised male insects to suppress wild populations. Male insects are preferentially released, as females of some species are responsible for crop damage when laying eggs, and in the case of malaria, females are responsible for disease propagation. SIT has been successfully used to eradicate populations of New World Screw-worm fly (*Cochliomyia hominivorax*) in large parts of the Americas, and has shown to be effective against Mediterranean fruit fly (*Ceratitis capitata*; Hendrichs, Franz, and Rendon, 1995), and Tsetse fly (*Glossina* spp; Abd-Alla et al., 2013). It has since been deployed against Anopheline mosquitoes (Laven, 1967; John, Birnstiel, and Jones, 1969; Davidson et al., 1970). Notably, SIT does have the advantage of being ecologically friendly, with limited effect outside the targeted population (Alphey et al., 2010).

Knipling's original methodology used a radiation source to induce sterility. Consequently, the fitness of modified individuals, and their subsequent ability to compete with wild-type individuals for mates may been unintentionally compromised. Indeed, in cases where SIT has failed to have a significant effect on a targeted population, it is typically attributable to poor production of sterile males, loss of fitness in sterile males, and the ingress of wild-type females from outside the target area (Benedict and Robinson, 2003). Moreover, when rearing large

numbers of insects for SIT, both sexes are produced, and females later removed and discarded. Various methods, both mechanical and genetic, have been developed for systematically removing females (Alphey et al., 2010). The former tends to be inefficient and results in the release of females, whilst the later results in significant fitness costs in selected individuals. Since then, these observations have led to the development of a broad range of transgenesis methods and control tools.

1.3.2 The Advent of Insect Transgenesis

Genetic engineering, as applied to insects, was first discussed in 1979 as a means to engineer the behaviours of agricultural pests (Levin, 1979). In 1982, Rubin and Spradling demonstrated the first successful germline transformation of an insect species using the *P*-element transposon in *Drosophila melanogaster* (Rubin and Spradling, 1982).The *P*-element is a 2907 bp class II transposon encoding a transposase (Engels, 2001). At the 5′ and 3′ termini are 31-bp inverted terminal repeats (ITRs) required for transposition. Should the transposase gene be defective, the *P*-element can still be mobilised by an external source of transposase. Such defective elements are known as non-autonomous *P*-elements.

Rubin and Spradling made use of these characteristics to mutate the wild-type eye colour of *Drosophila* using a two-component system: 1) A plasmid containing a non-autonomous *P*-element containing the DNA to be inserted, 2) A second plasmid expressing a functional transposase. This methodology had the advantage of stably integrating the transposon (and foreign DNA) as the transposase-expressing plasmid is lost with them. The *P*-element was consequently used as a genetic tool for transposon tagging (Searles et al., 1982) and enhancer trapping (Bellen et al., 1989) in *Drosophila*. The mobility of *P*-elements has been shown to be strain-dependent though highly mobile in *D.melanogaster* (Majumdar and Rio, 2015). *P*-element based vectors were later successfully employed in modifying *D.simulans* (Scavarda and Hartl, 1984) and *D.hawaiiensis* (Brennan, Rowan, and Dickinson, 1984). Neither species contains endogenous *P*-elements, and in both cases, *P*-elements retained their autonomy in ensuing generations (O'Brochta and Atkinson, 1996).

Unfortunately, it subsequently became apparent that the *P*-element was non-functional as a tool for genome engineering in non-drosophilid insect species. While *P*-element based vectors have successfully been used to transform various mosquito species (Miller et al., 1988; Miller et al., 1987; Morris, Eggleston, and Crampton, 1989), these events appear to be independent of *P*-element mediated transposition. As a result, a number of alternative transposon-based systems have been developed (Fraser, 2012). These systems are engineered around the *mariner* family tranposable elements *Minos* and *Mos1*, the *hAT* family *Hermes* transposon, and the *piggyBac* transposon. Additionally, the *Herves* transposon has been identified as a suitable element for expanding the existing repertoire of transposon-based transformation tools, although no non-drosophilid application has yet been published.

At the time of writing, all these transposon systems have been successfully deployed to facilitate germline-transformation of various mosquito species (Grossman et al., 2001; Coates et al., 1998; Jasinskiene et al., 1998; Catteruccia, Benton, and Crisanti, 2005), including *Anopheles gambiae*, the primary vector of malaria. These developments have had a major impact on the ability of researchers to manipulate vector genomes, allowing for greater fundamental understanding of insect biology, whilst also providing powerful tools in the fight against vector-borne disease and agricultural pests.

That is not to say that transposon-based systems for genetic engineering are without limitations. There are four main areas that require additional improvement: the random nature of their integration in the host genome, low transformation efficiency, instability following integration, and limits to the quantity of foreign DNA that can be integrated (Fraser, 2012). Given that integration of these engineered transposons is random, expression levels of inserted genes are variable depending on the genetic variation in the background of individual insects. This requires that multiple genetically modified strains must be generated and their performance compared across multiple generations in order to ensure optimal transgene expression.

Furthermore, remobilisation following insertion of the engineered transposon may result in unwanted modifications and undesirable consequences such as a fitness penalty. This also posits the risk of horizontal gene transfer between species, posing a potential public health risk and the possibility of unforeseen modifications in other vectors. To avoid this scenario, several methodologies have been devised to stabilise transposons post-integration. One approach, demonstrated in *D.melanogaster* and *C.capitata* respectively, utilises a nested transposon configuration that facilitates excision of one or both ITRs (Dafa'alla et al., 2006; Handler, Zimowska, and Horn, 2004). This permits for initial integration of the engineered transposon carrying the gene of interest, but ablates further mobilisation post-integration. An alternative methodology that has been devised makes use of prokaryotic transposons (Rowan et al., 2004). There is no native source of transposase for Tn5 in the insect genome, so following insertion using a stable synaptic complex containing an appropriate transposase, the transgene cannot be remobilised. It is worth noting, that this methodology still is poorly efficient in generating insertions.

As noted above, the performance of transposon-based vectors for germline transformation is sufficient for experimental efforts, but is highly variable and often low. For example, whilst the *P*-element and *Hermes* based systems have been demonstrated to exhibit transformation efficiencies in the range of 50% in *D.melanogaster* (O'Brochta and Atkinson, 1996), other systems, such as *piggyBac*, exhibited substantially lower efficacy (2.3-5.9%, and 1-3% respectively) (Li, Heinrich, and Scott, 2001; Handler, 2001). Providing a trans-acting source of transposase can aid in the increase of the frequency of integrations. However, it can also contribute to unwanted post-integration remobilisation and facilitate undesirable integrations, as discussed above.

Despite efforts to stabilise post-integration remobilisation of transposon-based insertions, the lack of specificity when generating transgenic lines makes them unreliable for vector control

strategies. To avoid the positional effects associated with random integration and facilitate site-specific modification, recombination systems were devised which allowed for the insertion of transgenes into predefined genomic sites. The first such systems were members of the tyrosine recombinase family - the FLP-*FRT* (Senecoff, Bruckner, and Cox, 1985) and Cre-*Lox* systems (Hoess and Abremski, 1985). Both systems make use of inverted 13-bp sequences flanking an 8-bp spacer. When recombinase is supplied, the spacer sequence facilitates recombination with an identical site. Sites are randomly integrated into the genome, strains are screened and subsequently selected as docking lines for transformation. Presence of a docking site in the genome permits recombination of DNA between a plasmid donor and the genomic target(s) in a process known as recombinase-mediated cassette exchange (RMCE). FLP-FRT was first demonstrated in 1997 in *Drosophila* to specifically generate *white* eye pigment mutants (Golic et al., 1997).

More recently, Cre-*Lox* recombination was used to rescue mutant phenotypes with high efficiency (Oberstein et al., 2005). The serine-catalysed ΦC31 integrase system has also been demonstrated in a range of insect vectors, including *A.gambiae*, for the purposes of disease control (Meredith et al., 2011). ΦC31 makes use of 34-bp *att*B and 39-bp *att*P recognition sites, which can be recombined together in the absence of cofactors. It does however only permit unidirectional integrations, unlike tyrosine-mediated systems, which permit both integration and excision of transgenic cassettes (Fraser, 2012). The ΦC31 system in particular mediates reliable, stable integrations into a target species. Whilst these systems remain a popular tool for routine transformation of a range of vectors, the requirement for a genomic docking site with associated fitness penalties remains an important consideration when generating transgenic lines for vector control methodologies.

All these early transgenic methods contributed greatly to the ability of researchers to construct more advanced genetic control systems, with greater efficacy and fewer costs. This in turn enhanced our collective ability to manage zoonotic pests, leading to novel vector control strategies.

1.3.3 RIDL Systems

RIDL (Release of Insects Carrying a Dominant Lethal Gene) makes use of an engineered dominant lethal gene as a means of vector control (Fig.1.5). When originally proposed (Thomas et al., 2000; Heinrich and Scott, 2000), the system consisted of two components - a promoter controlling expression of the tetracycline-repressible trans-activator fusion protein (tTa) (Gossen and Bujard, 1992), and a toxic gene under the control of a tTa-dependent promoter (here, *tet*O). The tTa fusion protein consists of the tetracycline repressor (*tet*R) and the C-terminal domain of VP16 from herpes simplex virus (HSV), which has shown to be a powerful transactivation domain in eukaroytes (Labow et al., 1990; Triezenberg, Kingsbury, and McKnight, 1988). In the absence of tetracycline, tTa is expressed and drives expression of the tTa-dependent promoter, in turn expressing a toxic gene product and causing lethality.

A single-component variant has also been designed (Gong et al., 2005), which uses tTa as both activator and effector. The transgene is placed under the control of a suitable promoter, and when tetracycline is withdrawn from modified individuals, tTa is expressed at low levels, binding the promoter. This results in an amplification loop and high levels of tTa expression, resulting in lethal transcriptional inhibition (Gill and Ptashne, 1988). Promoter elements may be selected such that toxicity is dependent on sex or developmental stage, depending on the intended application. Indeed, in cases where there are density-dependent mortality phases, delaying lethality to later developmental stages can enhance the efficacy of RIDL releases (Phuc et al., 2007; Frangioni and Borgioli, 1993; Yakob, Alphey, and Bonsall, 2008; Bax and Thresher, 2009; White, Rohani, and Sait, 2010; Alphey, Alphey, and Bonsall, 2011).

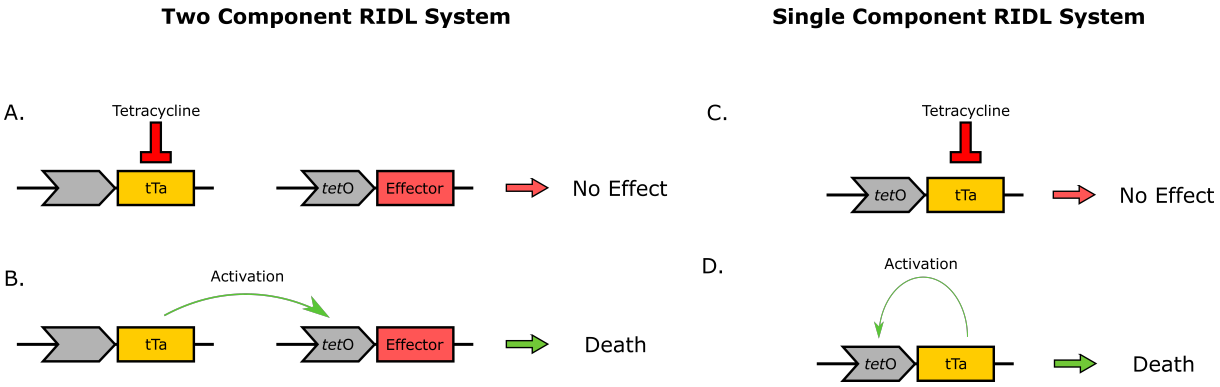


FIGURE 1.5: **RIDL System Topologies**. A) In two-component systems, tetracycine inhibits tTA activity, preventing activation of trans-expressed toxin. B) When tetracycline is withdrawn post-release, tTA becomes active, upregulating the trans-expressed toxin and causing lethality. C) In single-component RIDL systems, tetracycline inhibits tTA activity, but once withdrawn, D) tTA becomes active and starts an amplification loop, resulting in lethality.

RIDL systems were also demonstrated as a combined means of producing genetic sexing strains in mass-rearing conditions and vector control. Female-specific promoter elements were used to control expression of lethal effector genes. In both cases, promoters involved in expression of yolk proteins were chosen - *yp3* (Thomas et al., 2000) and *yp1* (Heinrich and Scott, 2000) respectively). As a result, in the absence of tetracycline, females were eliminated from populations. Female-specific RIDL (fsRIDL) systems thus can act as efficient genetic tool for systemic removal of females from mass-reared populations prior to release. RIDL males mate with wild-type females, producing heterozygous offspring - females die during development, whilst male progeny continue to mate with wild-type females, providing a diminishing, but continued degree of control (Alphey, 2014). Nonetheless, without sustained releases, the strong fitness cost associated with the female-specific lethality gene cause the system to be lost from the target population.

Bi-sex RIDL systems have been developed, allowing for the release of both male and female modified insects. This significantly simplifies the logistics of raising sufficient numbers of modified individuals to effectively suppress target populations, and removes the requirement for sex-specific screening methodologies. The heterozygous offspring of RIDL individuals will

all die as a result of the lethal transgene they carry. To date, bi-sex RIDL systems have been successfully constructed in a number of pest species, including Aedes mosquitoes (Harris et al., 2012; Phuc et al., 2007), pink bollworm (*Pectinophora gossypiella*), and diamondback moth (*Plutella xylostella*) (Jin et al., 2013).

A CRISPR-mediated variant of SIT termed precision guided SIT (pgSIT) (Kandul et al., 2019) has also been developed in *Drosophila*. The system involves release of homozygous Cas9 and gRNA lines, targeting male fertility genes and female viability genes. This results in a self-limiting control system that produces only sterile males as offspring, suppressing a localised population. While the precise nature of this strategy limits the fitness costs associated with traditional SIT methodologies, it is still susceptible to off-target Cas9 effects, and will not be suitable for targeting more broadly distributed vector populations. Despite advances in RIDL system technologies, the technique still has significant barriers to overcome if it is to extinguish malarial outbreaks on a global scale; as with SIT releases, the self-limiting nature of RIDL strategies means that frequent, high-volume releases are required to maintain suppression. Given the geo-political context in which malaria (and other mosquito-borne diseases) are most prevalent, it is likely that there is neither the logistical framework, nor the financial backing to make a substantial and lasting impact.

1.4 Introducing Gene Drive Systems

Conventional genetic control strategies have effectively been hamstrung by their self-limiting nature, preventing them from acting over a large area or persisting in endemic settings where mosquito populations may achieve extreme densities. Gene drive systems were conceived as a means to achieve super-Mendelian rates of inheritance of desired traits (Hamilton, 1967). They make use of selfish genetic elements, which are able to increase their frequency in each generation of a population, even in the face of an associated fitness cost. There are significant advantages to the use of gene drives, allowing for small releases of genetically-modified vectors with a persistent effect that can spread beyond the initial release site (Chan et al., 2011). Broadly speaking, there are two schools of thought when considering gene drive strategies: gene drives which crash a population (suppression gene drives), and gene drives which replace the wild-type field population with genetically modified vectors that carry a desirable trait (replacement gene drives). Gene drives of both kinds have been envisioned as a means to target relevant agricultural and disease vectors, and are the primary focus of this thesis. A number of gene drive systems have been evaluated below to illustrate their advantages and weaknesses -

1.4.1 Meiotic Drive Systems

The aim of meiotic drive systems is to subvert meiosis such that the inheritance of particular alleles is biased during meiosis, in turn increasing their frequency in ensuing gametes and future offspring (Buckler IV et al., 1999) (Fig.1.6). Relevant in the context of vector control are sex-linked meiotic drives, that are capable of skewing sex ratios in a targeted population, resulting in a dearth of one sex (typically females) and population suppression (Hamilton, 1967;

Hickey and Craig, 1966; Craig, Hickey, and Vandehey, 1960). A number of naturally occurring examples of sex-linked meiotic drive have been identified in *Aedes* and *Culex* mosquitoes - both important disease vectors (Hickey and Craig, 1966; Sweeny and Barr, 1978). Y-drive describes sex-linked meiotic drive systems that increase the proportion of males in a given population by biasing the inheritance of the Y chromosome (Lyttle, 1977). Theoretical work proposed the notion of an artificial Y-drive that would make use of an engineered endonuclease to target the X-chromosome in males during meiosis (Burt, 2003).

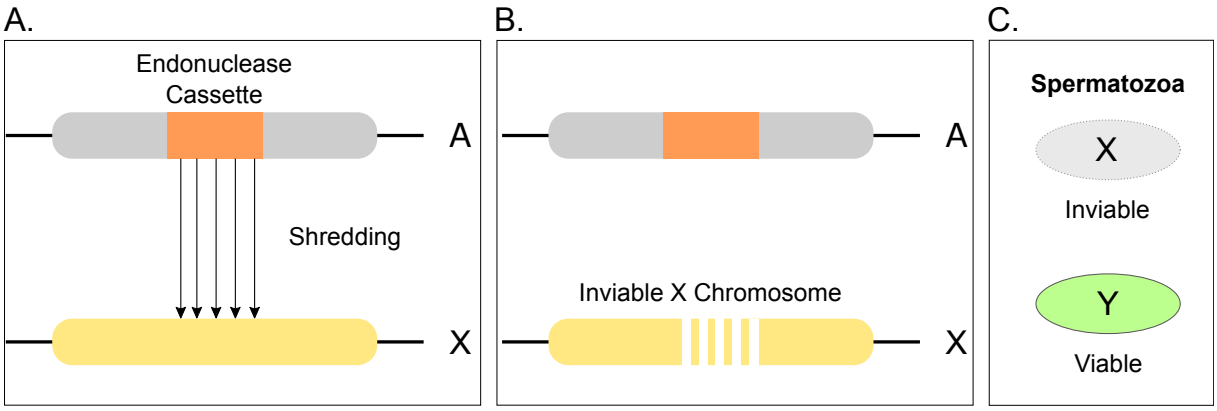


FIGURE 1.6: **X-Shredding Mediated by a Y-Drive System**. A) An autosomally integrated endonuclease cassette targets conserved repetitive sites on the X chromosome during spermatogenesis. B) Shredding results in the destruction of the X chromosome, and as a consequence C) the majority of spermatozoa bear Y chromosomes, introducing a male bias into the resulting offspring. Autosomal integration means that the meiotic drive is inherited at Mendelian rates, requiring a high inundative release frequency for this strategy to be effective. Adapted from Champer, Buchman, and Akbari, 2016.

The first example of an autosomal X-shredder was demonstrated in *A.gambiae*, using an engineered variant of the I-PpoI endonuclease, to target highly conserved rDNA repeats located exclusively on the X-chromosome (Flick et al., 1998). Initial experimentation found that use of wild-type I-PpoI resulted in cleavage of the X-chromosome in both the paternal and offspring, due to the persistence of I-PpoI in mature spermatocytes (Windbichler et al., 2007). To this end, work was undertaken to destabilise I-PpoI, reducing its persistence and preventing its interference in the early zygote. This resulted in production of fertile males that primarily fathered male offspring (Galizi et al., 2014). In this study, the endonuclease was autosomally located, and further work is now required to introduce Y-linkage such that all males inherit the drive element, allowing it to drive through a population (Galizi et al., 2016).

It is important to note that the repetitive rDNA sequences targeted by this drive are specific to *An.gambiae*, and so while this work has provided a foundation stone, other targets will need to be identified to expand the existing strategy to a platform technology. This work has since been expanded using the versatile CRISPR-Cas9 genome editing tool, making the designation of other target sequences effectively trivial (Galizi et al., 2016). An additional concern relates to the formation of resistance alleles, which are discussed in greater detail later in this thesis. When the target sequence is cleaved by an endonuclease, cellular repair machinery may

make use of the error-prone non-homologous end-joining (NHEJ) repair pathway, and introduce/delete bases from the sequence. As a result, the sequence will become resistant to further cleavage. Given the observation that there is natural selective pressure to maintain equal sex ratios (Hamilton, 1967), these resistance alleles are likely to propagate rapidly in a population, undermining the efficacy of the drive. Resistance may also arise in the form of epigenetic silencing or RNA interference (Hammond and Galizi, 2018), limiting expression of the endonuclease.

1.4.2 Underdominance Gene Drive Systems

Underdominance gene drives stem from the observation that heterozygotic chromosomal translocations can result in reduced fertility, which in turn could be used to cause a population crash in a target pest species by reducing the net fertility of the population (Serebrovsky, 1969; Davis, Bax, and Grewe, 2001). When two underdominant traits are present at a single locus, and both confer equal fitness when homozygotic, the initial frequency of each trait determines which is lost from the population (Sinkins and Gould, 2006). These population dynamics can be explained through the Hardy-Weinberg principle, as less common trait spend more time as reduced-fitness heterozygotes, versus the more abundant trait (Reeves et al., 2014). The use of chromosomal translocations as a basis for underdominance systems has had limited success, largely due to difficulties in linking cargo genes to translocations and high fitness costs (Robinson, 1976).

More recently, a number of strategies have been developed which make use of reciprocal transgenes expressing toxins and antidotes (Akbari et al., 2013; Reeves et al., 2014). Such systems consist of two constructs - each expressing a lethal transgene and an antidote for the opposing toxin. When both are present in an individual, the fitness cost of the system should be minimized. However, when separated in heterozygosity, each construct alone should result in lethality (Fig.1.7). It is possible to configure these underdominance systems such that both constructs exist at a single locus or at two separate loci. The spatiotemporal arrangement of these systems is critical for their efficacy and mode of action. To this end, two approaches have been proposed: The first makes use of antidotes which target the promoter of the reciprocal toxin transgene, suppressing its expression (Davis, Bax, and Grewe, 2001). The second more recent approach, termed maternal-effect lethal underdominance (UD^{MEL}), expresses the transgenic toxin maternally, whilst the antidote is expressed in the zygote of subsequent progeny (Akbari et al., 2013).

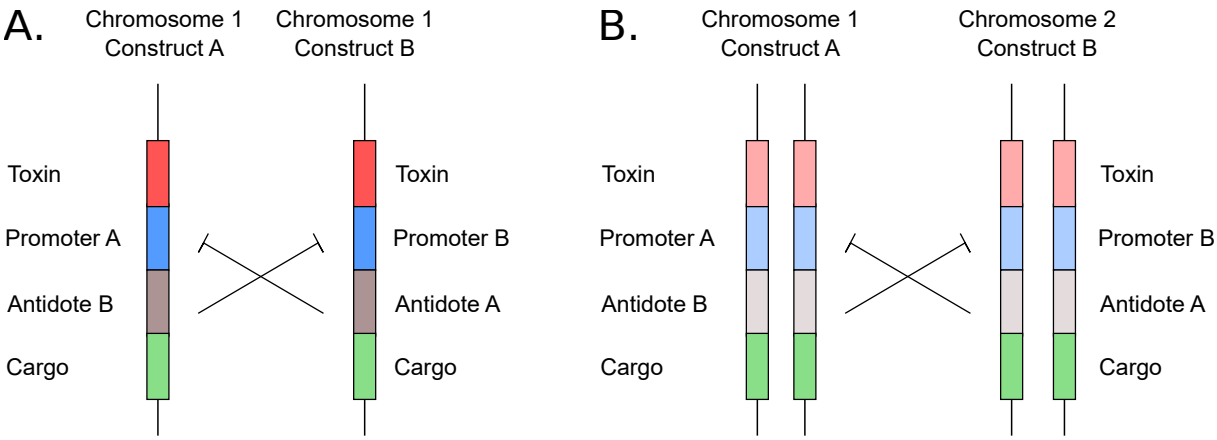


FIGURE 1.7: **One and Two-Locus Underdominance Drives**. A) In a single locus underdominance drive, two constructs are present at the same chromosomal locus, and reciprocate the activity of one another. B) In two-locus underdominance systems, two individual elements are present at separate loci. Each encodes a toxin, and the antidote for the corresponding element. Genotypes that inherit one or more copies of both constructs, or that are wild-type are viable. If underdominance individuals are introduced at a sufficient frequency into a target population, they replace the wild-type population. Figure adapted from Edgington and Alphey, 2019.

Underdominance systems present a number of advantages over other genetic control strategies. Two population models have indicated that underdominance drives can be confined to a single population, assuming there is limited migration between neighbouring populations and a strong negative selection pressure on heterozygotes (Altrock et al., 2010). Theoretical evaluations of underdominance systems have found that the high release threshold required to propagate such a drive system (Davis, Bax, and Grewe, 2001; Akbari et al., 2013) is ideal for preventing unwanted spread of these engineered transgenes, lending to the notion they represent the safest genetic drive strategy (Marshall, 2009). Moreover, introduction of wild-type individuals back into the population can be used to dilute underdominance transgenes and reverse release. Conversely, characteristics which make underdominance strategies suitable for confined, localised releases, may prevent its wider applicability. In the context of endemic malarial settings, it is likely to be impractical to release such large quantities of modified insects, and safety measures such as reversal will similarly be impractical. Other limitations that have been identified with underdominance systems pertain to their complexity and the potential for leaky expression to undermine their performance (Dhole et al., 2018)

1.4.3 Using *Wolbachia* - A Naturally Occurring Vector Control System

While *Wolbachia* are not considered as gene drives, their behaviour closely mimics the super-Mendelian inheritance of conventional drive systems, and are thus described here. This genus of highly invasive, gram-negative bacteria, were first discovered in 1924 (Hertig and Wolbach, 1924) and are estimated to have infected 66% of all insect species (Hilgenboecker et al., 2008). *Wolbachia* have been able to propagate to such a degree due to their ability to bias host reproduction to favour infected females. The reproductive phenotypes of this endosymbiont have been demonstrated to include a range of potentially useful traits for vector control including

cytoplasmic incompatibility (CI) (Champer, Buchman, and Akbari, 2016). CI is a mechanism whereby males infected with *Wolbachia* are able to produce progeny only when mating with infected females. The latter subsequently cytoplasmically transmit the bacteria to their progeny, whilst infected males make no cytoplasmic contribution to progeny, and thus do not transmit bacteria (Fig.1.8). Recent work has been undertaken to elucidate elements responsible for CI in *Drosophila melanogaster*, where the interactions of two adjacent genes, *cifA* and *cifB* was found to be critical for CI (Beckmann, Ronau, and Hochstrasser, 2017; LePage et al., 2017). These genes, involved in *w*Mel *Wolbachia* CI were identified to be part of the eukaryotic association module of WO phage (Bordenstein and Bordenstein, 2016), and the establishment of a Two-by-One model, whereby *cifA* and *cifB* together result in CI in males, and *cifA* rescues CI in infected females (Dylan Shropshire et al., 2018).

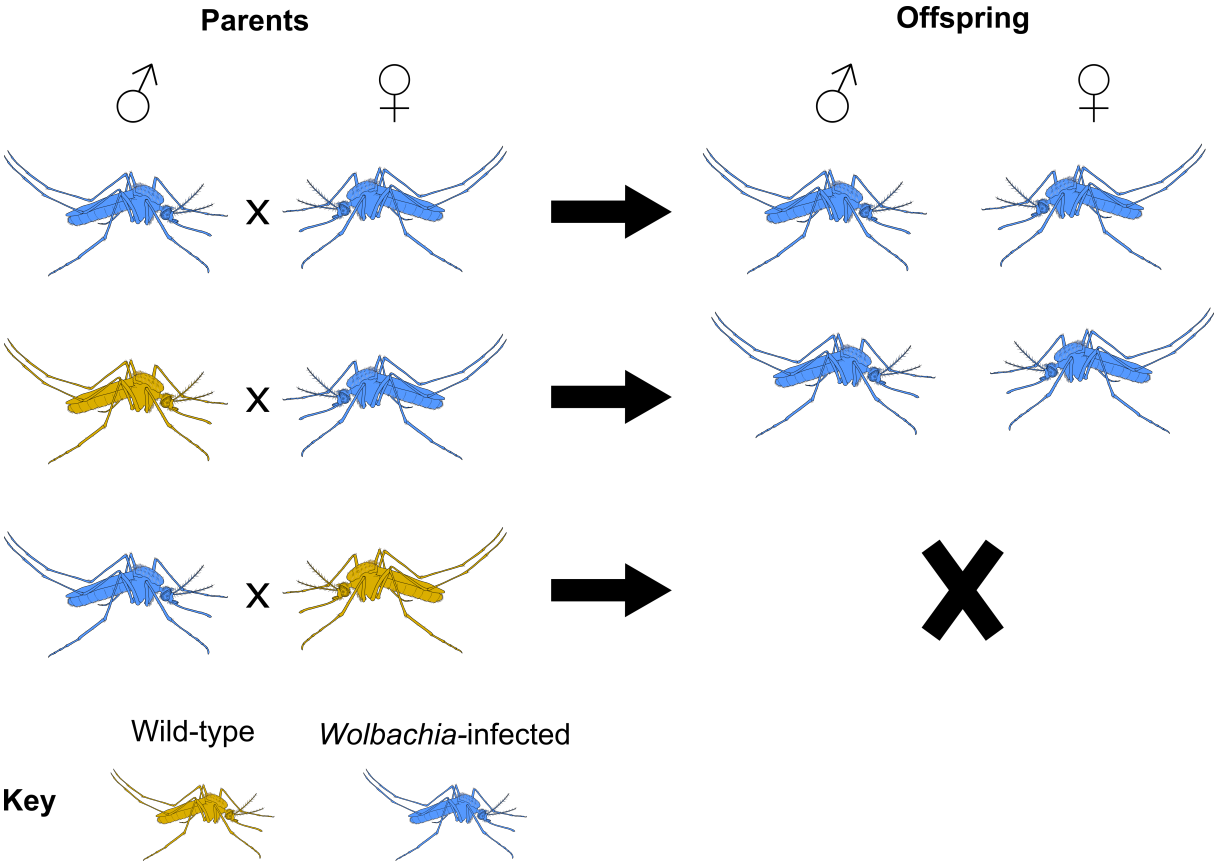


FIGURE 1.8: **Unidirectional *Wolbachia* Cytoplasmic Incompatibility**. When *wolbachia*-infected individuals mate, or unaffected males mate with infected females, infected progeny are produced. When *wolbachia*-infected males mate with unaffected females, the resulting offspring are inviable. As unaffected females are sterilised, and infected females maternally transmit *wolbachia*, the bacteria spreads rapidly through targeted populations.

In the context of vector control, the *w*Pip *Wolbachia* strain from *Culex pipiens Linneaus* has been shown to act to sterilize wild-type females, that harbour other *Wolbachia* strains in *Aedes albopictus* (Moretti and Calvitti, 2013) (Fig.1.9). The *w*Mel *Wolbachia* strain has also been adapted for use in *Aedes aegypti*, an important disease vector responsible for the spread of Dengue fever,

amongst other arboviruses. In this particular case, the use of *Wolbachia* had the dual advantages of being highly invasive, whilst simultaneously blocking transmission of a known dengue serotype (Walker et al., 2011).

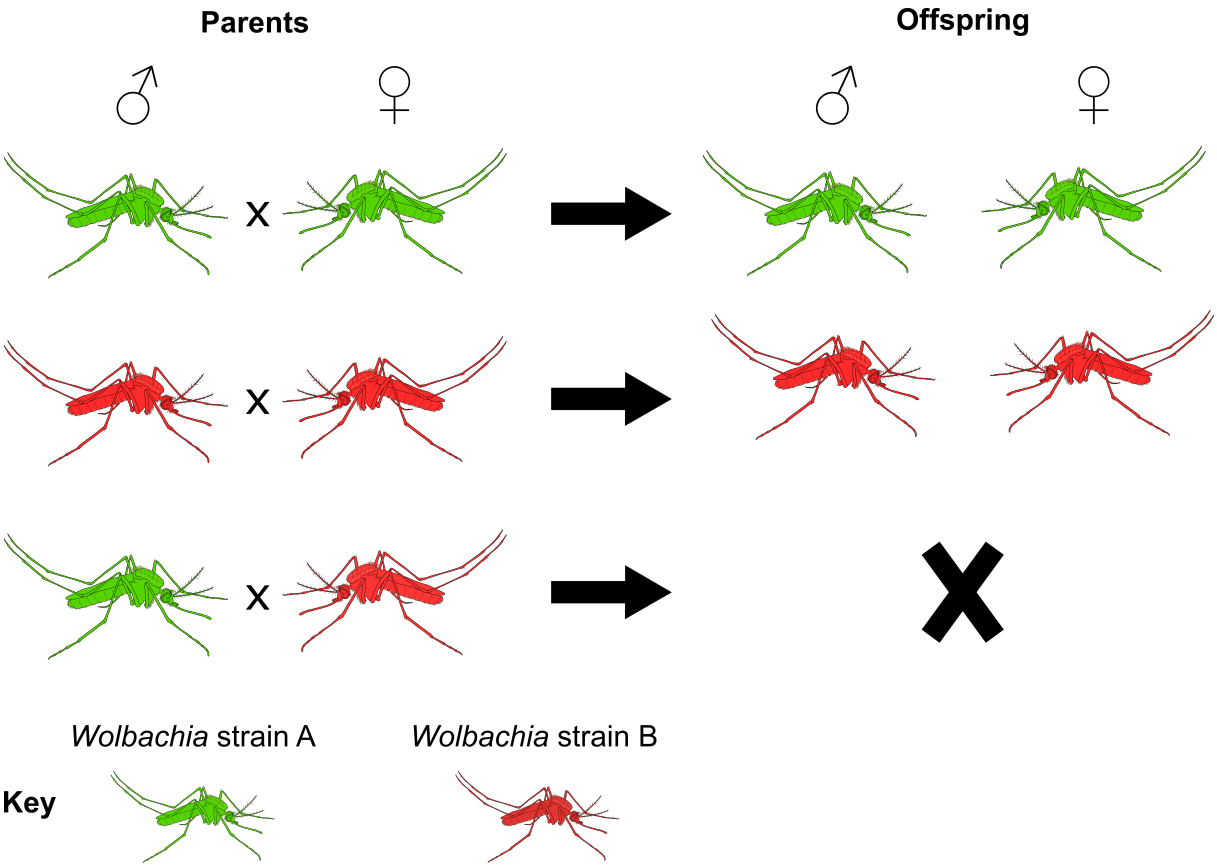


FIGURE 1.9: **Bidirectional *Wolbachia* Cytoplasmic Incompatibility**. A second type of cytomplasmic incompatibility is mediated by the use of multiple *wolbachia* strains. When mosquitoes infected with the same strain mate, they produce offspring infected with the same strain. When mosquitoes carrying incompatible strains mate, they produce no progeny. Figure adapted from (Alphey, 2014).

A number of *Anopheles* populations have been found to be naturally infected with *Wolbachia* strains (Shaw et al., 2016; Baldini et al., 2014), and evidence has shown that some such strains appear to have an anti-pathogen effect on *P.falciparum* suggesting their applicability for use in future vector control strategies (Kambris et al., 2010; Hughes et al., 2011; Bian et al., 2013). The notion of naturally occurring infections has however been brought into contention by meta-analysis of existing evidence (Chrostek and Gerth, 2019). As such, further evidence will be required for this claim to be substantiated.

A practical advantage of using *Wolbachia* as a vector control strategy to combat malaria and other mosquito-borne diseases, is that it is regulated as a microbial biopesticide (Macias, Ohm, and Rasgon, 2017). That said, sociological surveys have revealed negative connotations associated with 'biocontrols' (McNaughton, 2012), which may result in significant distrust and resistance to *Wolbachia* control methods being deployed.

A key concern when using *Wolbachia* in the context of malaria is that infection with the bacteria has been demonstrated to affect *Plasmodium* transmission, both decreasing and increasing the parasite load (Baton et al., 2013). Moreover, while natural *Wolbachia* infections have been detected in Anopheline mosquitoes, there is insufficient data to robustly understand the particular characteristics these strains. It is also worth noting that while versatile gene editing tools such as CRIPSR have become main-stays of molecular biology, there has been no published work pertaining to the engineering of *Wolbachia* as yet, suggesting that attempts to modify them to propagate specific traits or to correct issues with existing strains have yet to bear fruit.

1.4.4 Killer-Rescue Systems

Typically, killer-rescue systems consist of selfish-elements that self-propagate in affected populations by tight linkage of a lethal toxin that acts in *trans*, and a *cis*-acting antidote that diminishes its effect. Individuals carrying such a toxin-antidote system are unaffected by the effects of the toxin, whilst progeny that fail to inherit the system die as the toxin continues to take effect, but without an antidote to counter it. The first system to be identified was the *Medea* system, found naturally occurring in the flour beetle (*Tribolium castaneum*) (Beeman, Friesen, and Denell, 1992), and similar elements have been found in *Caenorhabditis elegans* (Ben-David, Burga, and Kruglyak, 2017), mice (Lyon, 2003), *D.melanogaster* (Larracuente and Presgraves, 2012), and in fission yeasts *Schizosacharomyces kambucha* and *Schizosarcharomyces pombe* (Nuckolls et al., 2017). Engineered *Medea* systems have consequently been demonstrated to function in *D.melanogaster* (Chen et al., 2007; Akbari et al., 2014).

Engineered *Medea* (maternal effect dominant embryonic arrest) systems that have been demonstrated in *D.melanogaster*, and more recently *Drosophila suzukii*, consist of a cassette expressing microRNAs during oogenesis (miRNA), which act to silence expression of an essential gene, expressed in the early embryo (Chen et al., 2007; Akbari et al., 2014; Buchman et al., 2018) (Fig.1.10). The antidote here is a modified variant of the target gene, which is not susceptible to miRNA-mediated silencing, expressed in the zygote. A desirable cargo gene can be linked to the *Medea* cassette, causing it to be driven through the targeted population. Variants upon this architecture exist, such as the inverse *Medea* system (Marshall and Hay, 2011) (Fig.1.10B), which makes use of a zygotically expressed toxin and a maternally expressed antidote, that suppress the heterozygous offspring of wild-type females.

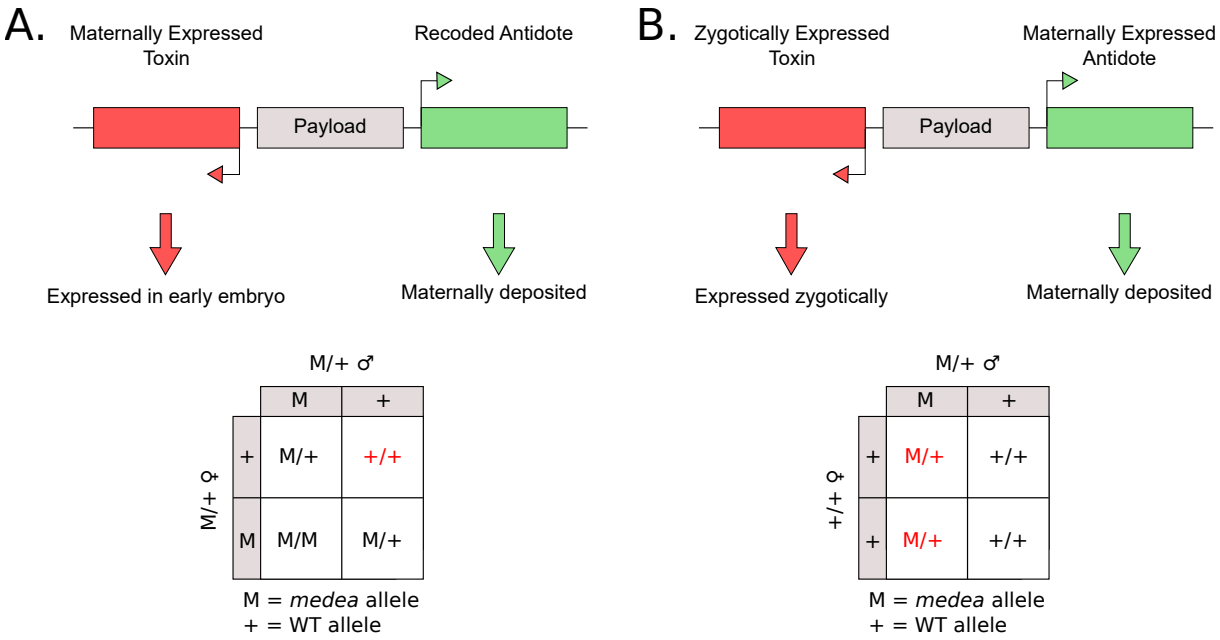


FIGURE 1.10: **Synthetic *Medea* Drive Systems**. A) *Medea* drives function by encoding a microRNA toxin that is active in oogenesis, and a recoded target gene that is expressed zygotically. Offspring inheriting the *Medea* inherit the recoded essential gene and survive, whilst those who do not are inviable. The Offspring of male individuals carrying a *Medea* element is unaffected, as males do not make a significant contribution to the embryo. B) An inverse *Medea* system functions by maternally depositing an antidote and zygotically expressing the toxin. Offspring of wild-type females will be unable to maternally deposit an antidote, rendering any offspring that are heterozygotic for the *medea* element to be inviable. Figure adapted from Chen et al., 2007 and Marshall and Hay, 2011.

Both approaches are suitable for localised population replacement but have limited applicability beyond this as they require a given threshold to be met before they can become invasive. As such they are more suited to the control of isolated island populations. Other more complex variants been devised such as the *Medusa* system (Marshall and Hay, 2014) making use of four components: on both the X and Y chromosomes there are sex-linked toxins, and the reciprocal antidote. The system is designed to select against females causing localised population suppression, though it has a high critical threshold (modelled as >41% of the population) and is susceptible to dilution if wild-type populations re-populate the release site. The *semele* system also makes use of sex-specific toxin and antidote elements, and is suitable for both localised replacement and suppression (Marshall et al., 2011). A more pervasive issue has pertained to the species-specific nature of existing systems. A significant degree of knowledge is required to construct such systems, and they are not broadly applicable to other vectors. This has the advantage of ensuring systems cannot spread beyond a target species, but hobbles the pace of progress. This has been remedied in part by the establishment of CRISPR-based killer-rescue systems, that selectively cleave lethal genes, whilst providing re-coded (and subsequently uncleavable) copies as a means of an antidote (Oberhofer, Ivy, and Hay, 2019); Champer et al., 2019a). Despite this, no such system has yet to have been demonstrated in any mosquito species, limiting its immediate applicability to the prevention of malaria transmission.

1.4.5 Homing Gene Drive Systems

Homing gene drive systems cleave genomic targets inside their hosts, and hijack the subsequently occurring natural repair mechanisms in order to bias their inheritance through populations. A range of endonucleases have been proposed as part of these systems, including homing endonuclease genes (HEGs), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and most recently, CRISPR-mediated endonucleases.

Homing endonuclease genes (HEGs) are naturally occurring selfish elements, that have no beneficial function to their host, but propagate themselves at super-Mendelian rates from one generation to the next (Burt and Koufopanou, 2004). The mechanism by which these homing endonucleases (also known as meganucleases) self-propagate relies on their distinct cleavage and homing behaviour. HEGs encode a homing endonuclease with a 15-30bp recognition sequence, with low frequency in the host genome. HEGs are normally situated within their own recognition sequence, preventing cleavage of the site within the chromosome. Upon expression, the homing endonuclease recognises an uncleaved site on an opposing chromosome and generates a double-stranded break (DSB). At this point the repair machinery of the cell will attempt to repair the break.

There are a number of repair pathways which can be used to repair the DSB, however the presence of an intact chromosome permits homology-directed repair (HDR), whereby the cell uses the uncleaved chromosome as a template. Given that the HEG is contained within this uncleaved template, it is duplicated into the DSB, preventing further cleavage, and increasing its copy number in the host genome in a process termed 'homing' (Fig.1.11) (Burt and Koufopanou, 2004). HEGs are able to insert themselves both between genes and within them. Where the latter occurs, HEGs are able to do this with the aid of inteins and self-splicing to permit them to invade regions with high sequence conservation, without disrupting the function of the locus, spreading with minimal burden to their host (Stoddard, 2005).

A number of examples exist of HEGs whose function is linked to self-splicing introns and inteins. The first characterised HEG, ω, was found to be contained within a self-splicing intron, within an rRNA gene in yeast mitochondria (Burt and Trivers, 2006). When the HEG mediates homing into an unconverted allele, the intron homes without it, allowing the HEG to replicate itself without disrupting the function of the rRNA gene. A similar effect is mediated by the activity of inteins, which allow for protein splicing (Anraku, Mizutani, and Satow, 2005), excising a section of a polypeptide, before annealing the flanking exteins to form a complete protein. Inteins were discovered whilst investigating vacuolar ATPase (*VMA1*) in *S. cerevisiae*. Whilst undertaking this work, a HEG (PI-*SceI*) was discovered within the intein (Moure, Gimble, and Quiocho, 2002). During translation of *VMA1*, PI-*SceI* is spliced from the polypeptide, producing separate functional proteins for both gene and HEG (Burt and Trivers, 2006).

Hosts which are hemizygous for a HEG will home into their unconverted counterpart allele resulting in homozygosity. If this process occurs in the germline of a host, over subsequent generations, it facilitates super-Mendelian transmission of the HEG.

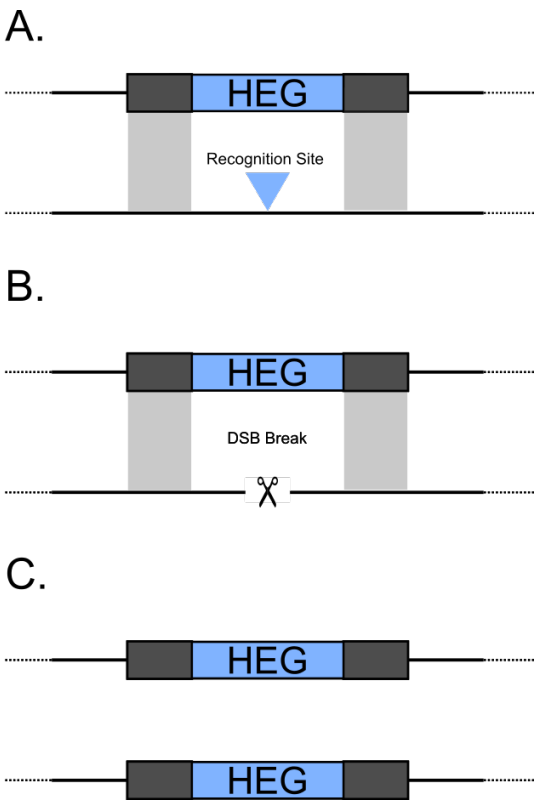


FIGURE 1.11: **Homing Endonuclease Gene Homing Activity**. A) The HEG element recognises its restriction site within the genome. It is flanked by regions homologous to the restriction site. B) The HEG mediates cleavage in the recognition site, generating a double-stranded break. C) Cellular repair machinery can then mediate repair via HDR, using the HEG as a repair template, causing it to home into the cleavage site. In so doing, it disrupts the recognition site, preventing further rounds of cleavage.

The theoretical basis for the use of HEGs as components of gene drive systems was established in 2003 (Burt, 2003). Burt proposed a number of (then) theoretical approaches to the use of HEGs for both population suppression and replacement:

1. For population suppression, a HEG could be targeted to a recessive essential gene in the host, resulting in the loss of function of said copy. Successful targeting of both allelic copies results in mortality or sterility. As the HEG spreads through the population, a 'genetic load' will be created, diminishing the population size until it crashes.

2. Similar to meiotic drive, HEGs can be targeted to the X chromosome, resulting in a male-bias in subsequent progeny. As the drive spreads over multiple generations, the proportion of males increases, lowering or crashing the population.

3. The Linking of a HEG to a desired payload gene that can render a mosquito population refractory to *Plasmodium* infection. The HEG is targeted to a locus with minimal functionality, such that upon insertion, the HEG and payload have a minimal fitness cost on the host, allowing it to be driven to fixation in the population.

4. Using HEGs to selectively knockout the mosquito genes required for malarial transmission. This strategy requires the disruption of the target gene to have a limited fitness cost, allowing the HEG to propagate itself amongst members of the population. Assuming a low fitness cost, resistance alleles will not have a particular selective advantage and should arise at lower rates.

Initial attempts at re-purposing HEGs for gene drive focused on the use of the naturally occurring I-SceI endonuclease in *An.gambiae* (Windbichler et al., 2011) and *D.melanogaster* (Chan et al., 2011). While homing rates were super-Mendelian at approximately 60%, mathematical simulations suggested that higher rates would be required to achieve a level of invasiveness commensurate with vector control (Deredec, Godfray, and Burt, 2011). This subsequently led to a number of attempts to improve gene drive performance by optimising the activity of a variety of HEGs. Work focused on improving HEG expression, as well as protein engineering to improve HEG cleavage efficiency (Chan et al., 2013a; Werther et al., 2017; Thyme et al., 2014; Takeuchi et al., 2011). A limitation of these initial feasibility studies was that HEGs were assayed against synthetic sequences, inserted into the genome of a given host organism. While it has been possible to successfully re-engineer HEGs to target a given sequence (Chan et al., 2013b), protein engineering is typically expensive and time consuming. To this end, other endonucleases have been evaluated as a means to construct homing gene drives.

One such alternative were the zinc-finger nucleases (ZFNs), a class of endonucleases constructed around zinc finger proteins (ZFPs) linked to *FokI* nuclease domains (Urnov et al., 2010) (Fig.1.12). The ZFP consists of an array of Cys2-His2 (Miller, McLachlan, and Klug, 1985), each capable of binding 3bp. It is subsequently possible to engineer these fingers to recognise (and subsequently cleave) a target sequence of choice. Transcription activator-like effector nucleases (TALENs) represent another synthetic endonuclease platform. TALENs contain a conserved, repetitive DNA binding domain derived from transcription activator-like effectors (TALEs) (Joung and Sander, 2013). These TALEs originate from *Xanthamonas* bacteria and were originally part of an invasive mechanism these microbes used to alter gene expression in plant hosts (Boch and Bonas, 2010). Each TALE consists of 33-35 amino acids, and each repeat is able to bind a single DNA base. Modification of the hypervariable 12th and 13th amino acid residue in each repeat permits a given base to be specified. Like ZFNs, TALENs contain a *FokI* nuclease domain, permitting cleavage upon binding of a specified sequence.

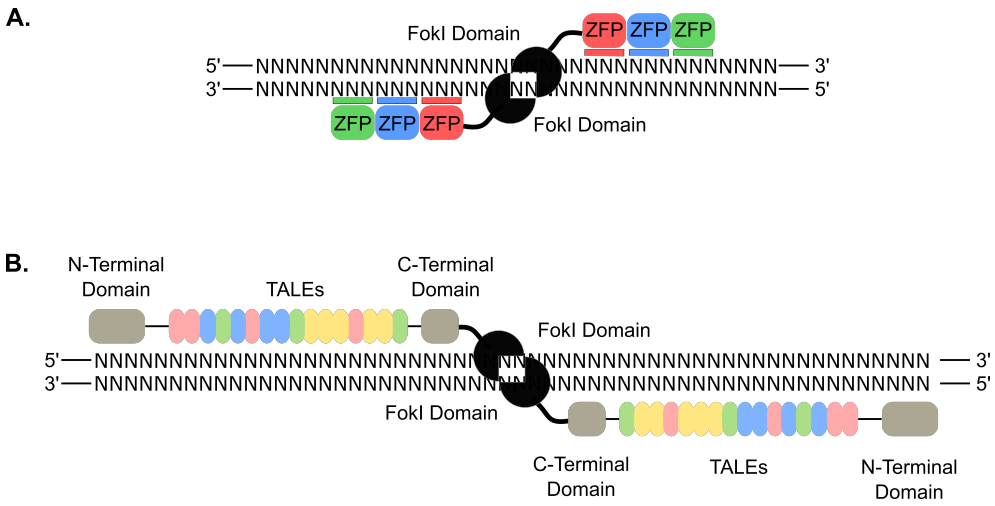


FIGURE 1.12: **ZFNs and TALENs Overview)**. A) ZFNs are constructed from ZFPs fused to a FokI endonuclease domain. Between 3-6 ZFPs direct site recognition, each binding 3 nucleotides. Upon dimerization, ZFNs are able to generate DSBs at a recognised target site. B) Similar to ZFNs, TALENs are constructed from TALE arrays fused to FokI endonuclease domains. Each TALE dictates binding to a single nucleotide, and dimerization of TALENs is required for cleavage to be affected. An N-terminal domain required for DNA binding, and which interacts with a conserved thymidine sequence is shown, as is a C-terminal region carrying nuclear localisation signals. Figure adapted from Joung and Sander, 2013

The homing capacity of ZFNs and TALENs have been trialled in *D.melanogaster*, where the authors generated 'synthetic selfish elements' targeted to an artificial recognition sequence (Simoni et al., 2014) under the control of a male-specific promoter. While both were able to elicit homing, high NHEJ rates were observed when compared to the naturally occurring *I-SceI* HEG, which is primarily repaired by HDR (Kuhar et al., 2014). An additional issue stemmed from the repetitive nature of ZFNs and TALENs - the SSEs that were constructed were prone to recombination, meaning that 60% of TALENs and 25% of ZFN mediated homing produced alleles that were unable to undergo a second round of homing. Thus, while these synthetic endonucleases were significantly more amenable to modification, they failed to perform in a manner that would be useful for vector control in homing gene drives. Many of these shortcomings have since been overcome by the advent of a novel genome editing tool - CRISPR systems.

1.5 Introducing CRISPR-Based Drives

1.5.1 The Fundamentals of CRISPR-Mediated Immunity

As previously described, a range of engineered endonucleases have been developed for precise genome editing in a range of organisms. The discovery of clustered regularly interspaced short palindromic repeats (CRISPR) has further served to expand this class of tools. While CRISPR repeats have only been developed into functional tools in recent years, they were originally observed in 1987, as researchers undertook work to understand the conversion of alkaline phosphatase in *E.coli* (Ishino et al., 1987). It was not until 2002 that this family of repetitive DNA

sequences was named (Jansen et al., 2002) and recognised as being widespread in a range of prokaryotes, including *Archaea* and both gram-negative and gram-positive bacteria (Mojica et al., 2000). The CRISPR system functions endogenously as an innate immune system to combat bacteriophage and plasmid invasion (Terns and Terns, 2011). CRISPRs consist of an array of 20-50bp repeats, each of which is separated by a unique spacer (Sorek, Lawrence, and Wiedenheft, 2013). Arrays are flanked by a leader sequence, which contain promoters and binding sites required for expression of system elements (Pougach et al., 2010). Functionally adjacent to the CRISPR arrays are cassettes of CRISPR-associated *cas* genes, also involved in mediating CRISPR immunity. A vast diversity of CRISPR systems have been identified (as recently reviewed by Makarova, Wolf, and Koonin, 2018), however, the mechanism by which immunity is generated can be broadly described in three stages:

1. **Adaptation** - This phase allows bacteria/archaea to build an arsenal of spacer sequences with which to recognise predators. Originally demonstrated in *Streptococcus thermophilus* (Barrangou and Marraffini, 2014), it was found that following infection with bacteriophage, some bacteria were able to counter their invaders and had incorporated elements of the phage genome as spacers into CRISPR arrays. Spacer acquisition is not random: sequences selected to be integrated into CRISPR arrays (termed proto-spacers) were found to be proximal to conserved motifs, referred to as proto-spacer motifs (PAMs). Different PAMs have been identified for different CRISPR subtypes (Mojica et al., 2009). It is important to note that the PAM site recognition is critical for the success of CRISPR immune systems - mutations in the PAM may render a previously identified phage/plasmid immune to interference.

2. **crRNA Biogenesis** - Having incorporated protospacers into CRISPR arrays, it is now possible for the host to express CRISPR-RNAs (crRNAs) complementary to the DNA sequences of a known predator, allowing them to be targeted for destruction. CRISPR arrays are transcribed as long pre-crRNAs, which are subsequently processed into shorter crRNAs by Cas proteins or RNases depending on the system type. Type II systems, which are the basis for the most commonly used CRISPR/Cas9 tools, make use of a trans-activating crRNA (tracrRNA) to aid in the maturation of crRNAs.

3. **Interference** - Upon maturation, crRNAs are able to form RNA:protein complexes with Cas proteins, that are capable of recognising and cleaving foreign DNA. In the case of the type II system, Cas9 endonuclease interfaces with the processed crRNA and tracrRNA that have formed an RNA duplex, which guides it to the PAM sequence in the target DNA. There, Cas9 generates a DSB, destroying the invader.

The key breakthrough in the use of endogenous CRISPR systems as genome editing tools came in 2012 (Jinek et al., 2012). The type II CRISPR system of *Streptococcus pyogenes* was adapted, and crRNA and tracrRNAs fused to form a single guide RNA (sgRNA). In so doing, it became simple to target any given genomic locus by synthesising a suitable sgRNA - assuming the target was adjacent to a suitable PAM site. In the case of SpCas9, this is an NGG motif found 3′ to the spacer sequence in the target DNA. DSBs are generated 3bp upstream of the

PAM within the spacer sequence. Since the initial establishment of CRISPR as a versatile and powerful genome editing tool, a range of variants have been developed to expand its remit beyond simple DNA cleavage. The Cas9 protein contains two endonuclease domains with homology to known HNH and RuvC domains. By introducing two point mutations into these domains (D10A and H840A in SpCas9), the cleavage activity of Cas9 is entirely ablated. However, it is still able to recognise and bind DNA as part of a Cas9:sgRNA complex. This variant is known as dead Cas9 (dCas9) and has a range of applications. These include CRISPR interference (CRISPRi) (Qi et al., 2013) and CRISPR activation (CRISPRa) (Bikard et al., 2013), whereby dCas9 can hinder or upregulate transcription of a targeted operon. DCas9 has also been fused to epigenetic factors, allowing researchers to systematically target and activate previously silenced genes (Hilton et al., 2015). It has similarly been used to visualise DNA sequences allowing one to study chromatin configurations in living cells (Roux et al., 2012) as well as as to study chromatin composition (Anton et al., 2014).

Limitations of CRISPR Editing

An important consideration when using CRISPR systems for genome editing is the possibility of off-target effects, which were illustrated by a number of early studies (Hsu et al., 2013; Pattanayak et al., 2013; Fu et al., 2013). Genome-wide analysis, conducted via chromatin immunoprecipitation (ChIP-seq) and next-generation sequencing (NGS), typically illustrated that mismatches in the sgRNA to the target site are tolerated more readily when they are distal to the PAM site, whilst proximal mismatches were poorly tolerated (Kuscu et al., 2014). These studies utilised catalytically inactive Cas9, and later work demonstrated lower tolerance for sgRNA mismatches for cleavage, when compared to target recognition and binding (Wu et al., 2014). Despite a range of studies aimed at understanding off-targets in the context of actual cleavage, the number of parameters involved has meant that a true appreciation has yet to be achieved (Adli, 2018).

In the context of homing gene drives, off-target effects may be highly detrimental to the fitness of a targeted individual, reducing the efficacy of the drive. A number of strategies have been devised to lower the frequency of off-target CRISPR activity. A number of computational methodologies have been devised to model potential off-target effects, and make *in silico* predictions about putative sgRNAs for a given locus. The inclusion of energetics parameters has since improved the quality of computational predictions further (Zhou et al., 2017; Alkan et al., 2018; Zhang et al., 2019; Young et al., 2019). As previously discussed, attempts have been made to experimentally validate predictions made by computational tools. For instance, genome-wide sequencing work using GUIDE-Seq and DISCOVER-Seq (Tsai et al., 2015; Wienert et al., 2019) has been attempted. However, due to differences in methodology, results have differed limiting current interpretations.

An alternative approach has been to limit the *in vivo* availability of of Cas9-sgRNA complexes, diminishing off-target binding. This has followed a hypothesis that posits off-target binding and cleavage occur once all desired target sites (which are preferentially targeted) have

been occupied (Fu et al., 2013; Hsu et al., 2013). Two studies have illustrated the means to achieve this goal: either by producing a Cas9:sgRNA complex *in vitro* and providing it directly (Kim et al., 2014), or by using an inducible Cas9 system, diminishing its activity in the absence of an activation molecule, resulting in a 25-fold increase in specificity (Davis et al., 2015). Both strategies reduced off-target effects, but at the cost of reducing cleavage frequencies.

As noted above, Cas9 proteins contain two enodnuclease domains: Firstly, by engineering a variant of SpCas9 with a D10A mutation in the RuvC domain, it was possible to generate a DNA nickase, capable of generating single strand DNA breaks (SSBs) (Cong et al., 2013). Secondly, by pairing Cas9-nickases, each targeted to make an SSB on an opposing strand at a given locus, it was possible to reduce off-target effects, though once again, at the cost of reducing on-target efficiency (Cho et al., 2014). Other similar approaches have focused on the rational design of novel variants of canonical Cas9, in an attempt to lower off-target effects, and were described recently in (Li, Zhou, and Zeng, 2019). As yet, an engineered Cas9 variant is to be produced which absolutely abolishes off-target effects.

A final approach focuses on modification of the sgRNA component. Truncation of the sgRNA to as few as 17bp to have fewer complementary bases has been shown to diminish off-target activity over 5,000 fold in U2OS cells (Fu et al., 2014). More recently, modifications have been introduced to the 5-terminus of the sgRNA spacer, introducing synthetic secondary structures that weaken interactions between the Cas9 endonculease and target DNA sequence. Subsequently, it is less energetically favourable for off-target effects to occur, reducing their frequency, though the current mechanism by which this occurs remains unknown (Kocak et al., 2019).

Since the advent of CRISPR, this versatile genome editing tool has been used to dramatically advance the efficacy of homing gene drives. Shortly after its establishment, conceptual work was published, positing CRISPR/Cas9 as a replacement for older endonuclease platforms within drive architectures (Esvelt et al., 2014). This in turn was followed by implementations of RNA-guided gene drives in *D.melanogaster* (Gantz and Bier, 2015), *S.cerevisiae* (DiCarlo et al., 2015), *An.stephensi* (Gantz et al., 2015), *An.gambiae* (Hammond et al., 2015), and mice (Grunwald et al., 2019). In Anopheline mosquitoes, transmission rates were observed at rates as high as 99.6%, making significant improvements on previously characterised homing gene drives.

1.5.2 Implementing CRISPR-mediated Homing Drives

An important barrier to the successful implementation of RNA-guided homing drives are behavioural or molecular traits, developed in either mosquito or parasite which confer some degree of resistance to the activity of the drive, be it for population suppression or replacement. In addition to those mechanisms noted above (see: Meiotic Drive Systems), assortative mating, parasite/vector mutation, and formation of cleavage-resistant alleles (Hammond et al., 2017). In particular, the generation of resistance alleles is a significant hurdle to the performance of

all homing gene drives, causing drives to reach a state of psuedo-fixation within a population before being lost. Gene drives function by biasing their own inheritance through germline cleavage and homing, using the cellular HDR mechanism as a copy-and-paste process. If repair occurs instead by the error-prone non-homologous end joining (NHEJ) or by microhomology-mediated end-joining (MMEJ) pathways, they are liable to producing indel mutations altering the sequence of the target site. This in turn may preclude further cleavage and the generation of a resistant allele. This differs considerably with HEG-based drives, as HEGs exhibit a higher degree of plasticity in their recognition of target sites, permitting them to cleave similar variants of their specific recognition sequence (Jurica and Stoddard, 1999). Given the fitness costs associated with homing gene drives, particularly for replacement strategies which carry a payload gene that must be expressed, resistant alleles with low cost can take over. This is further compounded by the prevalence of single nucleotide polymorphisms in field populations of Anopheline mosquitoes, which are estimated to occur every 2.2 bases in *An.gambiae* (Miles et al., 2017).

An important distinction to make at this juncture, is the difference between resistance alleles that render genes non-functional, versus those that leave gene function intact. Cleavage events resulting in repair by NHEJ may introduce in-frame point mutations, or frame-shifts that disrupt the function of the underlying target gene. This in turn is liable to introduce a fitness cost to that individual, causing it to be selected against on a population level. By contrast, repair events that disrupt the drive target sequence but do not disrupt the function of the target gene are a significant issue as they will not infer a fitness cost on the individual, allowing such alleles to accumulate and disrupt the continued function of a gene drive in that population.

A key mechanism for the formation of resistance alleles has been identified as maternal deposition of Cas9 (Champer et al., 2017). The aforementioned study made use of the well-characterised *vasa* promoter, used in the earliest iteration of a CRISPR homing drive in *Drosophila*, and the *nanos* promoter. The *vasa* promoter has been demonstrated to have leaky somatic expression (Champer et al., 2018b), whilst *nanos* is still expressed in the germline but has tighter spatiotemporal control. It was noted that gene drive lines using the *vasa* promoter encountered higher rates of post-fertilisation resistance, emphasising the importance of selecting a promoter element for Cas9 expression with a suitably narrow window of expression, limiting resistance allele formation.

There has been a number of strategies proposed to limit resistance. In the context of maternal deposition, the use of male-specific promoters may reduce resistance allele formation. This is because sperm makes a much smaller contribution to the egg, and so paternal deposition of Cas9 is less likely to occur. Recent evidence has suggested that by restricting drive conversion events to the male germline, it may be possible to suppress resistance arising by this mechanism entirely (Marshall et al., 2017). The choice of endonuclease may also beneficially affect resistance rates. For example, the use of endonucleases which cleave proximal to the target site but outside the target itself, such as the dCas9-FokI fusion protein. Such an approach is tolerant

to SNPs, and cleavage-induced mutations at the cleavage site will not prevent further rounds of targeting and cleavage by a hypothetical drive system.

Another approach is to target a locus which is highly conserved. Recently, the first laboratory mosquito populations were successfully eliminated using a suppression strategy that targeted the *doublesex* gene in *An.gambiae* (Kyrou et al., 2018). This gene has two transcript products that regulate sexual differentiation. The female transcript contains a highly conserved exon. By targeting this functionally constrained locus, females homozygous for the drive were rendered phenotypically intersex and sterile, resulting in eventual population collapse. As this locus is seemingly constrained by structural or functional considerations, it is intolerant to indel formation, limiting the creation of resistance alleles. It is important to caveat these observations - as yet there is no evidence to conclude that resistance alleles cannot arise when targeting loci of this kind.

The above approach is only suitable for suppression strategies. Replacement strategies will need to employ other solutions to overcome resistance allele formation. One such methodology has been in the targeting and rescue of a haplolethal gene (Champer et al., 2019b). Haplolethal homing drives target genes which require two functional alleles in order for the individual to be viable. If the drive should cleave a target site but fail to home, cellular repair machinery will attempt to resolve the DSB. If repair occurs by NHEJ, indel mutations may be introduced. These will either be in or out of frame, resulting in resistance alleles as the sgRNA target site is disrupted, with the latter resulting in frameshift of the haplolethal target gene. Such activity will render the individual non-viable and this type of resistance allele will be lost from the population, slowing the rate at which resistance alleles accrue. The haplolethal drive carries a functional copy of the target gene, meaning that it is able to rescue the gene functionality after cleaving and homing into the site.

A final approach that has been posited and since demonstrated in *S. cerevisiae* (Yan and Finnigan, 2018), is the use of multiplexed gRNA cassettes homing a drive into multiple loci within the host genome. In the case of suppression drives, mutations which prevent further cleavage of a locus must occur at all loci in a single organism, or else it will be eliminated from the population and with it the resistance alleles. In the case of replacement strategies, multiplexing offers a high degree of functional redundancy, increasing the longevity of a drive in a population. Drive elements and associated payloads may be propagated to multiple loci, requiring resistance alleles to arise at all loci before complete efficacy is lost.

Biosafety and Ethical Considerations

The very characteristics that make gene drive technologies particularly suitable for use as self-propagating, highly invasive genetic control tools, are the same that have raised significant

concerns with regards to biosafety. This is particularly true given that this work aims to expand existing replacement strategies, which aim to modify entire field populations. Recent work has been conducted examining the impact of gene drive technologies on four categories of organism considered to be at risk - targeted field populations, non-targeted field populations, cultivated crops and livestock, and humans (Oye et al., 2014). It was determined that non-targeted field populations, and cultivated fauna and flora are particularly sensitive to accidental/malicious releases of modified organisms carrying gene drives. This has emphasised the importance of undertaking thorough laboratory and field trials, and developing suitable biological controls to limit the impact of unwanted releases. The risk of gene drive technologies being weaponised to maliciously modify human populations or food sources has also been evaluated (Gurwitz, 2014). It is unlikely that gene drives would be able to propagate in commercial seed farms or livestock facilities in developed nations, due to the use of genetic screening, and the use of artificial insemination. That is not to say that malicious gene drive releases could not negatively affect food resources in developing nations that do not routinely observe such agricultural practices.

The invasiveness of gene drives, and their ability to self-perpetuate also lends to concerns regarding the spread of genetic modifications beyond geopolitical boundaries. Modelling has demonstrated that even low-level accidental releases can result in widespread propagation of the drive to non-targeted populations (Noble et al., 2017). Such observations have led to the development of containment systems for genetic drives, limiting the possibility of such an outcome. For example, the proposal of a split drive strategy, whereby components necessary for successful homing are confined to separate lines until the moment that a release is required, and the lines are crossed to generate a fully functional drive system (Amo et al., 2019). Self-limiting strategies have also been described, as part of 'Daisy Chain' drives - these linear systems consist of a chain of elements, the final element being a payload gene. Each element requires the previous element in the chain to be active for it to home. Over time, the anchor element in the chain is lost from the population, and the chain collapses, imposing spatiotemporal limitations on the propagation of the drive system.

A number of counter measures have been designed to limit the damage caused by accidental/malicious releases of gene drives. One such methodology, most applicable when countering suppression drives, is to release modified insects carrying a resistance allele (Vella et al., 2017). Hypothetically, this should result in the gradual decrease and pseudo-fixation of the gene drive, and its decline in the population over subsequent generations. This strategy is not without limitations however, as replacement gene drive systems are typically designed with low fitness cost in mind. Consequently, the released population carrying the synthetic resistance allele would have no particular competitive advantage over the drive population, diminishing the effect of their release.

Alternative measures include the deployment of a second, 'reversal' gene drive system. A number of strategies have been elucidated including Cas9-triggered chain ablation (CATCHA)

(Wu, Luo, and Gao, 2016), and Elements for Reversing Autocatalytic Chain Reaction (ERACR) (Castranova et al., 2016). CATCHA functions by expressing a gRNA cassette that is targeted to a site within the Cas9 of a released gene drive. The CATCHA transgene does not contain a Cas9 cassette of its own, so it cannot home in wild-type field populations of insects. In this way, CATCHA transgenes can disrupt run-away drive systems and neutralise them. ERACR transgenes function in a similar manner, consisting of several gRNA cassettes that are targeted to sites adjacent to the run-away gene drive in the insect genome. Upon encountering the run-away drive, its own Cas9 source facilitates complexing of ERACR gRNAs, which act to excise the drive from the host insect genome, replacing it with the ERACR cassette. It is worth noting that modelling has suggested scenarios in which release of such reversal drives is insufficient to clear the targeted drive (Vella et al., 2017). Moreover, such elements have their own associated risks, which must be evaluated when considering suitable countermeasures.

In addition to the reversal drives noted above are piggybacking systems such as Construct Hitchhiking on the Autocatalytic Chain Reaction (CHACR). In this system, the CHACR element utilises Cas9 expressed by the first drive to simultaneously excise and replace said drive, whilst also copying itself into wild-type alleles. The system can theoretically be used to correct off-target damage generated by the first drive at other sites, immunise potential target sites by modifying them, or correct resistance alleles as they accrue in a field-population (Gantz and Bier, 2016). CHACR can also be combined with ERACR (e-CHACR) facilitating removal of the original gene drive entirely.

Whilst genetically modified eukaryotes have existed for some years, there is little regulation specifically advising on the development and application of gene-drive organisms (Tang et al., 2017). In the context of genetically modified insects, regulatory guidelines in the USA have not been updated since 1992 (Holdren et al., 2015). Moreover, the regulatory framework is upheld separately by multiple federal agencies such as the Food and Drug Administration, US Department of Agriculture, and Environmental Protection Agency. A direct consequence is regulatory heterogeneity, and a tendency to pass the onus to another agency. By contrast the European Union employs a single organisation, the European Food Safety Agency, to oversee application of a regulatory framework for transgenic organisms. Unfortunately, outdated definitions of what constitutes a transgene mean that recent innovations using CRISPR/Cas9, and now deaminase base editors, limit the applicability of this regulatory framework. (Holdren et al., 2015). It is thus critical that a current and unified framework is developed to regulate the creation and use of genetically modified organisms, covering gene drive technologies and other more conventional applications.

As noted above, suppression gene drives may be argued to pose higher environmental risks, primarily due to their ability to cause population crashes of target species - though this has only been demonstrated in laboratory conditions. The removal of a key prey species required to sustain predator populations may lead to a change in behaviour, resulting in increased predation other potentially sensitive species. Similarly, targeting of a predatory species, may result in

devastating consequences further down the food-chain. A notable example of this pertains to the removal of wild goats and pigs on the Sarigan islands (Kessler, 2001). While these species posed a threat to local biodiversity as invasive herbivores, it transpired that they had been limiting the effects of an invasive vine species (*Operculina ventricosa*). Upon their removal, the vine quickly colonised large parts of the islands. It is also worth considering that alteration of insect population densities (even temporarily) may lead to the creation of novel ecosystems, or provide suitable prey for endemic species (Rode et al., 2013). These risks are not limited to homing gene drives, but are broad considerations that must be made when evaluating an environmental management strategy of any kind. This thesis focuses on the development of a novel replacement drive, bringing with it a different set of challenges.

1.5.3 Population Modification Approaches

Homing gene drive strategies against malaria that aim to modify a field population need to carry a cargo that will impart a trait beneficial to the reduction of *Plasmodium* transmission. The earliest concept of population modification was described by Christopher Curtis in 1968 (Curtis, 1968), and since then numerous antimalarial effectors have been characterised for use in genetic vector controls. A notable advantage of population replacement strategies, when compared to suppression strategies, is that their success is not predicated on the elimination of a target population or species. This is because they do not cause the collapse of a vector population that may occupy a crucial niche, potentially undermining the survival of other organisms or permitting another more dangerous species to take its place (Carballar-Lejarazú and James, 2017). Replacement strategies have been posited to be robust to low population density (Eckhoff et al., 2017). Moreover, modified populations should be able to persist, ensuring superior coverage, whereas suppression strategies may cause collapse of an effected population without reaching isolated wild-type populations. First described in 2003 (Burt, 2003), population replacement (also known as population modification) is arguably less drastic than its suppressive counterpart as it does not cause eradication of a field population, but instead converts it into a genetically modified population with reduced vectorial capacity. To achieve protection of the local human populus, population replacement drives propagate effector transgenes involved in modulating mosquito immune responses to parasite invasion, or mediating exogenous expression of effector molecules.

A number of potential anti-pathogen effectors have been investigated for use in the fight against malaria. These include peptides lethal to the parasite, peptides that block parasite invasion by interacting with mosquito proteins, antibodies, and immune system modulators (Wang and Jacobs-Lorena, 2013). An important consideration when introducing such effectors, is their spatiotemporal regulation. There are a number of key bottlenecks in the *Plasmodium* life cycle within the Anopheline mosquito. By selecting a promoter element with a suitable expression pattern, it is possible to limit the susceptibility of a modified mosquito to *Plasmodium* infection. Once intracellular stages of the parasite progress to infectious extracellular developmental stages for uptake by female Anopheline mosquitoes as part of a bloodmeal, they become vulnerable to elements of the human and mosquito immune system (Smith, Vega-Rodríguez, and

Jacobs-Lorena, 2014). *P.falciparum* numbers may decline from hundreds of gametocytes to as few as two oocysts once they have infected a mosquito host. Consequently, when considering suitable promoters for regulation of antimalarial effectors, correct spatiotemporal expression is important. Expression of effectors within the midgut or as part of the surrounding peritrophic matrix permits targeting of the parasite from the gametocyte stage through to oocysts. Expression in the salivary glands permits targeting of the sporozoite stage prior to vector-human transmission.

The development of tools for insect transgenesis has facilitated the conceptualisation of numerous vector control strategies that express anti-malarial effector genes. The aim is to use such genes to kill or block transmission of the parasite without imposing a fitness cost on the host. An important consideration when introducing such effectors is their spatiotemporal regulation. The carboxypeptidase (CP) promoter is a commonly used regulatory element for mediating expression of effector transgenes in the midgut following ingestion of a blood-meal (Moreira et al., 2000). Originally characterised in *Aedes aegypti*, promoters were derived from both *Ae.aegypti* and *An.gambiae CP* genes.

Additional midgut promoters have been characterized for use in *An.gambiae*, including those derived from *Antryp1* and *G12* genes (Nolan et al., 2011). The *An.gambiae VGT2* gene has similarly been used to generate a blood-feeding induced vitellogenin (Vg) promoter for expression of transgenes in *An.stephensi*(Nirmala et al., 2006). Work has also been conducted to use *Peritrophin1* to mediate midgut expression of effector transgenes in *An. gambiae* (Astrid Hoermann, Personal Communication). The efficacy of any given anti-malarial effector is intrinsically linked to its expression pattern, and thus promoter choice is as important as the choice of effector.

Transmssion Blocking and Anti-microbial Peptides

The earliest attempt to use an effector gene to block parasite transmission made use of SM1 (salivary gland- and midgut-binding peptide 1) (Ito et al., 2002). This 12-amino acid peptide, discovered by screening a bacteriophage library, binds the epithelium of the salivary glands and midgut surface - both of which must be traversed by the parasite for it to continue its life-cycle. SM1 competes with the parasite for important midgut and salivary gland ligands blocking their development and transmission (Ghosh, Ribolla, and Jacobs-Lorena, 2001). A synthetic gene expressing four linked copies of the peptide under the control of the *An.gambiae* carboxypeptidase promoter elicited an 81.6% reduction in oocyst formation, significantly blocking transmission. Other AMPs such as Vida3 (Meredith et al., 2011) and bee venom phospholipase A2 (*PLA2*) (Moreira et al., 2002) have similarly been demonstrated to inhibit parasite development in modified individuals. When evaluating the validity of these studies, it is important to take note of the specific parasite species being tested against. In the case of Vida3, robust transmission blocking was demonstrated against the murine parasite, *Plasmdoium yoelii*

nigeriensis, but inconsistent efficacy was observed against the human parasite, *Plasmodium falciparum*. Variations in the molecular activity of differing parasite species is liable to generate variations in the efficacy of any given effector gene.

Antibodies

Another approach to the use of effector transgenes to block of reduce parasite transmission revolves around the discovery that monoclonal antibodies (mAbs) are able to target parasite proteins, either on the surface of the parasite, or secreted during invasion of the host. A range of targets and mAbs have been assayed against *P.falciparum*. These include 4B7, which binds Pfs25, a sexual stage antigen, and which can block or reduce oocyst development in the mosquito midgut (Barr et al., 1991). A different approach utilises 1C3, preventing the formation of oocysts by targeting chitinase 1, a secreted enzyme required for the parasite to traverse the mosquito midgut (Li et al., 2004). Monoclonal antibodies have also been demonstrated to target circumsporozoite protein secreted by the parasite, interfering with their ability to invade liver hepatocytes in cell culture (Hollingdale et al., 1984). Unfortunately, the size and complexity of mAbs inhibits their usefulness in the context of generating transgenic mosquito strains refractory to parasite infection (Isaacs et al., 2011).

A potential alternative has since been realised in the form of single-chain antibodies (scFvs). These antibody fragments consist only of a variable light chain and heavy chain bound by a polypeptide linker (Raag and Whitlow, 1995). These recombinant fragments retain the binding ability of mAbs whilst having the advantage of requiring a single transcription unit for expression. Single-chain fragments derived from 1C3, 4B7, and 2A10, have since been expressed in *An.stephensi*, under carboxypeptidase, vitellogenin, and peritrophin promoters (Isaacs et al., 2011; Isaacs et al., 2012).

Anti-Malarial Immune Regulators

An alternative strategy to expressing transgenes, which directly interferes with *Plasmodium* infection, is the introduction of transgenes that are able to indirectly reduce *Plasmodium* transmission by modulating a response from the innate mosquito immune system. The reason that these peptides are normally insufficient to render the mosquito refractory to *Plasmodium* is that the parasite is able to invade tissues where such peptides are not expressed (Kim et al., 2004). By engineering mosquitoes to have non-native spatiotemporal expression patterns for these immune modulators it is possible to induce an effective immune response, and reduce parasite transmission.

A number of anti-parasite immune system modulators have been identified, and expressed under the CP and Vg promoters. These include: Cecropin A (Kim et al., 2004), thioester containing protein 1 (TEP1) (Volohonsky et al., 2017), the Akt signalling protein (Corby-Harris et al., 2010), phosphatase and tensin homolog (PTEN) (Hauck et al., 2013), Rel2 (Dong et al., 2011),

Anopheles *Plasmodium*-responsive Leucine-rich repeat protein 1(APL1) (Riehle et al., 2008), and fibrinogen domain containing immunolectin 9 (FBN9) (Simões et al., 2017).

An important consideration when considering antimalarial effectors is the possibility that the parasite may evolve resistance with time. Recently identified as a 'significant concern' for the success of population replacement interventions (James et al., 2020), parasite resistance to effector genes may arise in scenarios where drive systems fail to achieve fixation in a target population. Parasite resistance to effector genes is poorly understood, and is liable to remain so until larger scale releases are performed (James et al., 2020). The possibility of such an outcome highlights the importance of post-release surveillance and monitoring of parasite populations. The use of population replacement systems with multiple redundant components, or multiple independent systems may counter the onset of parasite resistance by limiting the ability for parasites to survive and propagate resistant alleles.

While extensive mathematical modelling has been conducted to evaluate key considerations required for high efficacy in suppressive drive systems (Deredec, Godfray, and Burt, 2011; Bull and Malik, 2017; Unckless, Clark, and Messer, 2017; North, Burt, and Godfray, 2019), and to investigate specific scenarios such as accidental release (Noble et al., 2018) or spatiotemporally tethered drives (Dhole, Lloyd, and Gould, 2019; Heffel and Finnigan, 2019), relatively little consideration has been given to replacement drive systems for propagating anti-malarial effectors through a target population. Nevertheless, a recent study (Beaghton et al., 2017) investigated key requirements for conventional gene drive constructs to propagate anti-malarial effectors through target populations.

This work identified a number of key parameters that had a high impact on the duration of the protection afforded by a conventional replacement drive construct when individually varying parameter values, whilst keeping all others at their baseline value. Initially, the target site was considered to be functionally neutral. Factors which negatively affected protection included: pre-existing resistance alleles within the target population, the frequency with which homologous recombination failed to occur, cost of expressing the nuclease and effector, the frequency with which the effector element was lost during homing, and the degree of protection afforded by the effector. These findings can be applied to resolving known issues in homing gene drive systems.

Beaghton et al., 2017 expanded on this initial work by modelling the targeting of an essential gene with a recessive lethal phenotype. Non-homologous recombination events were thought to either result in disruption of the target gene (with an associated fitness cost), or to form functional resistance alleles that left the target intact (and thus had no associated fitness cost). By targeting an essential gene, the duration of protection (here described by the frequency of the effector in the target population) remained at 95% for 90 generations, as opposed to 30 generations when targeting a neutral locus. The results from this study emphasise the importance of

maximising the efficacy of drive activity (i.e. homologous recombination rate, cost of expressing nuclease/effector), whilst leveraging the choice of host target site to reduce the propagation of resistance alleles.

1.5.4 Current Shortcomings in Gene Drive Designs

Gene drive research now primarily uses CRISPR genome editing tools as the basis for all homing elements. This is largely due to the versatility and ease with which target sites can be selected, and the high degrees of cleavage and homing that have been observed (Gantz et al., 2015; Hammond et al., 2016). Despite the existence of more complex architectures, the majority of gene drive constructs adhere to a 'classic' design, that has remained largely invariant since initial experimental proof-of-principle studies using HEGs (Chan et al., 2011). This paradigm makes use of pre-characterised promoter and terminator elements integrated in an arbitrary manner into (supposedly) neutral genomic loci. Regulatory elements must be characterised and optimised prior to use in a costly and time-consuming manner in order to ensure the desired spatiotemporal expression pattern. Such constructs tend to be large with a high degree of regulatory complexity in the spatiotemporal control of constituent elements.

Recent work has alluded to the potential challenges posed by more complex drive architectures. One example that made use of a Cas9 nuclease with multiplexed gRNAs (Oberhofer, Ivy, and Hay, 2018) found that whilst cleavage rates were high in heterozygous individuals carrying drive components, homing rates were low with a broad degree of variance. Moreover, incomplete homing events were observed resulting in loss of components essential to the continued success of the drive. Similarly, the only population replacement drive currently known to have been established in an Anopheline mosquito (Gantz et al., 2015) was found to have variation in homing efficiency between male G_1 founders and their G_2 progeny. This was despite accurate insertion of the drive component (confirmed by genotyping) and comparable homing rates between G_2 and G_3 progeny.

There are a range of explanations for these observations. When characterising promoters for use in gene drive (as with the majority of applications in molecular biology), regulatory elements are taken from an endogenous gene, and their behaviour characterised in order to determine regions critical to their activity (Papathanos et al., 2009; Chan et al., 2011; Chan et al., 2013a). It may be that these truncated elements do not fully replicate the activity of the loci from which they were derived, resulting in undesirable expression patterns. The interactions between the large number of elements within a drive construct may also result in abnormal behaviours as would their use in a non-native context in the genome of their host.

Furthermore, epigenetic effects have been speculated upon as a possible source of interference in drive behaviour, particularly when integrating constructs into 'neutral' loci without a known function (Hammond and Galizi, 2018; Gantz et al., 2015). However, epigenetic silencing as a consequence of methylation is unlikely to prove problematic in dipterans, given the low rates at which it has been found to occur (Provataris et al., 2018). The known issue of

maternally deposited Cas9 mRNA has also been cited as a source of aberrance in behaviour (Champer et al., 2018b; Champer et al., 2019b). Extensive modeling work has been performed to evaluate the theoretical requirements and limitations of both suppressive and replacement drive strategies. However, lesser consideration has been given to date of the resulting practical implications. The engineering of complex drive schema requires that all components (e.g. promoters, gRNA cassettes, markers, effector transgenes) must adhere to a set of regulatory requirements.

As noted above, a range of effector transgenes have now been assayed for their efficacy against laboratory strains of *Plasmodium*. However, the genetic diversity of these lab strains is unlikely to be representative of diverse field isolates, limiting the applicability of existing data sets as a measure of efficacy for use in vector control strategies. In order to accurately ascertain the efficacy of these effectors, they need to be assayed in a disease endemic environment with an afflicted human population. Existing population replacement drives in Anophline mosquitoes (Gantz et al., 2015) express both the CRISPR components required for homing and the effector from a single construct. This coupling simultaneously renders testing more hazardous as an accidental release could result in wanted propagation of the drive, and raises additional barriers when attempting to test the effector in the absence of the drive. In this paradigm, in order to conduct safe testing, effector strains must be first be generated, evaluated, and then re-engineered as part of a complete drive construct. This potential alters the behaviour of components and requiring extensive additional work.

1.5.5 Introducing Integral Gene Drives

An alternative to the 'classical' gene drive architecture, where all components are integrated at a single arbitrary locus within the genome of the host, is a split drive system, separating elements of the drive across two or more arbitrary loci. The earliest split gene drive was developed in yeast (DiCarlo et al., 2015) and was specifically developed as a means to provide additional molecular safeguarding during development. The drive consisted of two elements: an episomal Cas9-encoding plasmid, and a separate and independent element integrated into the host genome expressing a suitable gRNA. The episomal plasmid is rapidly degraded and provides a self-limiting barrier to homing activity. While this strategy may have limited applicability in vector control interventions against malaria, it provided initial proof-of-concept. Since then, a number of split drives have been produced in *Drosophila* (Champer et al., 2019b; Amo et al., 2019), and in the arbovirus vector, *Aedes aegypti* (Li et al., 2019). All such examples were found to have performance in terms of cleavage and transmission comparable with conventional drive architectures.

Split drives offer additional flexibility and applications when compared with conventional architectures. A recent study used a split drive design to investigate how maternal deposition affects conversion of a target allele by a homing event and resistance allele formation. As previously described, maternal deposition of Cas9 and a suitable gRNA can result in increased levels of resistance allele formation in the early embryo, compromising the efficacy of many

existing drive strategies (Marshall et al., 2017). When investigated in *An.stephensi*, maternal deposition was found to result in even higher rates of allele formation.

When originally positing the concept of a split gene drive, DiCarlo *et al.* provided the first demonstration of a genetic payload being propagated in a de-coupled fashion from the nuclease source. Existing work has focused on separating Cas9 nuclease and gRNA components, however it is clearly feasible to separate nuclease and effector transgene in a population replacement strategy. In doing so, it may be possible to overcome some of the shortcomings listed above, pertaining to field testing of effector transgenes: drive and effector lines can be engineered and tested safely and independently, before being combined to produce fully-driving replacement drives.

In order to overcome existing issues with the classical gene drive paradigm, we describe here a novel gene drive architecture for population replacement, termed Integral Replacement (Fig 1.13). We intend that integral gene drives (IGDs) will operate as split-drives, integrated directly into the coding sequence of endogenous genes, making use of host regulatory machinery. It is specifically designed to counter the previously discussed limitations in regards to engineering, molecular behaviour, implementation, regulation, and safety.

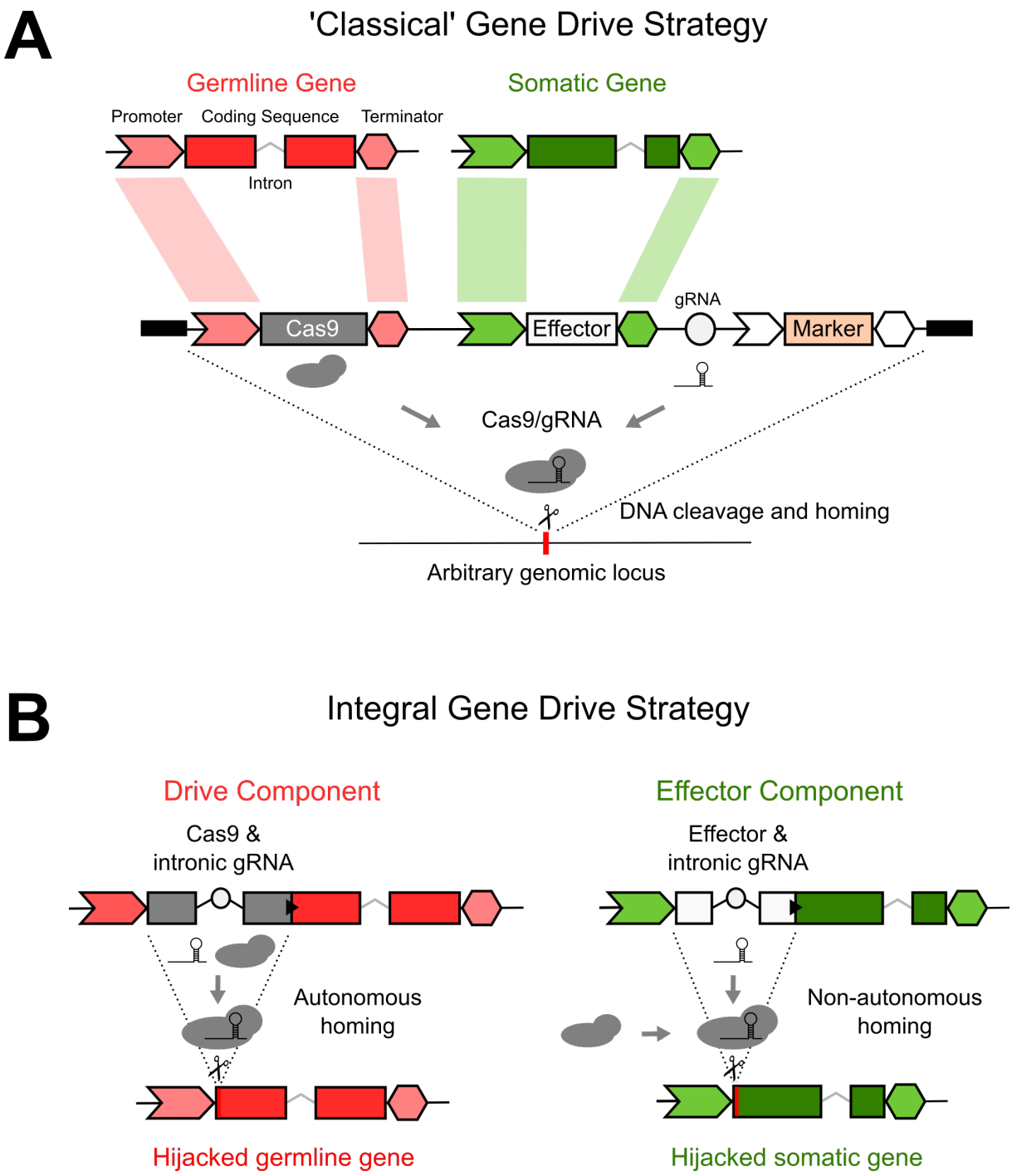


FIGURE 1.13: **Integral Gene Drive Overview**. A) A 'classical' replacement strategy requires all drive components regulated by pre-characterised regulatory elements to be integrated at a single arbitrary genomic locus. B) By contrast, IGD components operate in *trans*, integrated within appropriate germline and somatic loci using the endogenous regulatory machinery to control expression. A 2A sequence is indicated by the black arrow mediating seamless expression of the endogenous gene. Figure adapted from (Nash et al., 2019).

IGDs consist of at least two components - a drive component, encoding an endonuclease such as Cas9, and an effector component, encoding an anti-malarial effector. The IGD drive and effector components are intended to be integrated into germline and somatic loci respectively negating the need to characterise promoter elements for expression of drive components.

Consequently, the endonuclease activity of the IGD drive system is restricted to germline tissues where homing occurs, whilst the effector component's activity is restricted to the soma in order to combat *Plasmodium* infections. Both components carry an intronically expressed gRNA cassette proposed to prevent disruption of the endogenous host gene. Only the drive component carries an endonuclease cassette - the effector component can therefore only home in the presence of the drive component in the germline.

We hypothesise that this novel design paradigm may offer advantages over existing replacement drives. The minimal nature of the IGD components may facilitate low fitness costs and does away with the time-consuming need to characterise promoters for spatiotemporal control. Furthermore, by separating drive and effector components in a split-drive format, it may be possible to decrease regulatory barriers to field-testing. Effector lines may be imported separately for evaluation, and once validated, crossed to drive lines to produce fully driving systems. The modularity of the IGD strategy also permits multiple teams to conduct work in parallel and to rapidly combine different components.

1.5.6 Aims and Hypothesis

In this thesis, we worked to establish a proof-of-concept for a novel gene drive paradigm, Integral Replacement. Work was separated into three distinct chapters, where we aimed to answer key questions about this proposed strategy. Initially, we made a theoretical investigation of design considerations, and modelled IGDs against classical replacement drives, building upon existing frameworks for investigating population dynamics. We then proceeded to perform a pilot study to investigate the functionality of intronic gRNAs in *Drosophila* cell culture. Finally, we generated transgenic lines carrying IGD drive components in the model oragnism *Drosophila melanogaster* in order to assay the efficacy of the strategy *in vivo*.

Through these chapters we aimed to answer these questions:

1. How does the Integral Replacement paradigm compare to existing replacement drive architectures?

2. Is it possible to express gRNAs intronically, and what is the best method to do so?

3. Can IGD drive components be successfully integrated *in vivo*, and how do they behave?

In order to demonstrate a proof-of-principle of IGD, experimental work conducted in this thesis made use of *Drosophila melanogaster* rather than Anopheline mosquitoes. The shorter generational time, ease of rearing, abundance of genetic tools, and complete genome sequence render *D.melanogaster* an ideal analog for gene drive research.

Chapter 2

General Materials and Methods

2.1 Chemicals

Unless otherwise stated ,alkaline lysis preparation was conducted using Qiagen Spin Miniprep kits (Qiagen, CA, USA), Purelink HiPure Plasmid Midiprep kits (Invitrogen, Paisley, UK), and ZymoPure II Plasmid kits (Zymo Research, CA, USA). PCR reactions were performed using Phusion High-Fidelity PCR Master Mix with HF buffer (New England Biolabs, Hitchin, UK). Colony screening was performed using Qiagen Fast Cycling PCR kits (Qiagen, CA, USA). Gel electrophoresis was performed using ethidium bromide (Invitrogen, Paisley, UK), and GeneRuler 1kb DNA Ladder (Invitrogen, Paisley, UK). Gel extraction of PCR amplicons was performed using QIAquick Gel Extraction kits (Qiagen, CA, USA). Enzymatic digests of DNA were performed using FastDigest restriction enzymes (Invitrogen, Paisley, UK). Ligations were performed using Clontech DNA Ligation kit version 2.1 (Takara Bio Europe SAS, Saint-Germain-En-Laye, France). Gibson assembly of experimental constructs was performed using Gibson Assembly Master Mix (New England Biolabs, Hitchin, UK). Transfections of Drosophila S2 cells were performed using Effectene Transfection Reagent (Qiagen, CA, USA). Screening of candidate guide RNAs was conducted using the Guide-it Complete sgRNA Screening System (Takara Bio Europe SAS, Saint-Germain-En-Laye, France).

2.2 Bacterial Transformation

Plasmid amplification was conducted by bacterial transformation, and subsequent purification. All transformations were conducted in OneShot Top10 (Invitrogen, Paisley, UK). Cells were maintained at 80°C in a New England Scientific U570 freezer (New England Scientific Refrigeration, MA, USA). For transformation, 50 µl of cells was thawed and 1-2 µl of plasmid/ligation mix added. Cells were subsequently incubated for on ice for 30 minutes. Cells were heat-shocked at 42°C for 30 seconds in a Grant SubAqua Pro bath (Grant Instruments, Cambridge, UK), and transferred back to the ice for 5 minutes. 250 µl of SOC medium (0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 10 mM MgSO4, 20mM glucose) (New England Biolabs, Hitchin, UK) was added to each reaction, and cells were incubated at 37°C for 1 hour, at 225 rpm in a Stuart S1500 orbital shaking incubator (Stuart, Staffordshire, UK). Following incubation, between 50-100 µl cells were spread across an agar plate (1% tryptone, 0.5%

yeast extract, 1% NaCl, 1.5% agar) containing the appropriate antibiotic. Plates were incubated at 37°C in an incubation room, for 12-16 hours.

2.3 Alkaline Lysis Preparations of Plasmids

Having obtained transformed bacterial colonies, it was possible to isolate plasmid DNA, either via miniprep or midiprep.

2.3.1 Miniprep of Plasmid DNA

Single colonies were picked use a sterile pipette tip, and transferred to 5 ml tubes containing Luria-Bertani (LB) broth, supplemented with an appropriate antibiotic. Liquid cultures were transferred to a shaking incubator, and incubated for 12-16 hours at 225-250 rpm, and 37°C. Incubated liquid cultures were subsequently purified using Qiagen Spin Miniprep kits (Qiagen, CA, USA) according to the manufacturer's instructions, and eluted in 30 µl of pre-warmed Tris-EDTA buffer.

2.3.2 Midiprep of Plasmid DNA

Single colonies were picked use a sterile pipette tip, and transferred to 5 ml tubes containing LB broth, supplemented with an appropriate antibiotic to produce starter cultures. Said cultures were transferred to a shaking incubator, and incubated for 6-8 hours at 225-250 rpm, and 37°C. Incubated liquid cultures were subsequently transferred to 100 ml LB broth, supplemented with the same antibiotic. Cultures were incubated for 12-16 hours at 225-250 rpm, and 37°C. Incubated cultures were subsequently processed using either the Purelink HiPure Plasmid Midiprep (Invitrogen, Paisley, UK), or ZymoPure II Plasmid kit (Zymo Research, CA, USA), according to the manufacturer's instructions. Purified DNA was eluted in 100 µl of pre-warmed Tris-EDTA buffer or, when being used for microinjections, in RNase-free water.

2.4 Quantifying Plasmid and gDNA Concentrations

Plasmid DNA, and DNA fragments were quantified using 1 µl, against a RNase-free water blank, using a DS-11 NanoDrop Spectrophotometer (DeNovix, DE, USA). Genomic DNA samples were quantified for cloning, and amplicon sequencing using a Qubit Fluoremeter (Invitrogen, Paisley, UK), according to the manufacturer's instructions.

2.5 Restriction Digestions

Unless otherwise specified, all restriction-ligation reactions were performed using ThermoFisher FastDigest enzymes (Invitrogen, Paisley, UK). Approximately 1 µg of DNA was used per reaction. Reactions were conducted as follows: 1 µl enzyme, 1 µg DNA, 10 µl FastDigest Green buffer, 20-X µl RNAse-free water. Incubation was conducted at 37°C, for 1 hour in a water bath. Reactions were scaled appropriately where required.

Component	Volume (µl)	Final Concentration
T4 DNA ligase buffer (10x)	2	1x
Vector DNA	X	N/A
Insert DNA	Y	N/A
RNase-free water	10-X-Y	N/A
T4 DNA ligase	1	40,000 U/ml

TABLE 2.1: Ligation Reaction Components

2.6 Gel Purification of DNA

Where necessary, digested DNA products were visualised, and extracted using gel electrophoresis. Gels were prepared using agarose (Sigma-Aldrich, Dorset, UK) to 0.7-2.0% (w/v). Where samples lacked loading buffer, 5x loading buffer was added (Bioline, TN, USA). Samples were run at 100 V for 45-60 minutes, with a 1 kb hyperladder marker (Bioline, TN, USA). Gels were visualised using an Azure c600 gel imaging system (Azure Biosystems, CA, USA), and appropriate bands excised with a sterile scalpel. DNA was purified from bands using a QIAquick Gel Extraction kits (Qiagen, CA, USA), according to the manufacturer's instructions. Purified DNA was eluted in 10-30 µl of pre-warmed Tris-EDTA.

2.7 Ligation Reactions

Ligation reactions were conducted using digested DNA fragments, following gel purification. Reactions were set up as described (Table.2.1), and incubated at room temperature for at least 2 hours. Reactions were subsequently heat inactivated at 65°C for 10 minutes, and transformed as described above. The molar ratio of insert to vector was 3:1, ensuring at least 50 ng of vector backbone was used.

2.8 Colony PCR

Colony PCR screening was performed using a Qiagen Fast Cycling PCR kit (Qiagen, CA, USA). Single colonies were picked from plates with a sterile pipette, and resuspended in 30 µl of RNase-free water. Colony PCR reactions were set up as described (Table.2.2).

Component	Volume (µl)	Final Concentration
Forward primer	1	0.5 µM
Reverse primer	1	0.5 µM
Fast Cycle 2x mastermix	10	1x
RNase-free water	6	N/A
Resuspended colony	2	N/A

TABLE 2.2: Colony PCR Components

The general conditions were as follows; 95°C for 5 minutes, folowed by 35 cycles of 96°C for 5 seconds, 55°C for 5 seconds, 68°C for 3 seconds/100bp, and a final 72°C extension step for 1

minute, before a 4°C hold. Subsequently, PCR reactions were loaded directly onto 1% agarose gels, stained with ethidium bromide, and with 1kb hyperladder rulers at the end of each row. Gel electrophoresis was conducted for 60 minutes at 100 V, and the gels visualised as previously described for confirmation of the presence of a positive band. Bands were subsequently purified as described above, and sent for Sanger sequencing.

2.9 Sanger Sequencing of Constructs

When validating sequence veracity of constructs, plasmids purified by miniprep as described above were diluted to 50-100 ng/µl in a 20 µl reaction. The remainder of the reaction was made up by 1 µl of primer to a final concentration of 0.5 µM, and RNase-free water. Sequencing was performed by Eurofins (Eurofins Genomics, Ebersberg, Germany), or Genewiz (Genewiz, Leipzig, Germany).

2.10 Statistics

GraphPad Prism V8 (Graphpad Software, CA, USA), was used to plot and analyse all experimental data in this thesis. Analysis of modeling outcomes was performed using Wolfram Mathematica, as described in Chapter 3. Normality testing was conducted on data sets using Graphpad's normality function, using the D'Agostino K-squared test. Developmental and oviposition analysis was performed using the ordinary one-way ANOVA test, with a shared control; post-hoc analysis was performed using Dunnett's test. Data in the homing assay was found to be non-normal, and as such it was analysed using the Kruskal-Wallis test with a shared control; post-hoc analysis was performed using Dunn's test. In all cases, $p < 0.05$ was considered statistically significant.

Chapter 3

Results - Modelling Integral Gene Drive for Population Replacement

3.1 Acknowledgements

The contents of this chapter are summarised in a recent publication (Nash et al., 2019):
Initial conceptualisation was conducted by Andrea K. Beaghton, Nikolai Windbichler, Philipos Papathanos, and Astrid Hoermann. Methodological approaches and programming were conducted by myself, Giulia Mignini-Urdenata, and Andrea K. Beaghton. Formal analysis was conducted by myself, Giulia Mignini-Urdenata, and Andrea K. Beaghton. Writing was conducted by myself, Giulia Mignini-Urdenata, Andrea K. Beaghton, Philipos Papathanos, and Nikolai Windbichler. Supervision was provided by Nikolai Windbichler and George K. Christophides

3.2 Chapter Aims

In this chapter we conducted work to establish a theoretical basis for the Integral Replacement paradigm. We aimed to gain initial insights to help answer the following:

1. What traits should Integral gene drive components have?
2. How does Integral Replacement compare to a conventional replacement strategy?
3. What parameters are key for the success of Integral Replacement?
4. How does pre-existing resistance affect Integral Replacement?
5. Can we safely deploy an Integral gene drive to the field?

3.3 Introduction

Having proposed Integral Replacement as a novel design paradigm for the construction of replacement drive systems, we carried out our initial work which focused on theoretical development of the system and construction of a suitable mathematical model. Existing mathematical modelling of interventions against infectious diseases such as malaria is extensive and is particularly useful in forecasting the potential efficacy of a given control method. In the context

of gene drive systems population modeling is particularly useful as it can aid the elucidation of critical biological and demographic factors. In so doing, it is possible to understand their impact upon drive performance and reduction in disease transmission (Beaghton et al., 2017).

As previously stated, relatively little work has been done to investigate the requirements necessary for successful population replacement strategies. Initial evaluations have been made of replacement methods using underdominance (Champer et al., 2019a) and toxin-antidote systems (Champer et al., 2019c) allowing for some evaluation of critical parameters, such as initial release frequencies and the importance of minimizing drive-associated fitness costs. However these approaches differ significantly from Integral Replacement and conventional homing replacement drives, and their findings cannot be taken for granted.

In order to understand the requirements for Integral Replacement to act as a successful population replacement strategy, we began by defining the traits of drive and effector components in this chapter. We then proceeded to construct a model derived from an existing framework for conventional replacement drive constructs (Beaghton et al., 2017). This allowed us to evaluate Integral Replacement against an existing design paradigm, permitting evaluation of key parameters, as well as highlighting areas of improvement. We then proceeded to simulate scenarios in which two component IGD drives were released into field populations with varying levels of pre-existing resistance. We subsequently expanded the definition of our hypothetical IGD system to include two independently homing effector components in order to examine the concepts of staggered release and multi-loci resistance. Finally, we modelled a number of staged release scenarios so as to evaluate a potential deployment strategy for an IGD-based intervention.

3.4 Materials and Methods

The code for all models used in the published portion of this chapter (Nash et al., 2019) are available at https://github.com/genome-traffic/igd. The classical model used for comparison here was originally proposed by Beaghton et al., 2017 (Model I) and describes a conventional, single locus homing drive. In this classical model, the gene drive is considered to target a genomically neutral locus with no associated fitness costs. The model considers five possible alleles which may arise at the target site due to the molecular activity of the drive construct: a wild-type allele (w); the complete drive construct, with an active nuclease and effector component (c); a drive construct with an active nuclease but inactive effector component (n); a drive construct with an inactive nuclease but an active effector component (e); and a functional resistance allele (r). Resistance alleles are considered to be inherent within the target population and/or to result from host cellular repair machinery resolving DSBs generated by the activity of the gene drive. To this end, resistance alleles are considered to arise when NHEJ or MMEJ occurs as well as incomplete HDR. The parameter d describes the transmission rate of the drive. Parameter u describes the rate of resistance allele formation from NHEJ/MMEJ. Parameters l_n, l_e, and l_{ne} describe the effects of incomplete HDR resulting in loss of nuclease, effector, or both.

Given these definitions, the allele contributions that result from germline cleavage and homing from the complete construct (c) to the wild-type allele (w) are w:c:n:e:r according to ratio: (1-d)(1-u): d(1-l_n-l_e-l_{ne}: d(l_e: d(l_n) : (1-d)u + $l_{ne}d$. Allele contributions from germline cleavage and homing from the nuclease (n) to the wild-type allele (w) are w:n:r according to ratio: (1-d)(1-u): d(1-l_n): (1-d)u + $l_n d$.

The IGD model described here considers a simplified scenario with drive and effector components located on either two or three independent loci (i), where i = 1, 2, 3. Here, locus 1 represents the site for the drive component, and loci 2 and 3 are sites for effector components. At each of the loci, four possible alleles may arise: a wild-type allele (w); a transgene component, be that a drive or effector component (t_i), where i corresponds to the locus, and thus the presence of either the drive component or an effector component; a functional resistance allele (r_i); and a non-functional resistance allele (m_i). In the context of IGD, functional resistance alleles arise from incomplete HDR or in-frame EJ mutations that do not disrupt the functionality of the endogenous gene. They are considered to have no associated fitness cost, and are subsequently equivalent to the wild-type allele. Non-functional resistance alleles are the result of any mutation that causes a frameshift of the endogenous gene, disrupting its functionality. Here, it has been assumed that the target gene at each locus is essential, such that heterozygous m_i mutations confer a fitness cost, and homozygous m_i mutations result in lethality. As per the classical model, resistance can pre-exist in the target population, or arise via EJ mutations or incomplete HDR. Spontaneous mutations are not considered to occur, as we posit that rates are likely to be insignificant when compared to those caused by EJ events and incomplete HDR.

In the case of the IGD model, an individual is considered to have a complete drive construct (c) if there is an active nuclease allele present at locus 1, and at least one active effector component at locus 2 (and/or at locus 3, depending on the nature of the strategy). The nuclease allele is necessary for any effector allele to achieve super-Mendelian rates of transmission in the population. The field population being targeted by the drive in this model is considered to homogeneously be made up of wild-type alleles. The field population is not considered to have any pre-existing resistance alleles, though it is possible to alter the parameter values to introduce this if required for a given scenario. In both models, releases are simulated as the injection of homozygous drive individuals into the field population. Given that the populations simulated by both models are infinite, releases are considered to be relative proportions of the total population.

Homing is restricted to genotypes that are transgenic heterozygous for the nuclease allele and have a corresponding wild-type allele at locus 1 (w_1/n_1). Cleavage and homing at loci 2 and/or 3 is only possible in the presence of a nuclease allele at locus 1, and with the precondition that there is a suitable wild-type allele at locus 2 and/or 3 (w_2/e_2, and w_3/e_3 respectively). The probability of transmission of the nuclease allele at locus 1 is given by d_1. Transmission of the effector component at loci 2 and 3 are given by d_{2a}, and d_{3a} respectively if the nuclease

allele is heterozygous at locus 1. In the event that the nuclease allele is homozygous at locus 1, it is given by d_{2b}, and d_{3b} respectively. The model considers that resistance alleles, both functional and non-functional, arise as a consequence of homing. The ratio of functional:non-functional mutations is conservatively estimated as 1:2. It is assumed that the primary source of disruption resulting from mutations is frameshift of the endogenous locus. As a result, in 1/3 of cases, the frameshift is considered to result in an intact endogenous gene containing a functional resistance allele (r_i). In 2/3 of cases, frameshift is thought to disrupt the endogenous gene, resulting in loss of gene function (m_i).

Functional resistance alleles (r_i) are caused by incomplete HDR and EJ events at all three loci. l_1, l_2, and l_3 describe the probability of incomplete HDR occurring at loci 1,2, and 3. Similarly, EJ events result in resistance alleles (r_i) with probabilities u_1, u_2, and u_3. Loss of function resistance alleles occur with probabilities L_1, L_2, and L_3, due to incomplete HDR, and by EJ events with probabilities U_1, U_2, and U_3. Individuals that are heterozygous at locus 1 for the nuclease component are able to cleave, home, and produce resistance alleles. Their allelic contribution (w_1:n_1:r_1:m_1) is in the following ratio: (1-d_1)(1-u_1-U_1):d_1(1-l_1-L_1):(1-d_1)u_1+l_1d_1:(1-d_1)U_1+L_1d_1. Similarly, in the event that an individual is heterozygous or homozygous for the nuclease component at locus 1, and heterozygous for the effector component at locus 2, allelic contributions (w_2:e_2:r_2:m_2) are given in the ratio: (1-d_k)(1-u_2-U_2):d_{2k}(1-l_2-L_2):(1-d_{2k})u_2+l_2d_1:(1-d_{2k})U_2+L_2d_{2k}. Here, k is equal to either a or b, and describes the copy number of the nuclease component at locus 1. In the absence of a nuclease component at locus 1, germ-line homing does not occur, and inheritance of any alleles at locus 2 (or 3) is Mendelian. The above allelic contributions can be mirrored for locus 3 should an additional effector component be implemented.

The formulae describing the number of genotypes in either model can be given by $k - (\frac{k(k+1)}{2})^2$, where k = the number of alleles at each locus. The number of resultant gametes is given by the function k^l, where l is the number of loci. In the IGD model, given that there are four alleles possible at two loci, there are 16 gamete types and 100 diploid genotypes. For three independent loci, there are 64 gamete types and 1000 possible genotypes. The fitness of any given genotype is bootstrapped against the wild-type homozygote (w_1/w_1; w_2/w_2; w_3/w_3). This wild-type genotype is considered to have a perfect fitness value of 1. The fitness of each genotype is constructed using 10 parameters for a two-locus model, and 14 parameters for a three-locus model. An in-frame disruption of the endogenous host gene can be caused by insertion of the transgenic component or the presence of a functional resistance allele (r_i). When present at loci 1,2, and 3, the fitness cost is given by s_{d1}, s_{d2}, and s_{d3} respectively. The cost of disruption as a result of inactivation of the endogenous host gene by a loss of function mutation is given by (m_i). The fitness cost is assumed to be equal at all three loci and is given by s_m. The cost associated with expression of the nuclease component from locus 1 as part of the drive component is given by s_n, and the corresponding cost of expressing the effector component from loci 2 and 3, by s_{e2} and s_{e3}. The dominance coefficients for each of these fitness parameters is given by h_{d1}, h_{d2}, h_{d3}, h_m, h_n, h_{e2}, and h_{e3}. The fitness value assigned to each genotype is

scored from 0 (no cost; equivalent to the wild-type homozygote) to 1 (lethal). Correspondingly, dominance coefficients range from 0 (entirely recessive) to 1 (entirely dominant). In order to calculate the fitness cost for each genotype, the product of all costs is calculated. These costs are generated by the aforementionned site disruption (in-frame or out-of-frame) and transgene expression from each locus (including copy number of transgenes).

In both the classical and IGD model, allele frequencies and genotype abundances are modeled using deterministic discrete-generation recursion equations. All vectors are assumed to exist only as adults - other life stages are not considered. The field population is assumed to have no sex bias, and males and females are assumed to have no sex-specific variation in genetic and fitness parameters. All mating events are considered to be random and to be successful. The field population is considered to be of sufficient size such that stochastic effects are so small as to be negligible. The system of equations is solved numerically using Wolfram Mathematica (Wolfram Research, Inc., Mathematica, Version 11.3, Champaign, IL (2018)).

The efficacy of the IGD model is determined based on the frequency of each genotype within the modelled population and the resulting reduction in vector competence. Here vector competence is defined as the ability of an individual to transmit the *Plasmodium* parasite. In the two-locus IGD model, the reduction in vector competence is given by $h_{rc1}r_c$ if the effector is heterozygous at locus 2, and r_c if the individual is homozygous for the effector. When using multiple effector loci in the three-locus implementation of the model, reduction in vector competence is predicated on the total number of effector alleles at both locus 2, and locus 3. The reduction is given by $h_{rci}r_c$, where i = 1,2, or 3 effector alleles total over both effector loci, and r_c for 4 effector alleles total (i.e. homozygous at both loci). The values of r_c range between 0 (no protection), to 1 (total protection). As before, the dominance coefficient describing protection (h_{rci}, where i = 1,2, or 3) range from 0 (totally recessive) to 1 (totally dominant). Protection offered by reduction in vectorial capacity at a given discrete time point t can thus be given by 1-$V_{C[t]}$, where $V_{C[t]}$ is the vectorial capacity. $V_{C[t]}$ is determined by calculating the sum of genotype frequencies, multiplied by the vector competence for each group.

In the case of the IGD model, the two-locus implementation calculates vector competence, thus:

$$\begin{aligned} V_C[t] = \sum(\text{Frequency of genotypes lacking an effector component}) \\ +(1-h_{rc1}r_c)\sum(\text{Frequency of genotypes heterozygous for an effector component}) \\ +(1-h_{rc1}r_c)\sum(\text{Frequency of genotypes homozygous for an effector component}) \end{aligned}$$

In the case of the three-locus implementation, vector competence can be calculated thus:

$$
\begin{aligned}
V_C[t] = &\sum(\text{Frequency of genotypes lacking an effector component}) \\
&+(1-h_{rc1}r_c)\sum(\text{Frequency of genotypes with one effector component}) \\
&+(1-h_{rc2}r_c)\sum(\text{Frequency of genotypes with two effector components}) \\
&+(1-h_{rc3}r_c)\sum(\text{Frequency of genotypes with three effector components}) \\
&+(1-h_{rc1}r_c)\sum(\text{Frequency of genotypes homozygous for an effector component})
\end{aligned}
$$

3.5 Results

3.5.1 Designing IGD Drive Components

Existing gene drive constructs for population replacement are designed to exist at a functionally neutral locus within the host genome. This is done in order to prevent drive elements from incurring a fitness cost by disrupting important molecular interactions at the genetic level. By contrast, the IGD strategy looks to harness the potential for disruption as an advantage: as with existing drive strategies, for homing to occur, Cas9 needs to be expressed in the germline of an organism (whether expression is sex-specific or not). It is imperative that Cas9 expression has a minimal effect on expression and activity of the endogenous gene. To this end a suitable strategy must be employed to ensure adequate expression and activity of both. A well characterised methodology involves introduction of an internal ribosome entry site (IRES). Originally discovered in picornoviruses as a means to mediate cap-independent translation (Pelletier and Sonenberg, 1988; Jang et al., 1988), IRES sequences have since been developed as a tool for expressing polycistronic cassettes. An unfortunate limitation of the IRES is that within the polycistron, the expression levels of each consecutive element are diminished.

An alternative solution is a group of oligopeptides, derived from a variety of viral hosts, that facilitates 'ribosomal skipping'. These 2A peptides are named after the region of the viral genome in which they are found, in combination with a suffix designating the specific virus (Liu et al., 2017). Peptides are typically 18-22 amino acids in length and share a common conserved GDVEXNPGP sequence. The first 2A sequence to be characterised was found in foot-and-mouth disease virus (F2A) (Donnelly et al., 2001). Other commonly used 2A peptides include E2A (equine rhinitis A virus), P2A (porcine teschovirus-1 2A), as well as T2A (thosea asigna virus 2A) (Szymczak and Vignali, 2005). These peptides were previously believed to be self-cleaving, facilitating separation of a single peptide sequence into two separate peptides during translation. Since then, research has instead proposed a ribosomal skipping mechanism (Donnelly et al., 2001), though the exact mechanism remains unclear. When skipping, there are three potential results: 1) The skipping of glycyl-prolyl peptide bond by the ribosome during translation, resulting in the separate translation of two distinct peptides; 2) The 2A sequence triggers the ribosome to fall away, translating only the upstream peptide; 3) Skipping fails to occur, and instead a single fused peptide is produced (Liu et al., 2017).

The use of 2A peptides has been demonstrated successfully as a means to express fluorescent reporter genes endogenously from within the coding sequence of a host gene in tissue culture (Rojas-Fernandez et al., 2015). A key consideration is the ordering of the arrangement with regards to the Cas9 element and endogenous gene: it is possible to insert the transgene as both an N-terminal and C-terminal sequence within a hypothetical polycistron. In the case of an N-terminal insertion, it is of absolute importance that the expression of Cas9 does not confer a reduction in fitness cost in the host due to diminished expression of the downstream endogenous gene. 2A peptides have been demonstrated to confer high rates of downstream expression when compared to other strategies (Souza-Moreira et al., 2018). However, should the Cas9 coding sequence be disrupted by mutation, it is liable to produce a frameshift mutation that will result in the production of an aberrant fusion protein. To this end, it is likely that in this arrangement, mutations arising at the Cas9 locus will be poorly tolerated on a population level, generating a negative selection pressure against individuals harbouring such mutations. Point mutations which do not result in frameshift may still be tolerated however, resulting in knockout of the Cas9 element whilst leaving the endogenous gene intact.

To date, a range of promoter elements have been characterised to regulate gene drive expression in germline tissues. In *Drosophila* and Anopheline mosquitoes, these promoters include *vasa* (Papathanos et al., 2009), *nanos* (Calvo et al., 2005), and *zero population growth* (*zpg*) (Hammond et al., 2018; Kyrou et al., 2018), as well as sex-specific promoters such as *beta2-tubulin* (Catteruccia, Benton, and Crisanti, 2005) and *Rcd-1r* (Chan et al., 2013a). The endogenous genes, from which these promoters are derived, make for obvious choices when selecting a suitable host gene into which the drive component can be integrated. Any system using CRISPR-Cas9 systems require a gRNA to mediate targeting cleavage. Were the gRNA to be integrated into the endogenous gene alongside the Cas9 element, it would likely result in disruption of the endogenous reading frame. A potential solution, which is examined in more detail in Chapter 4 of this thesis, is the use of intronically encoded gRNA sequences. Cassettes expressing a gRNA of choice can be inserted within an intron, which in turn is inserted into the Cas9 coding sequence. Previous work has briefly proffered evidence that it is possible to express gRNAs intronically without a promoter (Kiani et al., 2014), in which case, spatiotemporal regulation would be dictated by the expression pattern of the endogenous gene and intron.

An alternative strategy is to provide an intronic gRNA with an RNA polymerase III promoter, facilitating ubiquitous and constitutive expression. In this way, a reservoir of gRNA will be readily available once the Cas9 element is expressed in the germline. As with a conventional gene drive strategy, the Cas9 and gRNA components will facilitate cleavage and homing - though in this case, homing will occur at non-converted copies of the endogenous gene rather than a neutral genomic locus. In this way, the drive component will represent an autonomously homing allele, designed to propagate through a target population with minimal fitness cost with the sole function of providing an invasive source of Cas9.

3.5.2 Designing IGD Effector Components

As previously discussed, expression of anti-malarial effector genes can inhibit or kill invading parasites taken up during blood-feeding, reducing malarial transmission to human hosts. A range of effectors have been evaluated for use against *Plasmodium* with varying degrees of effectiveness. In describing the effector component here (and in the following mathematical model), we consider the effector gene to express an exogenous polypeptide in the target tissue, reducing or entirely preventing parasite development in IGD mosquitoes.

Analagous to the drive component, IGD effector components are designed to be incorporated directly into endogenous host genes. The choice of host gene should correlate with the desired expression pattern in a process we term 'hijacking', exploiting the naturally occurring regulatory region of that locus. As discussed, a number of promoters have been characterised for tissue-specific expression of effector genes in Anopheline mosquitoes, including carboxypeptidase and vitellogenin. A number of other genes have been identified by genome-wide expression analyses that would be suitable for tissue-specific expression of effector genes (Giraldo-Calderón et al., 2015). A key advantage of the IGD strategy is that it dispenses with the need to experimentally validate the temporal and spatial expression patterns of synthetic promoter elements derived from genes of interest. Instead, components may be directly integrated into a host gene with a known expression pattern. A limitation of this approach is that there will likely be a restricted number of sites amenable to genome editing within the endogenous locus. For example, when using CRISPR/Cas9 to facilitate direct insertion of components, and to direct subsequent homing events, there will be a limited number of PAM sites adjacent to the target site. Moreover, it is likely that only some genes will tolerate insertion of IGD components, as they may cause undesirable alterations to expression levels.

As with the drive component, both N-terminal and C-terminal insertions of the effector component into a suitable somatic locus are possible. However, the former benefits from the same protections enjoyed by the drive component, whereby incomplete homing events resulting in end-joining mutations are liable to cause frameshift. This in turn causes knockout of the downstream endogenous gene and incurs a fitness cost to the individual. As with the drive component, each IGD effector transgene contains an intronically encoded gRNA cassette driven by a U6 RNAP III promoter. The effector component does not encode an endonuclease source, and is subsequently only able to home (in the germline) in the presence of Cas9 provided by the drive component. The effector component is thus a non-autonomous homing allele, which is designed to self-propagate through a population, increasing the frequency of the coupled effector trait. Here, non-autonomous gene drive refers to an allele which expresses a gRNA, permitting it to home in the presence of a suitable endonuclease provided by another genetic component. An alternative mechanism for hijacking involves the use of intein-facilitated protein splicing (Fig.3.1). Inteins have previously been identified which make use of HEGs to insert themselves into the coding sequence of host genes without causing disruption (Chevalier, 2001). Indeed, recent work aimed at providing Cas9 for gene therapies has demonstrated

it is possible to package a split Cas9 in a recombinant adenovirus, and using paired intein sites, reconstitute the entire Cas9 protein (Truong et al., 2015).

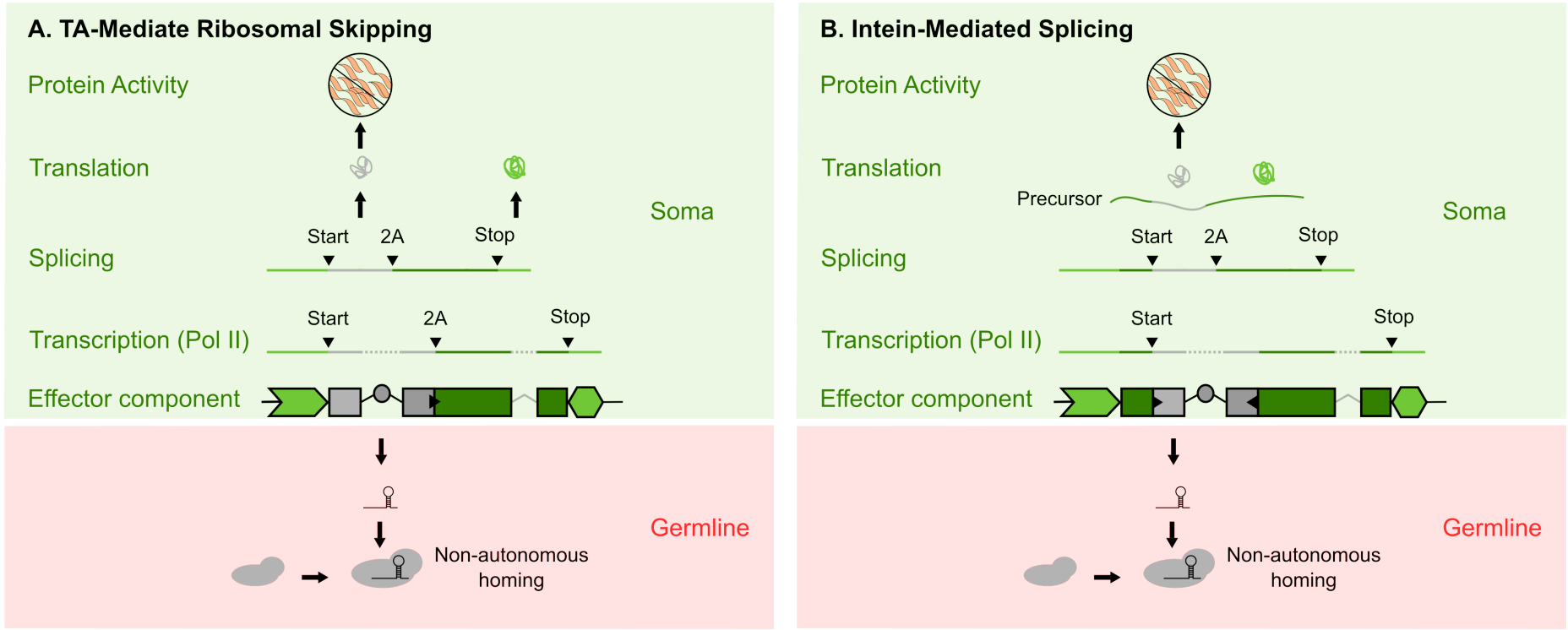


FIGURE 3.1: **Intein and 2A Sequence Mediated Splicing**. There are two possible strategies for mediating seamless integration of an IGD component into the coding sequence of an endogenous gene: A) The use of a 2A sequence (black triangle) that facilitates ribosomal skipping during translation, resulting in expression of both the transgene protein and the endogenous protein, and B) Intein sequences (inverted black triangles) that mediate excision of an intein from a precursor protein, ligating the N- and C- terminal exteins to form the spliced protein. Figure adapted from Nash et al., 2019.

3.5.3 Practical Development of IGD Components

The IGD strategy has the potential to offer a number of advantages in the way in which traits can be developed. Due to the modular nature of drive and effector components, it may be possible to conduct iterative construction and testing in an independent manner, even using different research teams. The drive component can be optimised in a laboratory setting for a range of parameters including: homing efficiency, Cas9 expression pattern, and the reduction of 'unsuccessful' cleavage events, resulting in repair via the non-homologous end joining, or microhomology-mediated end joining pathways. Furthermore, it would be possible to investigate means by which the costs of expressing the Cas9 nuclease from the drive component and hijacking the drive component locus can be reduced, in turn reducing the overall fitness cost of the IGD strategy in any modified individual carry such a component. At the same time, the effector component can be optimised for its ability to inhibit malarial transmission. This will include work to ensure an optimal spatiotemporal expression pattern - for instance, following ingestion of a blood-meal in the midgut lining (Moreira et al., 2000).

Other parameters which can be optimised include the costs of expressing the effector transgene, and the disruption mediated as a result of the hijacking of a suitable somatic host locus. Additionally, it will be possible to quantify the rate of non-autonomous homing for any combination of effector component(s), either in the presence of a non-driving source of Cas9, or in the presence of an IGD drive component. Having generated and tested individual components it would be possible to combine drive and effector components, by crossing transgenic strains under laboratory conditions. Fully-driving strains could then be assayed for their efficacy in smaller scale cage trials prior to large cage testing and subsequent field testing. The ability to isolate autonomous drive components from non-autonomous effector components may offer an additional level of safety over existing replacement drives, which are constructed in a such a manner that they are able to immediately propagate effector transgenes (Gantz et al., 2015). However, to maintain the integrity of this safety measure, it will be necessary to isolate laboratory strains - this is particularly important considering field-ready variants of components may not carry a reporter gene, which is described in further detail later in this chapter.

3.5.4 Modeling IGD Dynamics

In order to develop an understanding of the interactions between IGD components in a hypothetical population, a deterministic, discrete-time model was constructed. In order to compare the allelic dynamics and protection levels of IGD to existing population replacement strategies, an existing model (Beaghton et al., 2017) describing population dynamics of a 'classical' gene drive was modified from continuous time, to discrete time. While generations in field mosquito populations are non-discrete and are better reflected treating time as a continuous variable, this generates a significant computational load. In the context of the IGD model, which can simulate up to 1000 genotypes (and 2000 genotypes in a sex-specific implementation), this computational load is impractical and restrictive. Moreover, comparisons of the discrete-time implementation and continuous-time implementations of the classical model, (Beaghton et al., 2017);

Beaghton, Beaghton, and Burt, 2017), demonstrated no significant difference in allele frequency dynamics, nor protection over the parameter range that has been investigated here.

The model functions in a series of distinct steps

1. Initially, a matrix is generated to be populated by parameter values defining the framework of the model. The matrix defines the contribution of each genotype to each possible gamete
2. User-defined parameters are given and used to construct the matrix, generating gamete fractions
3. The model then proceeds to calculate the fitness of each genotype
4. The initial frequency of each genotype at $t = 0$ is defined
5. A system of recursion equations is then solved to calculate genotype frequencies at each timepoint. The calculated fitness values, gamete fractions, and total frequency of adults at each time point (this last value should be equal to 1) are used
6. These values are subsequently used to generate plots to visualise different outcomes, such as nuclease frequency, protection, etc...

The baseline parameter values for all results in this chapter are given below:

TABLE 3.1: I - Only present in three-locus IGD model; HDR - homology-directed repair; NHEJ - non-homologous end joining. Baseline parameter values used for all modeling in this chapter were obtained from existing work on mosquitoes (Hammond et al., 2016). These parameters describe: underlying molecular biology, fitness, vector competence, and initial genotype frequencies. Where parameters are shared by both the IGD and classical drive model, the same baseline values have been used to facilitate direct comparison.

Symbol	**Parameter$^+$**	**Baseline Value**
cc_0	Initial release frequency of homozygous drive and effector construct	10-4 x (1/3)
dd_0	Initial release frequency of homozygous drive construct	0
ee_0	Initial release frequency of homozygous effector construct	0
$rr1_0$	Initial frequency of resistance (locus 1)	0
$rr2_0$	Initial frequency of resistance (locus 2)	0
$rr3_0$	Initial frequency of resistance (locus 3)	0
d_1	Transmission rate of drive (locus 1)	0.99
d_{2a}	Transmission rate of drive (locus 2) if drive component present in heterozygosity	0.99
d_{2b}	Transmission rate of drive (locus 2) if drive component present in homozygosity	0.99
d_{3a}	Transmission rate of drive (locus 3) if drive component present in heterozygosity (I)	0.99
d_{3b}	Transmission rate of drive (locus 3) if drive component present in homozygosity (I)	0.99
l_1	Resistance arising from HR at locus 1	10-4 x (1/3)
l_2	Resistance arising from HR at locus 2	10-4 x (1/3)
l_3	Resistance arising from HR at locus 3 (I)	10-4 x (1/3)

Continued on next page

Table 3.1 – continued from previous page

Symbol	Parameter[+]	Baseline Value
u_1	Resistance arising from NHEJ at locus 1	0.5 x (1/3)
u_2	Resistance arising from NHEJ at locus 2	0.5 x (1/3)
u_3	Resistance arising from NHEJ at locus 3 (I)	0.5 x (1/3)
L_1	Loss of gene function arising from HR at locus 1	10-4 x (2/3)
L_2	Loss of gene function arising from HR at locus 2	10-4 x (2/3)
L_3	Loss of gene function arising from HR at locus 3 (I)	10-4 x (2/3)
U_1	Loss of gene function arising from NHEJ at locus 1	0.5 x (2/3)
U_2	Loss of gene function arising from NHEJ at locus 2	0.5 x (2/3)
U_3	Loss of gene function arising from NHEJ at locus 3 (I)	0.5 x (2/3)
s_m	Cost of loss of gene function	1
s_{d1}	Cost of hijacking target locus 1 (Drive)	0
s_{d2}	Cost of hijacking target locus 2 (Effector)	0
s_{d3}	Cost of hijacking target locus 3 (Effector) (I)	0
s_n	Cost of expressing nuclease at locus 1	0.05
s_{e2}	Cost of expressing effector at locus 2	0.1
s_{e3}	Cost of expressing effector at locus 3 (I)	0.1
h_m	Dominance coef. for loss of gene function	0.2
h_{d1}	Dominance coef. for hijacking at locus 1	0.5
h_{d2}	Dominance coef. for hijacking at locus 2	0.5
h_{d3}	Dominance coef. for hijacking at locus 3 (I)	0.5
h_n	Dominance coef. for expressing nuclease	0.5
h_{e2}	Dominance coef. for expressing effector (locus 2)	0.5
h_{e3}	Dominance coef. for expressing effector (locus 3) (I)	0.5
h_{rc1}	Dominance coef. for refractoriness (1 effector allele)	1
h_{rc2}	Dominance coef. for refractoriness (2 effector alleles) (I)	1
h_{rc3}	Dominance coef. for refractoriness (3 effector alleles) (I)	1
r_c	Homozygous degree of refractoriness	1

Initially, to make comparisons between the classical drive model and the IGD model, a two-locus model was constructed. It simulated one drive component homing into a germline locus (locus 1), permitting it to make use of endogenous regulatory elements. In parallel, an accompanying effector component hijacked a somatic locus (locus 2). This model was subsequently extended to a three-locus implementation, permitting investigation of two effector components located at separate and independent loci (loci 2 and 3). It was assumed in this scenario that both effectors had the same parameter values. In the context of protection within the model, a single effector component was considered to render a given individual refractory to *Plasmodium* infection, and thus refractory against malaria transmission. In order to evaluate and compare the efficacy of the IGD strategy to the classical replacement drive, the longevity of protection was compared. The endurance of protection is affected by a range of parameters, such as homing of components, frequency of resistance alleles, fitness costs incurred by expression of components, and efficacy of the effector components. In this work protection is given as the reduction in vectorial capacity, calculated from the sum of the genotype frequencies carrying at least one

effector component (at either effector locus), multiplied by the reductive effect they have on vector competence. Throughout this work, a baseline set of parameters (Table 3.1) was used in accordance with existing publications relating to gene drive in mosquitoes (Hammond et al., 2016), and to ensure effective comparison with the classical gene drive model (Table 3.2).

TABLE 3.2: Baseline parameter values for the conventional gene drive model, derived from Beaghton et al., 2017 and Beaghton, Beaghton, and Burt, 2017.

Symbol	Parameter$^+$	Baseline Value
cc_0	Initial release frequency of homozygous construct	10^{-4}
d	Transmission rate of drive	0.99
u	Probability of resistance (functional)	0.5
l_n	Probability of nuclease loss during homing	10^{-4}
l_e	Probability of effector loss during homing	10^{-4}
l_{ne}	Probability of nuclease and effector loss during homing	10^{-4}
s_d	Cost of target site disruption	0
s_n	Cost of nuclease expression	0.05
s_e	Cost of effector expression	0.1
h_d	Dominance coefficient for target site disruption	0.5
h_n	Dominance coefficient for nuclease expression	0.5
h_e	Dominance coefficient for effector expression	0.5
h_{rc}	Dominance coefficient for refractoriness	1
r_c	Homozygous degree of refractoriness	1

Comparison of IGD and a Classical Replacement Drive

In order to understand how Integral Replacement compares to the existing replacement paradigm, an initial comparison was made. It compared the allelic dynamics and protection offered by the two-locus IGD model when contrasted to the classical replacement model (Fig.3.2). When using identical baseline parameters, the IGD strategy was found to confer 95% protection for 81 generations, whilst the classical strategy provides the same level of protection for just 39 generations. Assuming a 20-day generational time (Depinay et al., 2004; Mordecai et al., 2013), this would result in approximately 4.5 years of protection in the case of the IGD drive and approximately 2.1 years in the case of the classical replacement drive. In both cases, protection decreases due to the generation of resistance alleles. As functional resistance mutations are equivalent to the wild-type allele, they carry no fitness cost and replace the nuclease and effector components. In the baseline parameter values used for this simulation, the value associated with nuclease expression from the drive component (s_n) was less than the cost for expressing the effector transgene (s_{e2}). As a result, the effector component was replaced by resistance alleles at locus 2 more rapidly than the drive component at locus 1. This contrasts with the classical approach, where the endonuclease and effector are linked and expressed from a single locus, compounding the fitness cost and causing the strategy to fail more rapidly. Additionally, in the IGD strategy, loss-of-function resistance alleles cause knockout of the endogenous host gene.

As a direct result, individuals carrying loss-of-function alleles are selected against on a population level. By contrast in the classical strategy where the target site is functionally neutral, such mutations are tolerated and accrue more rapidly. This initial comparison demonstrated the potential for IGD to offer significant improvements in longevity, and thus protection.

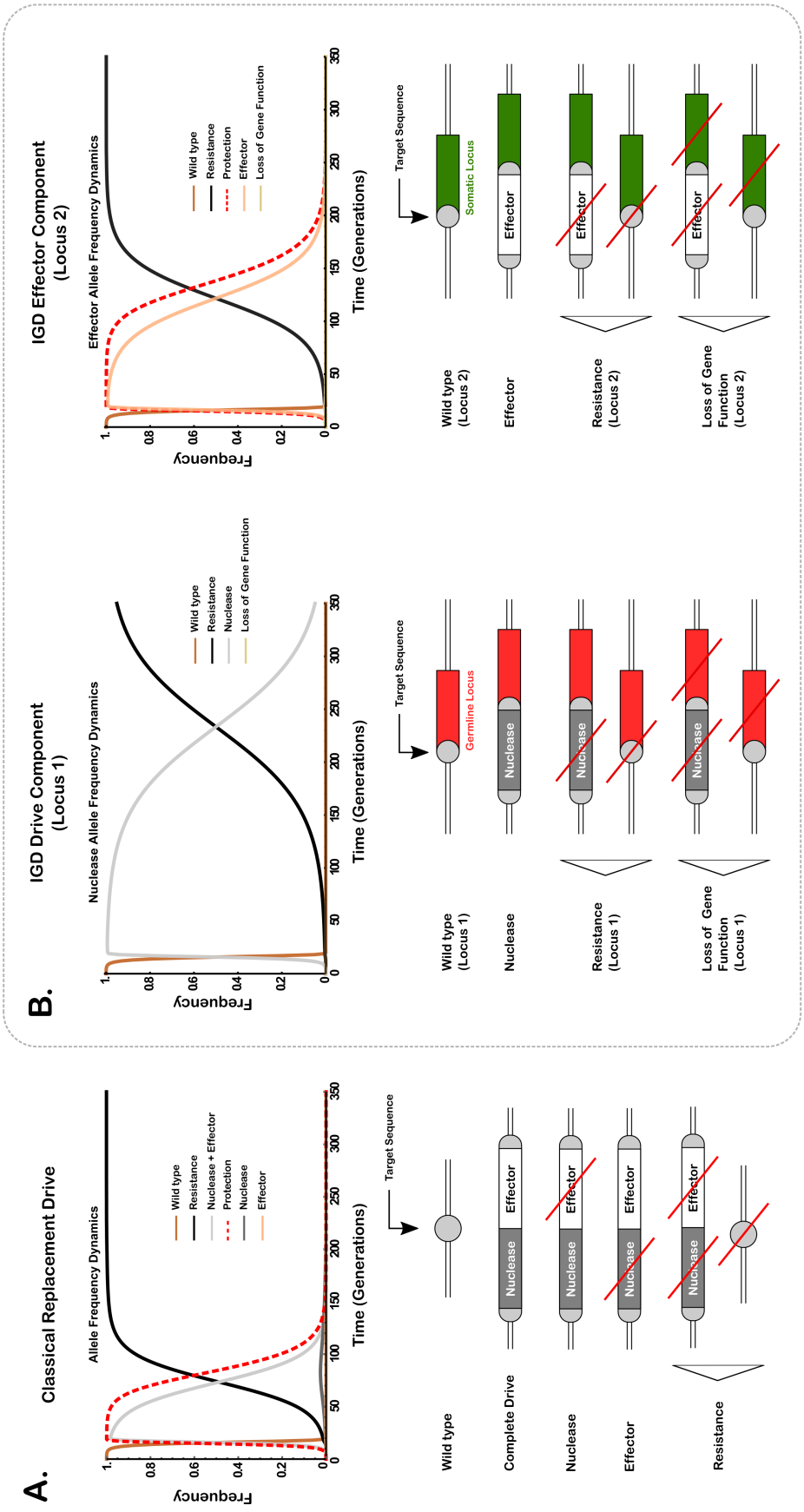


FIGURE 3.2: **Comparison of 'Classical' and Integral Gene Drive Dynamics.** Shown are the allele frequency dynamics for A) the 'Classical' drive configuration, and B) a two-component IGD strategy using baseline parameter values. Given that the IGD system acts as a split drive, modeled behaviours are depicted here for each component. The red-dotted line indicates protection defined as the drive-mediated reduction in vectorial capacity. Figure adapted from Nash et al., 2019.

In order to determine which underlying parameters are critical to the persistence of protection, each parameter was systematically varied whilst maintaining all other parameters at their baseline values (Fig. 3.3). Both the classical and IGD models demonstrated similar behaviours in this regard. However, it was possible to identify a number of parameters for which the IGD strategy is more robust. For example, it was found that increasing the proportion of resistance alleles (both functional and loss-of-function) at locus 1, within a biologically sensible range, does not impact the longevity of protection. This observation can be explained by examining the effect that the drive component (locus 1) has on locus 2. The drive component converts wild-type alleles to effector alleles in individual heterozygous for the effector component. This ensures rapid propagation of the effector component through the population, ensuring a high-allelic frequency before resistance at the drive locus (locus 1) can take over. Protection subsequently diminishes as functional resistance alleles take over at locus 2, replacing effector alleles in the population. Due to the rapidity with which the drive component is able to convert wild-type alleles into effector alleles, the loss of the drive component is of no consequence to the duration of protection, because conversion has already occurred at locus 2.

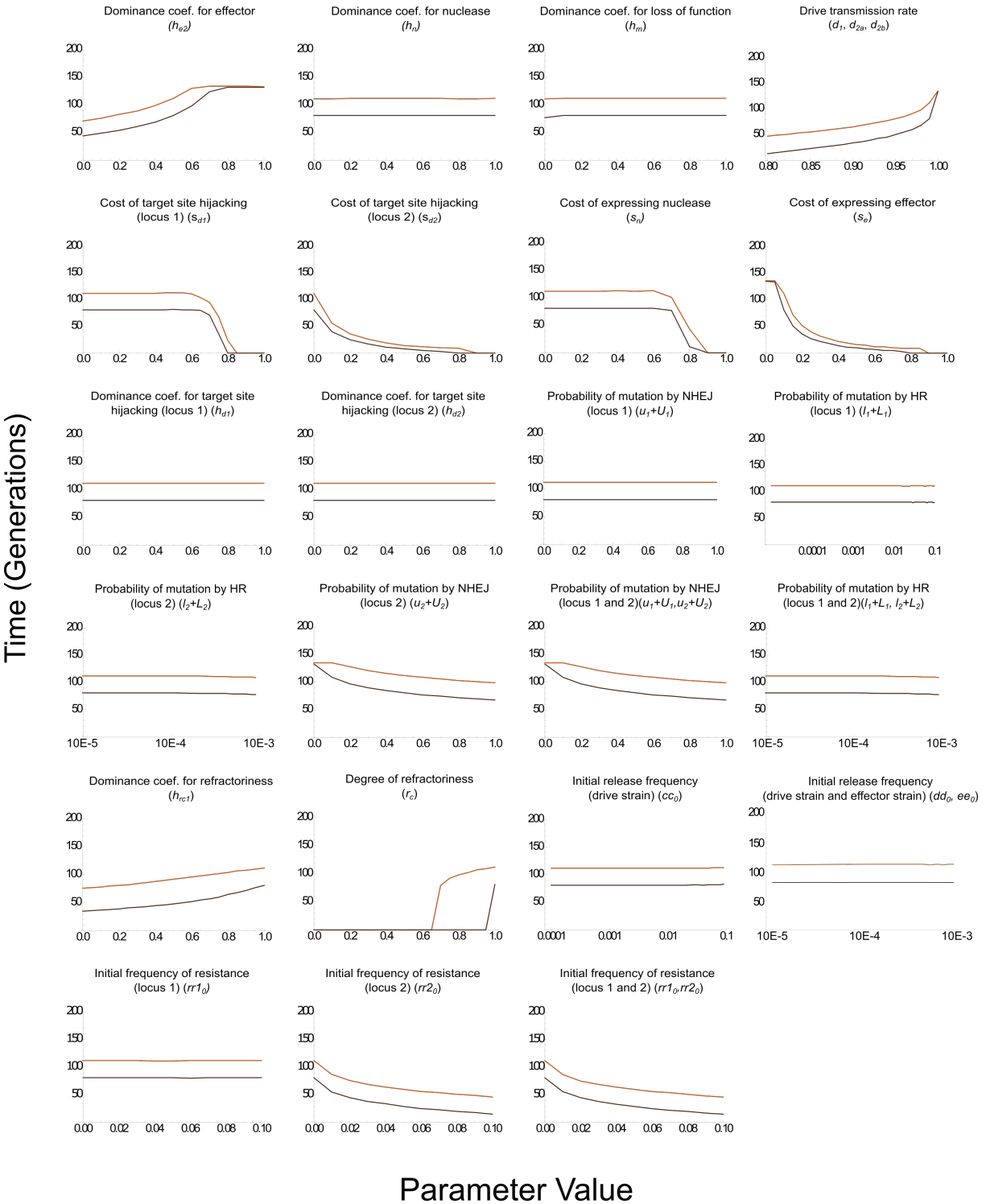


FIGURE 3.3: **Systematic Scan of IGD Model Parameters**. Depicted are parameter scans for all model parameters in the IGD model using a two-component (single drive component, single effector) implementation. The duration of protection at 67% (top line) and 95% (bottom line) are shown on each graph. In each graph, only the named parameter is varied, whilst all other parameters are mainted at baseline values. Figure from Nash et al., 2019.

The parameters most liable to reducing the longevity of protection were the cost of target site hijacking at the effector locus, and the cost of expressing the effector transgene. This is

intuitive given that the effector transgene is responsible for any protection offered by the IGD strategy, and fitness costs associated with the formation of new effector alleles, or expression of the effector result in individuals carrying the effector allele to be selected against, at a population level. Another parameter with a strong negative effect on the duration of protection, is the pre-existence of resistance alleles at the effector locus.

In order to further investigate the effects of pre-existing resistance in a target population, a number of iterations of the IGD model were performed with increasing frequencies of resistance alleles (Fig. 3.4). High frequency of pre-existing resistance alleles at the drive locus (locus 1) is well tolerated with significant reductions in protection only occurring once the initial frequency of resistance alleles increases above 80%, such that the target site is a minor allele within the population. By contrast, should the frequency of pre-existing alleles reach just 10% at both the drive (locus 1) and effector (locus 2) loci, the persistence of 95% longevity is decreased from 81 generations to 15 generations. It is reasonable to assume pre-existing resistance is present at significant frequencies in target populations (Anopheles gambiae 1000 Genomes Consortium et al., 2017), and our own work in implementing IGD drive components in *An. gambiae* in a laboratory setting has further suggested this (David Ellis, Personal Communication). Indeed, a recent study investigating pre-existing resistance has posited that it is existing genetic variation that is most liable to contribute to resistance and drive failure, as opposed to new resistance alleles generated by the activity of a drive component (Unckless, Clark, and Messer, 2017).

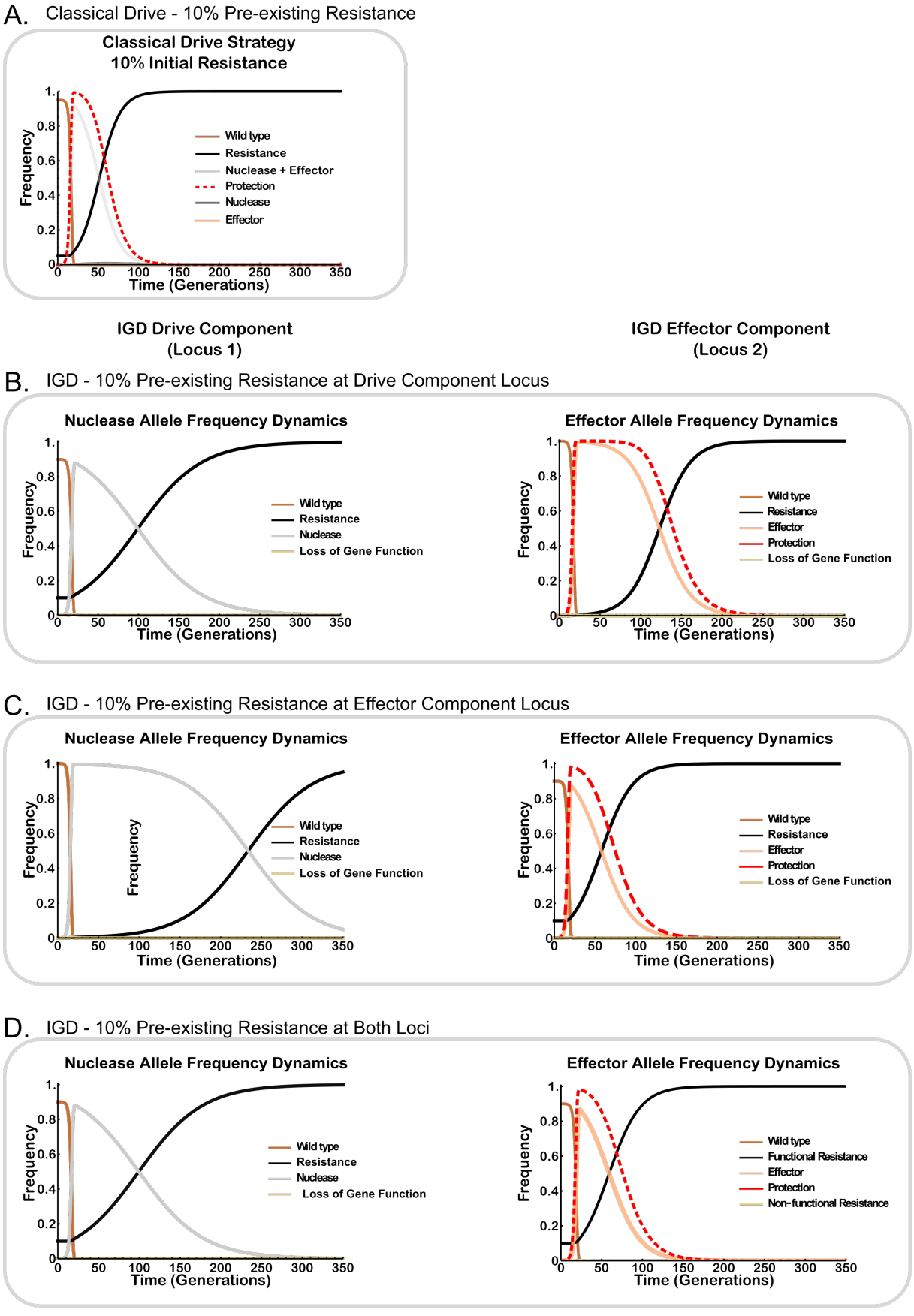


FIGURE 3.4: **Drive Dynamics in a Background with Pre-Existing Resistance**. Shown are modeled allele frequency dynamics when releasing drive systems into a simulated field population with 10% pre-existing resistance alleles at a given target locus. Resistance is modeled at A) the 'classical' drive, B) IGD drive component locus, C) IGD effector component locus, and D) both IGD loci. Figure from Nash et al., 2019.

3.5.5 Investigating a Three-Locus IGD Implementation

The architecture of the IGD model allows for multiple independent drive and effector components to be released and to interact. Given that findings relating to a two-locus IGD implementation emphasised the effector locus as being particularly vulnerable to pre-existing resistance, a three-locus, two-effector implementation was investigated. The deployment of multiple effector components should assist in the persistence of protection over a longer duration, increasing the efficacy of the strategy in reducing malarial transmission. This effect is due to the fact that resistance alleles must now be generated at multiple loci in significant frequencies for protection levels to be suitably diminished. Using baseline parameters, it was found that releasing a driving strain carrying two effectors into a naive field population resulted in an extended period of 95% protection at 103 generations (approximately 5.6 years) (Fig.3.5A). With pre-existing functional resistance alleles present at a frequency of 10% at all three loci (drive - locus 1; effector components - loci 2 and 3), 95% protection endured for 38 generations (approximately 2.1 years) (Fig.3.5B). This represents a significant increase in the robustness of protection when compared to release of a single-effector driving strain under identical conditions and to the classical two locus paradigm.

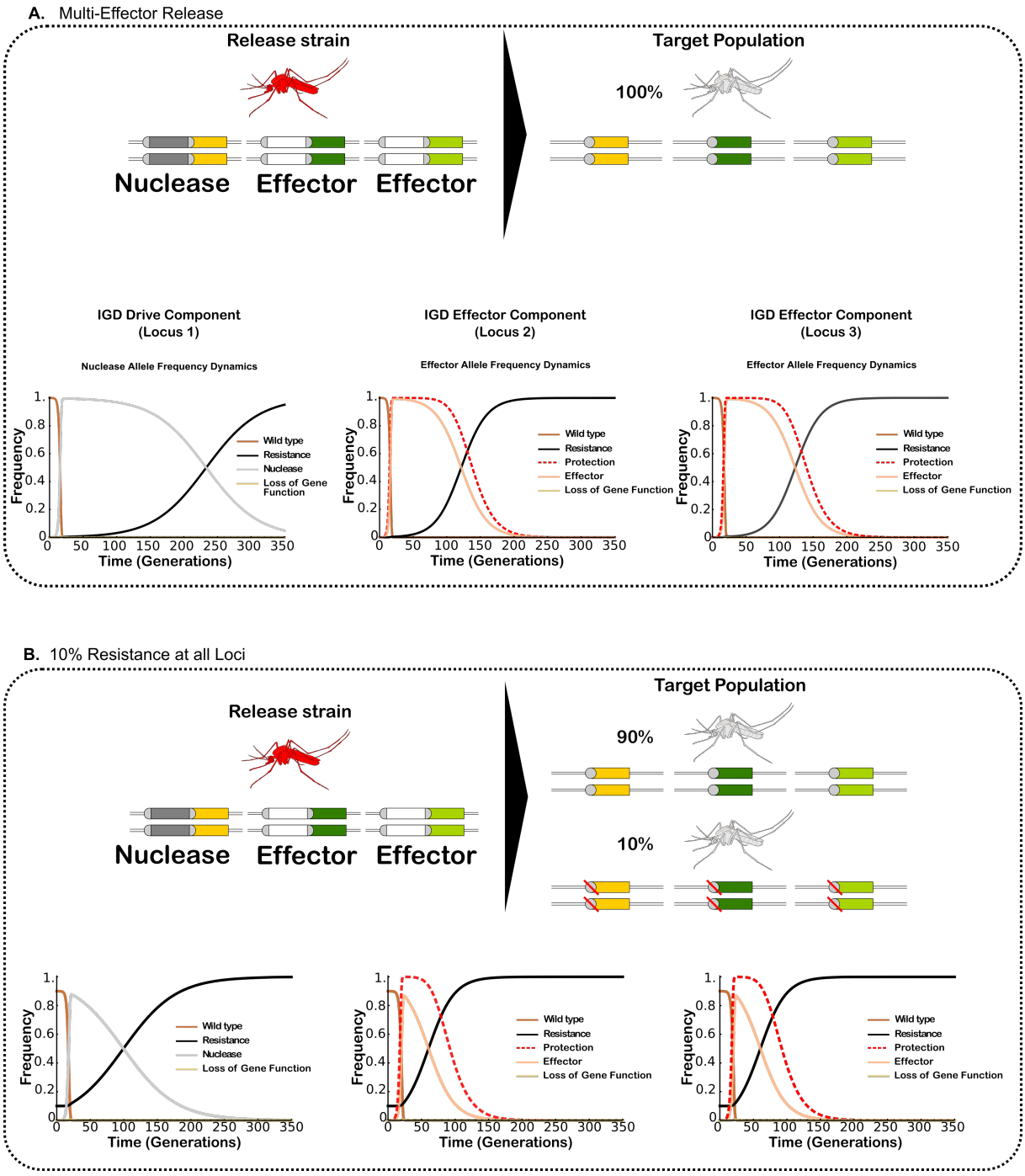


FIGURE 3.5: **Allele Frequency Dynamics of a Two-Effector IGD Implementation**. This implementation of the IGD model considers three components (one drive component and two effector components). Releases are simulated into A) an entirely susceptible field population, and B) a field population with pre-existing resistance alleles at 10% at all three loci. Figure adapted from Nash et al., 2019.

An alternative approach is to initially release a single-effector driving strain into a population, and subsequently release a second effector at a separate locus in a bid to increase the duration of protection (3.6). The secondary effector would be driven through the population by the activity of the drive component (locus 1) until it is replaced by resistance alleles. For the given baseline parameter values, it was demonstrated that release of a secondary effector when protection levels were reduced to 67% (Fig.3.5C) and 95% (Fig.3.5D) was sufficient to restore complete protection. Lowering the cost of nuclease expression (s_n) increases the duration over which the drive component is present in the population. This in turn extends the duration over which protection can be sustained by introduction of additional independent effector components. In order for full protection to be achieved again after release of a secondary effector component at locus 3, the allelic frequency of the nuclease must be $\geq$ 55% when it is released. By releasing the secondary effector component before protection has significantly decreased, it is possible to achieve the greatest extension to the duration of maximal protection. Here it was found that this optimum point is achieved when secondary release of an effector component at locus 3 occurs before protection decreases below 95%. These findings may aid in future release strategies, permitting researchers to maximise the efficiency of IGD strategies. From here, it is then necessary to consider other challenges to the deployment and success of IGDs.

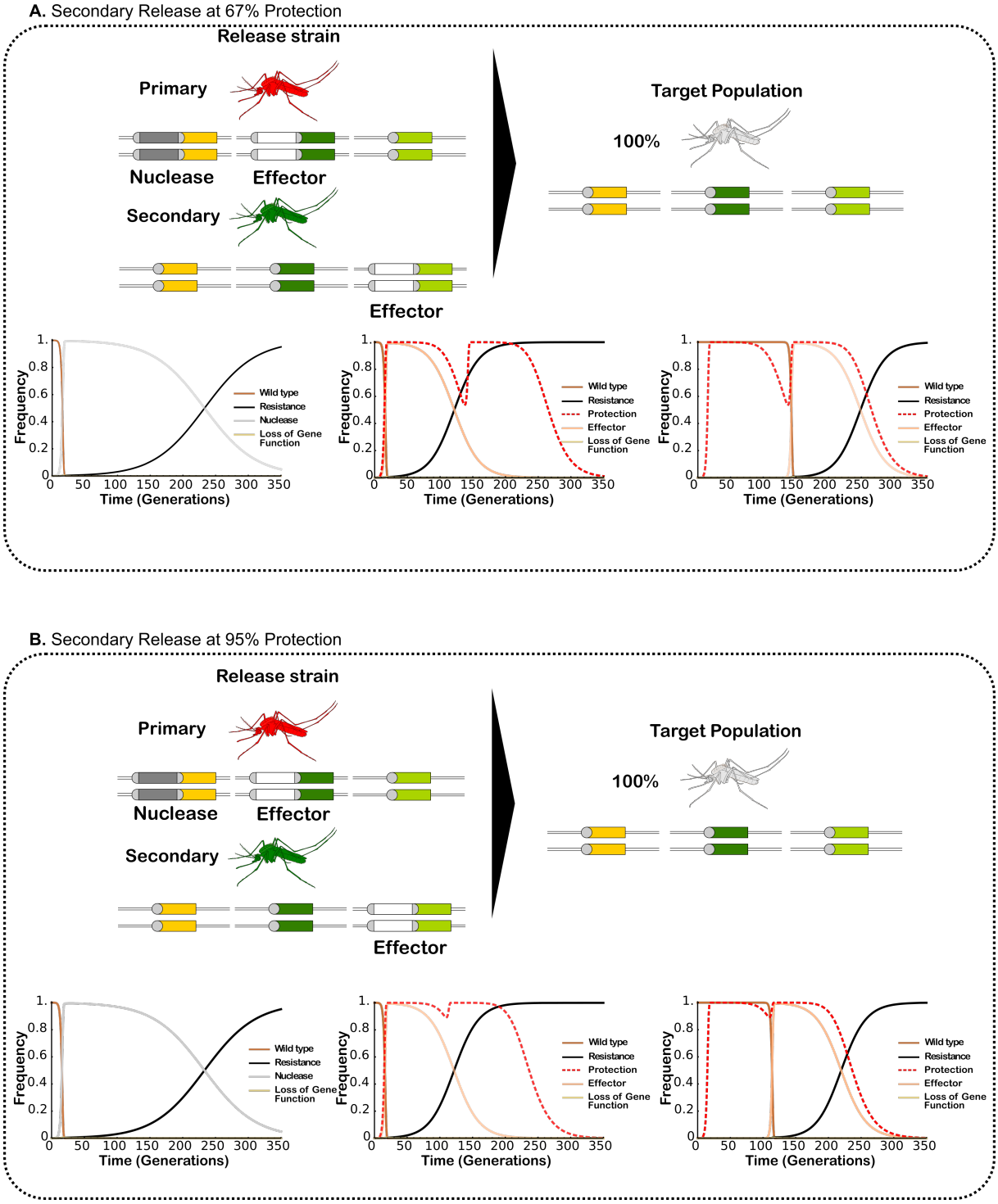


FIGURE 3.6: **Allele Frequency Dynamics of a Two-Effector IGD Implementation**. Releases A) and B) consider the staggered release of a secondary effector strain once protection drops below 67% and 95% respectively, with all other parameters maintained at their baseline values. IGD dynamics are depicted in terms of the modeled behaviour of all three components at their respective loci, and the red-dotted line indicates the relative reduction in vectorial capacity of the targeted field population. Figure adapted from Nash et al., 2019.

3.5.6 Field Deployment and Evaluation of IGDs

A major barrier to the deployment of gene drive mosquito strains in the field is the significant regulatory pipeline that must be navigated before import of GM vectors can occur. The first transgenic mosquito strain, carrying a system to cause male sterility (Windbichler, Papathanos, and Crisanti, 2008), was approved for import into Burkina Faso in partnership with the Target Malaria foundation in 2018 with small scale pilot releases performed in 2019. Importantly, this strain forms part of a self-limiting genetic control strategy: it is likely that self-perpetuating control systems, such as gene drives, will be faced with numerous regulatory barriers over an extended period before they can be approved for import and eventual field application. It is thus critical when designing and constructing gene drive systems to ensure that pragmatic considerations are made for potential regulatory and operational challenges. The IGD strategy offers a number of potential advantages over conventional gene drives, both suppressing and replacement, due to its design features.

IGD calls for components with minimal complexity in order to diminish potential fitness costs associated with disruption and expression. To this end, an effector component consists simply of an intronic gRNA placed within the coding sequence of an effector transgene, which is to be inserted into an endogenous gene in the host. All other regulatory features are specific to the chosen integration site. The simplicity of such effector components should aid in easing the process of field testing and in generating effector lines. The effector component, previously described here as capable only of non-autonomous homing, is not able to self-propagate. Therefore, it is likely that the regulatory scrutiny required for effector-only strains would be lower than that required of conventional fully-driving strains, carrying transgenes for population replacement (Fig.3.7). This poses the distinct advantage of increasing the speed with which effector transgenes can be assayed against field populations in malaria endemic settings. Modified strains harbouring IGD drive components are molecularly simple, but would be expected to incur a higher degree of scrutiny due to their ability to home in an autonomous fashion. However, these drive components are designed to have as small a phenotypic effect as possible and carry no effector transgenes. This is as opposed to existing replacement strategies (Gantz et al., 2015), or suppression drives (Kyrou et al., 2018; Hammond et al., 2016), which have a significant phenotypic effect in modified individuals. To this end, should an accidental release occur, it would be expected that consequences would be minimal. However, due to the posited low fitness costs associated with IGD drive components, a Cas9 source would be present in the affected population for a prolonged duration, and would preclude containment should accidental release occur of a non-autonomous drive component with a suitable gRNA element.

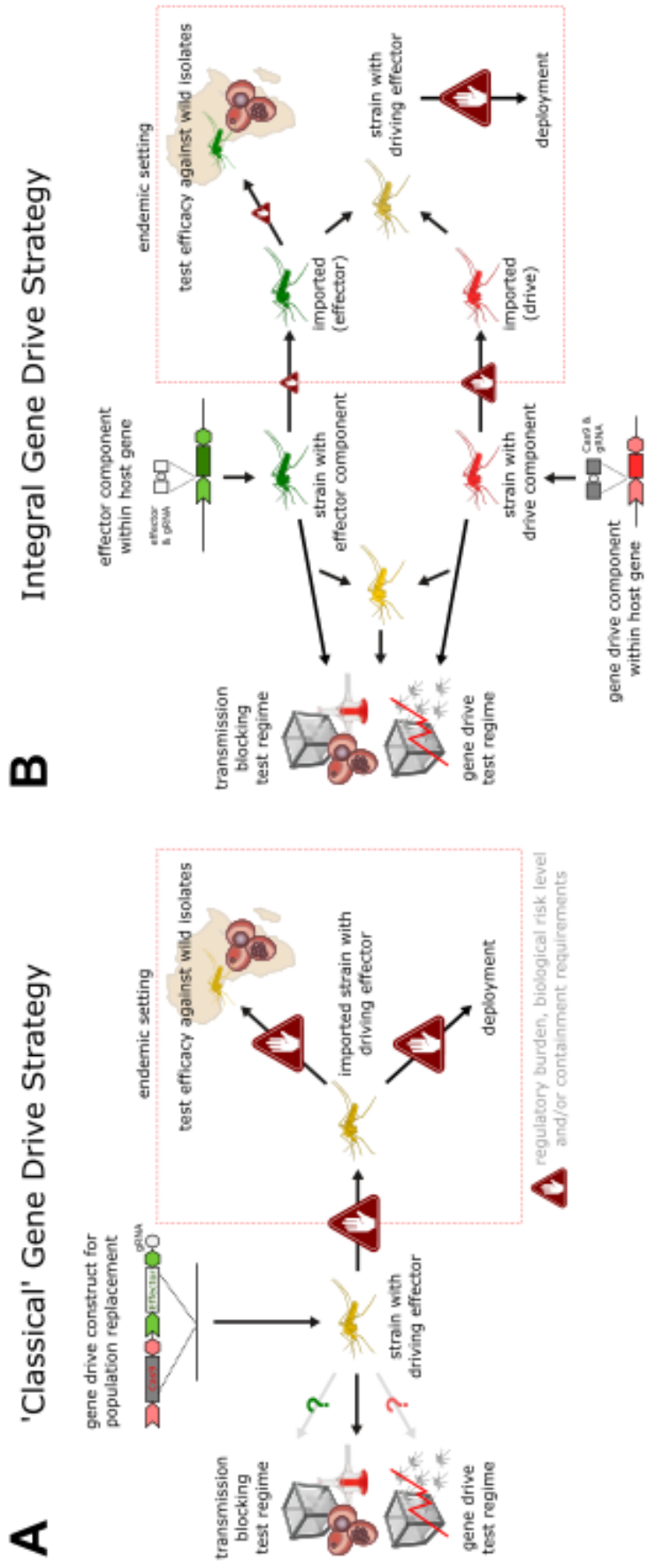


FIGURE 3.7: **Comparison of Theoretical Regulatory Pipelines for Conventional and IGD Drive Strains**. A comparison of possible regulatory and molecular steps required for laboratory and field testing. A) Conventional gene drives pose significant challenges when attempting to assay the efficacy of individual components in the laboratory and the field, and may face high regulatory burdens. By comparison, B) IGD drive components can be prototyped separately and imported into the field as individual strains, lowering the potential regulatory burden. Once *in situ*, strains carrying individual components can be tested and crossed to produce a single driving line for field deployment. Figure adapted from Nash et al., 2019.

An important consideration for biological monitoring of transgenes in field populations is the use of fluorescent reporter genes. Currently, the use of such reporters is recommended as a part of monitoring efforts, both prior to release, and following deployment. The scientific working group reviewing potential routes to the deployment of gene drive mosquitoes in Sub-Saharan Africa specifically stipulate: 'Monitoring will be simplified by the inclusion of a unique visual or easily identified molecular marker in the construct.' (James et al., 2018). However, it has been observed that in conventional gene drive strategies, the decoupling of drive components can occur as a consequence of incomplete HDR (Hammond et al., 2016; Oberhofer, Ivy, and Hay, 2018). In this scenario, there are two major consequences: 1) One may falsely assume that a gene drive is still present in a field population. This is a lesser risk, given that a decoupled reporter will no longer propagate at a super-Mendelian rate, and will be unlikely to persist in a target population for an extended period. Moreover, the reporter is unlikely to incur any particular phenotypic consequences of its own. The more serious scenario 2), occurs when decoupling allows a gene drive to spread undetected through a field population, inhibiting the ability of researchers to maintain proper biological monitoring. Furthermore, this scenario is liable to generate type II errors in data sets, falsely negating the efficacy of a given gene drive release. To this end, it is preferable to make use of molecular genotyping as a routine screening tool in monitoring gene drive activity. This is particularly true in the case of replacement gene drives, where it is possible to maintain stocks as complete driving strains in a laboratory setting prior to release.

When looking to deploy gene drives in the field, the current recommendation posited by the aforementioned scientific working group on gene drives is to develop a step-wise program for development and release (Fig.3.8). When looking to deploy conventional gene drives in the field, it is difficult to achieve a step-wise deployment as the small scale pilot study that is recommended is dependent on the ecological and geographic isolation of any given release site, and the underlying factors are not yet clear. By contrast, the architecture of the IGD system provides a clear pathway to step-wise testing. This is due to the modularity of its design and the independence of interacting drive components. To demonstrate this, a number of potential steps were modelled using the two-locus implementation of the IGD model.

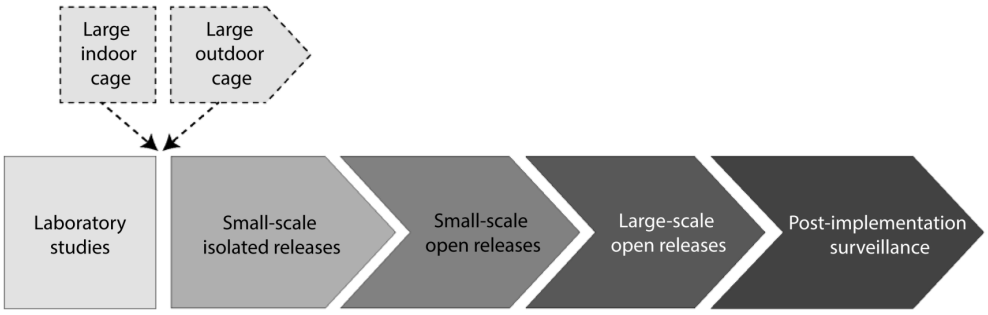


FIGURE 3.8: **Suggested pathway for the safe deployment of gene drive mosquitoes, in a step-wise manner**. Incremental steps are posited from laboratory bench to field deployment, and subsequent ecological surveillance to determine long-lasting consequences of gene drive release. Figure from James et al., 2018.

Initially, an inundative release was simulated (Fig.3.9) with different release frequencies (1%, 5%, 10%). This initial release of an effector-only strain would allow for quantification of mosquito fitness and efficacy of the effector transgene prior to a larger-scale release. The model predicted that even with a 10% release of an effector line, the frequency of the effector allele within the field population would converge towards 0 within 50 generations. As the effector component in the IGD strategy is not capable of self-propagation, this release is inherently self-limiting, providing a degree of safety and containment. The next step in this step-wise pipeline is the release of a strain carrying a single effector and a non-driving Cas9 source, that would facilitate limited localised spread of the effector component. Here, modelling demonstrated that varying the release frequency of this self-limiting strain would allow for contained testing of the homing efficacy of the effector component. Even when releasing effector and non-driving Cas9 alleles at 10% frequency, both alleles are lost from the population within approximately 200 generations. Given that both an inundative effector-only and a limited-drive release are self-limiting, it may be possible to allow for transgene frequencies to dissipate and the site be re-used. Following full laboratory investigation of the drive and effector components followed by self-limited field testing, it would be possible to combine drive and effector strains in the field, producing a fully-driving strain which is self-propagating. At this point, an inoculative release could be performed, and the transgenes would be allowed to spread through the target population at super-Mendelian rates. Modelling demonstrated that the release of nuclease and effector components at 0.0001% frequency was sufficient to achieve a prolonged period of protection before functional resistance alleles overcome the drive system.

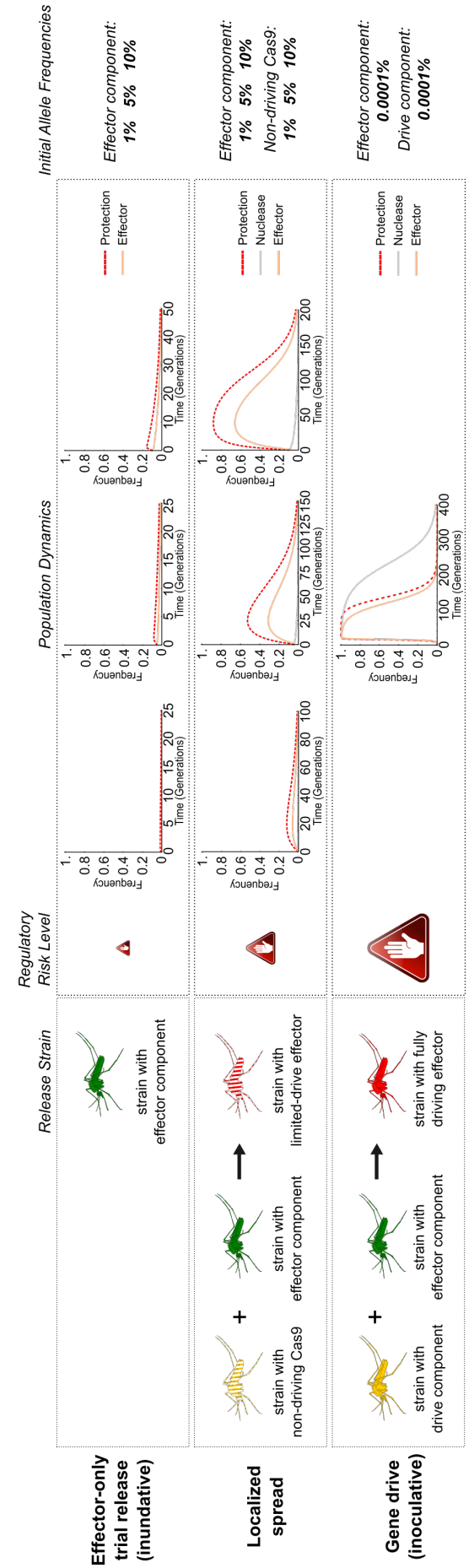


FIGURE 3.9: **Considering Allele Frequency Dynamics of Limited Release Strategies**. An advantage of the IGD strategy is that the split architecture facilitates incremental field-testing without the need for additional molecular engineering steps. A) Initially, inundative release of an effector-only strain permits quantification of the efficacy of a given anti-malarial transgene in a field setting. B) Subsequently, the effector strain may be crossed to a non-driving Cas9 strain, allowing for limited spread of the gene drive over a limited area. C) Finally, when key regulatory and biological queries have been fulfilled by staged testing, the effector-only strain may be crossed to a strain carrying a functional drive component. This would provide a fully-driving strain, which can subsequently be released as very low frequencies. Figure adapted from Nash et al., 2019.

A final important observation regarding the design features of the IGD system is that the traits of the effector component are unlikely to be altered in the presence of the drive component, as the two are distinct and, whilst cooperative, independent. By contrast, in a conventional drive strategy, multiple iterations of the drive would need to be constructed to test individual components (i.e. nuclease, effector), and to be appropriate for each step in the testing pipeline. The ability to separate components lends itself to a high-throughput design-build-test-learn (DBTL) product development cycle, facilitating fast and directed iterative design choices. The independent nature of IGD components allows for multiple modules to be developed in parallel and by separate teams, before being combined for testing. Modeling of the self-limiting release strategies suggests that it is possible to subsequently test these components in a field setting with low risk, allaying concerns typically held for gene drive testing. Moreover, a site may be used multiple times, assuming that the presence of the released components has been suitably reduced.

3.6 Discussion

The IGD strategy described here posits a versatile approach to vector control, allowing for release of multiple interacting components. This may offer a distinct advantage over inflexible single-construct population replacement strategies, permitting the strategy to be modified in real-time in reaction to observed phenomena in the field. This fluidity allows for ongoing refinement of genetic control post-release.

The modeling conducted herein provides the basis for a theoretical proof of the efficacy of a novel gene drive strategy for population replacement. The work conducted here suggests that, for the given parameters and associated values, the IGD strategy offers a significant improvement on existing population replacement designs. By integrating drive and effector components directly into the coding sequence of an endogenous host gene, undesirable mutations tend to result in frameshift and knockout of gene function. This, in turn, incurs a fitness cost which causes the individual to be selected against on a population level, reducing the rate at which resistance accrues, and increasing the duration over which the gene drive is effective. It is important to note that IGD cannot prevent resistance alleles from being generated, but it can significantly extend the degree of protection offered by coupled effector components. This approach may overcome existing challenges in target site selection: supposedly neutral genomic loci tend to be poorly conserved, whilst insertion into conserved sites tends to result in functional disruption. By contrast, the IGD model provides a means to exert a minimal burden on modified individuals, being seamlessly integrated into an endogenous coding sequence, whilst simultaneously utilising a locus that is highly conserved. A later chapter in this thesis discusses the practical challenges in achieving transgenesis of vectors with the IGD strategy.

It is important to consider significant assumptions made in the construction of the IGD model. In accordance with previous research published on this topic (Beaghton et al., 2017), the baseline parameter value given for nuclease expression (s_n=0.05) is lower than that of effector

transgene expression (s_{e2}=s_{e3}=0.1). This assumes use of somatic genes with stronger expression in order to achieve effective lethality/transmission blocking of the parasite. Germline loci with stronger expression could also be selected. With regards to the model's assumptions of the rate at which resistance alleles are generated, when using probabilistic models that simulate finite populations, resistance alleles incurred at low frequency are initially lost in a stochastic manner (Beaghton et al., 2017). This in turn would delay the initial onset of resistance and increase the duration over which protection should persist. It can be inferred from this observation that the deterministic model used here for IGD is relatively conservative in its predictions relating to protection.

By performing parameter scans with individual parameters, whilst maintaining all other parameters at baseline values, it was possible to identify parameters with the greatest effect on the duration of protection for the IGD strategy. These parameters, in a two-locus model, were: the cost of expressing the nuclease at locus 1 (s_n), the cost of expressing the effector at locus 2 (s_{e2}), the cost of hijacking the drive component target locus (s_{d1}), and the cost of hijacking the effector component target locus (s_{d2}). Longevity of protection is unresponsive to increases in fitness cost associated with expression of the nuclease from the drive component, or from hijacking the drive locus. However, changes to the cost of hijacking the effector locus (2 or 3), and to the cost of expressing the effector transgene, are poorly tolerated. This dynamic is explained by the high conversion rate of the drive component, facilitating the conversion of wild-type alleles into effector alleles. Therefore, the drive component is tolerant to high costs expressed with nuclease expression, target-site disruption, and protection is not strongly affected. This observation will be of particular importance when conducting molecular work to design and characterise effector components for use with the IGD system. It is clear that, given the parameters and associated values used here, it is critical to minimize the cost of effector transgene expression and to ensure seamless integration of the ensuing effector component into the target locus. The modular nature of the IGD strategy may aid in the testing of candidate effectors in a field setting, allowing effectors to be evaluated in the absence of other IGD system components, and then simply combined by crossbreeding.

In order to expand the applicability of this current IGD model, a number of additional features need to be implemented that are not currently considered. Spatial heterogeneity, climatic and meteorological features, as well as human geography, could be incorporated to render an increasingly realistic understanding of the effects of IGD. Currently, the simulated population is considered to exist in a homogeneous distribution across an undesignated landscape. Previously conducted work has examined the effect of spatial and density parameters on the efficacy of replacement drives (Eckhoff et al., 2017), and similar features would be valuable in the context of IGD. Work is currently being conducted by members of the Institute of Disease Modeling, to integrate IGD as a feature into the EMOD framework (Bershteyn et al., 2018) for disease modeling (Prashanth Selvaraj and Daniel Bridenbecker, Personal Communication). EMOD is a mechanistic, stochastic, agent-based framework that incorporates a broad range of features mentioned above, including geographic (natural and human), climatic, migration, and

intervention data, allowing for detailed predictions regarding complex vector-human interactions. An additional feature would involve the inclusion of sex-specific parameters, permitting investigation into the effect that sex-specific expression patterns for both drive and effector components (Beaghton, Beaghton, and Burt, 2017). This is particularly pertinent, given that it may become necessary to limit effector activity to females, whilst maintaining drive component activity in males to reduce fitness costs associated with expression of the nuclease and effector transgene. This approach may also prevent issues associated with maternal deposition of Cas9 in the early embryo (Champer et al., 2017). This feature was subsequently implemented in Chapter 5 of this thesis, and was used to investigate the effects of sex-specific drive expression.

The work conducted in the course of this thesis is limited to the analysis of a single drive component, and two-effector components. Development of more powerful iterations of this model or other frameworks based on this work could explore the dynamics of increasingly complex, multi-drive/effector strategies. It may be possible to evaluate the behaviours of IGD strategies using separate and independent drive target loci, ensuring a consistently high level of Cas9 expression, even in the face of resistance allele generation or pre-existing resistance in particular vector subpopulations. Similarly, the effects of releasing multiple, multi-effector strains integrated into different host genes could be evaluated to better understand how such an approach could be used to decrease the vectorial capacity of a target field population. By operating multiple effector components at multiple distinct loci, it may be possible to reduce the effects of resistance and standing variation in a field population. This is particularly true in the context of the IGD strategy, which is inherently modular and amenable to the use of multiple gRNAs. Previous studies investigating multiplexed gRNAs have posited a minimal effect on drive performance, though they incur a significant increase in design complexity (Champer et al., 2018a; Oberhofer, Ivy, and Hay, 2018), which can be avoided in the IGD strategy. A final approach that could be investigated through modeling is a hybrid strategy, using effector components in parallel with suppressive strains. In doing so, Cas9 could be provided to permit homing of the effector components, whilst simultaneously reducing the size of the target population. This would mean that in the event a suppressive strategy fails to eradicate a population, the remaining individuals would have a reduced vectorial capacity. The IGD strategy could offer an additional layer of security following a suppressive program, protecting against a resurgence in vectorial capacity once the original intervention has ended.

Chapter 4

Results - Engineering Intronic gRNAs

4.1 Chapter Aims

In this chapter, we conducted a pilot study to determine whether it was possible to express sgRNA cassettes from within introns. This was done in order to obtain an initial understanding of whether our proposed design for IGD drive and effector components was feasible. Work was conducted *in vitro* to permit rapid screening of possible intronic gRNA variants.

The work conducted had the following aims:

1. Construct a system for testing intronic gRNAs
2. Screen intronic gRNA cassettes for performance
3. Investigate whether poor performance of intronic gRNA cassettes is a consequence of low expression, or weak splicing efficiency

4.2 Introduction

The IGD paradigm proposed in Chapter 3 of this thesis proposes use of an intronic gRNA (igRNA) cassette in both drive and effector components. This is to permit seamless hijacking of a targeted germline or somatic locus without causing disruption to the host gene, which is liable to result in consequences for the fitness of the modified vector. Currently, there are few examples of intronic gRNAs being deployed in the existing literature, and to date both examples have originated in organisms outside the remit of this thesis - namely, rice plants (Ding et al., 2018) (published after this work), and mammalian cell culture (Kiani et al., 2014). As described in chapter 1, experimental work both *in vitro* and *in vivo* was conducted in *D.melanogaster*.

There are a number of limitations to using Anopheline mosquitoes for gene drive research - namely the complexity of their life cycle, the requirement to provide suitable blood sources for females, and the limited availability of genetic tools. By contrast, *D.melanogaster* is an organism that has been used to set the cornerstone of our current understanding of genetics. The complete genome sequence for *D.melanogaster* was first published in 2000 (Adams, 2000), and since numerous other *Drosophila* species have been sequenced. Its genome is 60% homologous to the human genome, with fewer functional redundancies, and approximately 75% of genes known to have a disease association in humans have a corresponding homolog in the fruit fly. The

entire genome sequence is a mere 180 Mb (Bosco et al., 2007), consisting of one pair of sex chromosomes, and three pairs of autosomes. Genomic annotations are available through Flybase, a curated knowledge base for genetic information pertaining to *D.melanogaster* (Thurmond et al., 2019).

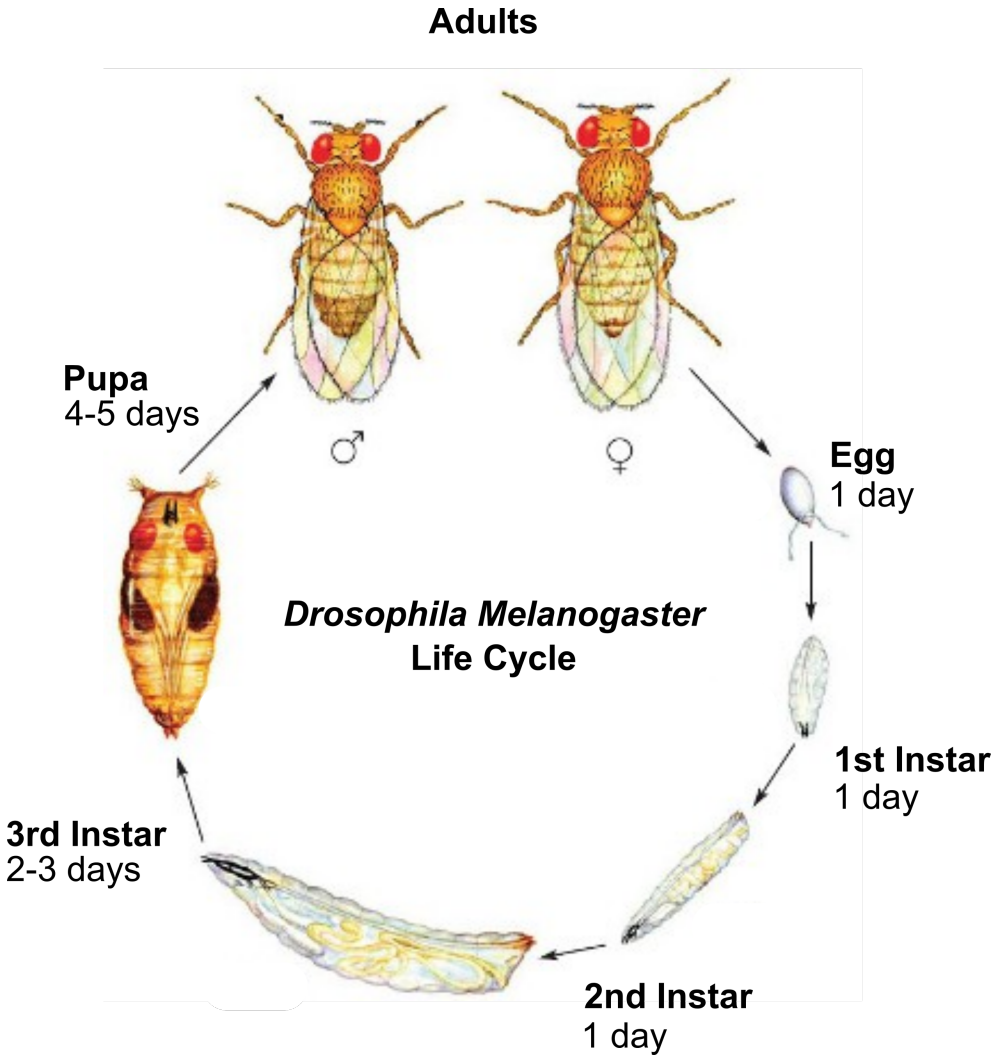


FIGURE 4.1: **Overview of the *Drosophila melanogaster* Life Cycle**. *Drosophila* undergo four terrestrial developmental stages, from to adult. Developmental time is temperature dependent, with times shown here reflective of a 25°C environment. Figure adapted from Flagg, 1988.

A key advantage of using *D.melanogaster* as a model organism lies in their short generational time, large generation size, and simple husbandry. This is particularly true when considering the extended duration, and complex nature of Anopheline mosquito's life cycle. *D.melanogaster* tend to be cultivated between 18°C and 25°C, with significant differences in generational time resulting from this temperature difference (Fig. 4.1). In the context of this work it is important to additionally bear in mind that Cas9 endonuclease activity is temperature-sensitive, and higher temperatures tend to increase cleavage rates (Xiang et al., 2017). At 18°C, development from embryo to adult typically takes approximately 18-20 days, whilst at 25°C it is as little as 9-10 days. *Drosophila* can be cultured from egg to adult in a single acrylic tube, containing a food

source, and plugged with a cotton or acetate cap. At 25°C, 1st instar larvae eclose from eggs, 24 hours post-laying. They then develop in the food for approximately 4 days, increasing in size to 2nd and 3rd instar larvae. When suitably mature, 3rd instar larvae crawl from the food, up the interior side of the tube and pupate. A metamorphic process occurs, and after 4-5 days, adults eclose from the pupal case. During metamorphosis, many larval structures are lysed, and adult structures develop from imaginal discs, producing legs, head, thorax, etc... Once adults have eclosed they become sexually mature within 6-12 hours, and mate to recommence the cycle. The small genome size, simple life cycle, low genetic complexity, and broad tool set available to investigate a range of topics makes *D.melanogaster* a deeply attractive organism for use in eukaryotic genome editing.

4.2.1 Intronic Splicing in *Drosophila*

In order to robustly design igRNA cassettes, it is important to consider how splicing functions in eukaroytes, and specifically in *Drosophila melanogaster*. In eukaryotes, DNA is transcribed to RNA, which is in turn translated into proteins. During transcription, intronic regions must be removed from the transcript, allowing for the pre-messenger RNA (pre-mRNA) to be processed into mRNA , consisting of concatenated exonic transcripts. The majority of splicing occurs concurrently with transcription, though alternatively spliced exons, and long non-coding RNAs tend to be spliced afterwards (Tilgner et al., 2012). There are three major groups of introns; group I introns are primarily found in the genomes of bacteria and organelles, while group II are found in protista and fungi (Roy and Gilbert, 2006). A third group consists of spliceosomal introns, found in the nuclei of eukaryotes, and the group that is most relevant here.

Spliceosomal introns are so named because their processing is mediated by the spliceosome, composed of five small nuclear ribonucleoproteins (snRNPs) - U1, U2, U4, U5, and U6, as well as several spliceosome-associated proteins (SAPs) (Matera and Wang, 2014). Splicing is catalysed by RNAs, and occurs by recognition of splice site motifs that are present at the junctions between exonic and intronic features. Such features are broadly conserved between eukaryotes, though there is a degree of species specificity (Fig. 4.2) A donor site is present at the 5′ end of the intron, and an acceptor site at the 3′ end. Upstream of the 3′ donor site, is a pyrmidine rich region, and an adenosine branch point site. In mammals, genes tend to be large, and additional regulatory features are responsible for the regulation of alternative splicing, by which multiple products may be produced from a single transcript, depending on the choice of splice site. Importantly, mutations at the DNA level, or errors in transcription can result in the formation of cryptic splice sites, that are recognised by the spliceosome and result in the formation of non-standard transcripts. In the context of this work, cryptic splice sites are important to avoid so as to prevent the generation of truncated Cas9 proteins, misexpression of the downstream endogenous gene, and ensure correct igRNA expression.

Splicing proceeds by the formation of a series of complexes (Will and Max, 2002). The U1 snRNP binds the 5′ splice site in the transcribed pre-mRNA, while Splice Factor 1 (SF1) binds the branch point, forming the E complex. U2 snRNP displaces SF1, forming the A complex,

also known as the prespliceosome (Herold et al., 2009). Meanwhile U4, U5 and U6 snRNPs assemble into a trimer, and interact with the A complex, to form the B complex - the U5 snRNP interacts with the 5′ site, while the U6 snRNP interacts with the U2 snRNP. The U1 and U4 snRNPs are removed, and a nucleophilic attack occurs from the branch point, cleaving the 5′ splice site, and producing a lariat structure. The 3′ splice site is subsequently cleaved, and the adjacent exons ligated to produce a complete mRNA transcript (Herold et al., 2009).



FIGURE 4.2: ***D.melanogaster* splice site consensus sequences**
Letter height corresponds to the relative frequency of a given nucleotide at each position. Splice sites are given by arrows. Adapted from Lim and Burge, 2001.

Two candidate introns were selected as a basis to express sgRNA cassettes - a modified variant of the *fushi tarazu* (*ftz*) intron, and a synthetic intron originally produced for expression of sgRNAs in mammalian cell culture (Kiani et al., 2014). During embryonic development in *Drosophila*, maternal-effect genes in the mother produce RNA and protein products, which are subsequently maternally deposited into the early embryo (Gilbert, 2000). These products form morphogen gradients, and regulate expression of zygotic gap genes. These genes are broadly expressed in three segments in the embryo. Differential expression levels of each gap gene in turn causes expression of pair-rule genes, establishing pairs of segments known as metameres. Pair-rule genes in turn produce transcription factors, that regulate expression of segment polarity genes, producing 14 morphologically identical metameres. The products of the gap, pair-rule, and segmentation genes subsequently interact to regulate the homeoctic selector genes, that decide the fate of each segment (Gilbert, 2000).

The *ftz* gene is one such pair-rule gene, and encodes a transcription factor with a homeodomain motif (Laughon and Scott, 1984). *ftz*$^{-}$ mutants are recessive lethal, and embryos have half the number of segments (Wakimoto, Turner, and Kaufman, 1984), leading to the name 'fushi tarazu' or 'segment deficient'. In *D.melanogaster ftz* is initially expressed segmentally during the blastoderm stage (Hafen, Kuroiwa, and Gehring, 1984), and is involved in the developmental hierarchy of the early embryo where it regulates expression of *engrailed*, a segmentation gene (Howard and Ingham, 1986). In the later stages of embroygenesis, *ftz* is implicated in development of the hindgut (Krause, Klemenz, and Gehring, 1988). An MCS was inserted into the intron to mediate insertion of gRNAs. Additionally the intron was evaluated for predicted splice efficacy using the NNSPLICE neural network (Reese et al., 1997), in conjunction with data pertaining to canonical splice site sequences in *D.melanogaster* from modENCODE (Celniker et al., 2009). The intron splice sites were modified to increase their predicted splice efficacy (see methods for sequence).

The minimal mammalian intron was constructed just from canonical mammalian splice motifs and a MCS for insertion of gRNA cassette. It was previously employed in a multi-layered

repression device, where it was used to express an igRNA cassette from within the mKate and iRFP reporter genes (Kiani et al., 2014). Using flow cytometry, it was inferred that the mKate reporter spliced successfully when containing the minimal intron in mammalian cell culture. When compared to similar devices making use of exonically encoded gRNA cassettes, it was demonstrated to have 1,000 fold weaker repressive ability. Analysis of the intron (with MCS) using the NNSplice neural network predicted high splice efficiency in *D.melanogaster*, though its performance in invertebrate organisms was then unknown.

4.3 Materials and Methods

4.3.1 Construction of Experimental Plasmids

All experimental constructs used for assaying igRNA expression in S2 cell culture, were derived from pQUAST, a gift from Liquon Luo (Addgene plasmid #24349). Cassettes containing a 5xQUAS motif, minimal promoter, split eGFP reporter containing the *ftz* or minimal intron, and SV40 terminator were chemically synthesised and cloned into a high-copy cloning vector (Atum, CA, USA) (Fig. 4.3). Following synthesis the constructs were extracted from filter paper by centrifugation at 13000xg for 1 minute, following application of 100µl TE buffer (Qiagen, CA, USA). The elutant was subsequently used to transform Escherichia coli TOP10 cells (Invitrogen, Paisley, UK). The construct was transformed and miniprepped as previously described, and transformed cells plated on ampicillin agar plates. Transformant colonies were utilised to inoculate 5ml of LB-miller broth (Sigma-Aldrich, Dorset, UK), which was supplemented with 5µl of ampicillin. This inoculate was incubated in a 37°C shaking incubator at 220rpm overnight. The incubated inoculate was then subjected to alkaline lysis preparation, and the purified DNA quantified by nanodrop.

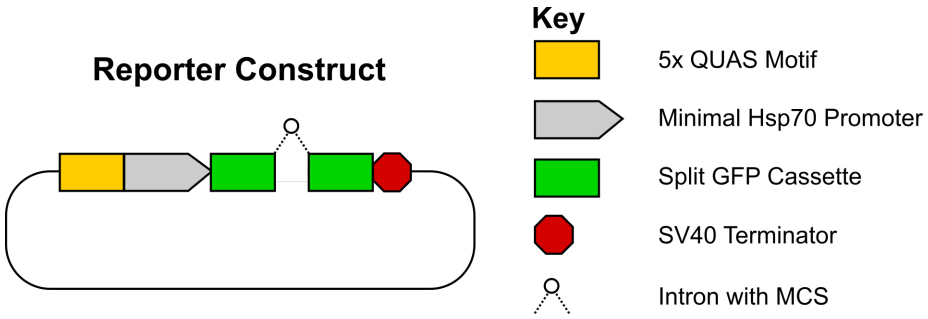


FIGURE 4.3: **Reporter Construct Schematic**. The reporter construct consists of a 5xQUAS motif, which regulates expression of a minimal Hsp70 promoter. Downstream of the promoter is a split eGFP cassette, containing the intron of interest, adjacent to an SV40 terminator sequence. Both introns have been modified to include MCS, suitable for cloning of igRNA cassettes.

Cloning gRNA Spacers into pCFD3-dU6:3-gRNA Expression Vector

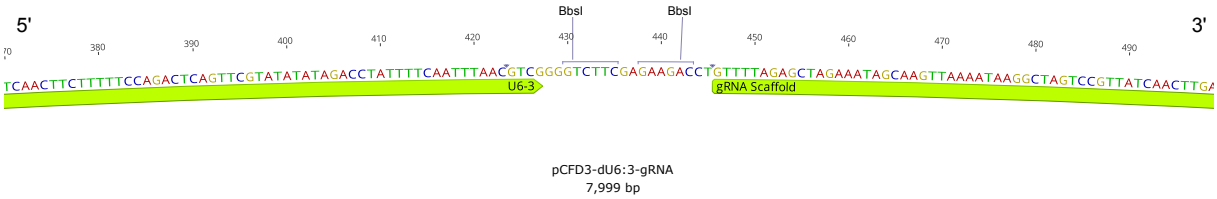


FIGURE 4.4: **pCFD3-dU6:3-gRNA gRNA Spacer Cloning Site**. A BbsI spacer sequence is present in the unmodified pCFD3-dU6:3-gRNA backbone. Digestion with BbsI yields 4bp overhangs, compatible with annealed gRNA oligonucleotides, facilitating direct insertion.

To generate U6:3-gRNA-T6 gRNA cassettes to be integrated into introns, spacers were synthesised as two, complementary ssDNA oligos (Table.4.1) (Eurofins Genomics, Ebersberg, Germany), with 4bp overhangs, compatible with a BbsI restriction site in the pCFD3-dU6:3-gRNA expression vector(Fig. 4.4). The 5′ overhang was 5′-GTCG-3′, whilst the invariant 3′ overhang was 5′-AAAC-3′. Three sgRNAs were chosen, having been previously experimentally validated against the QUAS motif (Lin et al., 2015). The pCFD3-dU6:3-gRNA plasmid was a gift from Simon Bullock (Addgene plasmid #49410) (Port et al., 2014), and was digested as previously described using BbsI, and gel purified.

Spacer ID	Sense Strand	Antisense Strand
gRNA 1	GTCG CTCGGGTAATCGCTTATCCT	AAAC GAGAGCTCCCAACGCGTTGG
gRNA 2	GTCG CGGATAAACAATTATCCTCA	AAAC TGAGGATAATTGTTTATCCG
gRNA 3	GTCG CCAACGCGTTGGGAGCTCTC	AAAC GAGAGCTCCCAACGCGTTGG

TABLE 4.1: **Intronic gRNA Spacers**

The ssDNA oligos were annealed and phosphorylated in Applied Biosystems Veriti 96-well Thermal Cyclers (Applied Biosystems, CA, USA) as follows: 37°C for 30 minutes, 95°C for 5 minutes, ramp down to 25°C at 5°C/minute. Reactions were setup in 10 µl as described in Table. 4.2.

Reagent	Volume (µl)	Final Concentration
T4 PNK	0.5	0.5 units
Sense oligo	1	10 µM
Anti-sense oligo	1	10 µM
RNase-free water	6.5	N/A
10x T4 ligation buffer	1	1x

TABLE 4.2: **Guide RNA Ligation Reaction Reagents**

A subsequent ligation reaction was performed at room temperature for 30 minutes, using the annealed and phosphorylated dsDNA oligos. Reactions were setup in 15 µl as described in Table. 4.3.

Reagent	**Volume (µl)**	**Final Concentration**
Digested pCFD3	x	3.3 ng/µl
Annealed oligonucleotides	1	1:200
10x T4 ligation buffer	1.5	1x
RNase-free water	x	N/A

TABLE 4.3: **Guide RNA Vector Assembly Conditions**

The resulting ligation mix was then transformed as previously described, using ampicillin agar plates. Single colonies were isolated, and pepared by mini-prep to produce purified plasmid stocks of the pCFD3-dU6:3gRNA vector, expressing each of the three experimental sgRNAs. These constructs were used as PCR templates to produce three variants of each gRNA sequence, for integration into the *ftz* and minimal introns. PCR reactions were conducted in 20 µl reactions as follows; 1 µl forward primer, 1 µl reverse primer, 7 µl RNase-free water, 10 µl 2x Phusion mastermix (NEB UK, Hertfordshire, UK), 1 µl DNA template (0.5-10 ng/µl). PCR conditions were performed as follows; 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, 55-72°C for 20 seconds, 72°C for 30 seconds per kb, with a final 7 minute extension step at 72°C, and a 4°C hold. Primers used for these PCR amplification reactions are listed in Table. 4.4.

Primer ID	**Description**	**Sequence**	**Forward/Reverse**
		Minimal igRNA Variant	
pC1	gRNA1 Minimal	TTATGGCTAGCCTCGGGTAATCGCTTATCCTCGGG	Forward
pC2	gRNA1 and 3 Minimal	CATAATTCGAAGCACCGACTCGGTGCCAC	Reverse
pC3	gRNA 2 Minimal	TTATGGCTAGCCGGATAAACAATTATCCTCACGGGTTTTAG	Forward
pC4	gRNA 2 Minimal	CATAATTCGAAGCACCGACTCGGTGCC	Reverse
pC5	gRNA 3 Minimal	TTATGGCTAGCCCAACGCGTTGGGAGCTCTC	Forward
		Full igRNA Variant	
pC11	gRNA Full (All)	TTATGGCTAGCTTTTTTGCTCACCTGTGATTGCTCC	Forward
pC12	gRNA Full (All)	CATAATTCGAAAAAAAAGCACCGACTCGGTGC	Reverse
		Poly-T igRNA Variant	
pC13	pQUAST	GGCACCGAGTCGGTGCTTTTTTCTTTGCTTATGCTTACAGAGCG	Forward
pC14	pQUAST	AGCAATCACAGGTGAGCAAAAAATGGGTTAGCTAATGAGTTTTATAACTGC	Reverse
pC15	gRNA Poly-T (All)	GCAGTTATAAAACTCATTAGCTAACCCATTTTTTGCTCACCTGTGATTGCT	Forward
pC16	gRNA Poly-T (All)	CGCTCTGTAAGCATAAGCAAAGAAAAAAGCACCGACTCGGTGCC	Reverse

TABLE 4.4: Primers are listed describing the igRNA cassette they were used to amplify. In the case of the Poly-T igRNA, an additional pair of primers is listed, that were used to linearise the reporter backbone via PCR, ahead of insertion.

Insertion of igRNA cassettes into the *ftz* intron occurred by two means - restriction ligation for Full and Minimal variants, and Gibson assembly for Poly-T variants. Restriction ligation was performed using NheI and BstBI on igRNA cassettes, and the reporter construct backbone. Fragments were purified and ligated as previously described. Colony PCR was performed on transformant colonies, and ligations verified by Sanger sequencing. Gibson assembly was performed to insert Poly-T igRNA cassettes using a mastermix that was produced in accordance with the protocol listed by OpenWetWare, with all components provided by NEB (New England Biolabs, Hitchin, UK). The general protocol for Gibson assembly is given here -

1. Take purified DNA fragments and thaw
2. Thaw a single 15 µl mastermix per reaction on ice
3. Assemble DNA fragments in 5 µl total volume, diluted with RNAse-free water where required. DNA should be present in equimolar quantities, and in a 3:1 ratio relative to the backbone (if one is present). Use 75 ng per fragment
4. Add DNA mix, and incubate reactions at 50°C for 60 minutes
5. Take 2 µl direct from the reaction, and 2 µl of a 1:100 dilution in water, and transform each using an appropriate antibiotic

Transformant colonies were screened via colony PCR using a corresponding primer pair (Table. 4.5), and successful insertion confirmed by sanger sequencing (Table. 4.4).

Primer ID	**Sequence**	**Forward/Reverse**
pC19	CTCGGGTAATCGCTTATCCTCGGG	Forward
pC20	AAAAAAGCACCGACTCGGTGC	Reverse

TABLE 4.5: Reporter Construct - igRNA Cassette Colony PCR Primers (*ftz* Intron

In order to replace the existing *ftz* intron with the minimal mammalian intron derived from Kiani et al., 2014, the reporter backbone was linearised, with the targeted excision of the intron from within the eGFP reporter. Full sequences for the *ftz* and Minimal introns are given in A.3. Two complementary ssDNA oligonucleotides corresponding to the minimal intron were synthesised (Eurofins Genomics, Ebersberg, Germany), and annealed. These oligonucleotides were synthesised with 15bp overhangs, complementary to the 5′ and 3′ segments of the eGFP reporter. Fragments were assembled in a 2-part Gibson reaction, and as before transformants were screened by colony PCR, and insertion confirmed by Sanger sequencing (Table. 4.6).

Primer ID	**Sequence**	**Forward/Reverse**
pC45	CTGCACCACCGGCAAGC	Forward
pC46	TTGTCGGGCAGCAGCAC	Reverse

TABLE 4.6: Reporter Construct - Minimal Intron Colony PCR Primers

The cloning process was repeated for the minimal intron using the primers described here (Table. 4.7). Minimal igRNA cassettes were amplified using the same primers (pC1-5) as were used for insertion in the *ftz* intron. They were inserted by the same digestion and ligation procedure as described above. Full and Poly-T igRNAs were inserted by Gibson assembly, in 2-part reactions. In each case the vector backbone, and igRNA cassette were amplified as linear fragments, gel purified and assembled. As before, transformants were screened by colony PCR, and insertion confirmed by Sanger sequencing using the same primers as used for the *ftz* intron (Table.4.4).

Primer ID	Description	Sequence	Forward/Reverse
		Full igRNA Cassette - Minimal Intron	
G46	igRNA Cassette	agggtaagtggtccgctagcTTTTTTGCTCACCTGTGATTGC	Forward
G47	igRNA Cassette	gactcgagagttagtattcgACAAAAAAGCACCGACTCGG	Reverse
G48	Vector Backbone	aatcacaggtgagcaaaaaaGCTAGCGGACCACTTACC	Reverse
G49	Vector Backbone	ccgagtcggtgctttttttgtCGAATACTAACTCTCGAGTCTTC	Forward
		Poly-T igRNA Cassette - Minimal Intron	
G42	igRNA Cassette	ttcgaatactaactctcgagtcttcTTTTTTGCTCACCTGTGATTGCTCC	Forward
G43	igRNA Cassette	tggtgcgctctgtgaaaaaaaaaaaGCACCGACTCGGTGCCAC	Reverse
G44	Vector Backbone	ggagcaatcacaggtgagcaaaaaaGAAGACTCGAGAGTTAGTATTCG	Reverse
G45	Vector Backbone	tgaaaaagtggcaccgagtcggtgcTTTTTTTTTTTCACAGAGCG	Forward

TABLE 4.7: **igRNA Cassette Assembly Primers - *ftz* Intron**. In each case primers are listed for amplification of igRNA cassettes from the pCFD3 expression plasmid, and linearisation of the reporter construct, with resulting 20-25bp overhangs required for Gibson assembly.

The resulting cloning produced nine constructs for each intron, producing eighteen reporter construct variants in total (Fig. 4.5).

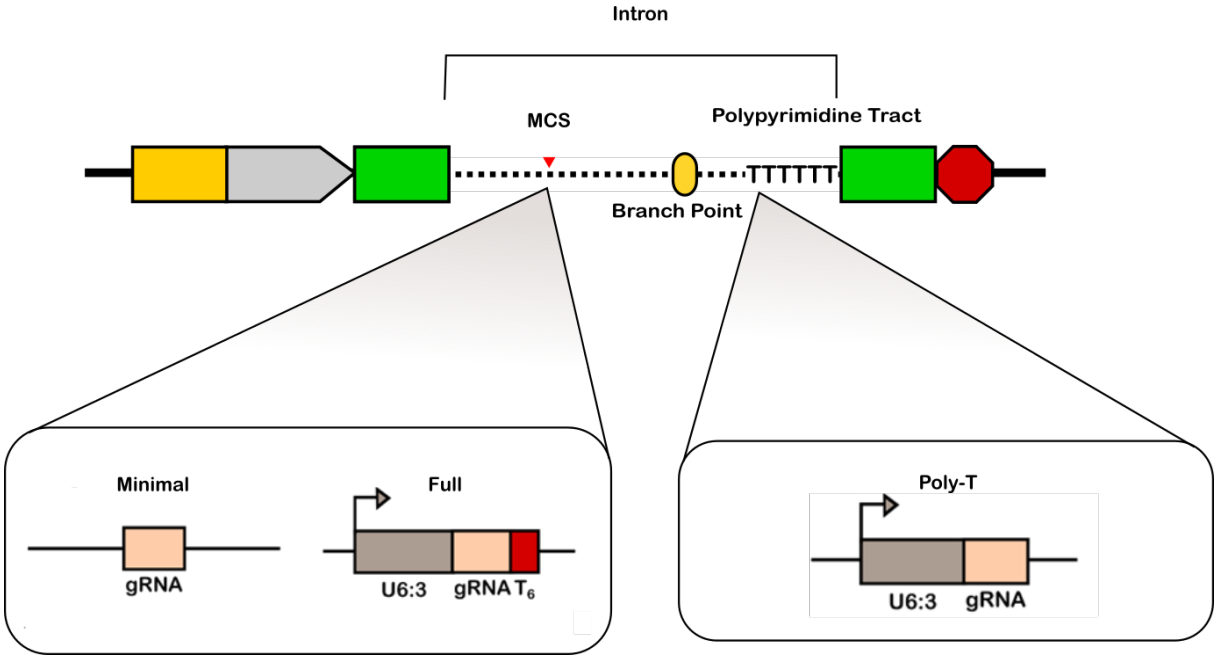


FIGURE 4.5: **igRNA Variant Schematic**. As previously described, each igRNA cassette was cloned into either the *ftz* or minimal intron. Full and Minimal igRNA cassettes were inserted into an MCS upstream of the intron branchpoint. Poly-T igRNA cassettes were inserted adjacent to the beginning of the polypyrimidine tract in each intron.

4.3.2 Construction of Control Constructs

The constructs acting as negative experimental controls were derived from pAct:dCas9-VP64 (Addgene #78901), and pAct:dCas9-VPR (Addgene plasmid #78898) - both were gifts from Norbert Perrimon (Chavez et al., 2016). Each construct expresses a dCas9 protein, fused via glycine linkers to an activation domain - either VP64, or VPR. To produce the negative control (hereafter referred to as the dCas9 control), pAct:dCas9-VP64, and pAct:dCas9-VPR were amplified

in two parts by PCR, with primers conferring 15-30bp of homology to an mCherry reporter cassette, creating part of a 5'-T2A sequence and a 3'-MCS. An mCherry cassette was amplified from pAct:mCher-mTurq-Puromycin, a gift from Manjunatha Kogenaru, with primers conferring corresponding 15-30bp homology regions, and completing the T2A sequence. All parts were gel purified, and assembled in two independent 3-part Gibson assembly reactions, yielding the VP64, and VPR dCas9 control constructs (Fig. 4.6).

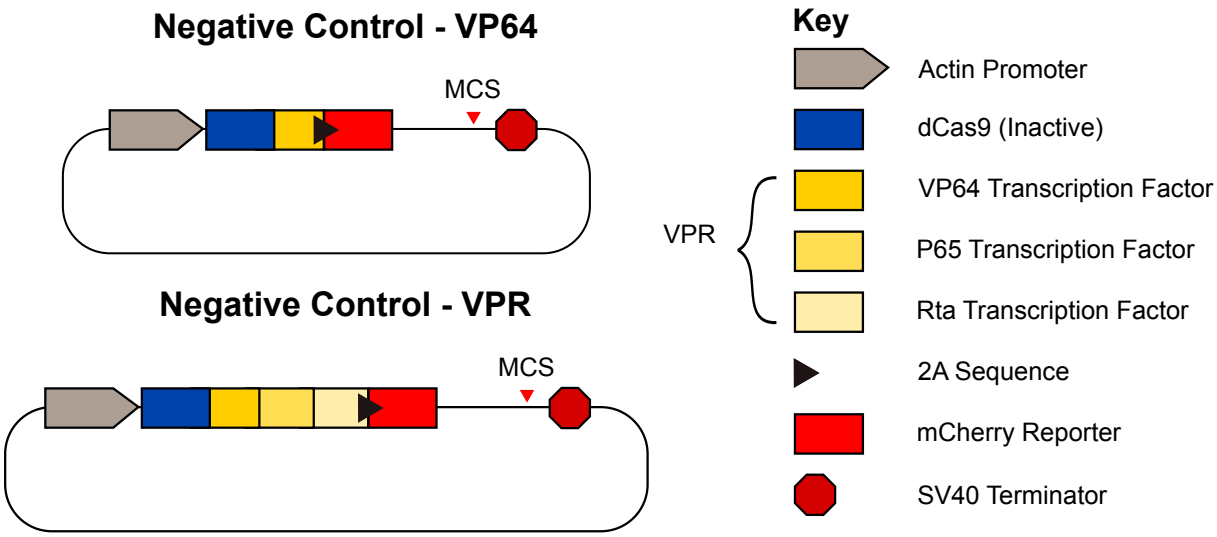


FIGURE 4.6: **dCas9 Control Constructs**. An mCherry reporter, with 5'-T2A sequence was inserted by Gibson assembly, downstream of a dCas9 cassette, fused to either a VP64, or VPR activation domain. An additional MCS was inserted downstream of the mCherry reporter.

In order to produce positive experimental controls (hereafter referred to as activation controls), the dCas9 control plasmids were digested with KpnI and SalI, and purified as previously described. A U6:3-gRNA-T6 gRNA cassette, expressing gRNA 1 was amplified from the pCFD3 expression construct. The primers (Table.4.8) used conferred 5' and 3' KpnI and SalI restriction sites. Following restriction, and gel purification, igRNA cassettes had sticky overhangs corresponding to the cut site in the dCas9 control vector, permitting ligation (Fig.4.7). Transformant colonies were screened via colony PCR.

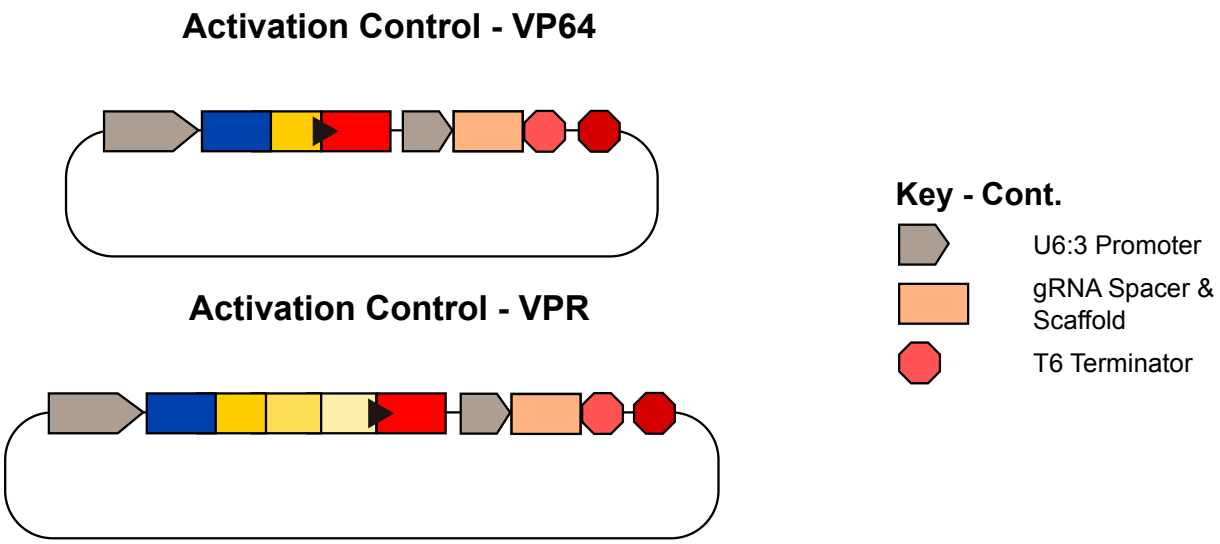


FIGURE 4.7: **Activation Control Constructs**. A U6:3-gRNA-T6 cassette, containing gRNA1 was inserted by restriction and ligation, into the dCas9 control vectors. The result are constructs that constitutively express dCas9, fused to an activation domain, and a constitutive source of gRNA.

Primer ID	Sequence	Forward/Reverse
pC27	TGACCGGTACCTTTTTTGCTCACCTGTGATTGCTCC	Forward
pC28	CTTAGGTCGACAAAAAAGCACCGACTCGGTGC	Reverse

TABLE 4.8: **Activation Control - gRNA PCR Primers**

4.3.3 Compensation Control Constructs

Compensation constructs were required to generate a compensation matrix, to account for spectral overlap between the mCherry and eGFP fluorescent proteins used for two-colour flow cytometry analysis. As with the activation control constructs, an eGFP reporter gene was inserted into pAct:dCas9-VP64, with a 5′ T2A sequence. An identical cloning scheme was used, with primers specific for eGFP (Fig. 4.8). To ensure commensurate levels of mCherry expression, the VP64 activation control construct was used, ensuring that both reporters were expressed from the same genetic background.

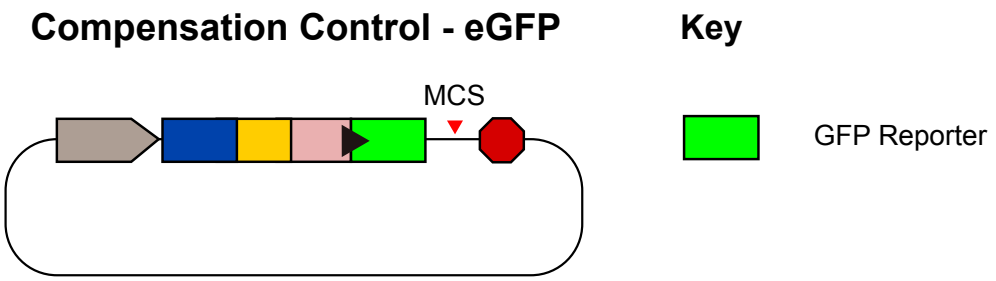


FIGURE 4.8: **Compensation Control Construct - eGFP**. An eGFP reporter with a 5′-T2A sequence was inserted by Gibson assembly into the pAct:dCas9-VP64 backbone.

Primer ID	Description	Sequence	Forward/Reverse
g11	Colony PCR	TTTGAGTTCTTGTGCTGTG	Forward
pC67	Colony PCR	GCATGCCCCAAGTGAATATC	Reverse
		dCas9 Control - Knockout	
g16	Vector Linearisation	CGGTAGTGATATGAAAACTG	Reverse
pC32	Vector Linearisation	CCTACTGACTTGCGGAGATGTCGAAGAGAAC CCTGGCCCTATGGTGTCCAAGGGCGAAGAGG	Forward

TABLE 4.9: **dCas9 Knockout Constructs - Primers**. Primers are listed for colony screening of the dCas9-Mutant construct, and vector backbone linearisation to produce the dCas9-Knockout construct.

4.3.4 Additional Controls

Additional constructs were generated by restriction digestion of the pAct:dCas9-VP64-mCherry dCas9 control. Targeting of the PAM recognition lobe of dCas9 produced a partial knockout, that lacked the ability to successfully recognise the gRNA target. A second construct was designed that lacked the dCas9 cassette, and VP64 domain entirely. The first construct (Fig. 4.9) was produced by digesting the pAct:dCas9-VP64-mCherry dCas9 control construct with BstXI and SacI, and ligating in a small spacer sequence, to ensure the downstream mCherry reporter remained in-frame. The spacer sequence was synthesised as ssDNA oligonucleotides (Eurofins Genomics, Ebersberg, Germany), and annealed as previously described. The resulting dsDNA spacer was ligated into the cut-site, and the resulting construct was transformed, and screened by colony PCR (Table. 4.9). The resulting construct is hereafter referred to as dCas9-Mutant.

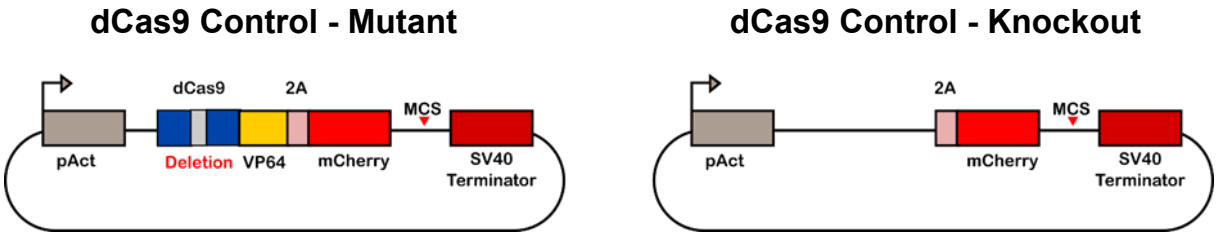


FIGURE 4.9: **dCas9 Knockout Constructs**. The pAct:dCas9-VP64-mCherry dCas9 control was used to produce two construct, lacking the PAM site recognition lobe, and the dCas9-VP64 cassette respectively.

The second construct was produced by linearsing the pAct:dCas9-VP64-mCherry dCas9 control construct via PCR (Table.4.9), and in so doing excising the dCas9-VP64 cassette. The resultant construct was ligated, transformed, and transformants verified by gel electrophoresis to validate the size of the resulting construct, hereafter referred to as dCas9-Knockout (Fig. 4.9).

4.3.5 Cultivating *Drosophila* S2 Cell Culture

Maintenance of Cells

Drosophila S2 cells (Invitrogen, Paisley, UK) were utilised to interrogate intronic gRNA expression. Cells were incubated at 25°C in a Binder BD 56 incubator (Binder, Tuttlingen, Germany) with an ambient atmosphere. All handling of cells was performed in sterile conditions in a

laminar flow hood. Cells were maintained in complete growth medium consisting of 90% Schneider's Drosophila Medium (Invitrogen, Paisley, UK) and 10% fetal bovine serum (FBS) (Sigma-Aldrich, Dorset, UK). No antibiotics were used in maintaining this cell line. Cells were cultured in T25 flasks with 0.2 µM filter caps. Cultures were monitored daily using a Nikon TMS inverted microscope (Nikon UK Ltd, Kingston Upon Thames, UK) equipped with a 10x confocal lens.

Passaging Cells

Cells were passaged biweekly, at which point they had achieved a density of $\geq$1x106 cells/ml - cell count was quantified using a Scepter 2.0 Cell Counter (Millipore, Watford UK). Prior to passage, cells were agitated with a cell scarper to displace any that had become adherent, and 4ml of room temperature growth medium was transferred to a sterile T25 flask. A 1ml volume of each resuspended culture was transferred by pipetting to this sterile T25 flask. Cultures were then replaced in the incubator, and excess cells disposed of in 1% virkon.

Cryopreservation of Cells

Cells were cryopreserved throughout this project. Prior to cryopreservation, cells were permitted to reach a density of $\geq$1.5x106 cells/ml. Cultures were resuspended as described above, and cell count quantified. Cells were then pipetted into a 15ml falcon tube, and spun at 500xg for 5 minutes. The supernatant was discarded and the cells were resuspended in 1ml of complete growth medium supplemented with 5% sterile Dimethyl sulfoxide (DMSO) (Sigma-Aldrich, Dorset, UK) to a final density of 2x106 cells/ml. Aliquots were transferred to cryovials, and placed at -80°C in a Mr Frosty freezing container (Invitrogen, Paisley, UK). After 24 hours, they were transferred to liquid nitrogen for storage.

4.3.6 Thawing Cells

To seed initial cultures, aliquots of frozen cells were thawed from storage in liquid nitrogen. Aliquots were wiped with 70% ethanol and transferred to a laminar flow hood. They were then thawed at room temperature, by careful application of complete growth medium to the vial using a pipette. As the vial thawed, the contents were transferred to a T25 flask, which was supplemented with 4ml of complete growth medium. The flask was then transferred to the incubator and maintained as described above.

Mycoplasma Contamination Testing

In order to detect and prevent mycoplasma contamination, regular testing was performed via PCR. Both the primers and protocol were adapted from Uphoff and Drexler, 2002, Uphoff and Drexler, 2011, and Praetorius, 2016, with additional help kindly provided by Dr. Eduardo Alves dos Santos. Primers (Table.4.10) were prepared to a final concentration 100 µm in RNase-free water. Primers were then prepared in a 100 µl working stock, consisting of 10 µl of each primer, and 40 µl RNase-free water.

Primer ID	Sequence	Forward/Reverse
Myco-5-1	CGCCTGAGTAGTACGTTCGC	Forward
Myco-5-2	CGCCTGAGTAGTACGTACGC	Forward
Myco-5-3	TGCCTGAGTAGTACATTCGC	Forward
Myco-5-4	TGCCTGGGTAGTACATTCGC	Forward
Myco-5-5	CGCCTGGGTAGTACATTCGC	Forward
Myco-5-6	CGCCTGAGTAGTATGCTCGC	Forward
Myco-3-1	GCGGTGTGTACAAGACCCGA	Reverse
Myco-3-2	GCGGTGTGTACAAAACCCGA	Reverse
Myco-3-3	GCGGTGTGTACAAACCCCGA	Reverse

TABLE 4.10: **Mycoplasma Screening PCR Primers**

Prospective cell culture samples to be tested had 100 µl taken from them, and incubated for 5 minutes at 95°C, before being centrifuged at >13000 x*g* for 2 minutes. PCR reactions were subsequently set up as follows (Table.4.11). PCR reactions were run with the following conditions; 95°C for 2 minutes, followed by 5 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds, followed by a further 30 cycles of 94°C for 15 seconds, 56°C for 15 seconds and 72°C for 30 seconds, before a 4°C hold. A positive control (a gift from Dr. Alves dos Santos), and negative control were included in each gel. PCR products were visualised on 1% agarose gel, run at 100 V for 60 minutes. An 500bp band corresponded to contamination. No contamination was observed in experimental samples.

Component	Volume (µl)	Final Concentration
RNase-free water	5	N/A
FastCycle 2x mastermix	10	1x
Forward primers	1	0.5 µM
Reverse primers	1	0.5 µM
Sample	2	N/A

TABLE 4.11: **Mycoplasma Screening PCR Setup**

4.4 Transient Transfection of *Drosophila* S2 Cell Culture

Constructs used for experimental work in S2 cells were prepared for transfection by midi-prep using E.coli TOP10 cells. Following transformation, individual colonies were selected and used to inoculate 5ml LB-Miller starter cultures, supplemented with ampicillin. Starter cultures were incubated for 8 hours at 37°C in a shaking incubator at 220rpm. They were then used to inoculate 500ml falcon flasks containing 100ml of LB-Miller broth, supplemented with ampicillin. Falcon flasks were incubated overnight at 37°C in a shaking incubator, at 220rpm. Liquid cultures were then harvested and purified by midiprep.

S2 cells were passaged every 24 hours, for 72 hours prior to transfection. Cells were subsequently resuspended as previously described, and cell count was quantified. Cells were

aliquotted from T25 flasks directly into 24 well plates such that each plate was seeded with 3.5x104 cells/well. Each well was supplemented with complete growth medium to 700µl. Transfection was then immediately performed using Effectene, as per the manufacturer's instructions, using 0.2µg of each construct per well. Transfected cells were incubated for 48 hours prior to flow cytometry analysis.

4.5 Flow Cytometry Conditions

Fluorescence in S2 cells was analysed by a three-laser BD LSR Fortessa analyser (BD Biosciences, Oxford, UK). For each condition analysed, at least 500 gated events were recorded, with at least three replicates per condition. The following voltages were used – Forward Scatter (FSC): 170V, Side Scatter (SSC): 253V, 530nm channel: 390V, and 610nm channel: 591V, Excitation of eGFP was performed using a 488nm blue laser, using a 530/30 bandpass filter. Excitation of mCherry was performed using a 561nm yellow laser, using a 610/20 bandpass filter.

4.6 Results

4.6.1 Constructing a Novel igRNA Test Platform

In order to interrogate candidate sgRNA cassettes and introns, a two component trans-activation system was designed. The system makes use of elements of the Q-system (Potter et al., 2010); a binary transactivation system analogous to the well characterised GAL4-UAS (Brand and Perrimon, 1993) and LexA/*LexAop* systems (Lai and Lee, 2006). As with the aforementioned GAL4-UAS, one component consists of a chosen promoter expressing a transcription factor, and the other component is regulated by a promoter induced by the same transcription factor. Binary expression systems offer a number of advantages for expressing a transgene of interest, when compared to direct expression from a promoter at a single locus. Aside from enhanced expression as a consequence of transcription factor-mediated activity, such systems allow for the expression of transgenes that may otherwise be lethal to the organism if expressed directly, preventing the creation of transgenic individuals for experimentation (Potter et al., 2010). By using a binary system, it is possible to express the transgene only when the transcription factor is introduced to the transgenic individual, typically through crossing of strains. Transcription factors can themselves be subjected to regulation, permitting temporal and spatial control of transgene expression. Using this methodology, it is also possible to generate large libraries of transgenic organisms carrying different transcription factors, and target transgenes, allowing for systematic investigation of the activity of transgenes in different spatiotemporal patterns.

The Q system is comprised of regulatory genes derived from the bread mold, *Neurospora crasas*. The genes in question are a cluster involved in quinic acid catabolism, and include two regulatory genes - *QA-1F* and *QA-1S*, which are a transcriptional activator and suppressor, respectively. *QA-1F*, commonly referred to as *QF*, binds a conserved 16bp site within the gene cluster, and autoregulates the cluster by also controlling expression of the suppressor (Baum,

Geever, and Giles, 1987), *QA-1S*, commonly known as *QS*. In the Q system, this 16bp binding site has been arranged in a 5-copy array, termed QUAS. When assayed in *Drosophila* Schneider 2 and human HeLa cell culture, the Q system was able to elicit luminesence levels comparable with that of the GAL4-UAS system, and significantly greater degrees of suppression (Potter et al., 2010). When originally designed, the QF transcriptional activator was found to have toxic properties, though the underlying cause was not identified (Riabinina and Potter, 2016). To this end, a second generation variant was engineered by codon optimising the *QF* gene for codons with a moderate level of expression in *Drosophila* (Riabinina et al., 2015), yielding QF2. The Q system can be used intersectionally with the GAL4-UAS system, as no cross-activation or suppression occurs (Potter et al., 2010). Moreover, it offers a number of advantages over existing binary activation systems that have been characterised for use in *Drosophila*; 1) the LexA/*LexAop* system lacks its own suppressor, and cannot be independently repressed when used in conjunction with Gal4-UAS, 2) The GAL4-UAS suppressor is temperature dependent. The Q system has a powerful suppressor, and is robust to ambient temperature (Potter et al., 2010;Riabinina and Potter, 2016). Here, the Q system has been selected due to several practical considerations, namely the availability of experimentally validated gRNAs for use against the *QUAS* motif for the purpose of transactivation with a CRISPRa system.

Previous work conducted with the QUAS system has made use of a catalytically inactive dCas9 endonuclease, fused to trans-activation domains in place of the QF activation molecule (Lin et al., 2015). The dCas9 endonuclease differs from canonical SpCas9, in that it contains two mutations (D10A and H841A) that remove the cleavage activity mediated by the RuvC1 and HNH endonuclease domains (Jinek et al., 2012). Earliest attempts to achieve dCas9-mediated transactivation made use of dCas9 fused to known transactivators such as VP64, which consists of four fused copies of the VP16 transactivation domain (Gilbert et al., 2014). In order to improve the efficiency of activation when using dCas9, a novel chimeric activator was additionally devised, consisting of VP64, p65 and Rta transactivation domains (Chavez et al., 2015). Both dCas9-VP64, and dCas9-VPR were utilised here to investigate the efficacy of igRNA expression cassettes.

An analog of the Q system was initially designed, consisting of two constructs - a reporter construct, and an activation construct. The reporter was derived from the pQUAST plasmid (Potter et al., 2010), itself a derivative of the original pUAST construct designed for use as part of the GAL4-UAS system. The construct contains a 5xQUAS motif, upstream of a minimal Hsp70 promoter. Downstream of the promoter is a multiple cloning site (MCS), and an SV40 termination signal. An eGFP reporter gene was cloned into the MCS, containing either the *ftz* or minimal intron, which in turn contained each of the igRNA cassettes to be assayed. Hsp70 has been demonstrated to have significant basal activity in S2 cells (Di Nocera and Dawid, 1983). Normally, this would be problematic and result in leaky expression of a given downstream gene, however here it is essential. As with many naturally occurring and synthetic feedback loops, leaky expression is required to mediate the system through different activation states.

Given that this platform was specifically intended to interrogate igRNA expression, the canonical activation component of the Q system (QF) was replaced with a construct expressing the aforementioned dCas9, fused to either the VP64 or VPR transactivation domains (AD) (Fig.4.10). These activation components each have a downstream mCherry reporter. When a reporter construct is co-transfected with an activation construct, dCas9-AD is expressed and (if igRNA are expressed), forms a dCas9-AD:igRNA complex. There are three gRNA sequences used in this work, which were obtained from previously validated work (Lin et al., 2015). Each sequence was cloned into each of the three igRNA cassettes (Minimal, Full, Poly-T) and when expressed guide the dCas9-AD protein complex to the 5xQUAS motif of the reporter construct, where the transactivation domains upregulate expression of eGFP. Cells are co-transfected with both the activation and reporter constructs, and subpopulations double positive for mCherry and eGFP fluorescence (when compared to an untransfected sample) are then selected. As the introns are located within the eGFP coding sequence, failure to splice will result in low fluorescence output, as will poor igRNA expression. This fluorescence can then be measured as a readout of igRNA and intron efficacy using a flow cytometer. The negative control consisted of the reporter construct containing each intron, but without the presence of an igRNA cassette, co-transfected with each dCas9 construct. In order to construct a positive control, hereafter referred to as the 'activation control', a U6:3-gRNA cassette was cloned into a site downstream of the mCherry reporter, providing a constitutively expressed source of gRNA. This was subsequently cotransfected with reporter constructs lacking an igRNA.

There are three igRNA designs that were tested, which have been termed 'Minimal', 'Full', and 'Poly-T' respectively (Fig.4.10); 1) The Minimal variant consists of a gRNA spacer and gRNA scaffold, 2) The Full variant consists of a gRNA spacer, scaffold, U6:3 promoter to drive expression, and a T6 terminator sequence, 3) The Poly-T variant consists of a gRNA spacer, scaffold, and a U6:3 promoter, but the cassette is cloned into each intron such that the polypyrmidine-rich tract at the 3' end of the intron act as a surrogate terminator sequence (Fig.4.10) . Whilst the Full and Poly-T variants have their own U6:3 promoters to drive strong constutive expression of the gRNA, the Minimal variant relies entirely on the efficient splicing of its host intron for expression - following splicing introns are typically degraded (Talhouarne and Gall, 2018), though there are suggestions that some introns may persist for a time with roles in processing of other RNAs and translation (Hesselberth, 2013). Following splicing, the excised lariat structure of the intron is linearised by a debranching enzyme (Dbr1) (Ooi et al., 2001), and subsquently degraded by cellular exonucleases. It is during this final step that the gRNA and scaffold may be liberated from intronic transcript. It then is able to act as a guide RNA, further increasing levels of eGFP expression and spliced transcript levels. This is largely dependent on the nature of exonuclease activity and the half-life of the intronic transcript following linearisation. Theoretically, if this process occurs successfully, any igRNA cassette with a functional gRNA spacer will be able to induce an amplification loop in this system.

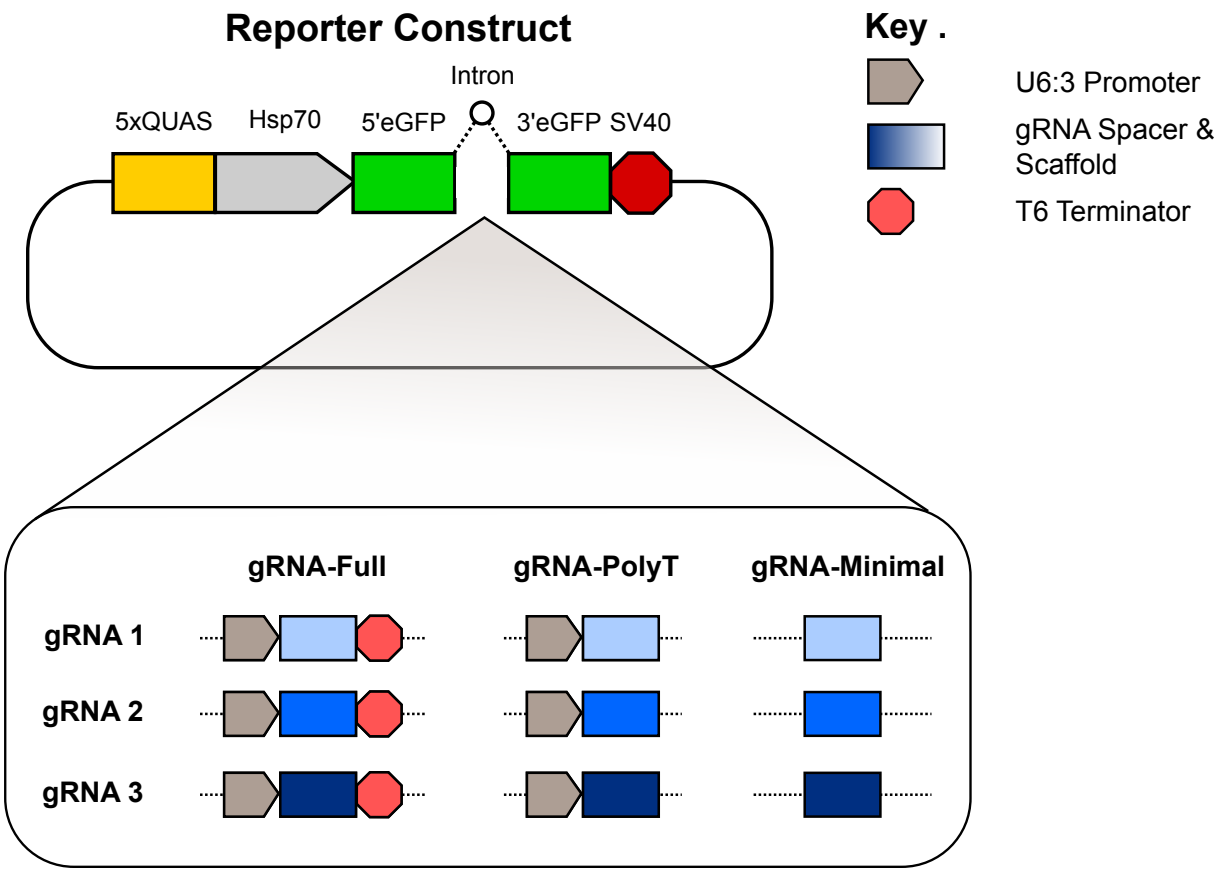


FIGURE 4.10: **Overview of Experimental Reporter Constructs**. Three variants of each gRNA spacer sequence were implemented in the *ftz* and minimal intron, within the reporter construct backbone. Each igRNA cassette contains a gRNA spacer, and scaffold for interfacing with dCas9. The Full igRNA cassettes include a U6:3 promoter, and T6 terminator. The Poly-T igRNA cassettes include a U6:3 promoter, but no terminator. These variants were inserted inside the polypyrimidine tract of each intron, with this thymidine region acting as a T6 terminator sequence.

The Schneider 2 (S2) cell line was used for all transfections. Originally isolated from 20-24 hour-old *D.melanogaster* embryos (Schneider, 1972), this robust cell line can be maintained in the laboratory with minimal care. S2 cells have been shown to be amenable to transfection, with not just one, but multiple plasmids without significant reduction in transfection efficiency (Yang and Reth, 2012). Moreover, they are typically used for heterologous expression of proteins, making them ideal for evaluating ciruits producing large quantities of fluorescent protein. As with all cell culture experiments there are limitations to the degree with which findings can be translated to the original organism.

4.6.2 Flow Cytometry Compensation

When using multiple fluorophores to conduct flow cytometric analysis, there is the possibility that the spectral emissions of one fluorophore overlap with that of another; this phenomenon is known as spillover. A simple methodology to minimize this spillover is to select fluorophores which have suitably distinct excitation and emission spectra, such that fluorescence stemming each can be distinguished. It is nonetheless necessary to apply a compensation matrix, which

can subtract this overlap. To do this, single-colour constructs are required for each fluorophore being measured in the experimental design. Here, a variant of the dCas9-VP64 activation control with an eGFP reporter was constructed. The existing dCas9-VP64 activation control with the existing mCherry reporter was used as the second compensation control (Fig.4.11). S2 cells were transfected separately with each construct, during each experiment, and suitable gating was applied. A compensation matrix was subsequently calculated using FlowJo V7.0, and applied to experimental samples.

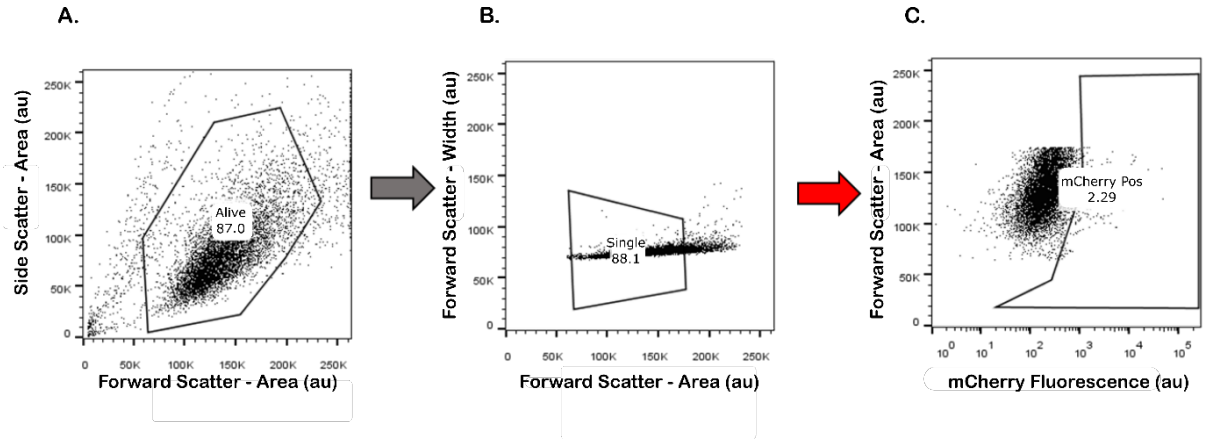


FIGURE 4.11: **Flow Cytometry Compensation Gating**. A) Forward scatter is used as a measure of cell size, whilst side scatter gives an indication of the internal complexity of cells. This is used to determine which cells within the sample being analysed are viable. This subpopulation was gated. B) The resulting subpopulation was further gated to ensure that only single cells were passing being measured for fluorescence, based on size. Where cells aggregate, and multiple cells are measured as single cells, the resulting measurements can provide false positive results. This subpopulation was also gated. C) The subpopulation is gated (against a wild-type sample) for the fluorophore being detected - the example given is the mCherry compensation control. The resulting fluorescence readings are used to construct a compensation matrix.

4.6.3 Establishing the Feasibility of igRNA Expression

Initially, cells were transfected with single constructs in triplicate to ensure all fluorescent proteins were being expressed as designed. It was subsequently possible to begin co-transfections of experimental plasmids against activation components, in order to evaluate the efficacy of different sgRNA cassettes in the *ftz* and minimal intron backgrounds. Cells were gated to isolate those that had been successfully co-transfected with both system components (Fig.4.13). It was possible to then evaluate the fluorescence of each sgRNA cassette, relative to the dCas9 control. Initially the igRNA cassettes were investigated in the *ftz* intron.

ftz cotransfections suggested that it was possible to elicit increased levels of eGFP fluorescence, relative to a dCas9 negative control. The greatest relative change in fluorescence was mediated by the Full-1 igRNA, resulting in a 7.8-fold change using the VP64 activation domain, and a 6.5-fold change using the VPR activation domain, against the dCas9 negative control (Fig.4.12).

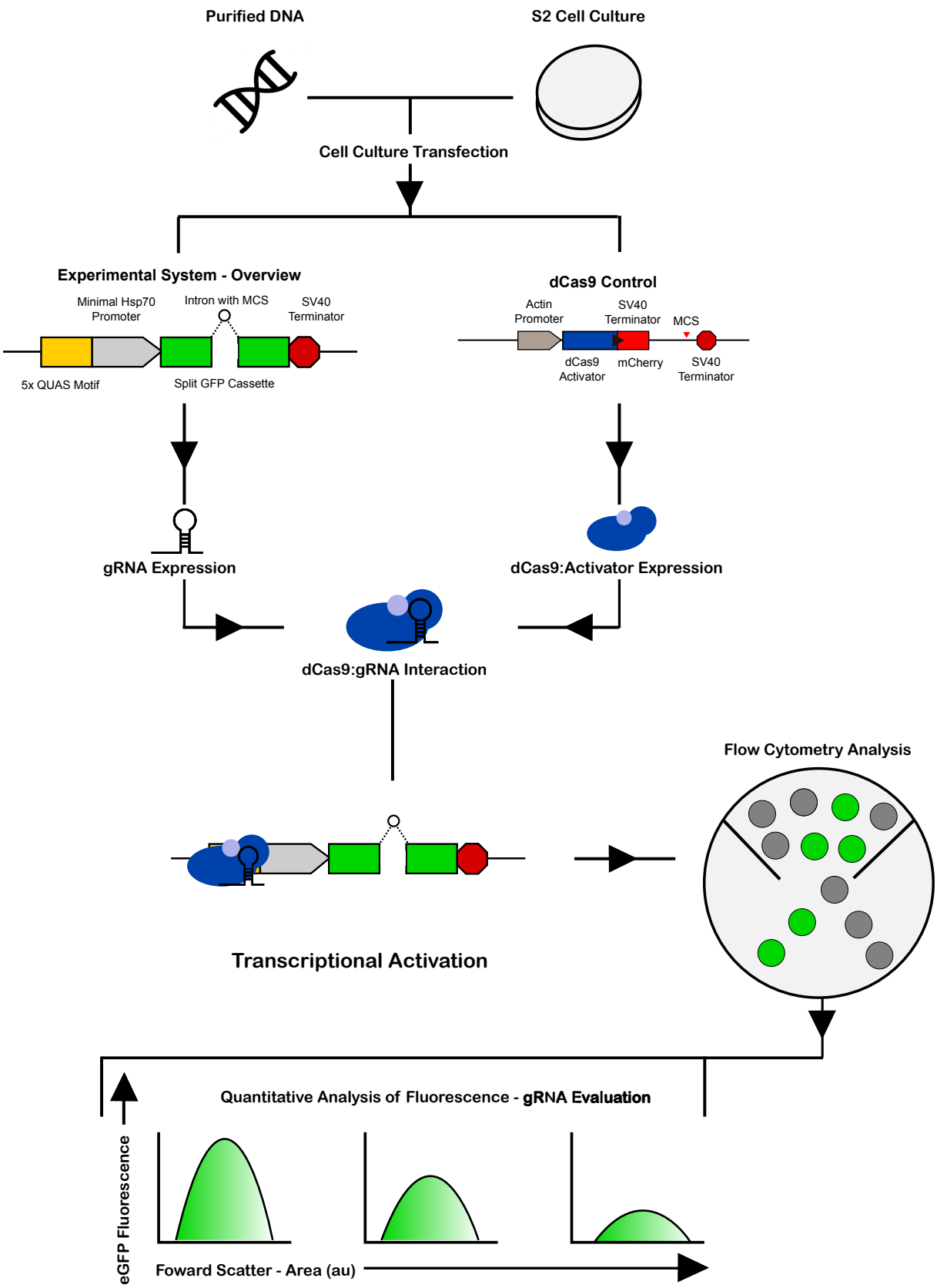


FIGURE 4.13: **Overview of igRNA Platform Workflow**. S2 cells are seeded, and co-transfected with two plasmids - a reporter construct, and the dCas9 control construct. The dCas9 protein, fused to an activation domain is expressed, and assuming an igRNA cassette is successfully expressed, a dCas9-AD:igRNA complex forms. This complex targets a 5xQUAS motif in the reporter construct, upregulating eGFP fluorescence - this may also begin an amplification loop. The resulting eGFP fluorescence can be quantified using flow cytometry, relative to a negative control consisting of the dCas9 control construct, and a reporter construct lacking an igRNA cassette.

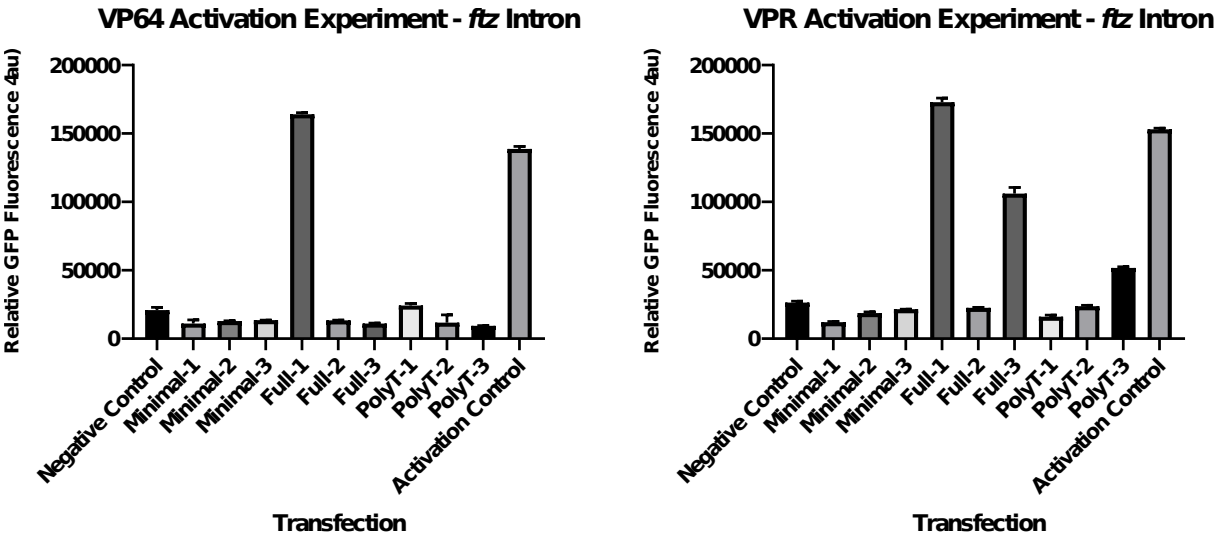


FIGURE 4.12: **Activation Experiments - *ftz* Intron**. Reporter constructs, containing the *ftz* intron were co-transfected with the dCas9-VP4, and dCas9-VPR control constructs. Horizontal bars represent the average of the geometric fluorescence means for three replicates, with SEM.

Subsequently, igRNA cassettes were investigated using the minimal intron (Fig.4.14). Once again, the Full-1 igRNA variant elicited the greatest degree of eGFP expression relative to the dCas9 negative control. Additionally, a number of other variants were able to elicit similarly enhanced levels of fluorescence, when assayed against the dCas9 negative control.

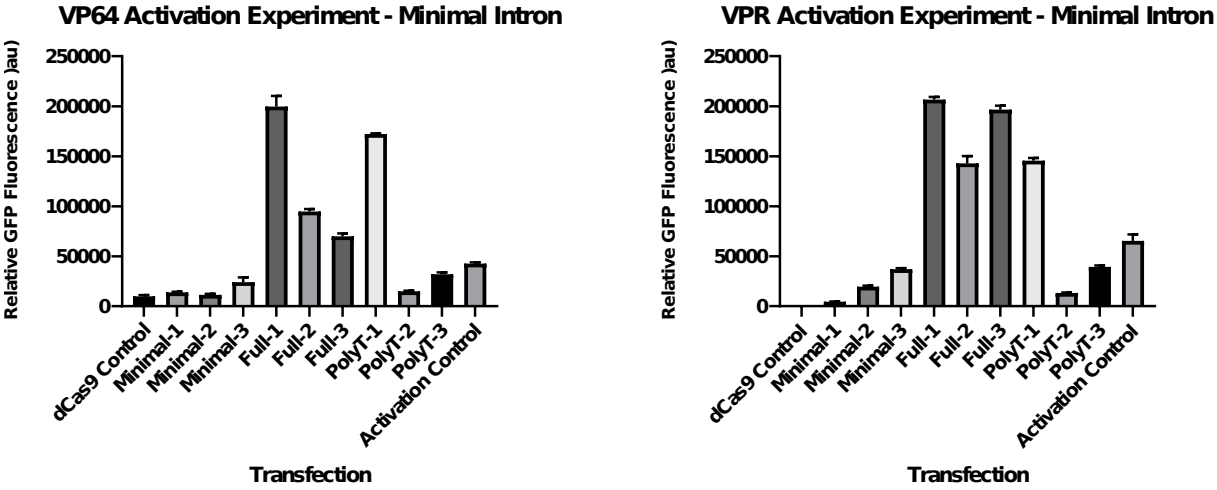


FIGURE 4.14: Reporter constructs, containing the *minimal* intron were co-transfected with the dCas9-VP4, and dCas9-VPR control constructs. Horizontal bars represent the average of the geometric fluorescence means for three replicates, with SEM.

4.6.4 Investigating Inactive igRNA Variants

Having initially established that it was possible to elicit increased levels of eGFP expression (and thus igRNA expression), using both the *ftz* and minimal introns, further screening was performed to investigate why some sgRNA cassettes appeared functional, whilst others were not. It was posited that there were two likely reasons for this observation: 1) the presence of

the sgRNA cassette in question was introducing cryptic splice sites inside the intron that were inhibiting efficient splicing, and 2) poorly performing gRNA cassettes were being expressed at low levels, resulting in low levels of the dCas9-Activator:sgRNA complex from being formed. To investigate these hypotheses, experimental constructs containing the gRNA1-Minimal, and gRNA1-Full cassettes were co-transfected with both the VP64, and VPR activation control constructs. The activation control constructs each constitutively expresses gRNA 1, and evidence from the previous experiment demonstrates that this is sufficient to elicit eGFP expression in the absence of another sgRNA source. Where the co-transfection is unable to produce increased levels of eGFP expression when compared the dCas9 negative control, this suggests that cryptic splicing is occurring. Should the construct elicit increased levels of eGFP expression, the experiment suggests that poor gRNA expression is occurring. These experiments were conducted using constructs engineered in both the *ftz* (Fig.4.15A,B) and minimal introns (Fig.4.15C,D). In all conditions, it was shown that it was possible to elicit positive fold change in eGFP fluorescence when compared to the negative control.

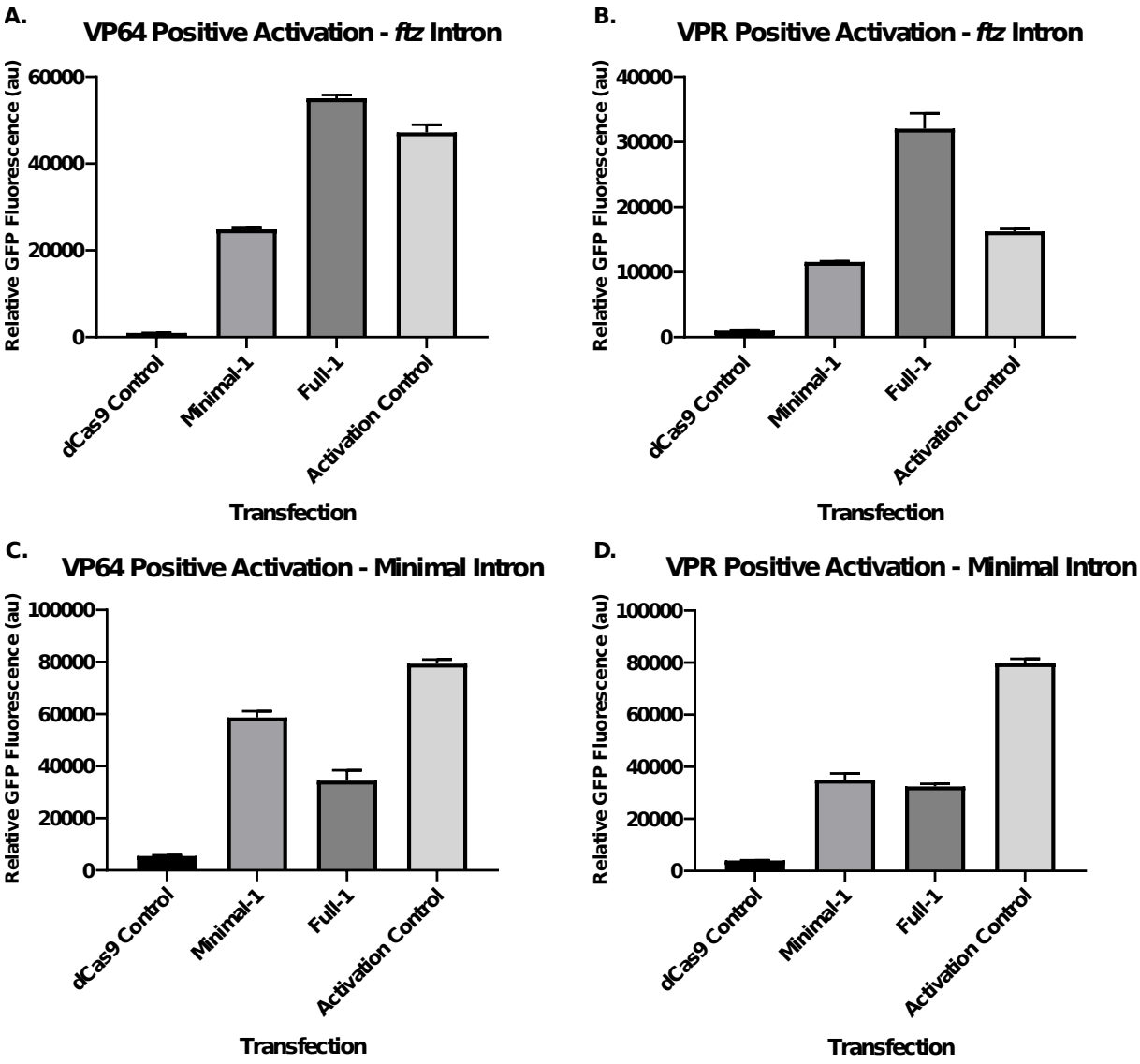


FIGURE 4.15: **Positive Activation Experiment - *ftz* and Minimal Intron.**. Reporter constructs were co-transfected with VP64, and VPR activation constructs. Experiments were performed in triplicate, with the geometric mean of eGFP fluorescence recorded. Horizontal bars represent the average of the geometric fluorescence means for three replicates, with SEM.

4.6.5 Demonstrating Inactivity of dCas9-AD

As an additional control, it was necessary to demonstrate that the activation constructs do not have non-specific activity that results in increased eGFP fluorescence levels. To this end, two additional constructs were generated from the dCas9-VP64 activation construct. The first consisted of a knockout of the entire dCas9 coding sequence, mediated via digestion and re-ligation of the plasmid. The second was derived from the dCas9-VP64 plasmid by restriction to remove the PAM recognition lobe of dCas9, preventing dCas9-VP64 from interacting with the 5-QUAS motif in the reporter construct. In both cases, the downstream elements remain in-frame, ensuring that mCherry expression still occurs. Constructs were co-transfected against

reporter constructs lacking an igRNA, and compared to the VP64 dCas9 negative control. Neither construct was able to elicit increased levels of eGFP fluorescence compared to the dCas9 negative control, in the absence of a gRNA source (Fig.4.16).

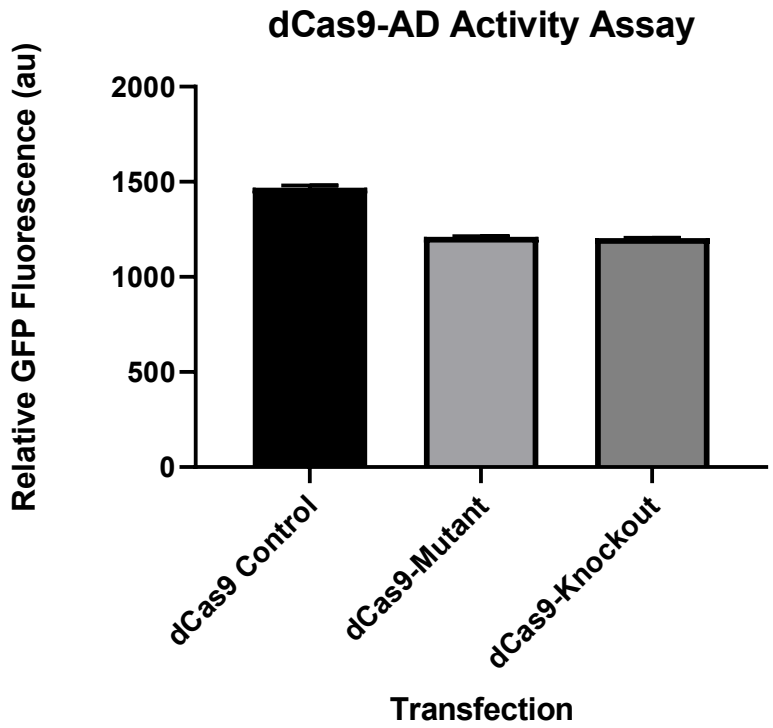


FIGURE 4.16: **dCas9 Activity Assay**. dCas9 knockout constructs were co-transfected with the dCas9-VP64 control construct. Experiments were performed in triplicate, with the geometric mean of eGFP fluorescence recorded. Horizontal bars represent the average of the geometric fluorescence means for three replicates, with SEM.

4.7 Discussion

In this chapter, a pilot study was undertaken in order to assess whether it was feasible to successfully express igRNA cassettes from within the well characterised *D.melanogaster ftz* intron, as well as within a synthetic mammalian intron. This was done as igRNA cassettes are a necessary design feature of the IGD strategy, required to prevent disruption of endogenous coding sequences. In so doing a novel platform was designed, constructed, and tested to allow for the interrogation of any combination of intron and igRNA.

When using both the *ftz* and minimal introns, and with both VP64 and VPR, it was possible to elicit enhanced levels of eGFP fluorescence relative to the relevant dCas9 negative control. It is important to caveat these findings with the observation that experiments were conducted only once in technical triplicate, due to time and cost constraints. Nevertheless, in all cases the Full igRNA variant was able to elicit the greatest fold change in fluorescence, when compared to each respective dCas9 negative control. It is thus the obvious variant to take forward for use in IGD drive and effector components. It is true that the poor performance of Poly-T variants in

the *ftz* intron could be attributable to their position, which result in the generation of context-specific cryptic splice sites that inhibit eGFP fluorescence. An alternative explanation would pertain to disruption of the activity of the spliceosome. Previous work has indicated that the *ftz* intron is able to splice independent of the U1 snRNP, relying solely on recognition of the 3' splice site and cis-acting elements (Crispino et al., 1996). One such element was found proximal to the branch site - given the proximity of the Poly-T igRNA cassette to this region of the intron, it is possible it may disrupt spliceosome interactions, diminishing splice efficacy.

When conducting these initial experiments to determine the most efficient igRNA variant, it was found that in the *ftz* intron, the Minimal igRNA variant was unable to elicit a positive increase in fluorescence. By contrast, in the minimal intron,increased expression was observed. In order to determine whether poor performance in the *ftz* intron was a result of inefficient splicing, or poor igRNA expression, reporter constructs were co-transfected with the positive control construct. It was demonstrated that it was possible to elicit increased levels of eGFP fluorescence in the *ftz* intron with the Minimal igRNA 1 variant. Previously the Minimal 1 igRNA variant had been able to elicit eGFP fluorescence representative of 0.44 and 0.45 fold of the dCas9 negative control, using VP64 and VPR respectively. With rescue provided by the positive control construct, a 26.0 and 11.6 fold increase in fluorescence were observed, relative to the negative dCas9 control, using VP64 and VPR respectively. This suggests that the limiting factor with regards to performance was poor igRNA expression, and not poor splice efficiency. Given that the larger Full igRNA cassette was functional in the initial round of experiments, this is not surprising - the Full variant would have been more likely to introduce cryptic splice sites into the intron.

Concurrent with these experiments, work was being conducted to begin design and construction of IGD drive and effector components, for use in *Drosophila melanogaster*, and by other team members in *Anopheles gambaie*. Whilst the minimal intron appears to be able to mediate higher fold change in fluorescence, it has never been evaluated for use in whole organisms. As such, its expression pattern is unknown, and given that it has been constructed with mammalian splice site motifs, it is unclear whether its performance would be maintained *in vivo*. By contrast, the *ftz* intron is strongly spliced, and homologs have been identified as part of the Hox gene cluster in *An.gambie* (Ahanger et al., 2013). Moreover, the *ftz* intron is known to be robust to changes in size *in vivo*, whereas other naturally occuring *D.melanogaster* introns have shown variability when expanded in *in vitro* vs. *in vivo* comparisons (Talerico and Berget, 1994). To this end, the *ftz* intron represents the most biologically suitable solution to the expression of igRNA cassettes in these invertebrate hosts, and was taken forward for future work.

It has been previously posited that a downstream splicing enhancer is required for the efficient splicing of the *ftz* intron (Yue and Akusjärvi, 1999). However, this work was not conducted in an insect cell line, but instead in mammalian HeLa cells, and so its applicability may be limited. No such considerations were made in this work due to the limited scope of these previously established findings, and there is no indication that this has generated any issues

from a design perspective. Nonetheless, it may be important to consider the presence of cis-acting enhancer sequences required for efficient splicing when designing constructs.

It is important to consider how splice site pairing occurs in *Drosophila melanogaster*. Previous work has demonstrated that for smaller (<200-250bp) introns in *D.melanogaster*, splice site recognition occurs across introns. For larger introns, splice site recognition occurs across exons, though the consequences of this are strongly debated - previous studies have indicated that exonic recognition tends to be prone to slower splicing, and may give rise to alternative splicing (Fox-Walsh et al., 2005). However, a more recent study suggested the contrary, finding longer exon-defined introns to be spliced more accurately and quickly, than shorter intron-defined splice sites (Pai et al., 2017). It is important to consider that the size of the intron is liable to result in changes to splice site recognition mode, and in turn have an effect on splice efficiency when deployed as part of IGD components. To date, analysis conducted in the wake of this work in *An.gambiae* has demonstrated efficient splicing when employing this *ftz* intron inside a number of effector components (Astrid Hoermann, Personal Communication).

An additional consideration for future improvements to the platform pertains to the use of fluorescent standards. Currently, as with most flow cytometry experiments, the experimental design necessitates that fluorescence is quantified in relative terms against a negative control. In the context of this work, this is not a limiting factor, however were the platform to be used to evaluate genetic components in order to build standardised parts for a toolbox, absolute fluorescence readings would allow for proper quantification of activity. This is achievable by using MESF beads (Molecules of Equivalent Soluble Fluorophore), and converting fluorescence output into a unit representative of molecules of fluorescein, a common fluorescent standard. The regulatory components utilised in components of this platform have been designed specifically for use in *D.melanogaster* cell culture, however the generalised schematic could be applied to a variety of organisms. It may be feasible to expand the remit of the platform by exchanging the 5xQUAS-Hsp70 regulatory element for other elements that may be useful as part of CRISPRa strategies, allowing for high-throughput prototyping before deploying systems into whole organisms.

Chapter 5

Results - Implementing Integral Gene Drive in *D.melanogaster*

5.1 Acknowledgements

Many thanks to Tong Zhu for her support in generating the transgenic lines described in this chapter, and for helping the setup and analysis of initial developmental, oviposition and homing assays. Thanks also to Kadris Oras and Dr Alla Madich at the Fly Facility in the Department of Genetics at the University of Cambridge for performing a number of the microinjections described here.

5.2 Chapter Aims

In this chapter, we aimed to construct IGD drive components and subsequently generate transgenic *Drosophila melanogaster* lines. Using these lines, we aimed to characterise relative fitness and fecundity, before investigating the transmission rate of the IGD drive. As per the IGD paradigm, we further investigated a markerless drive component and its own transmission rate. We were trying to investigate the following points -

1. Is it possible to generate IGD lines?
2. Do transgenic lines have significant costs to fitness or fecundity?
3. Do transgenic lines exhibit homing activity?
4. If lines exhibit homing activity, do markerless lines differ in their activity?

5.3 Introduction

Having inferred that it is feasible to express gRNAs intronically in an initial pilot study and having selected the most promising gRNA expression cassette for use, it was possible to proceed to construction and testing of IGD gene drive components in *Drosophila melanogaster*. Historically, a range of germline promoters have been characterised for use in *D. melanogaster* and Anopheline mosquitoes for the specific purpose of regulating gene drive components. In the context of IGD, spatiotemporal control of the drive component is conferred entirely by the

expression pattern of the host gene. A number of germline promoters have already been characterised for use in *D.melanogaster* and the Anopheline mosquito to regulate gene drive constructs, with an extensive review conducted by Chan et al., 2013a, combining a variety of 5' and 3' UTR sequences. Initially, six germline genes previously used as a basis for germline promoters were selected as candidate sites.

The *β2-tubulin* promoter was originally characterised for use in *An.stephensi* (Catteruccia, Benton, and Crisanti, 2005) and used as part of a genetic sexing strategy due to its male-specific expression. It was later tested as part of a sex-distortion strategy in *An.gambiae* (Windbichler, Papathanos, and Crisanti, 2008). The *D.melanogaster* ortholog has been demonstrated to be testes-restricted (Michiels et al., 1989). Given the recent discovery that maternally deposited Cas9 mRNAs result in increased rates of resistance allele formation (Champer et al., 2017), the use of male-specific promoters such as this may reduce this issue. In *D.melanogaster*, use of a *β2-tubulin* promoter derived from the *βTub85D* gene has yielded low cleavage and homing rates, though this represents a truncated version of the full 5'UTR, which may result in changes to expression efficiency and patterning. Targeted disruption of the *βTub85D* gene through site-specific mutagenesis has been shown to be capable of inducing sterility in males (Fackenthal et al., 1995) incurring the fitness costs that may be useful in the context of the IGD strategy in limiting resistance allele formation.

Another male-specific promoter has been assayed in *D.melanogaster* - *Rcd-1r* (also known as *zeus*) (Chan et al., 2013a; Simoni et al., 2014). Originally identified as a retrogene located in a male sterility locus (Bai et al., 2007), *Rcd-1r* is a gene with testes-specific expression, implicated in late-stage spermatogenesis or fertilisation and which has been identified as an important component of male fertility (Chen et al., 2012). RNAi knockdown of *Rcd-1r* resulted in abnormal testes morphology and poorly developed sperm, while male mutants were found to be sterile or semi-sterile when compared to a control. Use of part of the *Rcd-1r* 5' UTR to drive expression of HEGs in *Drosophila* was shown to mediate reasonable rates of cleavage and homing. The importance of *Rcd-1r* for male fertility renders it a suitable locus for IGD hijacking.

The *vasa* regulatory region has been used to drive expression of *D.melanogaster* germline genes (Sano, Nakamura, and Kobayashi, 2002; Bischof et al., 2007), having been identified as being expressed during oogenesis and embryogenesis. The *vasa* regulatory region has since been used in both *Drosophila* (Chan et al., 2013a; Gantz and Bier, 2015) and Anopheline mosquito (Gantz et al., 2015) to drive expression of gene drive constructs. Noticeably, in *D.melanogaster*, the *vasa* regulatory region was initially found to mediate low levels of homing (Chan et al., 2011), though later iterations demonstrated increased performance (Chan et al., 2013a), with some work indicating tranmission rates above 95% (Gantz and Bier, 2015). An important short coming in using the *vasa* regulatory region (and thus the endogenous regulatory elements) to drive expression of drive constructs, is that expression is not restricted to the germ-line. Previous work (Port et al., 2014) has established that use of the *vasa* regulatory region to mediate Cas9 expression resulted in genetic mosaicism as a consequence of somatic expression of Cas9.

Such activity is liable to induce toxicity in transgenic *vasa* lines and has been implicated in increased rates of maternal deposition (Champer et al., 2018b). Given that *vasa* has been identified as being an essential maternal effect gene, required for germ-line development, disruption of the *vasa* locus is liable to result in sterility. Thus, despite issues with its expression patterning, *vasa* is a suitable candidate for an IGD drive component proof-of-concept study.

An alternative to the *vasa* regulatory region for germ-line expression of gene drive constructs is *nanos*. The functions of *nanos* are numerous: it is implicated in regulating abdominal segment formation (Wang and Lehmann, 1991), in promoting pole cell migration (Kobayashi et al., 1996), and ensuring survival and development of germ-line stem cells in both sexes (Forbes and Lehmann, 1998; Bhat, 1999). Importantly in the context of the IGD strategy, *nanos* has been shown to be essential for adult viability through P-element mediated insertion (Spradling et al., 1999). The *nanos* regulatory region has been used to drive constructs in *D.melanogaster* and shown to perform with commensurate (Champer et al., 2018b) or improved homing efficiency (Champer et al., 2017). High efficiency expression of Cas9 has also been mediated in *An.gambiae* (Hammond et al., 2018). The important improvement that *nanos* offers as a promoter element, is germ-line restricted expression as compared with *vasa*, resulting in lower toxicity due to reduced somatic activity (Champer et al., 2018b).

A further alternative stems from the *zero population growth* gene (*zpg*). Implicated as a gap junction gene essential for differentiating germ cells during gametogenesis (Tazuke et al., 2002), expression of *zpg* is limited to the germ-line, and promoters derived from the locus have been shown to exhibit high homing ratings in *An.gambiae* when using its ortholog (AGAP006241) (Kyrou et al., 2018; Hammond et al., 2018). As compared with a *nanos* promoter, *zpg* was shown to have an even greater reduction in maternal deposition, suggesting it is the most appropriate germ-line promoter currently identified when attempting to reduce the rate of resistance allele formation. Null mutants for *zpg* are viable but sterile, and should thus limit resistance allele formation when used as an endogenous locus in the IGD strategy.

A final candidate germ-line gene that was selected for investigation here was *bag of marbles* (*bam*), a gene previously used to drive HEG expression in *D.melanogaster* (Chan et al., 2013a). The role of *bam* has been identified as being essential to gametogenesis in both males and females (McKearin and Spradling, 1990). Null mutants exhibited stunted egg development and males produced aberrant spermatocytes. Historically, the *bam* regulatory region has been used to investigate HEG-mediate gene drive in males (Chan et al., 2011; Chan et al., 2013a) due to its expression in spermatogonia, though high-throughput expression data (Celniker et al., 2009), and biological studies (Parma, Bennett, and Boswell, 2007) suggest a role in early oogenesis. A *bam* promoter was found to mediate poor HEG activity with low cleavage rates, though it mediated a high homing rate. When initially assayed, the *Rcd-1r* regulatory region was selected to supersede *bam* due to its stronger expression in the spermatagonia (Chan et al., 2013a), though its native expression pattern when expressed from an endogenous locus may differ.

5.4 Materials and Methods

5.4.1 *In vitro* Screening of Putative gRNA Targets

Candidate gRNAs for use in gene drive constructs were screened *in vitro* using the Guide-it Complete sgRNA Screening System (Takara Bio Europe SAS, Saint-Germain-En-Laye, France). Candidate gRNAs were synthesised as ssDNA olignonucleotides (Eurofins Genomics, Ebersberg, Germany) and annealed and amplified to a proprietary scaffold template (Table.5.1). PCR conditions were as follows: 33 cycles at 98°C for 10 seconds, and 68°C for 10 seconds, and then held at 4°C indefinitely. PCR products were verified on a 2% agarose gel, using a 1kb hyperladder marker (Bioline, TN, USA), run at 100 V for 60 minutes.

gRNA Target	**T7 Primer**
vasa	CCTCTAATACGACTCACTATAGGATATCAATATGTCTGACGACGTTTAAGAGCTATGC
Rcd-1r	CCTCTAATACGACTCACTATAGGATGAGTGCGGAACCAAGTCGTTTAAGAGCTATGC
bam	CCTCTAATACGACTCACTATAGGGCAACGACGACCAGCAGTGTTTAAGAGCTATGC
β2-tubulin	CCTCTAATACGACTCACTATAGGCAATGCGGTAACCAGATCGGGTTTAAGAGCTATGC
nanos	CCTCTAATACGACTCACTATAGGTGTTGCAAATCCCCCCTCGTGTTTAAGAGCTATGC
zpg	CCTCTAATACGACTCACTATAGGCTGCACTCCAAAGTCACAGGTTTAAGAGCTATGC

TABLE 5.1: T7 primers required for assembly of sgRNA cassettes

The PCR product for each gRNA was subsequently used in an *in vitro* transcription reaction to generate purified gRNA. Reactions were setup as follows: 5 µl PCR template, 7 µl transcription buffer, 3 µl T7 polymerase mastermix, 5 µl RNase free water. Reactions were incubated at 37°C for 4 hours in a thermal cycler. After incubation, 2 µlof DNase I were added to each reaction and reactions were incubated at 37°C for a further 15 minutes. Purification of RNA was performed according to the manufacturer's instructions with Qiagen RNeasy mini kits (Qiagen, CA, USA).

Target amplicons to be cleaved by the synthesised gRNAs were obtained by genomic DNA extraction and PCR amplification from w^{1118} flies. Genomic DNA was extracted according to the manufacturer's instructions using a Qiagen DNeasy Blood and Tissue Extraction kit (Qiagen, CA, USA). PCR amplification of desired amplicons was performed as described (Table.5.2), using the following conditions: 98°C for 2 minutes, 35 cycles at 98°C for 10 seconds, 60°C for 20 seconds, 72°C for 2.5 minutes, a 72°C extension step of 7 minutes, and then held at 4°C.

Component	**Volume (µl)**	**Concentration**
10 µM Forward Primer	1	0.5 µM
10 µM Reverse Primer	1	0.5 µM
2x Phusion mastermix	10	1x
Template gDNA (∼50ng)	1	2.5ng/µl
RNase-free water	7	N/A

TABLE 5.2: PCR conditions for amplification of target amplicons from gDNA

Suitable target amplicons pertaining to each assayed gRNA were selected such that Cas9 cleavage would occur asymmetrically within the amplicon, resulting in bands of different sizing when resolved by gel electrophoresis (Table.5.3). Cas9 protein (500 ng/µl) was incubated with each gRNA (50 ng/µl) in a ratio of 1:2 at 37°C for 5 minutes, to form individual Cas9: gRNA complexes. These complexes were in turn incubated with the aforementioned amplicons, along with a control gRNA and amplicon as follows: 37°C for 1 hour, 80°C for 5 minutes, 4°C hold. Reactions were then resolved on a 2.0 % agarose gel, run at 100 V for 60 minutes.

Target Gene	Target Amplicon Size (bp)	Fragment Size (bp)	Amplicon Primer (Forward)	Amplicon Primer (Reverse)
vasa	1923	524, 1398	CCGTCGAAATGTGCTGCTTT	GGCCCTTTGGGTGGATTTTTC
Rcd-1r	1903	478, 1425	GTCAGACAGTGAGGCACGAC	AGTCCTGAGCCAGCTGCTAT
bam	1901	438, 1463	CGCCGTGCAGACCAATCATA	TCTTTCTTTCAGGCTGCTCCT
β2-tubulin	1908	482, 1426	GGGTACTCCTCGCGGATCTT	CCGTGATAATAAGGCGGGCA
nanos	1942	498, 1426	ATTTTCCATTGCGCGGTGTC	ACCTCTTTGGCCTTGCTGTT
zpg	1929	701, 1228	CGCCATTGTAAAGGGACAGC	CCTCCTGCGGTGGAAGTATC

TABLE 5.3: Amplicon primers for the *in vitro* gRNA activity assay, with corresponding fragment sizes following successful cleavage

5.4.2 Cloning of Experimental Constructs

Cloning gRNA Spacers into pCFD3-dU6:3gRNA Expression Vector

To generate U6:3-gRNA-T6 gRNA cassettes, spacers were synthesised as two, complementary ssDNA oligos (Table.5.4) (Eurofins Genomics, Ebersberg, Germany), with 3-4bp overhangs, compatible with a BbsI restriction site in the pCFD3-dU6:3gRNA expression vector. Where spacers began with a guanine nucleotide, the additional 5′ overhang was 5′-GTC-3′, otherwise the 5′ overhang was 5′-GTCG-3′. The invariant 3′ overhang was 5′-AAAC-3′. The pCFD3-dU6:3gRNA plasmid was a gift from Simon Bullock (Addgene plasmid #49410) (Port et al., 2014).

Spacer ID	Sense Strand	Antisense Strand
vasa	GTCGATATCAATATGTCTGACGAC	AAACGTCGTCAGACATATTGATAT
β2-tubulin	GTCGCAATGCGGTAACCAGATCGG	AAACCCGATCTGGTTACCGCATTG
nanos	GTCGTGTTGCAAATCCCCCCTCGT	AAACACGAGGGGGGATTTGCAACA
bam	GTCGGGCAACGACGACCAGCAGT	AAACACTGCTGGTCGTCGTTGCCC
Rcd-1r	GTCGATGAGTGCGGAACCAAGTC	AAACGACTTGGTTCCGCACTCAT
zpg	GTCGCTGCACTCCAAAGTCACAG	AAACCTGTGACTTTGGAGTGCAG

TABLE 5.4: Single-stranded DNA oligos used to construct sgRNA sequences for insertion into pCFD3-dU6:3gRNA

All subsequent annealing and ligation steps were performed as reported in Chapter 4.

Assembling IGD Drive Components

Initially, a drive component for hijacking of the *vasa* locus was constructed by Gibson assembly (Gibson et al., 2009). The backbone was derived from pBID, a gift from Brian McCabe (Addgene Plasmid #35190) (Nissim et al., 2014). The drive consisted of four fragments; two 1kb homology

arms derived from regions adjacent to the *vasa* locus target site, an eGFP cassette with a 3′ furin cleavage site and T2A sequence, and a split SpCas9 cassette containing an *ftz* intron and with a 3′ recoded T2A sequence (see Table.5.5 for primers). All fragments were amplified such that they had 5′ and 3′ overhangs of 15-30bp, giving them homology to either a backbone vector, or an adjacent fragment. Gibson assembly mastermix was produced in accordance with the protocol listed by OpenWetWare, with all components provided by NEB (New England Biolabs, Hitchin, UK). The general protocol for Gibson assembly is given here:

1. Take purified DNA fragments and thaw
2. Thaw a single 15 µl mastermix per reaction on ice
3. Assemble DNA fragments in 5 µl total volume, diluted with RNAse-free water where required. DNA should be present in equimolar quantites, and in a 3:1 ratio relative to the backbone (if one is present). Use 75 ng per fragment
4. Add DNA mix, and incubate reactions at 50°C for 60 minutes
5. Take 2 µl direct from the reaction, and 2 µl of a 1:100 dilution in water, and transform each using an appropriate antibiotic

The SpCas9 cassette was obtained from pBS-Hsp70-SpCas9, a gift from Melissa Harrison, Kate O'Conner-Giles, and Jill Wildonger (Addgene plasmid #46294). This plasmid was split into two fragments in the SpCas9 coding sequence by PCR amplification. A 575bp synthetic fragment containing the modified *ftz* intron was produced by ATUM (ATUM, CA, USA), with 200bp overhangs homologous to the split SpCas9 cassette. The entire plasmid was reassembled by Gibson assembly, with the additional inclusion of the *ftz* intron. The construct was digested using BaeI, and gel purified. The *vasa* gRNA cassette was amplified by PCR from the pCFD3 expression construct with suitable homology arms, and ligated in a 2-part Gibson reaction into the *ftz* intron.

Primer ID	Sequence	Amplification Site	Strand
	pBID-*vasa*-eGFP-SpCas9-*ftz*		
G8	GGCCGCGGCAGCCTGCTGACCTGCGGCGATGTGGAGGAGAAC CCCGGCCCCATGTCGGACGATTGGGATGATGAGGTTCGTATTG	Right homology arm	Sense
G9	ATTACGTGGCCAAGATCTTCTAGACCTTGCTGTAAATTCCTCCATTG	Right homology arm	Antisense
G33	CCATCTACAATAGTATCTACTGAGGGCCGCGGCAGCCTG	Right homology arm T2A addition	Sense
G2	GCGGCCGCGGCTCGAGAAGGTACCGAATATTGGATCCTGCGTTC	Left homology arm	Sense
G35	GTAGATACTATTGTAGATGGTACGGCCCTTGCTCACGTCA	Left homology arm	Antisense
G4	CAATATGTCTGACGTGAGCAAGGGCGAGGAG	eGFP	Sense
G5	CTCCACATCGCCACAGGTGAGCAGGCTTCCGCGAC CCTCCCTCTTACGGCGCTTGTACAGCTCGTCCATGC	eGFP	Antisense
G6	CCTGCTCACCTGTGGCGATGTGGAGGAAAA CCCAGGTCCCATGGACTATAAGGACCACG	Cas9-T2A	Sense
G7	CGCAGGTCAGCAGGCTGCCGCGGCCCTCG CGCTTGCGGCGCTTTTTCTTTTTTGCCTG	Cas9-T2A	Antisense
G62	ATACACCGGTAAGTAGCGATTTTTTTGCTCACCTGTGATTGCT	U6:3-gRNA-T6	Sense
G63	GAAGGAGTAGAATTAGTACCTGCAAAAAAAGCACCGACTCGGTG	U6:3-gRNA-T6	Antisense
	pBID-*vasa*-eGFP-SpCas9-*ftz*-CFP		
G80	TGAAAATATGGCTAGCTTTTTTGCTCACCTGTGATTGCT	U6:3-gRNA-T6	Sense
G81	ATGTATACAAATTCGAAAAAAGCACCGACTCGGTG	U6:3-gRNA-T6	Antisense
G54	ATATTGAAAATATGGATAACTTCGTATAGCATACATTA TACGAAGTTATAGCTTTAAGATACATTGATGAGTTTG	3xP3:CFP:SV40	Sense
G55	AATTCGAAACTAGAGATCGGCGTCGTGTCACTTCTATAACTTCGT ATAATGTATGCTATACGAAGTTATCCGCTCTAGAACTAGTGGATC	3xP3:CFP:SV40	Antisense
	pEX-258-Drive-SpCas9-*ftz*-CFP		
G134	ATGGACTATAAGGACCACGACGGAG	SpCas9-*ftz*-CFP	Sense
G135	CTTTTTCTTTTTTGCCTGGCCGG	SpCas9-*ftz*-CFP	Antisense

TABLE 5.5: Primers used for cloning of fragments required for construction of IGD drive components

The SpCas9-*ftz* fragment was subsequently amplified with the addition of a 3' furin cleavage site and T2A sequence. The eGFP fragment was amplified from pAc-STABLE2-neo, a gift from Manjunatha Kogenaru, and gel purified. The pBID backbone was linearised by digestion with XbaI, and gel purified. The left and right homology arms were cloned with homology to the pBID backbone vector, and with the inclusion of a BaeI restriction site between one another. They were subsequently gel purified and assembled with the linearised pBID vector in a 3-part Gibson reaction. This intermediate plasmid was digested using BaeI, gel purified, and Spcas9 and eGFP cassettes were ligated into the linearised intermediate in a 3-part Gibson reaction. Confirmation sequence veracity was provided by Sanger sequencing. This construct will hereafter be referred to as pBID-*vasa*-eGFP-SpCas9-*ftz* (Fig.5.1).

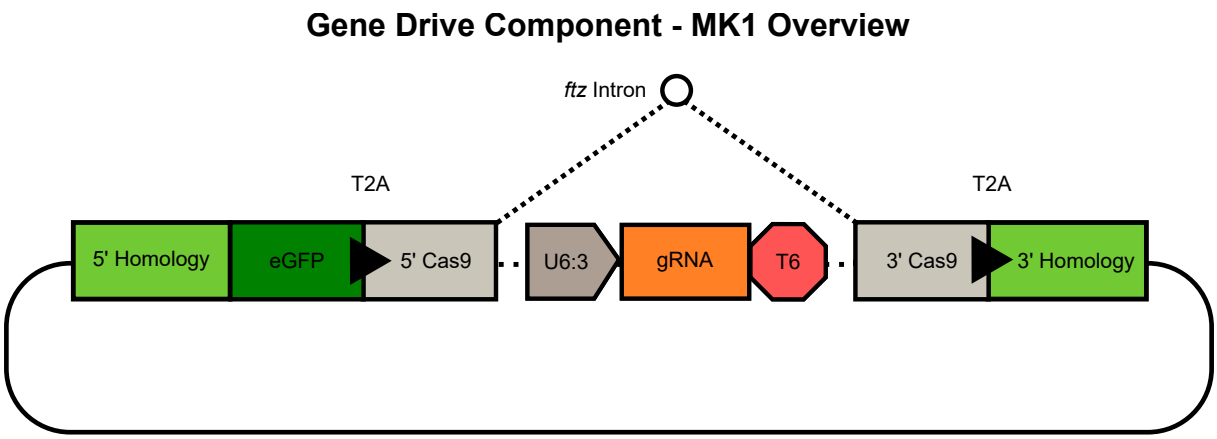


FIGURE 5.1: **Gene Drive Construct MK1**. The initial drive construct, targeted at the *vasa* locus, was constructed from a single SpCas9 cassette, containing a modified variant of the *ftz* intron. Within the intron, was a U6:3 promoter, driving expression of a sgRNA. Flanking the SpCas9 cassette were T2A sequences, the 3' variant recoded to prevent unwanted homology events. An eGFP cassette was included 5' to SpCas9, and 1kb homology arms flanked the entire insert.

A further iteration of this initial IGD drive construct was assembled (Fig.5.2) by the additional insertion of a 3xP3:CFP:SV40 cassette (a gift from Andrew Hammond). The U6:3-gRNA-T6 cassette was amplified from the pCFD3 expression construct using primers G80 and G81, and gel purified. pBS-Hsp70-SpCas9-*ftz* was digested with Bsp119I and purified. The new gRNA cassette was ligated into the site in a 2-part Gibson reaction. The CFP cassette was cloned into the resulting pBS-Hsp70-Cas9-*ftz*-gRNA intermediate using an NheI restriction site. Sanger sequencing was performed to ensure veracity of the insertion. SpCas9 was amplified from the resulting plasmid and used to construct the drive construct as before. The resultant plasmid was sequenced as before.

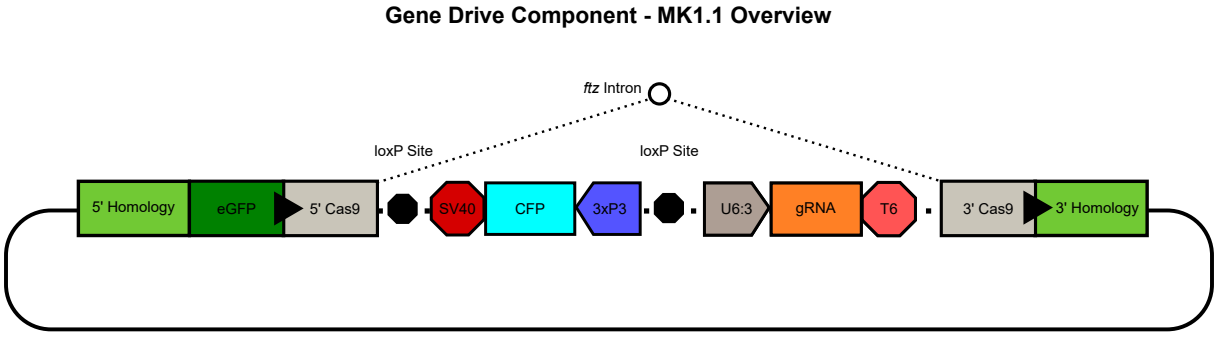


FIGURE 5.2: **Gene Drive Construct MK1.1**. In order to mitigate possible issues with observing the germline-expressed eGFP cassette, a 3xP3:CFP:SV40 cassette was inserted within the *ftz* intron, permitting constitutive expression of CFP in the developing imaginal disc, eyes, and ocelli.

The final IGD drive iteration used in this work (Fig.5.3), and which was successfully used to produce transgenic *D.melanogaster* lines, included constructs targeted to the *bam*, *nanos*, *Rcd-1r*, *β2-tubulin*, and *zpg* loci. In this design, the eGFP cassette was abandoned, leaving the CFP

reporter present in the *ftz* intron within SpCas9. 1kb homology arms for each locus were synthesised in a pEX-258 backbone with ampicillin cassette (ATUM, CA, USA). A BbsI, or BaeI site was included in an excisable sequence between the two homology arms. Upon receipt of the synthesised constructs, DNA was purified by centrifugation from provided filter paper according to the manufacturer's instructions. The constructs were concentrated by mini-prep, and digested using BbsI or BaeI as appropriate, and gel purified to produce a linearised backbone for each drive construct. Appropriate gRNA cassettes were amplified from the pCFD3 expression construct as before, and cloned into a Bsp119I site of the *ftz* intron within pBS-Hsp70-SpCas9-*ftz* as before, followed by subsequent insertion of the 3xP3:CFP:SV40 cassette. The resulting SpCas9 fragments were amplified using primers G134 and G135. The resultant SpCas9 fragments had homology to the insertion site in all backbones. Each variant was purified as before, and assembled into the linearised gene backbones in a 2-part Gibson assembly. All resultant drive constructs were verified by Sanger sequencing. All primers used for sequencing can be found in A.1 and A.2.

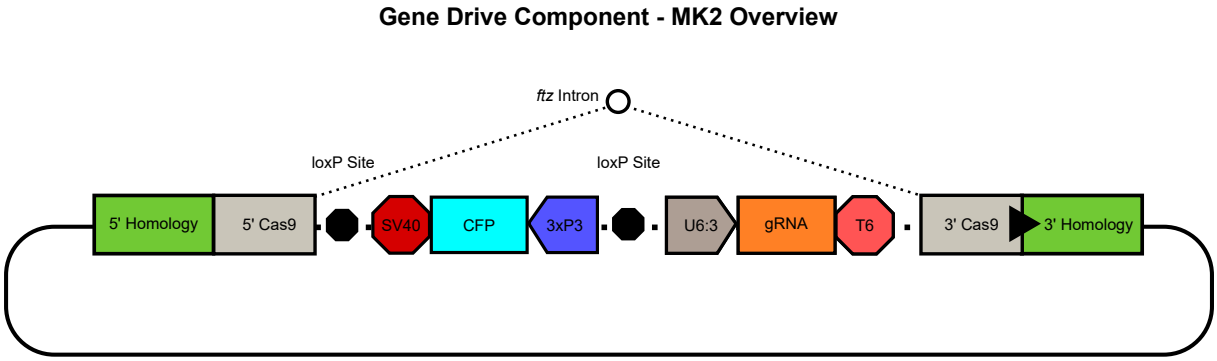


FIGURE 5.3: **Gene Drive Construct MK2**. The final variant of the gene drive, which was expanded to target a range of germ-line loci, was simplified, removing the germ-line regulated eGFP cassette, retaining only the constitutively expressed 3xP3:CFP:SV40 cassette, contained within the *ftz* intron.

5.4.3 Line Generation and Biosecurity

Microinjections were performed both on site at Imperial College London (ICL), and at the Fly Microinjection Facility at the University of Cambridge. Injections at ICL were performed using a Leica Labovert FS microscope (Leica Microsystems, UK). Needles were obtained using 1.0mm thin-wall capillary tubes (WPI, UK), and pulled using a P97 Sutter Flaming/Brown Micropipette Puller (Sutter, UK).

In order to collect embryos for microinjection, *D.melanogaster* 13-20 stock were reared in a 250 ml conical bottle, plugged with cellulose acetate plugs (VWR, UK) at 25°C. 35 mm x 10 mm petri dishes were filled with grape agar (Genesee Scientific, CA, USA), smeared with yeast paste, and placed into the neck of the conical bottle. After 24 hours, plates were discarded and replaced. On the day of microinjection, plates were replaced, and after 6 hours embryo collection began.

Embryos were aligned on 10 mm x 10 mm microscope slide covers and held in place with heptane glue. Aligned embryos were laid on microscope slides and microinjected. Needles were filled with 3 µl of drive construct, concentrated at 400 ng/µl. Following injection, embryos were retained on the coverslip and placed in vials and maintained at 22°C in an LMS Series 1A 80NP incubator with fruit fly modification (LMS, UK)

Injections performed at the University of Cambridge were performed by K. Oras and A. Madich. Embryos were dechorionated in 50% bleach, prior to microinjection. Embryos were aligned as before on coverslips. Needles were pulled using a P1000 Sutter Micropipette Puller (Sutter, UK). Injections into w^{1118} lines were conducted at 500 ng/µl, consisting of 400 ng/µl drive construct, and 100 ng/µl Cas9 helper plasmid. Following microinjection, embryos were retained on coverslips, and placed onto 100mm x 100mm agar plates for transportation. Upon hatching, 1st instar larvae were placed into yeast paste and transferred to plugged vials to mature at 22°C. To prevent the possibility of transgenic flies escaping and propagating drive alleles in field populations, all microinjected stocks were maintained in an insectary compliant with Arthropod Containment Guidelines Level 2. All stocks were maintained with six degrees of separation from the external environment. In the event of escape, flies were to be destroyed using ethanol. All lines obtained from external sources for work in this thesis are described in TableA.6.

5.4.4 Cultivating *Drosophila* lines

All *Drosophila* stocks were maintained at 18°C in an LMS 1A 80NP incubator (LMS, UK). All experiments were conducted at 25°C in a JENCONS LMS incubator (Jencons Scientific, UK). Unless otherwise stated, all flies were reared in vials, fitted with acetate plugs (Fisher Scientific, NH, USA), filled with a standard yeast food. Yeast food was produced as follows:

1. Add 240 g of polenta to a 15 L pot, with 10 L of water
2. Bring to a boil, while stirring continuously
3. Premix 960 g of fructose and 1200 g of Brewer's yeast
4. Once the solution is boiling, add fructose and yeast
5. Simmer for 10 minutes on low heat
6. Turn off heat, and add an additional 2 L of water
7. Allow to cool to below 70°C before adding 60 ml of Nipagin (15% w/v in ethanol)
8. Add 90 ml of Propionic acid
9. Aliquot into vials using a Watson-Marlow 501 peristaltic pump (Watson-Marlow, Falmouth, UK)
10. Allow to cool for 12-16 hours, and close with acetate plugs

Where necessary, flies were anaesthetised using CO_2 on Flystuff Flypads (Genesee Scientific, CA, USA), fitted with water bubblers to prevent dehydration. Flies were disposed of into 70% ethanol traps, which were routinely emptied. Stock vials were disposed of securely by refrigeration at -20°C for 24 hours, followed by disposal in an established CL2 waste bin. All 18°C stocks were flipped into fresh vials of food on a fortnightly basis. Where stocks were maintained at 25°C for experimental purposes, they were flipped every 10 days into fresh food.

5.4.5 Crossing and Genotyping of *Drosophila* lines

The molecular screening of transgenic *Drosophila* was conducted by genomic DNA extraction, PCR amplification, and visualisation by gel electrophoresis. Genomic DNA extraction was performed using Chelex resin (BioRad, Hertfordshire, UK). An extraction solution was produced by suspending 1.25 g of Chelex resin in 100 ml of RNase-free water. This was supplemented with 4 ml of proteinase K solution (20 mg/ml) (Invitrogen, Paisley, UK). This solution was added in 100 µl aliquots 1.5 ml microcentrifuge tubes containing flies, ensuring to regularly agitate the solution whilst pipetting to avoid Chelex resin from dropping out of suspension. Microcentrifuge tubes were incubated using an Eppendorf Thermomixer-C fitted with a 24-tube adaptor (Eppendorf UK, Stevenage, UK) . Tubes were incubated for 2 hours at 55°C, and 700 rpm, and then proteinase K was inactivated by incubation for 20 minutes at 99°C. Tubes were stored at -20°C when not in use.

PCR mastermix was constructed per reaction as follows, and scaled appropriately: 7 µl RNase-free water, 1 µl forward primer (0.5 M final concentration), 1 µl reverse primer (0.5 M final concentration), 1 µl gDNA sample, 10 µl REDTaq DNA polymerase mastermix (VWR, UK). All reactions were setup in 8-well PCR strip tubes (VWR, UK). PCR reactions were conducted in Applied Biosystems Veriti 96-well Thermal Cyclers (Applied Biosystems, CA, USA) as follows: Denatured at 98°C for 2 minutes, followed by 35 cycles of 98°C denaturation step for 20 seconds, 59°C annealing step for 30 seconds, and 72°C extension step for 30 seconds, followed by a final 72°C extension step for 7 minutes, and then a 4°C hold. Primers are listed in Table.5.6.

Primer ID	Sequence	Amplification Site	Amplicon Size (bp)
pC166	CGCAATAGATCGGACCTCCC	*bam* 3′ junction	1706
pC167	CCCCCGAGGATAATGAGCAG		
pC168	ACAGCTGGTTGTAGGTCTGC	*bam* 5′ junction	1874
pC169	TTCCCCGATAAGAAACCGCC		
pC176	GATCCTCCGCACTGATTCCA	*bam* WT target site	2165
pC177	GTGAGGCTTCTAATGATTGGGC		
LJFV2	TAGCAAAGTCAGGGCGTAGC	*Rcd-1r* 3′ junction	609
LJRV2	CACCGGGATAAGCCCATCAG		
WTF1V2	ACGGGCGTCGTTATTATGGG	*Rcd-1r* WT target site	517
WTR1V2	ACGAATTCCAAGCTGCCAATC		
RJFV2	AGAAGCTGTCGTCCACCTTG	*Rcd-1r* 5′ junction	586
RJRV2	ACGTGCTTTCGGTCCTTTCT		
pC160	CTTAAGCCGGCCTATGCTCA	Whole *Rcd-1r* insertion	9073
pC163	CACACAGATGACGCACCAAC		

TABLE 5.6: Primers used for genotyping of the $Rcd\text{-}1r^{d}$ and bam^{d} drive insertions. Here *d* denotes the presence of the drive component.

Amplified PCR products were subsequently visualised using 0.7% agarose gel, stained with SYBRSafe stain. An O'GeneRuler Express ladder was used in 10 µl volumes. Images were taken using an Azure c600 gel imaging system (Azure Biosystems, CA, USA). Gels were run at 100V for 60 minutes, using a BioRad PowerPac Power supply (BioRad, Hertfordshire, UK).

5.4.6 Cre:Lox Recombination of *Drosophila* lines

Drive lines are denoted by the use of a $^{'d'}$. Where drive components contain a 3xP3:CFP:SV40 reporter cassette, they are denoted by $^{'d,CFP'}$.
w;Rcd-1r^{d}*;* lines were generated by performing 10 singles crosses, of ♂*w;Rcd-1r*^{d,CFP} with ♀*y[1] w[67c23]; sna[Sco]/CyO, Pw[+mC]=CrewDH1* virgins to excise the 3xP3:CFP:SV40 cassette contained within the Cas9-located *ftz* intron, of the drive component. Subsequent G_1 progeny were all reporterless, and were crossed to produce homozygous *w;Rcd-1r* G_2 progeny. Accurate excision was confirmed by molecular genotyping of G_1 parents, as previously described using the primers listed (Table.5.7), followed by sanger sequencing confirmation. In the event that excision is unsuccessful, a 2250bp amplicon is produced, whilst success result in an approximately 694bp amplicon.

Primer ID	Sequence
pC182	ATGTTTTTGTATTTGAGTAGGAGCA
pC183	CCAGAGCTTCATCGAGCG

TABLE 5.7: Primers used to confirm successful Cre:Lox mediated excision of the 3xP3:CFP:SV40 reporter cassette

5.4.7 *Drosophila* Development Studies

Development studies were conducted to investigate developmental success of transgenic lines, following a protocol published by Fasulo et al., 2012. Transgenic *Rcd-1r*^{d,CFP}, *Rcd-1r*^{d}, and *bam*^{d,CFP}/TM6B, Tb[1] males were crossed to w^{1118} virgins to produce heterozygote offspring for use in these studies. Offspring were subsequently mated with w^{1118} individuals. Whether using homozygote or heterozygous offspring, all females were collected as virgins, with all flies being aged for 1-3 days prior to commencement of each experiment. Mass crosses were performed in Flystuff 60 mm oviposition cages (Genesee Scientific, CA, USA), using 30 males, and 30 virgin females. Cages were equipped with grape agar plates (Genesee Scientific, CA, USA), and Allinson easy bake yeast paste applied to the surface (Allinson Flour, Peterborough, UK). The following crosses were performed;

♂*w*; *Rcd-1r*^{d,CFP}/+ ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d,CFP}/+ ;	x	♂w^{1118};;
♂*w*; *Rcd-1r*^{d,CFP}/ *Rcd-1r*^{d,CFP} ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d,CFP}/ *Rcd-1r*^{d,CFP} ;	x	♂w^{1118};;
♂*w*; *Rcd-1r*^{d}/+ ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d}/+ ;	x	♂w^{1118};;
♂*w*; *Rcd-1r*^{d}/ *Rcd-1r*^{d} ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d}/ *Rcd-1r*^{d};	x	♂w^{1118};;
♂*w*; ;*bam*^{d,CFP}/+	x	♀w^{1118};;
♀*w*; ;*bam*^{d,CFP}/+	x	♂w^{1118};;

Flies were mated for 24 hours, prior to commencement of each experiment. Crosses were transferred to Flystuff 35 mm oviposition cages (Genesee Scientific, CA, USA) at the beginning of each egg collection period, once again equipped with grape agar plates with yeast paste. Eggs were collected every 3 hours, and aligned on 18 mm x 18 mm glass microscope cover slips (VWR, UK), using heptane glue. Aligned eggs were subsequently transferred to vials containing standard yeast food. Between 50-150 eggs were transferred to each vial, with at least 6 collections per cage. At the end of each 12 hour egg collection period, flies were returned to the Flystuff 60 mm oviposition cages overnight. Eggs were collected for four days, until at least 500 eggs had been collected per cage. The number of eggs, hatched eggs, pupae, and adults (male and female) was recorded per vial. Statistical analysis was performed using the Kruskal-Wallis test, assuming non-parametric distribution due to sample size, using a w^{1118} cross as a control.

5.4.8 Oviposition and Hatching Rate Assay

To determine changes in fecundity due to insertion and expression of the IGD drive component, *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} homozygous lines were assayed against *w1118* flies, using a protocol adapted from Simoni et al., 2014. Single virgin females from each line were collected, and crossed singly for 48 hours. Mated females were subsequently allowed to lay eggs in a vial

containing standard yeast food for 3 consecutive days, with flies flipped into fresh tubes at the end of each 24 hour period. At least 8 replicates were performed per line. The total number of eggs was quantified as was the number of eclosed adults from each vial. Statistical analysis was performed using the Kruskal-Wallis test.

5.4.9 *Drosophila* Homing Assays

Homing assays were conducted for all viable and fertile transgenic lines. For *Rcd-1r*^{d,CFP} crosses, heterozygotes were obtained by crossing single homozygous *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} males to w^{1118} virgins. It was not possible to produce a fertile *bam* homozygote line. To this end, balanced *w*; ;*bam*^{d,CFP}/TM6B, Tb[1] males were crossed to w^{1118} virgins. Subsequent G_1 heterozygotic progeny were screened such that only *w*; ; *bam*^{d,CFP}/+ individuals were selected to act as parents for homing assay crosses. G_1 heterozygotic progeny were collected such that all females were virgin, and all individuals were 1-3 days old. Single males, were crossed to three females in each cross, with at least 20 replicates established for each line. Males and female heterozygote transgenics were assayed separately to determine sex-specific effects. Where replicates failed to yield offspring, they were discarded.

Crosses were performed thus;

♂*w*; *Rcd-1r*^{d,CFP}/+ ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d,CFP}/+ ;	x	♂w^{1118};;
♂*w*; *Rcd-1r*^{d}/+ ;	x	♀w^{1118};;
♀*w*; *Rcd-1r*^{d}/+ ;	x	♂w^{1118};;
♂*w*; ;*bam*^{d,CFP}/+	x	♀w^{1118};;
♀*w*; ;*bam*^{d,CFP}/+	x	♂w^{1118};;

G_2 progeny were collected, and screened using an Olympus MVX10 fluorescent microscope (Olympus, UK) for CFP fluorescence.

5.4.10 Genotyping Markerless *Rcd-1r*^{d} Progeny

Having collected G_2 progeny of *w*; *Rcd-1r*^{d}; crosses, it was necessary to perform molecular screening to quantify homing rates. Upon eclosing, G_2 progeny were collected, and frozen using liquid nitrogen. Male offspring from each successful crosses were separated and individually placed into the wells of a 96-well PCR plate (BioRad, Hertfordshire, UK). Flies were manipulated using tweezers, which were routinely cleaned with RNAse*Zap* decontamination solution (Invitrogen, Paisley, UK). This solution was also applied to the work area prior to the commencement of work. Control flies (w^{1118} from existing stocks of a commensurate age) were added to each plate to act a biological controls for cross-contamination. Genomic DNA extraction was performed using Chelex resin (BioRad, Hertfordshire, UK). An extraction solution was produced by suspending 1.25 g of Chelex resin in 100 ml of RNase-free water. This was supplemented with 4 ml of proteinase K solution (20 mg/ml) (Invitrogen, Paisley, UK). This

solution was added in 100 µl aliquots to each well containing a fly, ensuring to regularly agitate the solution whilst pipetting to avoid Chelex resin from dropping out of suspension. Plates were sealed using adhesive plate foils (Invitrogen, Paisley, UK), and placed in an Eppendorf Thermomixer-C, fitted with a lid and 96-well plate adaptor (Eppendorf UK, Stevenage, UK). Plates were incubated for 2 hours at 55°C, and 700 rpm, and then proteinase K was inactivated by incubation for 20 minutes at 99°C. Plates were stored at -20°C when not in use.

Following genomic DNA extraction, PCR reactions were set up in 96-well plates, corresponding to each gDNA plate. The primers used were RJFV2, and RJRV2, previously listed in table.5.6. PCR mastermix was constructed per reaction as follows, and scaled appropriately: 7 µl RNase-free water, 1 µl forward primer (0.5 M final concentration), 1 µl reverse primer (0.5 M final concentration), 1 µl gDNA sample, 10 µl REDTaq DNA polymerase mastermix (VWR, UK). Prior to beginning PCRs, each gDNA plate was centrifuged for 5 minutes, at 4300 x*g* to limit Chelex interference with PCR efficiency. Mastermix was dispensed into wells using an Eppendorf Multipipette E3 equipped with Eppendorf 0.5 ml combitips (Eppendorf UK, Stevenage, UK). Two additional template-controls were added to each PCR plate, consisting just of PCR mastermix, devoid of gDNA. Genomic DNA was added last to each reaction, using an 8-channel pippette, equipped with sterilised pipet tips. PCR reactions were conducted in Applied Biosystems Veriti 96-well Thermal Cyclers (Applied Biosystems, CA, USA) as follows: Denatured at 98°C for 2 minutes, followed by 35 cycles of 98°C denaturation step for 20 seconds, 59°C annealing step for 30 seconds, and 72°C extension step for 30 seconds, followed by a final 72°C extension step for 7 minutes, and then a 4°C hold. Plates were sealed, as before, with adhesive foil to prevent cross-contamination between wells.

PCR reactions were visualised using 1% agarose gels, run at 100 V for 55 minutes, and stained with ethidium bromide to provide higher sensitivity. An O'GeneRuler Express ladder was used in 7 µl quantities at either end of each row of wells. Gels were visualised, and bands corresponding the transgenic *Rcd-1r*^{d} amplicon were quantified - the presence of any band of the correct size was considered to be a positive indication of a transgenic individual. Where contamination of any of the negative control lanes occurred, PCR reactions were repeated up to a single time, and discarded if contamination persisted.

5.4.11 Amplicon Sequencing

In order to investigate Cas9 activity and indel formation as a consequence of the drive component's activity, amplicon sequencing was performed on CFP-negative G_2 from the ♂*w*; *Rcd-1r*^{d,CFP}/+ ; homing assay. 50 CFP-negative G_2 progeny were selected at random from a given biological replicate, and pooled genomic DNA extraction was performed - this was conducted for 12 replicates. Genomic DNA extraction was performed using the Wizard Genomic DNA Purification kit (Promega UK, Southampton, UK). Simultaneously, 200 w^{1118} flies of a commensurate age were selected, and pooled genomic DNA extraction was performed - this was conducted for 3 replicates. A Qubit was used to quantify yield, and each sample was normalised to 50 ng/µl using Tris-EDTA. PCR amplification was then performed using primers (Table.5.8)

for the wild-type *Rcd-1r* target site, producing a 320 bp amplicon. Primers included a partial scaffold compatible with Illumina NGS sequencing.

Primer ID	Primer Sequence	Strand
pC186	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACGCTTTTCCACAGCATGG	Sense
pC187	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTCGGTCCTTTCTCGCTTGA	Antisense

TABLE 5.8: Primers used for amplification of the *Rcd-1r* gRNA target site, for amplicon sequencing

Reaction setup was performed as described (Table.5.9).

Component	Volume (µl)	Concentration
10 µM Forward Primer	1.25	0.5 µM
10 µM Reverse Primer	1.25	0.5 µM
2x CloneAmp HiFi mastermix	12.5	1x
Template gDNA (∼50ng)	1	2.5ng/µl
RNase-free water	9	N/A

TABLE 5.9: PCR reaction conditions for high-fidelity amplification of the *Rcd-1r* locus for amplicon sequencing

PCR conditions were as follows: 98°C denaturation step for 2 minutes, followed by 23 cycles of 98°C for 10 seconds, 57°C for 10 seconds, 72°C for 30 seconds, followed by a final 72°C extension for 7 minutes, and a 4°C hold. Successful amplification of PCR products was confirmed by loading 5 µl of each sample onto a 1% agarose gel, running at 100 V for 60 minutes, and visualising. PCR products were subsequently purified using a Qiagen MinElute PCR Purification kit (Qiagen, CA, USA), and purified DNA products eluted in 10 µl Tris-EDTA. Purified DNA was quanitifed using a Qubit fluoremeter as before, and normalised to 20 ng/µl, in at least 20 µl total volume using Tris-EDTA. Samples were processed by Genewiz (Genewiz, Leipzig, Germany), and SNP analysis performed. Resulting data files were analysed for indel formation, and frameshift mutation rates using CRISPResso2 (Clement et al., 2019).

5.5 Model Generation and Parameters

The existing IGD model was expanded to permit consideration of sex-specific traits. The existing framework was expanded by duplicating existing parameters, and specifying each set as being male or female specific. As such the model expanded from 22 to 44 parameters. The underlying fitness and genotyping equations were similarly reproduced to permit consideration of males and females individually. Initial frequencies, fitness equations, gamete contributions, and genotype frequencies are calculated separately for males and females. Initial frequencies and genotype frequencies are halved such that the total population size is equal to 1, regardless of the proportion of males and females. Male and female parameter names and values are listed in A.4 and A.5.

5.6 Results

5.6.1 Engineering IGD Drive Components

As mentioned in Chapter 3, once an IGD component has hijacked a target site (whether germline, or somatic) it is essential that expression of the endogenous locus is not disrupted, as this is liable to incur a fitness cost. Given that all integrations described here occur at the 5′ end of host loci, it is necessary to introduce a 2A sequence at the 3′ terminus of the drive component. In this way, the endogenous gene will still be efficiently expressed under its own regulatory elements. Historic literature has been conflicting as to the most efficient 2A sequence - both P2A and T2A have been identified to mediate downstream transcript expression with high efficiency in human cell lines, zebrafish and mice (Szymczak et al., 2004; Donnelly et al., 2001; Kim et al., 2011). In *Drosophila*, P2A and T2A have been found to mediate high efficiency expression of bicistronic constructs, with no significant difference found in performance (Daniels et al., 2014). Given the success of using T2A peptides in *Drosophila* cell culture (Diao et al., 2015), and whole organism (Lee et al., 2018), it was selected for use as part of the IGD drive component. It is important to note that the hijacking of the endogenous gene occurs after the ATG translational start site, meaning that part of the endogenous coding sequence is truncated by insertion of the drive component. This truncated sequence was recoded to prevent recombination, and inserted at the 3′ end of each drive, after the final T2A sequence. This ensured that the entire endogenous gene was expressed.

The SpCas9 cassette used as the core of the IGD drive component was derived from pBS-Hsp70-Cas9, and though the SpCas9 gene is codon-optimised for human expression, it has been used extensively in *Drosophila* (Gratz et al., 2013). The Cas9 cassette includes 5′ and 3′ nuclear localisation signals to ensure transport of the Cas9 protein product to cellular nuclei. Work conducted in Zebrafish has indicated that the specific nature of the NLS in the context of Cas9 editing is of secondary importance as compared with the gRNA used, and the locus targeted (Hu et al., 2018). A 3xFlag tag is also included to permit purification by affinity chromatography in order to validate production of Cas9 protein in an organism or tissue. Having established that it is possible to efficiently express an igRNA cassette from within the *ftz* intron, SpCas9 was bifurcated by insertion of the intron. Aside from permitting expression of the U6:3-gRNA-SV40 cassette without disrupting the endogenous locus, insertion of an intron has been demonstrated to increases transgene expression in *Drosophila*, independent of enhancer elements (Duncker, Davies, and Walker, 1997).

In initial iterations of the drive component designed for the *vasa* locus, an eGFP reporter gene was included 5′ of the Cas9 coding sequence and 3′ of the left homology arm (Fig.5.4). Two T2A sequences were utilised to separate transcripts - the second T2A sequence was recoded to avoid recombination events during homing. However, revisions were made to the design following observations from RNAseq data available in the modENCODE database (Celniker et al., 2009), namely that expression of eGFP was likely restricted to the gonads, and would restrict ease of screening. To this end, drive components were revised, and the T2A-eGFP cassette removed.

In its place, a 3xP3:CFP:SV40 cassette was inserted into the *ftz* intron within SpCas9, flanked by loxP sites (Fig.5.4). The 3xP3 promoter has been demonstrated to be highly efficient in a range of insects (Horn et al., 2002), and in *Drosophila* mediates expression in the imaginal disc of larvae, and the eyes and ocelli of adults. LoxP sites allow for the reporter cassette to be excised at a later stage by crossing to a line expressing Cre-recombinase. The Cre/Lox system has been demonstrated to function with high efficiency in both *D.melanogaster* (Siegal and Hartl, 1996), and *An.gambiae* (Volohonsky et al., 2015). Constructs with suitable homology arms and gRNA sequences were generated for the *vasa*, *bam*, *Rcd-1r*, *zpg*, *nanos*, and *β2-tubulin* loci.

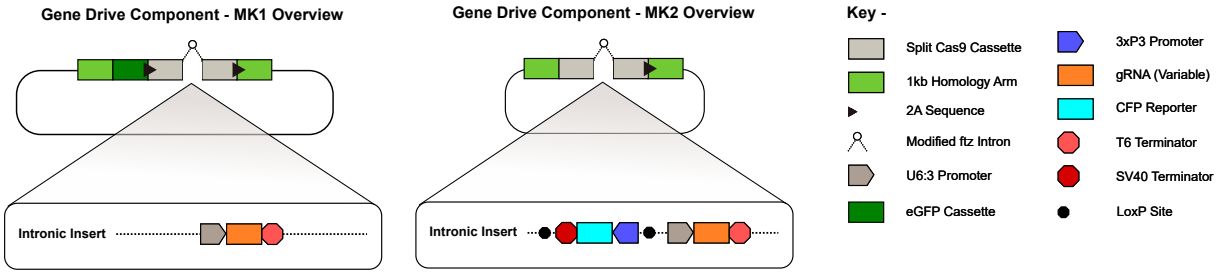


FIGURE 5.4: **Overview of Gene Drive Constructs**. Initially, an eGFP cassette was inserted upstream of the Cas9 coding sequence, to act as a reporter. A recoded T2A sequence was inserted between the two, to ensure efficient expression of each transcript, independent of one another. Subsequently, modifications were made, and the eGFP cassette was replaced with a 3xP3:CFP:SV40 cassette, flanked by LoxP sites, within the *ftz* intron of each construct.

5.6.2 gRNA Design and Selection

Having established potential candidates for hijacking in the *D.melanogaster* germ-line through literature review, it was necessary to assay candidate gRNAs. Candidate gRNAs were generated using flyCRISPR, an online target finding tool for *Drosophila* gene editing (Gratz et al., 2014). Target sites were selected at the 5′ end of the coding sequence of each locus, such that cleavage sites were distal to the translational start site of the gene. All candidate loci were screened using an optimal target finder (Gratz et al., 2014). Candidate gRNAs with high off-target activity were discounted, and a gRNA spacer selected for each of the six target germ-line loci (Fig.5.5).

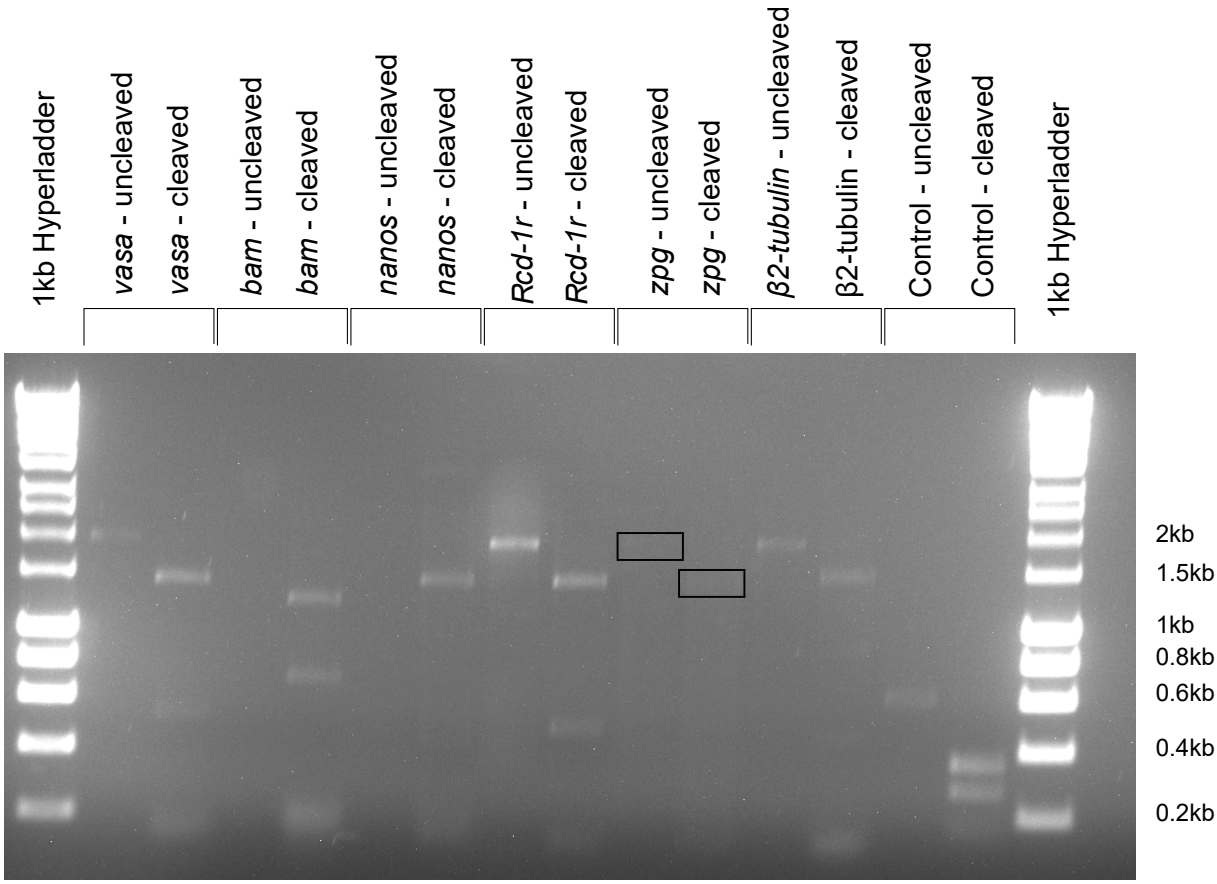


FIGURE 5.5: ***In Vitro* sgRNA Screen**. In each case, an uncleaved amplicon containing the gRNA target site was ran adjacent to a cleaved amplicon. In the case of *bam* and *nanos* the uncleaved amplicon is not visible, however fragments of the correct size (see methods) are clearly visible, indicating that the gRNA was still successful. In the case of *zpg*, bands are visible when the image is adjusted to high contrast. Boxes indicate the approximate position of each band.

Subsequently, gRNAs were synthesised and screened *in vitro* to determine whether, in ideal conditions, they were sufficient to facilitate SpCas9-mediated cleavage of their target site (Fig.5.5). An *in vitro* gRNA assay was performed, against synthetic PCR fragments, representative of the target sequence in the *D.melanogaster* genome. All gRNAs were able to mediate cleavage of their target sequence *in vitro*. As a result they were incorporated into the drive constructs described above. A figure illustrating the targeted insertional site for each gRNA sequence is given in the appendix, showing where each drive component was designed to be integrated within the gene transcript (Fig.A.1).

Target ID	**gRNA Spacer Sequence (5′-3′)**	**PAM Site**	**Spacer Length**	**Strand**
vasa	ATATCAATATGTCTGACGAC	TGG	20	Sense
bam	GGGCAACGACGACCAGCAGT	TGG	20	Anti-sense
Rcd-1r	GATGAGTGCGGAACCAAGTC	CGG	20	Anti-sense
β2-tubulin	CAATGCGGTAACCAGATCGG	TGG	20	Anti-sense
nanos	TGTTGCAAATCCCCCCTCGT	TGG	20	Sense
zpg	GCTGCACTCCAAAGTCACAG	TGG	20	Anti-sense

5.6.3 Generating Transgenic Lines

Several strategies were pursued in order to generate transgenic lines, the results of which are recorded in Table.5.10. Initially, the MK1 variant of the drive construct was injected into embryos from line 51323, or 54591, stably expressing Cas9 under a *vasa* or *nanos*-derived promoter, integrated on the X-chromosome.

Target Locus	Construct	Line	No. of Eggs Injected	No. of G0 Adults Recovered	No. of Founders
vasa	vasa-MK1	51323	1400	359 (25.6%)	0
vasa	vasa-MK2	54591	1250	211 (17.5%)	0
zpg	zpg-MK2	54591	600	188 (31.3%)	0
vasa	vasa-MK2 (with gRNA helper plasmid)	54591	400	132 (33%)	0
vasa	vasa-MK2	54591	499	127 (31.8%)	0

TABLE 5.10: Details of initial microinjections, with corresponding targets and recovery rates

Whilst microinjections provided substantial numbers of surviving G_0 adults, when crossed to w^{1118} flies as single crosses, MK1 drives were thought to be impractical, as eGFP reporter cassettes would be only expressed in the germline and difficult to screen. MK2 drives, containing additional 3xP3:CFP:SV40 reporter cassettes did not give rise to transgenic individuals expressing a fluorescent reporter. It was thought that by modulating the promoter controlling expression of Cas9, it would be possible to achieve higher levels of transgenesis, hence the decision to move from line 51323 and a *vasa*-derived promoter to line 54591, with a *nanos*-derived promoter. Additionally, a set of injections was performed with the addition of a pCFD3 derived sgRNA helper plasmid, stably expressing the *vasa* locus gRNA. This was performed with the notion that despite initial proof of concept testing in cell culture, igRNA expression was insufficient to mediate successful transgenesis.

An alternative approach was subsequently taken, providing a transient plasmid source of Cas9, under a *nanos* promoter. Microinjections were performed in w_{1118} individuals, offering a line with minimal background interference, whilst being amenable to the detection of low-level fluorescence from weakly expressed reporter proteins (Table.5.11).

Target Gene	Construct	Line	No. of Eggs Injected	No. of G0 Adults Recovered	No. of Founders
vasa	vasa-MK2 (+ pnos-Cas9)	w^{1118}	400	127 (31.8%)	0
vasa	nanos-MK2 (+ pnos-Cas9)	w^{1118}	400	135 (33.8%)	0
Rcd-1r	Rcd-1r-MK2 (+ pnos-Cas9)	w^{1118}	400	152 (38%)	13 (3.25%
bam	bam-MK2 (+ pnos-Cas9)	w^{1118}	400	134 (33.5%)	1 (0.25%
zpg	zpg-MK2 (+ pnos-Cas9)	w^{1118}	400	140 (35%)	0

TABLE 5.11: Improved microinjections with corresponding targets and recovery rates

As before, G_0 individuals were crossed to w^{1118} flies, and the resulting G_1 screened for CFP fluorescence, by fluorescence microscopy. In this instance, a small number of putatively transgenic individuals were observed for the *Rcd-1r* and *bam* target sites. Where this occurred, G_1 individuals were segregated by founder, and subjected to pooled genotyping to determine the nature of any putative transgenic insertion.

Primers were selected corresponding to the left and right junctions between the transgene and the surrounding insertion sites, at the *Rcd-1r* and *bam* loci. Additionally, primers were designed, corresponding to the unmodified target locus, binding to sites of the homology regions. For both *Rcd-1r* (Fig.5.6), and *bam* (Fig.5.7), all resulting G_1 lines were found to have the correct transgenic insertion. In all cases, this was confirmed by Sanger sequencing amplicons to determine sequence validity. In the case of *Rcd-1r* transgenics, a line derived from a single founder was taken forward for further characterisation, whilst other candidate lines were maintained as reserves. The single *bam* transgenic line was also taken forward for further characterisation. Importantly this result demonstrated that the igRNA cassette screened *in vitro* in the previous chapter do indeed function in whole *D.melanogaster*, providing additional validation.

In addition, genotyping was performed to correlate intensity of observed fluorescence with copy number. Fluorescence patterning and intensity has been previously used in insect models as a measure of transgene copy number (Kyrou et al., 2018). Pools of five individuals, expressing CFP brightly or dimly were analysed by the same genotyping. Both bright and dim pools were shown to contain individuals carrying correctly integrated copies of the *bam* drive transgene, however the dim pool gave rise to an amplicon corresponding to the wild-type target locus, whereas the bright pools did not, indicating the heterozygosity and homozygosity respectively.

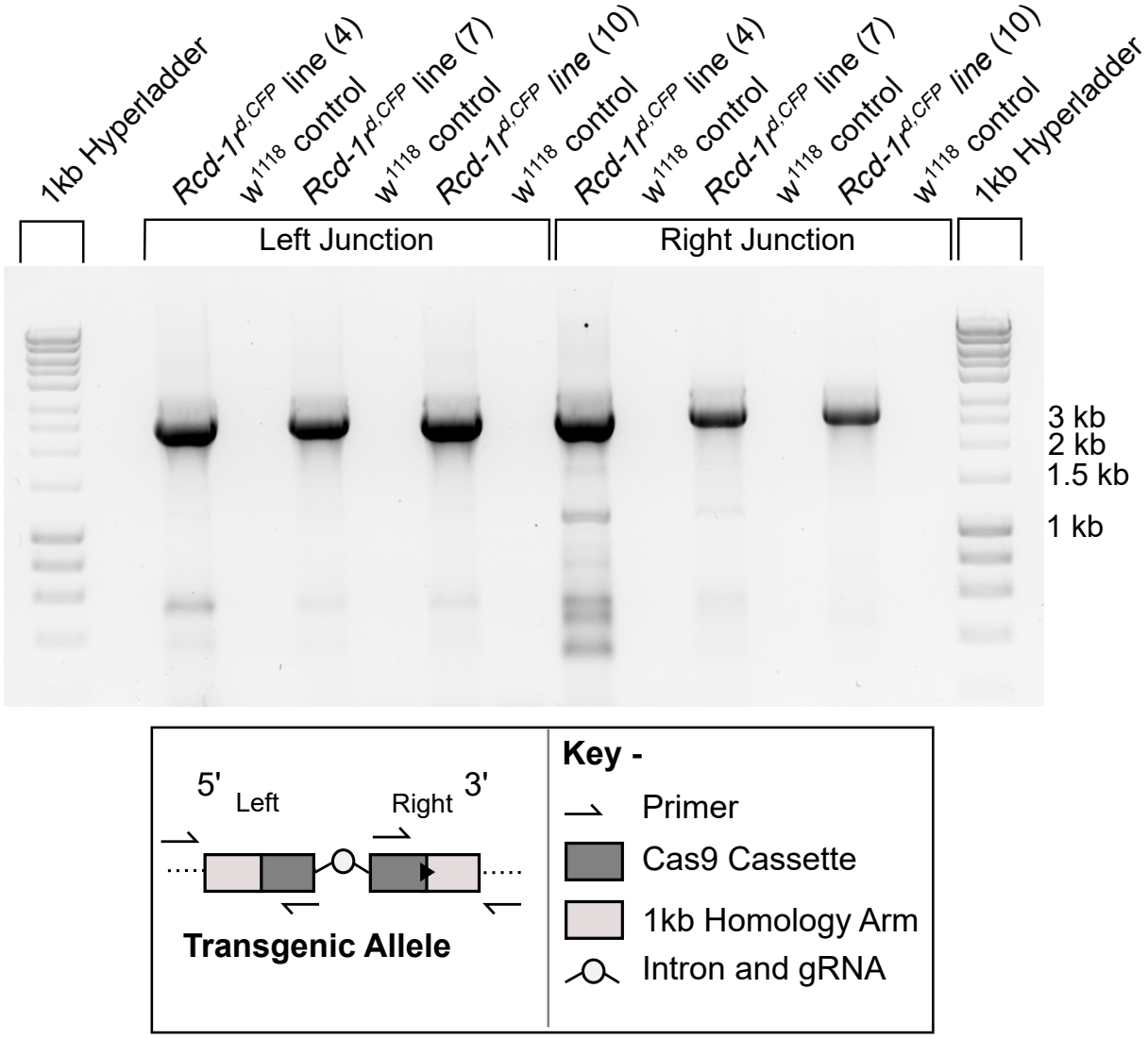


FIGURE 5.6: **Genotyping - *Rcd-1r*^{d,CFP} Founders**. Having recovered G_0 progeny from microinjections, surviving individuals were crossed to a w^{1118} line in single crosses. Resulting G_1 offspring were screened for CFP expression. Where positive expression of CFP was identified, lines were sampled and genotyped, using genomic DNA extracted from w^{1118} individuals as a negative control. Here, the left and right insertional junctions of the *Rcd-1r*^{d,CFP} drive component were assayed, including the homology arm, with the correct amplicon being produced (2209bp and 2658bp respectively). By using one primer in each pair that binds outside the homology arm, it was possible to discount false positives generated by amplification of injected drive construct, which may have persisted.

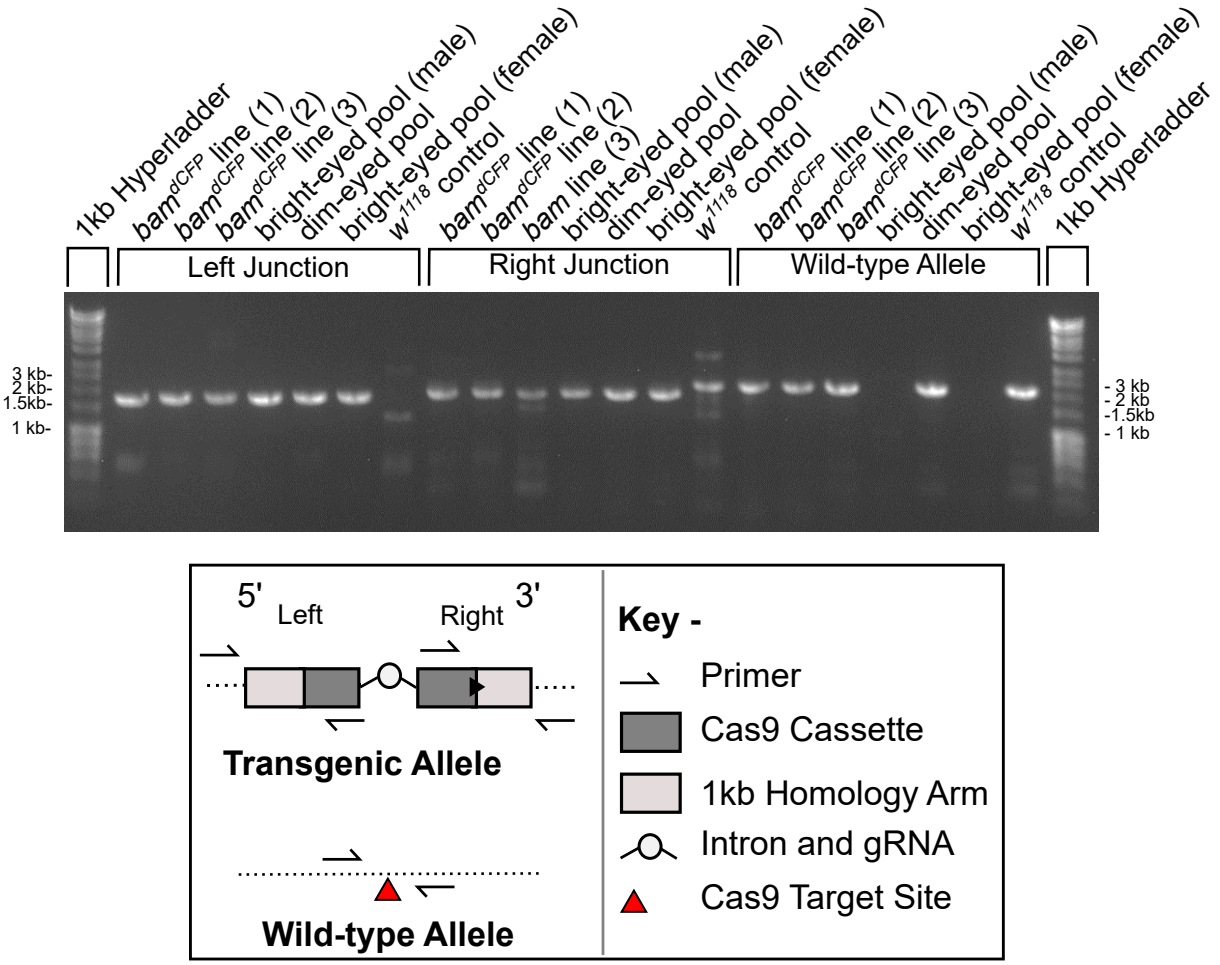


FIGURE 5.7: **Genotyping - $bam^{d,CFP}$ Founders and Homozygote Pools**. Three pools of five individuals, exhibiting the CFP reporter were taken from the *bam* G_1 population and genotyped. Amplicons for the left junction, right junction, and wild-type locus were 1716bp, 2165bp, and 1834bp respectively. Analysis was also performed on G_1 pools to determine a link between fluorescence intensity of the CFP reporter, and copy number of the transgene. As with *Rcd-1r*, by using one primer in each pair that binds outside the homology arm, it was possible to discount false positives generated by amplification of injected drive construct, which may have persisted. In the case of the unexpected band in the wild-type control lane of the right junction PCR set, it is likely this is overflow from the adjacent wild-type PCR set.

In order to obtain stable intermediary stocks, $Rcd\text{-}1r^{d,CFP}$ and $bam^{d,CFP}$ transgenic G_1 males were crossed to balancer lines (Fig.5.8). $Rcd\text{-}1r^{d,CFP}$ males were crossed to a w^{1118}; *if*/SM6a balancer line, whilst $bam^{d,CFP}$ males were crossed to a w^{1118}; MKRS/TM6B, Tb^1 balancer line (Fig.5.10). Resulting offspring were identifiable by *curly* and *tubby* phenotypes respectively.

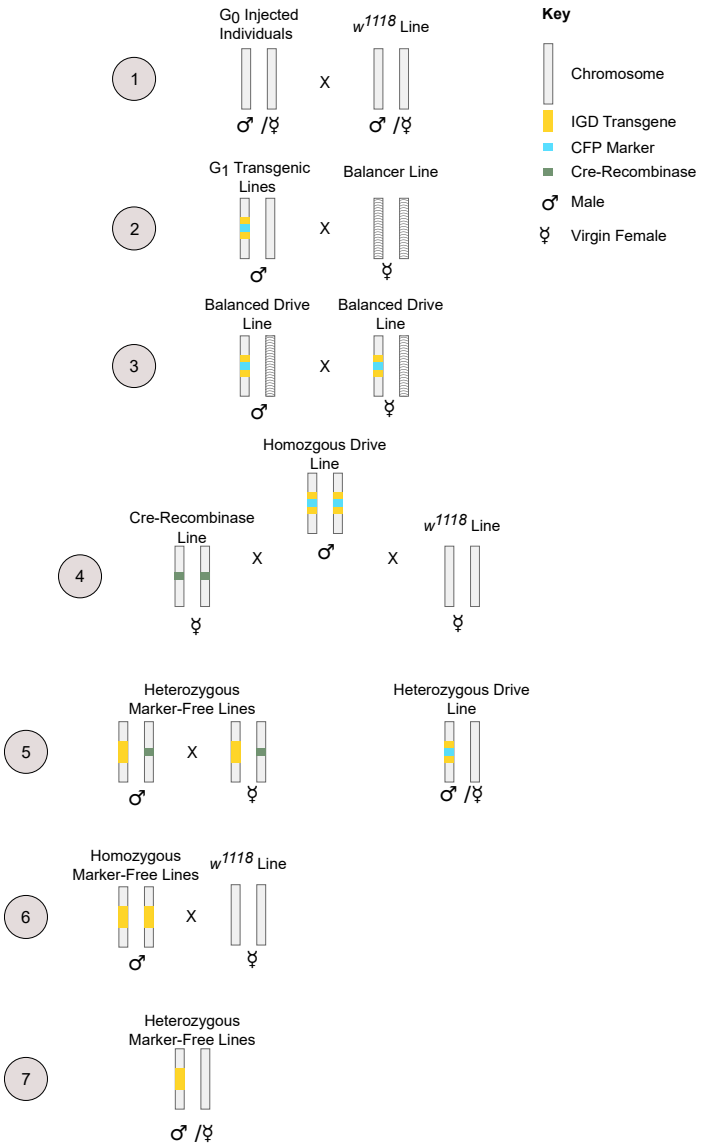


FIGURE 5.8: **Crossing Scheme - Overview**. For both *Rcd-1r*^{d,CFP} and *bam*^{d,CFP} lines, the aim was to establish balanced stocks, with appropriate dominant markers for the balancer chromosome. Subsequently, this would permit generation of homozygous lines bearing the CFP reporter, whilst simultaneously permitting additional steps to excision of it. The markerless line was to then be maintained as a homozygous line. Importantly, this was not feasible for the *bam*^{d,CFP} due to sterility. Steps were performed as follows: 1) Generation of transgenics; 2)Balancing to produce stable stocks; 3) Crossing of balanced heterozygotes to produce transgenic homozygotes; 4) Crossing of homozygotes with a Cre-Recombinase line to excise the CFP reporter cassette; 5) Crossing of heterozygous marker-free lines, screening for the absence of a dominant marker associated with the Cre-locus; 6) Subsequent generation of markerless homozygotes and 7) markerless heterozygotes.

A detailed schematic of all crosses undertaken in this chapter is given in Fig.5.9. Note that whilst further crosses were originally proposed for the *bam*^{d,CFP} transgenic line, sterility observed in homozygotes limited the future potential of the line for use in an IGD strategy, and *Rcd-1r* drive lines were prioritised instead. This was observed when 12 independent single-crosses of homozygous *bam*^{d,CFP} individuals was conducted, and no offspring were recovered.

Given that the CFP reporter cassette represents a substantial insertion within the *ftz* intron within the $bam^{d,CFP}$ drive component, it was hypothesised that cryptic splicing events may have been disrupting endogenous *bam* expression. To circumnavigate this issue, $bam^{d,CFP}$ males were crossed to a line expressing a Cre-recombinase (Bloomington stock 1501), in order to excise the reporter element. Markerless progeny were crossed as before in 10 independent single-crosses, however once again no offspring were recovered. This observation suggestions that the *bam* locus may not be amenable to modification in this manner.

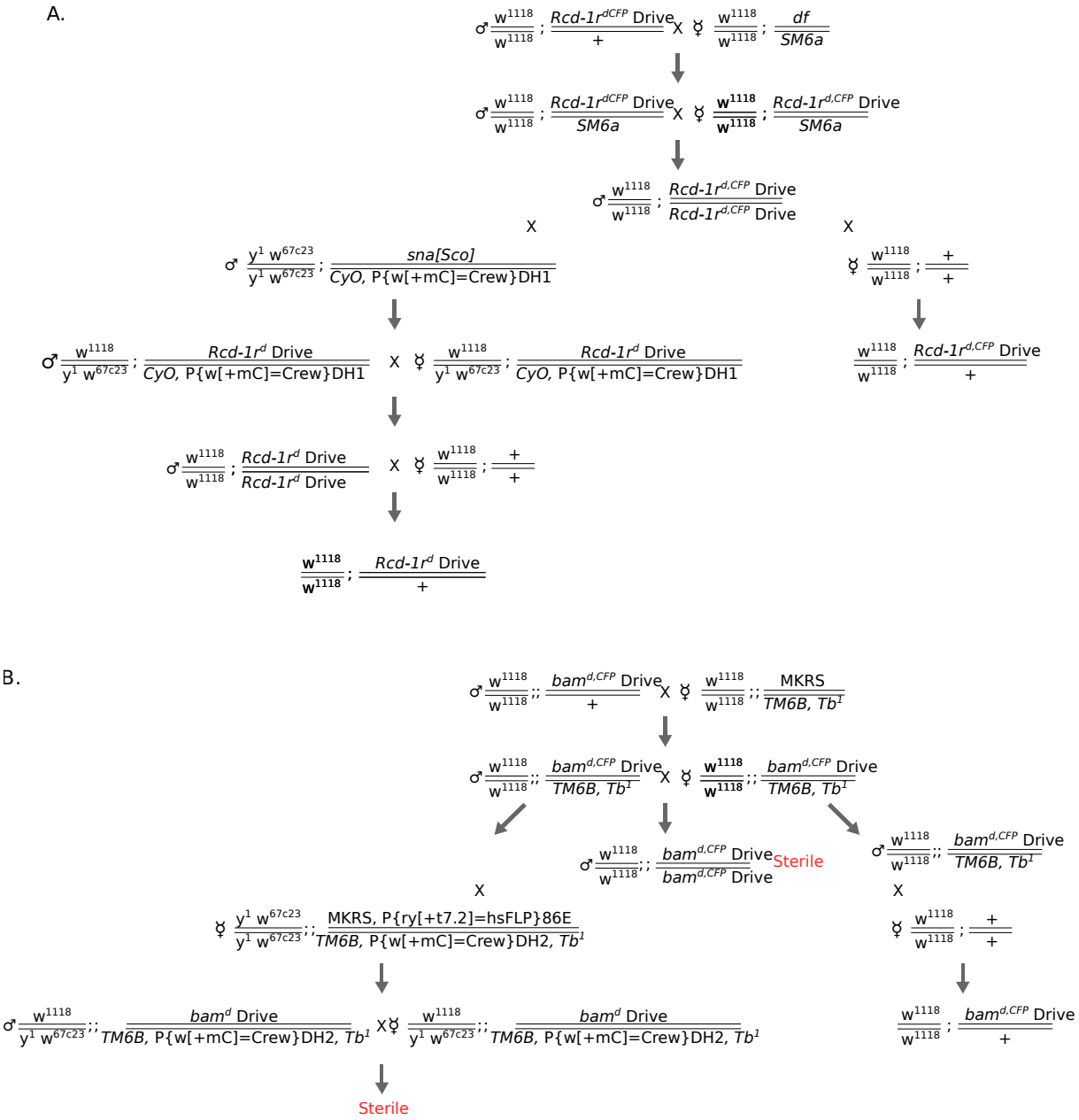


FIGURE 5.9: **Detailed Crossing Scheme - *Rcd-1r* and *bam* Transgenic Lines**.Expanding on the general scheme detailed in Fig.5.8, genetic details of proposed crossing schema are detailed here.

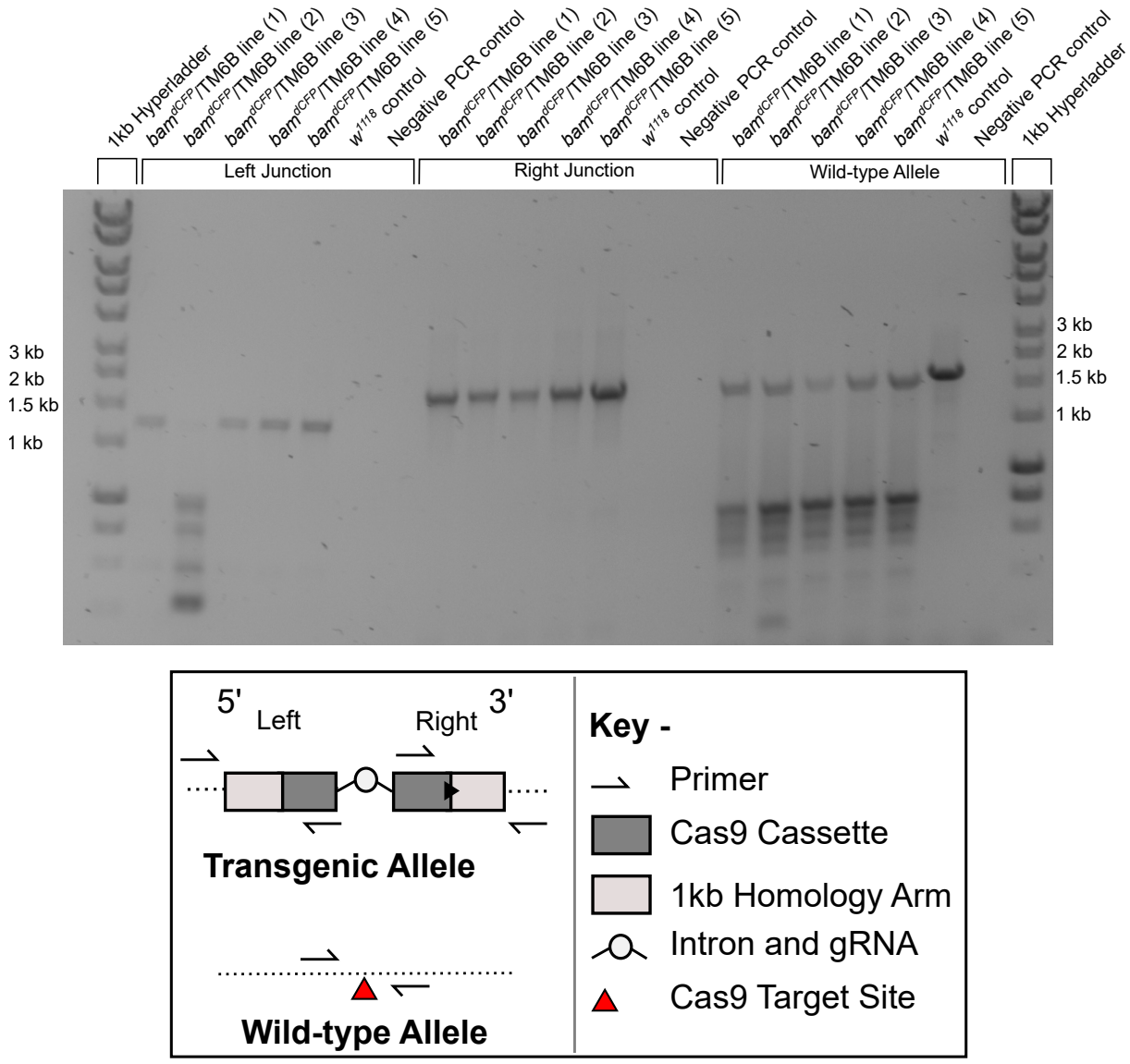


FIGURE 5.10: **Genotyping - *bam*/TM6B Balanced Lines**. Following discovery of sterility, lines were genotyped to confirm insertion of the *bam* IGD drive component, in the *bam* locus. Five single crosses were performed of $w^{1118;;bam_{d,CFP}}$ and $w^{1118};;$ TM6B, Tb[1 , and phenotypically correct progeny were screened from each in pools of 5 individuals. Aside from line (2), correct amplicons were identified - cross (2) was discarded. Desired bands were from left to right - 1716bp, 1834bp, and 2165bp.

By contrast, it was possible to phenotypically screen *Rcd-1r*^{d,CFP}/Sm6a offspring, and cross them in order to produce *Rcd-1r*^{d,CFP} homozygotes lines. Homozygotes were identifiable by screening offspring for individuals displaying CFP fluorescence, but lacking the Curly phenotype associated with the SM6a balancer chromosome. Putative homozygotic *Rcd-1r*^{d,CFP} individuals were crossed individually, and recovered for genotyping (Fig.5.11) as an additional screening step. *Rcd-1r*^{d,CFP} homozygote cross (7) was selected as the basis for establishing an *Rcd-1r*^{d,CFP} homozygotic line.

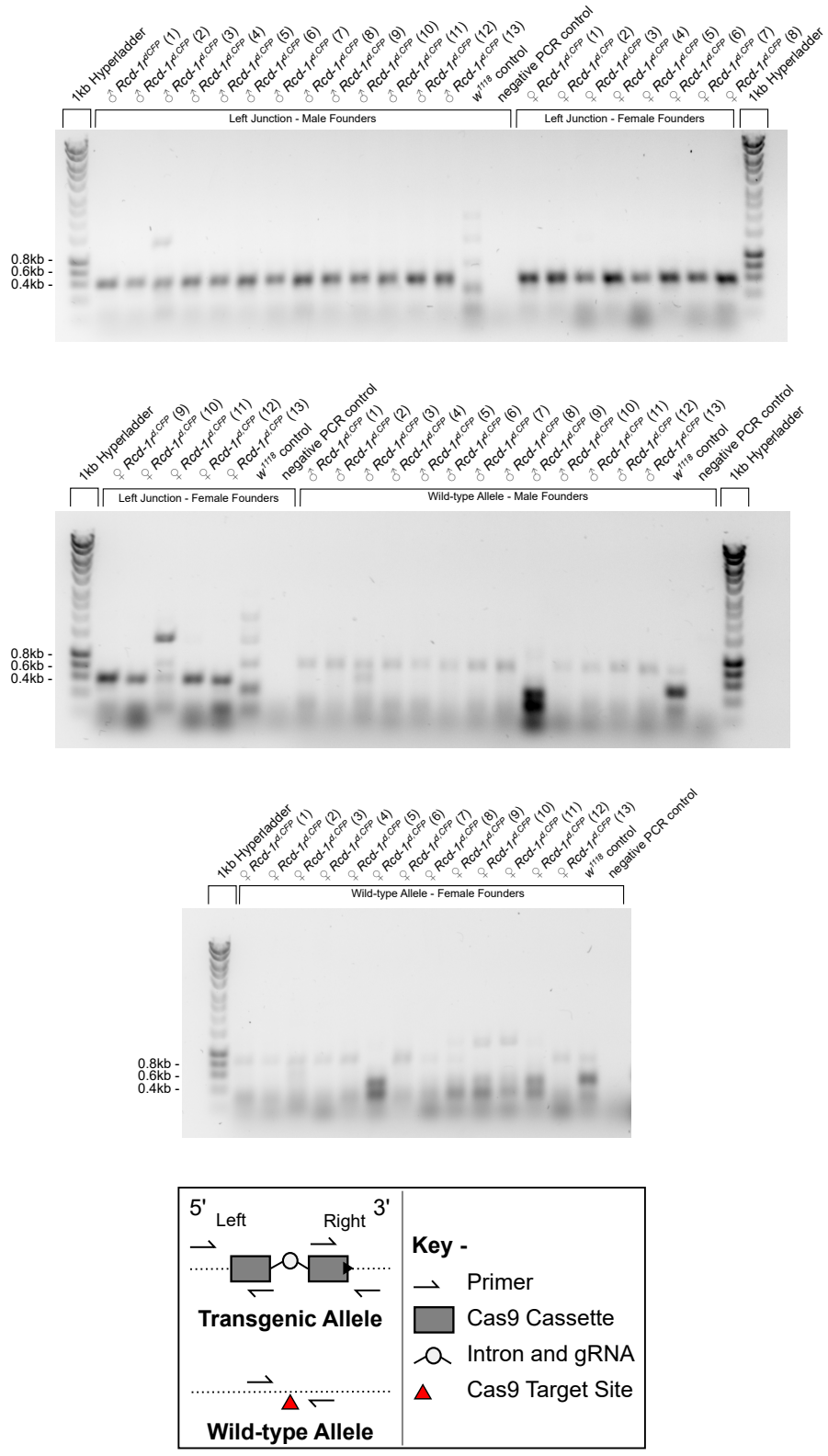


FIGURE 5.11: **Genotyping - *Rcd-1r^{d,CFP}* Homozygote Drive Lines**. Balanced *Rcd-1r^{d,CFP}*/Sm6a individuals established from each of the 13 founding G_0 individuals recovered from microinjection, were crossed to one another in single crosses. Subsequent progeny were screened for the absence of *curly* phenotype, and recessive Sm6a markers. All progeny exhibited intense CFP fluorescence. Parents were individually genotyped; screening of the left and right insertional junctions yielded clear results, however amplification of the wild-type locus offered a higher degree of ambiguity. A single line (7) was selected due to the clear absence of the wild-type allele, and phenotype was revalidated. The desired amplicons, from left to right were as follows - Left - 609bp; Right - 586bp; Wild-type - 517bp. Nonspecific banding is particularly visible when screening for the wild-type allele, and where bands corresponded to approximately the same size as the wild-type (517bp) lines were immediately discarded.

5.6.4 Cre-Lox-Mediated Removal of a Reporter Cassette

Having established that it was possible to mediate super-Mendelian levels of drive transmission in the male *Rcd-1r*^{d,CFP} drive line, it was necessary to excise the CFP reporter cassette contained within the *ftz* intron of the drive component. As we have previously suggested (See Chapter 3), we predict that reporter genes will be insufficient means to track the spread of gene drives through wild-type populations in the field, posing the real threat of gene drives acting unnoticed. To this end, we posit that molecular screening is necessary to drive down the risk of false negatives. Additionally, reduction of the intron size, by excision of the cassette, is liable to reduce the presence of cryptic splice sites that may negatively affect splice efficiency. This observation has been confirmed in *An.gambiae* IGD effector lines, where the presence of the CFP cassette has resulted in diminished expression of a downstream endogenous gene, only for expression to be restored upon excision of the cassette (Astrid Hoermann, Personal Communication).

Briefly, *Rcd-1r*^{d,CFP} homozygotes bearing the CFP reporter cassette were crossed to Bloomington line #1092. The G_1 virgin females, and freshly eclosed males from this cross were subsequently crossed to one another generating *Rcd-1r*^{d}*/Cre* heterozygotes. The resulting G_2 progeny were screened for the absence of any marker, indicative of successful excision of the CFP reporter cassette, with individuals of this lineage hereafter referred to as *Rcd-1r*^{d}. Aside from phenotypic screening, G_1 and G_3 offspring were genotyped to ensure successful establishment of *Rcd-1r*^{d}*/Cre* heterozygotes (Fig.5.12), and *Rcd-1r*^{d} homozygotes (Fig.5.13).

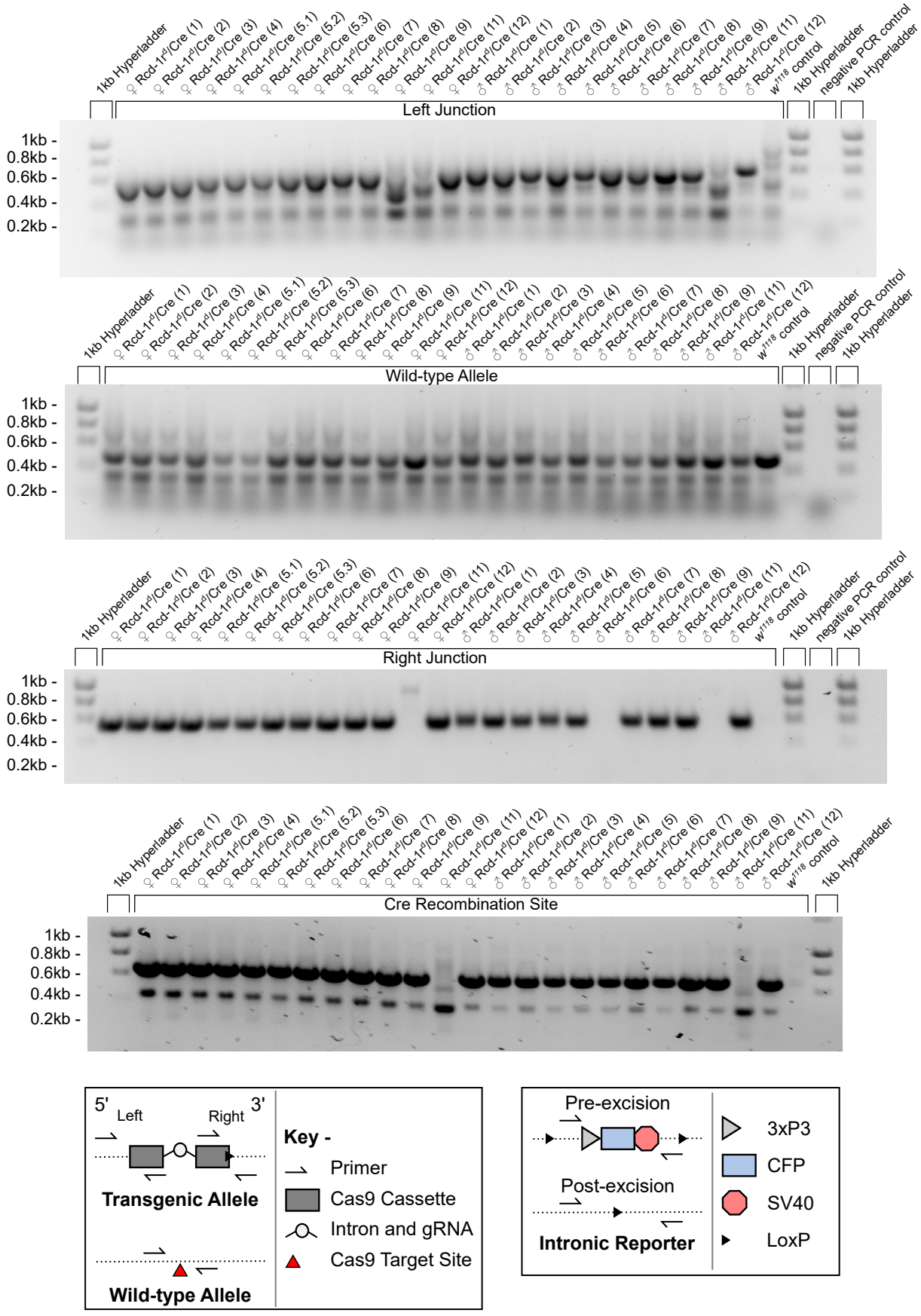


FIGURE 5.12: **Genotyping - *Rcd-1r^{d}*/Cre Heterozygote Lines**. Genotyping was performed of *Rcd-1r^{d}* /Cre individuals in order to confirm generation of the line. 12 individual crosses were performed, and parents were genotyped for each, confirming presence, or absence of the 5′ and 3′ insertional junctions, as well as an amplicon corresponding to the excised CFP cassette. Where bands were not produced, lines were discarded. The desired amplicons from left to right were: Left - 609bp; Right - 586bp; Wild-type - 517bp; and Cre Recombination Site - 694bp.

It is important to note that both Bloomington lines #1501, and #1092, carried the recessive y^1 mutation. It has been long established that pigment mutations can result in altered phenotypes and behaviours (Sturtevant, 1915). Recent work has identified that *Drosophila* carrying null mutants for the recessive *yellow* gene undergo melanisation of the male sex combs (Massey et al., 2019). This results in lowered mating success, as the y^1 phenotype causes males to have difficulty in mounting females during copulation. To avoid the y^1 allele affecting the mating ability of lines in this work, G_2 *Rcd-1r*^{d} individuals with wild-type pigmentation were selected for, and crossed to maintain *Rcd-1r*^{d} lines that are phenotypically wild-type in this respect. All fly lines employed in this thesis carry *white* mutations, resulting in the loss of the characteristic brick red eye pigmentation - again, the *white* phenotype has been linked to behavioural changes in mating (Krstic, Boll, and Noll, 2013). However, on a practical level, fluorescent reporters are more difficult to observe in the wild-type, red-eyed phenotype. In this work these behavioural changes, which are liable to negatively affect mating performance, have been accounted for by consistently ensuring all lines are w^-, effectively bootstrapping performance changes.

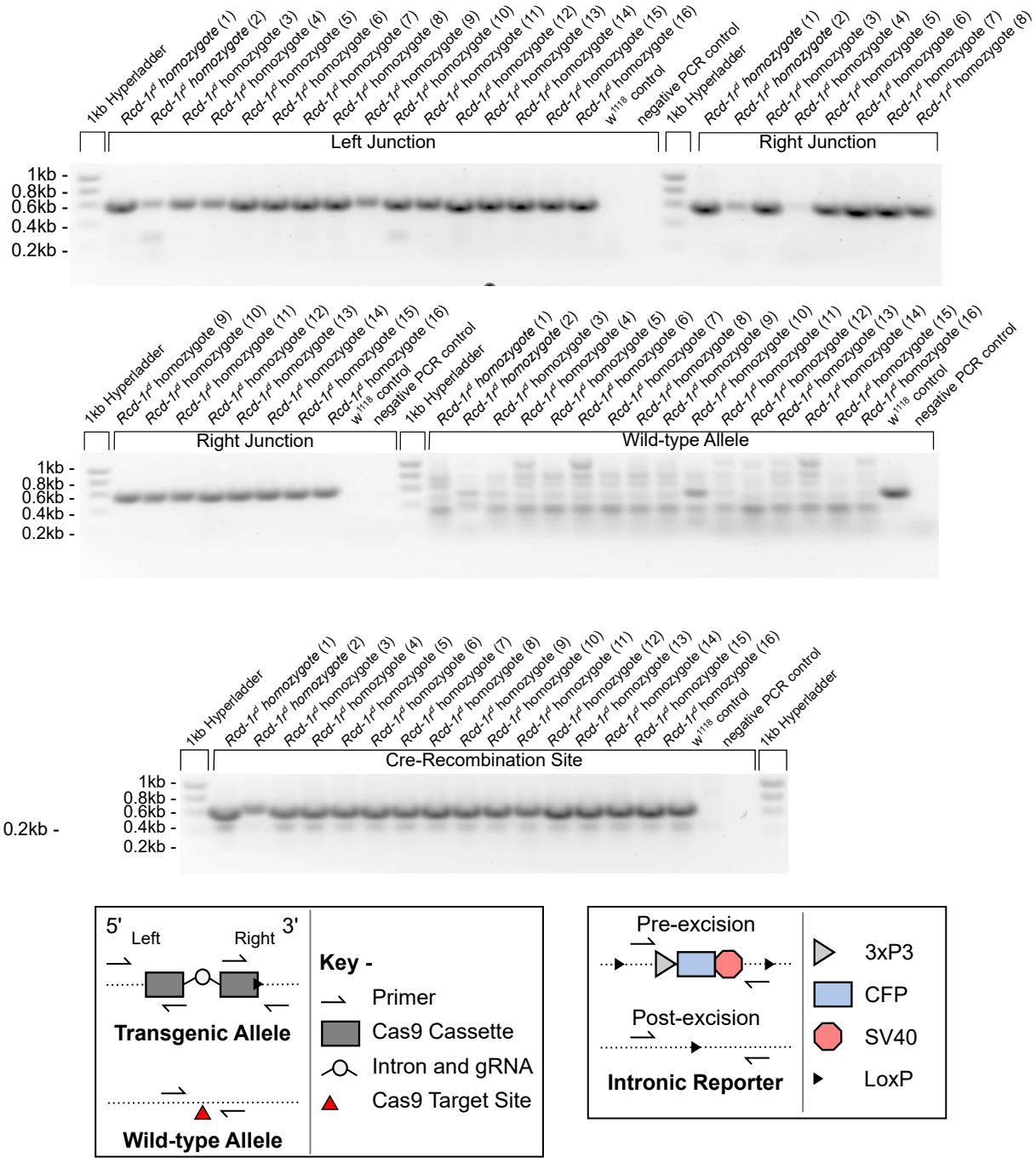


FIGURE 5.13: **Genotyping - *Rcd-1r^d* Homozygote Lines**. In order to generate homozygous *Rcd-1r^d* lines, following excision of the CFP reporter cassette, individual crosses of *w*;*Rcd-1r^d* /Cre flies were performed. Resultant offspring were screened for the *curly* phenotype, associated with the Cre allele. Putative *Rcd-1r^d* homozygotes were crossed singly, and were genotyped following egg laying to confirm the homozygosity of each line. Non-specific banding when screening the wild-type allele was validated by confirmation of the phenotype of each given line. Line (7) was found to have a clear genotype (as well as being phenotypically correct), and was taken forward for further experimentation. The desired amplicons from left to right were: Left - 609bp; Right - 586bp; Wild-type - 517bp; and Cre Recombination Site - 694bp.

5.6.5 Development Assays

In order to determine whether *Rcd-1r*^{d,CFP}, *bam*^{d,CFP}, and *Rcd-1r*^{d} lines had been impacted by the insertion and activity of their respective drive components, development assays were performed, comparing survival rates at selected developmental stages to those observed in w^{1118} individuals. Transgenic lines were assayed by crossing 30 male or virgin females to an equal number of w^{1118} individuals of the opposing sex, and of the same age. Resultant eggs were collected and allowed to hatch in individual vials, in groups of 50-150. The number of eggs, eclosed eggs, pupae, and adults (both male and female) was quantified and recorded (Fig.5.14). It was subsequently possible to determine developmental differences between transgenic lines, in both males and females. In the case of *bam* transgenics, only marker-bearing heterozygotes were assayed. As previously noted, when crossed in 12 independent single-matings, homozygotes produced no offspring, and so were excluded due to time constraints. Similarly, it was not possible to generate markerless lines as *bam*/Cre individuals, whilst viable, were unable to produce offspring when crossed in 10 independent single-matings.

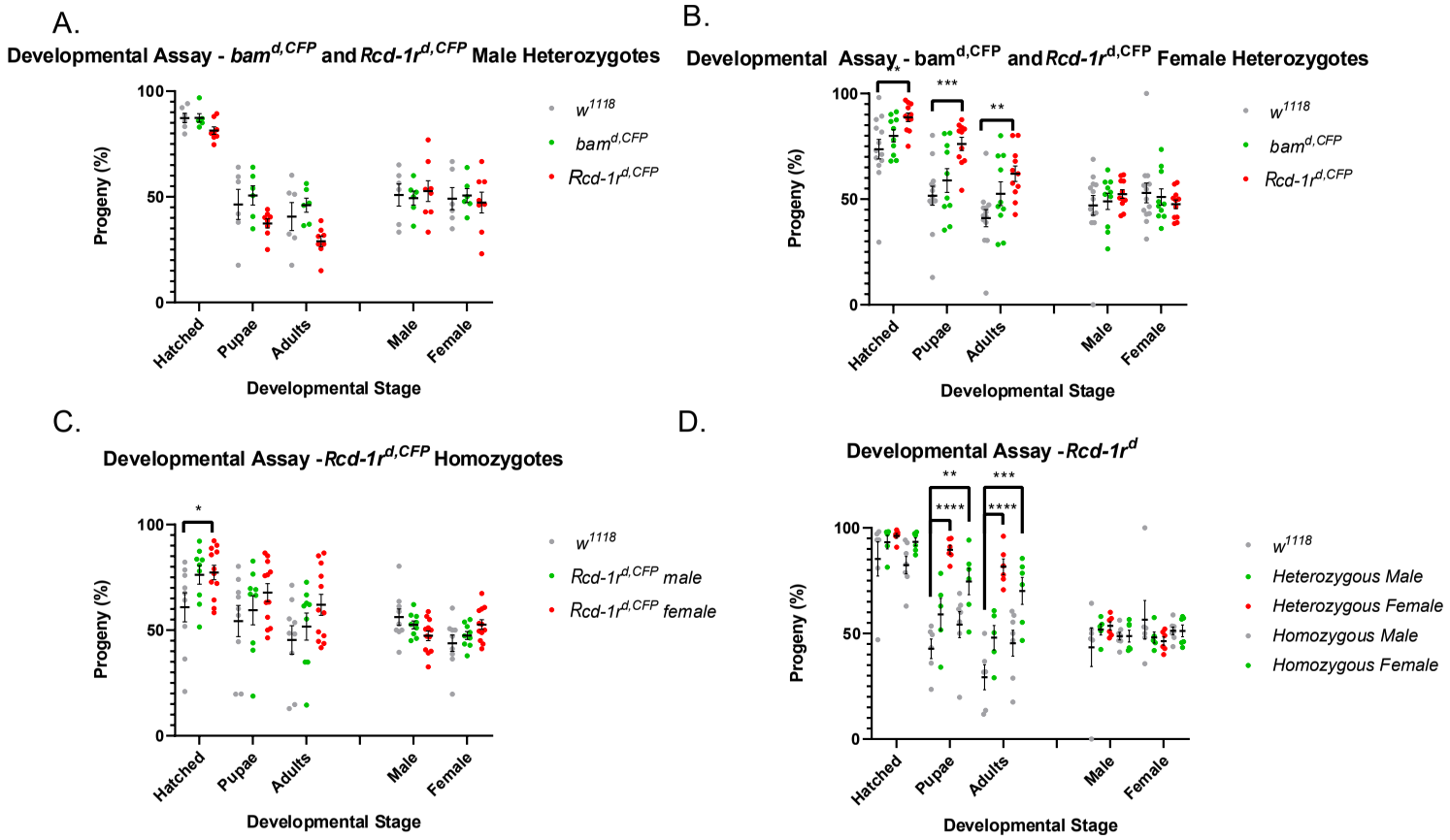


FIGURE 5.14: **Development Assays - Transgenic Lines**. Assays were conducted against a w^{1118} control, comparing rates of hatching, pupae formation, and eclosion as adults. Sex ratio was also quantified to determine if the transgene in each case, was introducing bias. A) No significant differences in development was noted in progeny from $bam^{d,CFP}$ or $Rcd\text{-}1r^{d,CFP}$ male heterozygote crosses. B) The same was found for $bam^{d,CFP}$ progeny. By contrast, $Rcd\text{-}1r^{d,CFP}$ female heterozygote progeny exhibited a significant increase in survival at the hatching, pupal and adult developmental stages (P=0.0.0093, 0.0008 ,and 0.0032 respectively). C) No significant differences were found for progeny of $Rcd\text{-}1r^{d,CFP}$ male and female homozygotes, other than a small improvement in hatching rate for $Rcd\text{-}1r^{d,CFP}$ females (P= 0.0403). D) $Rcd\text{-}1r^{d}$ female heterozygote progeny were found to exhibit statistically increased survival rates for the pupal (P= <0.0001) and adult stages (P= 0.0017). The same was found for female $Rcd\text{-}1r^{d}$ homozygote progeny (P= <0.0001,and 0.0001 respectively). Statistical analysis was conducted using one-way Anova, with Dunnett's post-hoc test. In each case, the wild-type was used as a shared control. ($\star$ $P < 0.05$; $\star\star$ $P < 0.01$; $\star\star\star$ $P < 0.001$; $\star\star\star\star$ $P < 0.0001$).

In every development assay conducted, neither $bam^{d,CFP}$ nor $Rcd\text{-}1r^{d,CFP}$ exhibited a statistically significant reduction in survival rate at any developmental stage. Nor was there a statistically significant difference in the proportions of male and female adults amongst the eclosed offspring. By contrast, heterozygous $Rcd\text{-}1r^{d,CFP}$ and $Rcd\text{-}1r^{d}$, and homozygous $Rcd\text{-}1r^{d,CFP}$ and $Rcd\text{-}1r^{d}$ females exhibited statistically significant increases in survival versus the w^{1118} control line in reaching pupal and adult developmental stages.

Oviposition and Hatching Rate Assays

To further ascertain the fertility of transgenic lines, phenotypic assays were performed, quantifying egg output and hatching rate. Given that the $bam^{d,CFP}$ drive line was previously shown to confer sterility when homozygous in both females and males, oviposition assays were only performed for the $Rcd\text{-}1r^{d,CFP}$ and $Rcd\text{-}1r^{d}$ homozygous lines. Individual homozygotes were crossed in vials with standard yeast food, and allowed to mate for 24 hours. A w^{1118} line was used to establish control crosses. At least 8 replicates were performed per line, with assays performed over 72 hours, flipping mated females into a fresh vial every 24 hours. Egg output was quantified per female, per day for each line (Fig.5.15A). Tubes continued to be maintained at 25°C, and eggs were allowed to hatch, and mature. Three days after eclosion of pupae began, the number of adults in each vial were quantified, and the hatching rate determined as the percentage of eggs that successfully yielded an adult fly (Fig.5.15B). No statistical difference was found between the $Rcd\text{-}1r^{d,CFP}$, and $Rcd\text{-}1r^{d}$ drive lines, when compared to a w^{1118} control line in terms of fertility, or hatching rate.

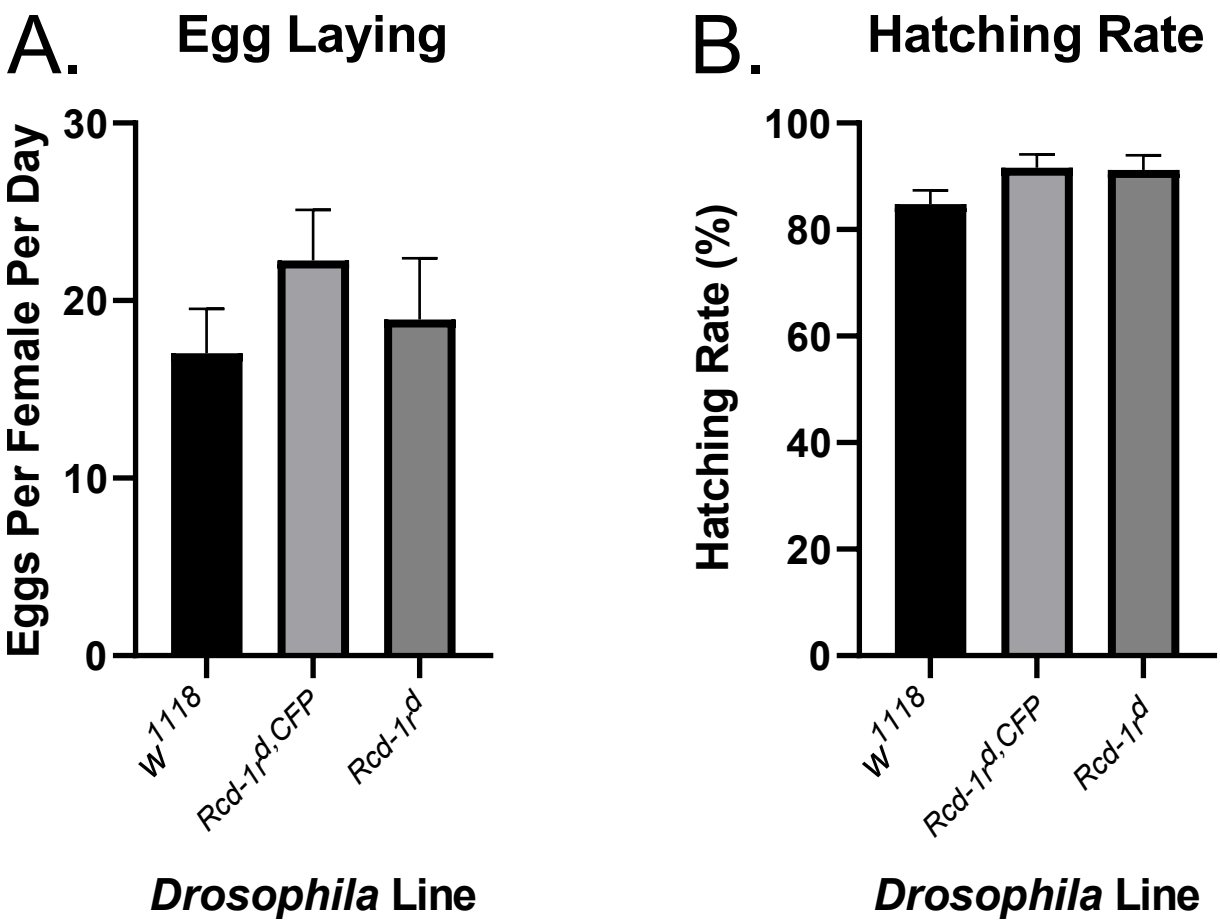


FIGURE 5.15: **Fertility and Hatching Rate Analysis**. A) Oviposition analysis for the *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} homozygous lines, using w^{1118} as a control. Bars represent the mean number of eggs laid per female, per day. Error bars are indicative of SEM between experimental replicates. Statistical testing was performed using the Kruskal-Wallis test, with multiple comparisons. No statistical difference was found compared to the w^{1118} control. B) Hatching rate analysis, given as the percentage of eggs that yield adults. Normality testing was performed, and data was found to be normally distributed. Statistical testing was performed using one-way ANOVA using Dunnett's post-hoc test, with the w^{1118} line as a shared control. No statistically significant difference was found. Error bars are indicative of SEM between experimental replicates.

5.6.6 Homing Assays

To investigate germline homing in both the *bam* and *Rcd-1r* drive lines, a series of homing assays were performed (Fig.5.16). Homozygotes of the *Rcd-1r*^{d} line were crossed to w^{1118} flies, to generate G_1 heterozygotes.It was not possible to obtain fertile *bam*^{d,CFP} homozygotes, so balanced *bam*^{d,CFP}*/TM6B* lines were used instead to produce *bam* heterozygotes. These were subsequently crossed to w^{1118} flies, and homing was determined by CFP expression in G_2 progeny. In order to observe sex-specific expression of the drive construct, heterozygote crosses were performed separately for males and females in both lines.

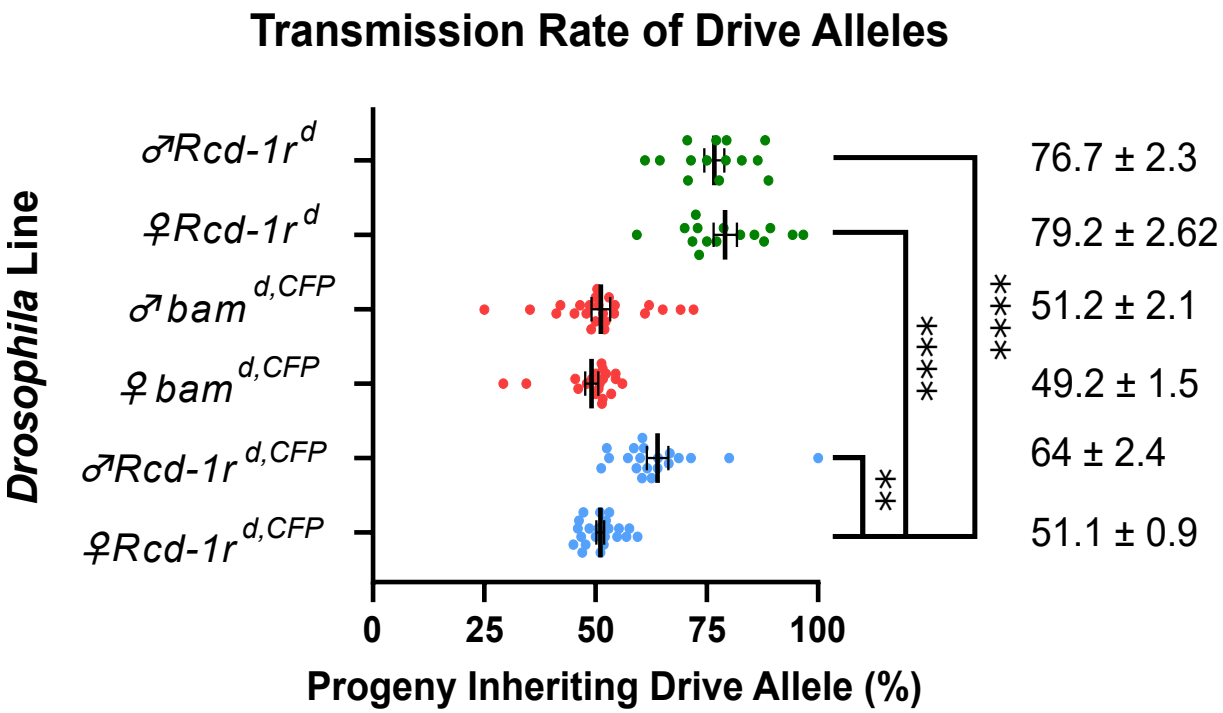


FIGURE 5.16: **Transmission Rate of Drive Alleles**. Transmission rates of IGD drive alleles amongst progeny of $bam^{d,CFP}$, $Rcd\text{-}1r^{d,CFP}$, and $Rcd\text{-}1r^{d}$ transgenic lines. Values were recorded for $\geq$ 20 crosses for $bam^{d,CFP}$ and $Rcd\text{-}1r^{d,CFP}$ lines, and $\geq$ 14 crosses for the markerless $Rcd\text{-}1r^{d}$ line. Vertical bars represent arithmetic averages, with SEM given. These values are given to the right of the graph for each set of crosses. Normality testing was conducted, and data sets were found to be non-normally distributed. Consequently, statistical analysis was conducted with the Kruskal-Wallis test; $\star$ $P < 0.05$; $\star\star$ $P < 0.01$; $\star\star\star$ $P < 0.001$; $\star\star\star\star$ $P < 0.0001$.

The $Rcd\text{-}1r^{d,CFP}$ female heterozygote line was used as a control, given that there is existing evidence that *Rcd-1r* is male specific. Both it, and both male and female $bam^{d,CFP}$ lines exhibit Mendelian levels of transmission. By contrast, Super-Mendelian transmission rates were recorded for the male $Rcd\text{-}1r^{d,CFP}$ heterozygote line, as well as for both male and female $Rcd\text{-}1r^{d}$ heterozygote lines. These findings provide an initial proof-of-principal for the Integral Replacement paradigm, and demonstrate it is feasible to elicit super-Mendelian inheritance of IGD drive components using the *Rcd-1r* locus. Moreover, they validate initial findings from Chapter 3 that it possible to use igRNA cassettes to mediate drive activity by demonstrating success *in vivo*. It is surprising that female $Rcd\text{-}1r^{d}$ elicited super-Mendelian transmission given that, as previously discussed, *Rcd-1r* is thought to be male specific. Further discussion of this is given in the discussion of this chapter.

Genotyping of a Reporterless Transgenic Line

Once markerless $Rcd\text{-}1r^{d}$ lines had been established and genotyped, it was possible to conduct a homing assay to understand how excision of the CFP reporter affected homing performance. $Rcd\text{-}1r^{d}$ homozygote males were crossed to w^{1118} females, to produce $Rcd\text{-}1r^{d}$/+ heterozygotes. Resultant virgin female, and nascent male G_1 heterozygotes were crossed to w^{1118} separately. At least 20 crosses were established for each sex, with a single male crossed to three females.

As before, crosses that failed to produce offspring were discarded - these represented on average less than 10% of all crosses. G_2 progeny were harvested, and frozen in liquid nitrogen to prevent degradation of genomic DNA.

In some cases it was not possible to collect females whilst they could still be confirmed to be virgins. In this case, it is possible that transgenic males may have mated with them, resulting in a source of the transgene (in spermatozoa) being deposited in the spermathecae of these females. It would not therefore be possible to determine if said females were carrying the IGD drive component themselves, or had merely mated with a male carrying the transgene. Subsequently, only male G_2 progeny were used for genotyping, with genomic DNA extraction and PCR reactions being performed in 96 well plates to facilitate rapid screening (Fig.5.17). Having said this, it is unlikely that a single dose of sperm would be sufficient (approximately 1,000 cells) to cause contamination against the background of an entire adult *Drosophila*, consisting of many millions of cells. Molecular genotyping was performed by targeted amplification of the 5' junction of the *Rcd-1r*^{d} drive transgene and the wild-type genome . To limit type II errors, arising as a result of cross-contamination between samples during gDNA extraction, and then during setup of PCR reactions, three negative controls were inserted into the 96-well gDNA extraction plate. Samples were arranged in columns, with the controls positioned at the first, middle, and last wells, and consisting of stock w^{1118} individuals. The same controls were subsequently used as negative biological controls for PCR setup, with two additional controls consisting solely of PCR mastermix added in the final two wells of the plate, acting as template-free negative control. PCR reactions were visualised using gel electrophoresis. Where contamination was observed in any one of the five controls (three biological, two template-free), the PCR reaction for that biological replicate was repeated up to one time, whereupon contaminated replicates were discarded.

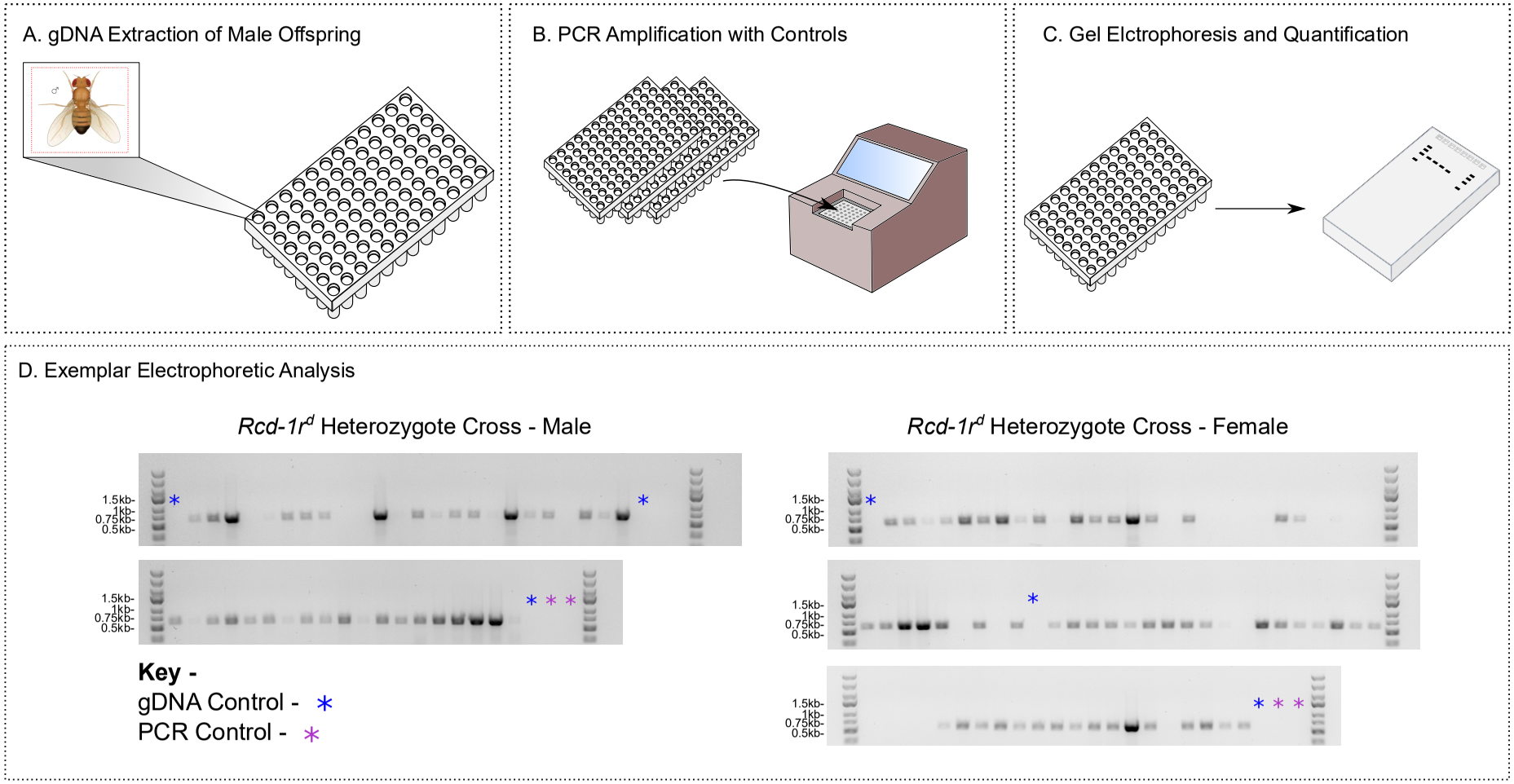


FIGURE 5.17: **Experimental Pipeline for Determining Transmission Dynamics of a Markerless Line**. In order to genotype G_1 progeny, resulting from *Rcd-1r*^{d} x w^{1118} single crosses, a three stage experimental workflow was devised. A) Male flies were selected, and individually stored in 96 well plates, with appropriate wild-type controls. Genomic DNA extraction was performed, using a rapid, low cost methodology using Chelex resin. B) Extracted genomic DNA was subsequently used to act as template in PCR reactions, screening for the 5' *Rcd-1r*^{d} insertional junction. Additional PCR controls were added to the plates. C) Amplified plates were visualised via gel electrophoresis for the presence or absence of the desired 586bp amplicon. An example of such a gel is given D), with appropriate controls labelled.

An observed shortcoming in this pipeline, related to PCR inhibition as a result of Chelex resin remaining in suspension prior to selection of gDNA template for PCR amplification. A consequence is that some reactions failed, resulting in an increase in the false negative rate (Fig.5.18).

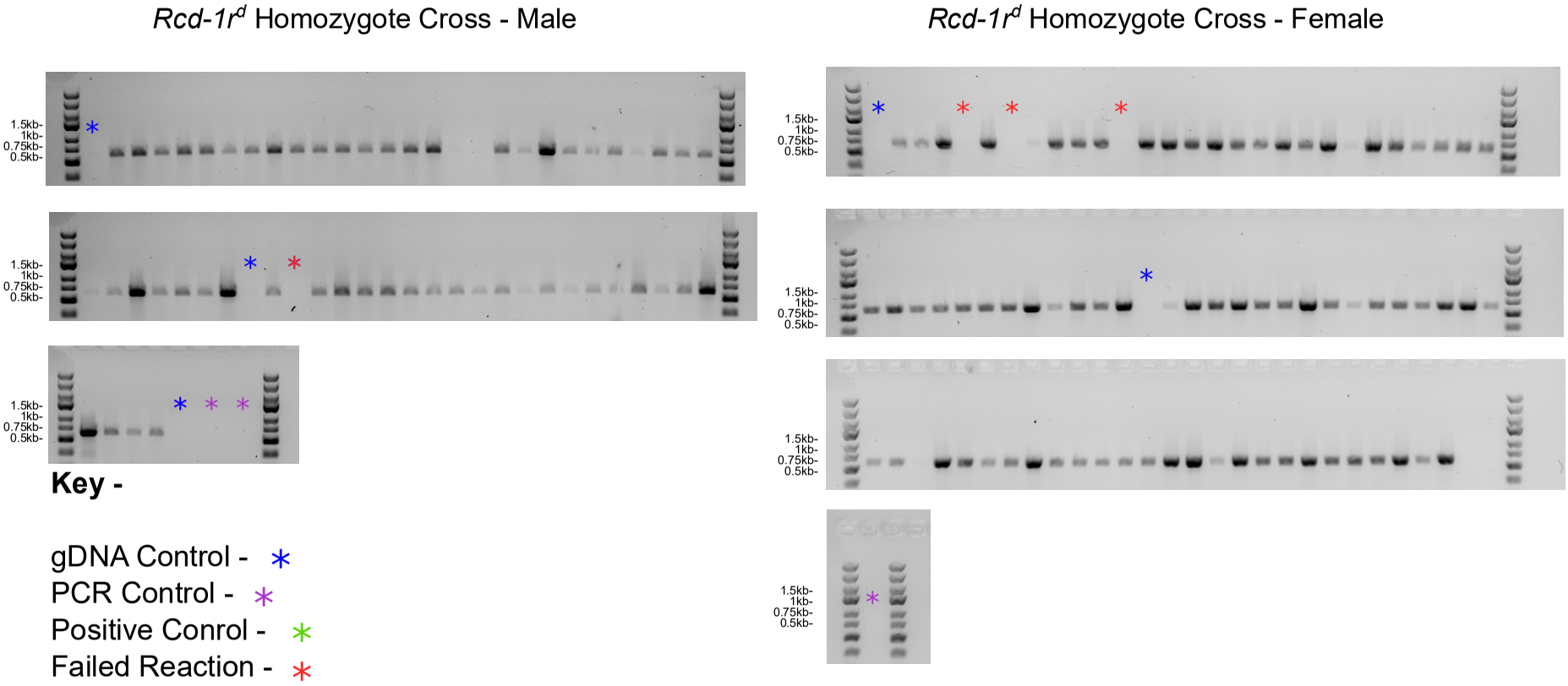


FIGURE 5.18: **Limitations in Genotyping Markerless Individuals**. As before, *Rcd-1r*^{d} homozygote male and females were individually crossed to a w^{1118} control line. Progeny were recovered and genotyped, as previously described. While it was possible to recover amplicons of the correct size (586bp) from $\geq$95% of progeny, some reactions failed. A w^{1118} control gel is shown for comparison.

5.6.7 Investigating Indel Formation

Having established that the *Rcd-1r*^{d,CFP} transgenic line was capable of mediating homing in male transgenics, amplicon sequencing was performed on their wild-type G_2 progeny obtained from the homing assay. This was done in order to gain an understanding of what events occur when the drive component fails to home, and the rate at which indel mutations are formed. The wild-type allele was amplified from twelve pools of G_2 progeny (50 flies per pool), obtained from the *Rcd-1r*^{d,CFP} homing assay, and compared against three pools of w^{1118} flies (200 flies per pool). *Rcd-1r*^{d,CFP} progeny were used because the presence of the CFP reporter allows progeny to be screened without the need for genotyping. Resulting reads were analysed using CRISPResso2 (Clement et al., 2019) in order to determine the rate of indel formation (Fig.5.19A), and qualitatively assess the position of indels relative to the sgRNA target site. Given that each experimental pool contained 50 individuals, that are the G_1 offspring of *Rcd-1r*^{d,CFP} heterozygotes crossed to *w*^{1118} virgins, at least 50% of alleles are liable to be wild-type. Reads representing modified reads would thus be relatively few in comparison, especially if they stem from a single individual. CRISPResso2 analysis was performed to determine the percentage of indels formed at the sgRNA target site (Fig.5.19A). Experimental replicates were subsequently pooled and a qualitative description of the relative positioning and frequency of indels found in all reads was produced (Fig.5.19B). Additional analysis is described in A.2.

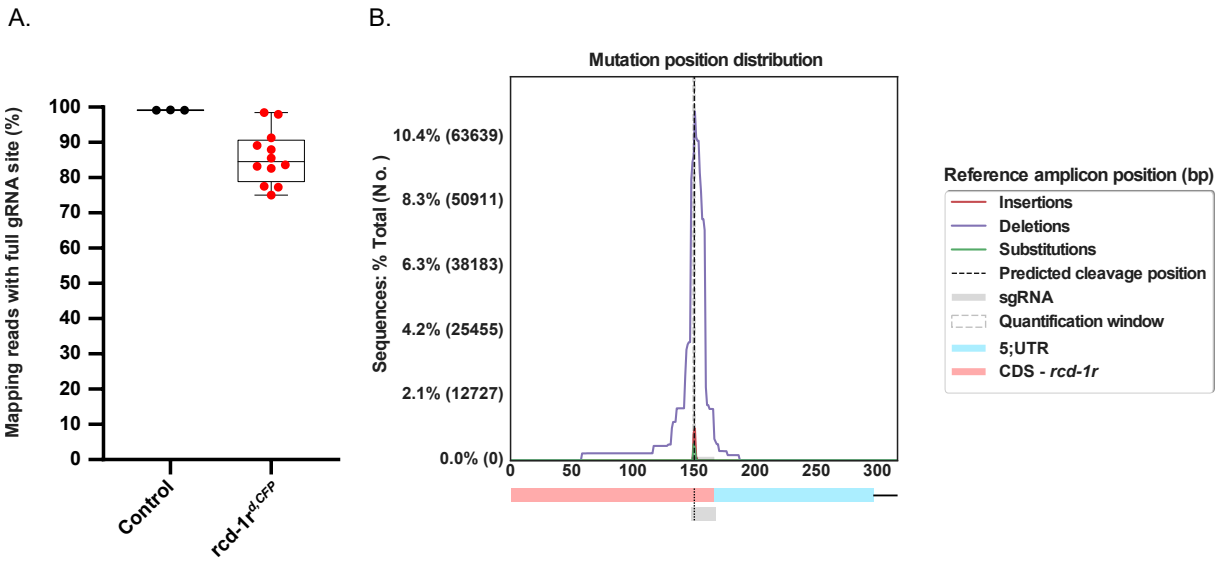


FIGURE 5.19: **Amplicon Sequencing Analysis of *Rcd-1r*^{d,CFP} Offspring.** Progeny of *Rcd-1r*^{d,CFP} x w^{1118} crosses were screened for absence of the drive allele. These individuals were pooled into groups of 50 by cross, and analysed by amplicon sequencing (total n_s = 600). Offspring from w^{1118} mass crosses were used as a negative control, in three pools of 200 (n_s = *600)*. Primers were designed to interrogate the drive gRNA site, in the endogenous *Rcd-1r* locus. A) Reads were analysed for presence of the intact gRNA site. The wild-type control exhibits minimal modification of the target site (99.15%), compared to the experimental samples (85.79%). The number of unique alleles representing disruption of the gRNA target site, represented by B) Analysis was conducted on the total pooled reads from experimental replicates, in order to give a qualitative description of the relative position and density of indel mutations. It is clear that deletions are most favoured, with more than 10% of all pooled reads containing a deletion adjacent to the target site window. Horizontal boxes indicate the inter-quartile range with arithmetic mean. Horizontal bars represent minimum and maximum values.

In wild-type control replicates, on average >99% of reads corresponded to an unmodified gRNA target site, implying high homogeneity. F_1 of the *Rcd-1r*^{d,CFP} displayed an average 13.36% reduction in cleavable gRNA target sites. The allele with the highest average representation across the experimental replicates (4.29% of reads) is a 12-bp, in-frame deletion, with deletions representing the most common mutation as a consequence of gene drive activity. Analysis of the gRNA target site, using a microhomology prediction tool (Bae et al., 2014), identifies this allele as a potential result of microhomology-mediated end joining. This infers that, though the drive-mediated Cas9 activity is still occurring, the true indel formation rate may be lower if this allele is discounted. The fact that cleavage is evidently occurring provides additional validation for the IGD paradigm and supports the notion that it is possible to successfully express igRNAs (as piloted in Chapter 3), and cassettes such as Cas9 using an endogenous germline locus.

5.6.8 Integrating Sex-specific Parameters into the IGD Model

A number of sex-specific promoters have previously evaluated for use in regulating gene drive components (Chan et al., 2013a; Catteruccia, Benton, and Crisanti, 2005). In recent years, their

use has come to the fore again due to the discovery that maternal deposition of Cas9 results in increased rates of resistance allele formation (Marshall et al., 2017; Champer et al., 2019b). By making use of promoters that are only active in the male germline it may be possible to entirely avoid these effects. Moreover, when implementing the IGD strategy in the primary malaria vector, *Anopheles gambiae* sex-specific fertility costs were observed at selected target sites. In order to suitably model sex-specific effects in either scenario, it was necessary to expand the existing IGD three-locus, two-effector implementation. To this end, the model was expanded from 1,000 to 2,000 genotypes, with all parameters duplicated such a parameter sets could be assigned to male and female individuals (Tables A.4 and A.5).

A future step for this work is the execution of a population trial, using the *Rcd-1r*^{d} drive line, in order to determine drive dynamics over multiple generations. In order to better understand a suitable time frame (described in generations), the transmission rate (male - 76.7%; female - 79.2%) and frameshift values (25%) established in this chapter were attributed to appropriate parameters, and the model was simulated for 25 generations (Fig.5.20). Release frequencies were simulated at 5%, 15%, and 25% of the total population, using heterozygous *Rcd-1r*^{d} males, in a background without any pre-existing resistance.

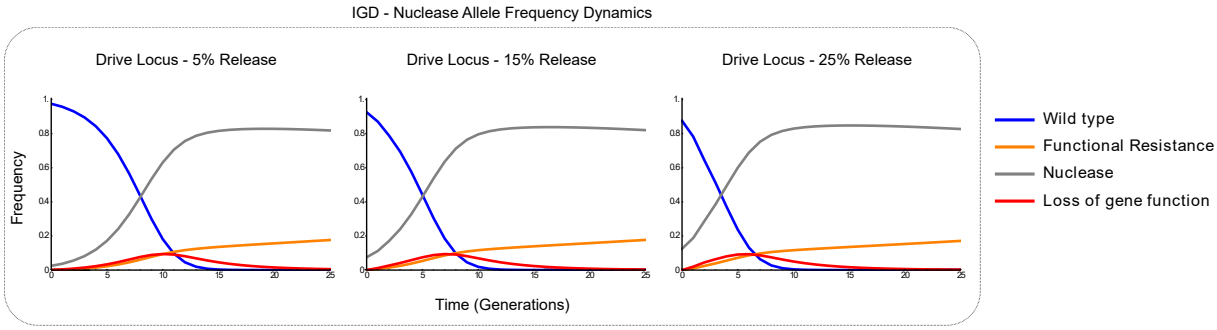


FIGURE 5.20: **Sex-Specific Modelling Predictions**. Experimental transmission and frameshift rates were implemented in a sex-specific iteration of the model. All other parameters were maintained at baseline values (Nash et al., 2019). Simulations were given for $t = 25$ (generations), and for release of drive heterozygotes as 5%, 15%, and 25% of the total population.

In all three of the modelled scenarios, nuclease allele frequency stabilises, and begins to decrease after between 10-15 generations. As the trial commences, and results are analysed, it may be possible to revise these parameters, to increase the quality of fit to better project results from extant generations.

5.7 Discussion

Here we have demonstrated the first proof-of-concept of a minimal integral gene drive, integrated within an endogenous germline locus, and regulated by endogenous regulatory elements. At the time of writing, this is the only experimentally characterised gene drive that makes use of its host's regulatory machinery to mediate expression patterning, and as a consequence represents one of the most minimal design architectures yet described.

Feasibility of Generating Transgenic Lines

The aim of this chapter was the design and construction of IGD drive components against a number of candidate germline loci, to characterise the resulting transgenic lines, and evaluate homing efficiency. The generation of transgenic lines posed a significant challenge, resulting in a number of design iterations against a range of candidate sites, using a variety of injection mixes. It is interesting that no transgenics were recovered from microinjections made into Cas9-expressing *drosophila* lines, despite the use of lines using both the broadly active *vasa* promoter, and more tightly regulated *nanos* promoter (Champer et al., 2017). It is possible that the sites targeted when microinjecting into these lines are not easily amenable to modification - sterile mutations have been identified for both *vasa* and *zpg* (Schüpbach and Wieschaus, 1986), and it is possible that even successful in-frame insertions may have negatively affected the fitness of modified individuals, reducing the probability of success such that transgenics were never observed.

However, it may be also be possible that maternal deposition of Cas9 (especially when using a *vasa* promoter to mediate Cas9-expression) resulted in the generation of indels in the target germline locus. Thus, even if a copy of the transgene was successfully integrated into a target locus, maternal deposited Cas9 (in conjunction with the constitutively expressed gRNA source) may have generated detrimental mutations at the unmodified locus. If this is the case, the use of a transient Cas9-expression plasmid in a neutral genetic background may prove beneficial as was shown to work here.

An alternative explanation pertains to the efficacy of the selected igRNA spacers. Whilst all the chosen igRNAs were able to mediate cleavage of their corresponding target site when assayed *in vitro*, it is a known biological phenomenon that the *in vitro* performance of gRNAs often outstrips the performance of the same gRNAs *in vivo* (Thyme et al., 2016). This has been attributed to cross-reactivity between gRNAs, limiting the formation of Cas9:gRNA complexes, and the genomic context of certain target sites, similarly limiting Cas9:gRNA activity.

5.7.1 Assessing Fitness Costs to Transgenic Lines

The results of the development assays conducted on the $bam^{d,CFP}$, $Rcd\text{-}1r^{d,CFP}$, and $Rcd\text{-}1r^{d}$ drive lines indicate that there was no statistically significant reduction in survival at each developmental stage assayed, when comparing progeny of homozygote and heterozygote transgenic populations against a w^{1118} control line. It is important to caveat this with the understanding that the $bam^{d,CFP}$ homozygote line was found to be infertile, indicating recessive sterility, limiting the use of the line as part of a population replacement strategy. Moreover, as homozygote and heterozygote transgenic populations were being crossed to w^{1118} lines, all progeny would have received at least one rescue copy of the wild-type *bam* or *Rcd-1r* allele. This limits what can be concluded regarding the effects of the IGD drive component integration, as progeny were never homozygous for the transgene. However, from the crosses performed, it was possible

to conclude that when the IGD component is present in heterozygosity, it does not negatively impact developmental fitness.

While it is possible to identify that there are statistically significant increases in survival for the progeny of *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} heterozygous and homozygous females when compared to the control, the data available here is insufficient to resolve the underlying biological phenomenon that causes this effect. It is possible that when establishing the *Rcd-1r*^{d,CFP} line, the founder or an intermediate mate exhibited high-fitness, and those traits are subsequently present in the line and the derivative *Rcd-1r*^{d} line. An alternative mechanism may pertain to the action of the *Rcd-1r*^{d} drive component upon the endogenous *Rcd-1r* locus. Modulation of *Rcd-1r* may result in upregulation of genes implicated in female fecundity - however, this hypothesis seems unlikely given the results of the homing assay that suggest *Rcd-1r* is not active in the *Rcd-1r*^{d,CFP} line.

It is worth noting that existing publications suggest male-mate preference for high-fitness females can impose a fitness cost on these individuals (Long et al., 2009). This may have negative implications for the fitness of transgenic females in a release scenario, where field populations of males harass said females, reducing their fitness towards the mean fitness for the entire population. Nonetheless, the results posited here have positive implications for the *Rcd-1r*^{d} drive line as a vector for the IGD drive component.

The results of the oviposition assay indicate no significant differences in fecundity when comparing the *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} drive homozygous lines to a w^{1118} control line, evaluating average egg laying and eclosing rates, per female, per day. While there is ambiguity as to the fertility of the heterozygous *Rcd-1r*^{d,CFP} line, it is clear from both the development and oviposition assays, that there is no statistically significant reduction in fitness or fecundity for the homozygous lines. This is a positive indication given that these parameters are influential in determining drive efficacy (Beaghton et al., 2017; Beaghton, Beaghton, and Burt, 2017).

Development and oviposition assays were not conducted using homozygous *bam*^{d,CFP} individuals. This is because, when crossing *bam* homozygotes, no progeny were recovered. Phenotypic data available from FlyBase (Thurmond et al., 2019) demonstrated that point mutations in the *bam* locus resulted in infertility in both males and females. The line had previously been sequenced and the *bam* IGD drive component insertion found to be accurate. It is possible that inclusion of the drive component diminished expression levels of *bam*, resulting in infertility in the homozygotes. This may have been a consequence of cryptic/inefficient splicing of the *ftz* intron contained within the SpCas9 cassette of the drive. Given that the *bam*^{d,CFP} line was shown not to home, the *Rcd-1r*^{d} and *Rcd-1r*^{d,CFP} lines were prioritised instead.

5.7.2 Investigating Indel Formation Due to Drive Activity

Results from the amplicon sequencing infer successful Cas9-mediated cleavage of the *Rcd-1r* target site. These findings support evidence from the homing assay that the IGD drive component can function, and is being successfully expressed from within the *Rcd-1r* endogenous locus, with expression regulated by the endogenous promoter elements. When using an unpaired t-test to compare percentages of unmodifed gRNA target sites, there is a significant difference between the control and *Rcd-1r*^{d,CFP} (p = 0.0109). As described above, the most prevalent mutant allele amongst the *Rcd-1r*^{d,CFP} replicates was predicted to be a likely result of microhomology-mediated repair. Formation of such alleles is liable to be site-specific and may infer a lower true indel formation rate for the drive component in this context.

The subsequent pooling and analysis of indels and substitutions in the experimental samples is not without caveats; given that each experimental replicate gave rise to different numbers of reads, some will more represented in the analysis than others. However, it is possible to gain a qualitative understanding of the frequency of each mutation type (insertion, deletion, and substitution) proximal to the gRNA target site. It is clear that regardless of normalisation, deletions are the most common outcome of Cas9:sgRNA mediated activity at the *Rcd-1r* target site. Individual analysis for each experimental replicate is available in A.3. The high proportion of deletions amongst all amplicon reads suggests that improved endonuclease efficacy may significantly decrease the rate at which non-functional alleles occur. This in turn may positive effect the durability of the IGD system in a target population.

A clear limitation of this data set is that while it is possible to infer that cleavage is occurring, it is not possible to quantify whether indel mutations are resulting in decreased survival of G_1 progeny, resulting in a lower than expected proportion of mutations corresponding to frameshifts. As discussed above, one might expect frameshift mutations to occur with a frequency of 2/3, versus in-frame mutations. Given the relatively low rate of indel formation in G_1 progeny, this may be a consequence of low Cas9 activity, poor fertility of progeny with frameshift mutations at the *Rcd-1r* locus, or another biological mechanism that is not distinguishable with this assay.

5.7.3 Analysis of Homing Efficacy and *Rcd-1r* Expression

Transmission rates of the *bam*^{d,CFP}, *Rcd-1r*^{d,CFP}, and *Rcd-1r* IGD drive lines was conducted using male and female transgenic heterozygotes, crossed against a w^{1118} line. The *Rcd-1r*^{d,CFP} cross set was used as a control, as it had been hitherto believed that *Rcd-1r* expression was entirely male-specific. No statistically significant deviation in transmission rates were observed with the *bam*^{d,CFP} lines (in either males or females). This is in keeping with previous studies, where *bam* has been shown to have poor performance driving HEG expression and homing (Chan et al., 2011; Chan et al., 2013a).

By contrast, the *Rcd-1r*^{CFP} line demonstrated super-Mendelian transmission of the drive component in resultant progeny, when using males (64 ± 2.4%). As expected, here females exhibited a near-perfect Mendelian rate of transmission (51.1 ± 0.9%). This is in keeping with our initial understanding of the *Rcd-1r* locus, which had been previously described as being testes-specific (Quezada-Díaz et al., 2010; Chen et al., 2012). The markerless variant however, exhibited unexpected drive activity in both males (76.7 ± 2.3) and females (78 ± 2.5). These results establish an initial proof-of-concept for the IGD strategy, demonstrating that it is possible to mediate homing of a drive component from within an endogenous germline gene, at super-Mendelian rates. The fact that the *Rcd-1r*^{d,CFP} females do not exhibit drive activity, but the *Rcd-1r* females do, may be due to the presence of the reporter cassette in the former, with stronger expression of *Rcd-1r* in males. With the removal of the reporter cassette, expression of correctly spliced drive RNA increases, driving up expression in both males and females up to an upper threshold, beyond which differences in expression of *Rcd-1r* no longer influence drive efficacy.

An important finding from this homing assay was inferred expression of *Rcd-1r* in females of the *Rcd-1r*^{d} drive line. Previous work characterising *Rcd-1r* had determined the gene to be expressed specifically in the testes, with this finding supported by micro-array expression data and phenotypic analysis using RNAi knockdown and *p*-element disruption of the gene (Chen et al., 2012; Quezada-Díaz et al., 2010). The evidence gathered from our homing assay appears to contradict these observations. However, *in-situ* hybridisation and FISH imaging (Lécuyer et al., 2007; Wilk et al., 2016) depicts notable expression of *Rcd-1r* in the embryo, visible from stages 1-3 of development. This positive staining may indicate maternal, or zygotic expression.

It is also of note that while the existing FlyAtlas micro-array data set is poorly indicative of expression of *Rcd-1r* in the adult brain, modENCODE transcriptome data indicates expression in the L3 larval imaginal disc (Brown et al., 2014). Moreover, several studies have made use of *Rcd-1r* (referred to by its alternate name, *zeus*) as a suitable marker for observing mitotic spindles during cellular division (Connell, 2013; Mauri et al., 2014). This suggests that *Rcd-1r* expression may not be fully characterised, and that further studies are required to resolve this discrepancy. A key limitation in this undertaking relates to the genomic context of *Rcd-1r*; it overlaps with a putative gene (*CG13102*), which exists in parallel on the sense strand. While high-throughput data exists, indicating expression of the putative gene in testes (Quezada-Díaz et al., 2010) and midgut (Tomancak et al., 2002a), little is known of its function. Due to the overlapping nature of these two genes and the nature of their expression patterns, it may be difficult to further dissect their expression patterns - further experimentation required to distinguish the two is discussed in detail in Quezada-Díaz et al., 2010. High-throughput RNA-seq data available through FlyBase suggests a possible mistake in annotating the CG13102 locus. If this is the case, gene expression of CG13102 can be discounted, and further expression studies performed, which are expanded upon in Chapter 6 of this thesis.

With regards to the evidence of expression of *Rcd-1r* in the early embryo, stemming from *in situ* hybridisation experiments, consideration must be made for sexual differentiation. In

D.melanogaster, sex is determined by the ratio of X chromosomes to austosomes inherited by an individual. The X:A ratio mediates expression of *Sex-lethal* (*Sxl*), generating an activation loop in embryos with an X:A ratio of 1 (e.g. 2X and 2A chromosomes) that are destined to become female. Embryos with an X:A ratio of 0.5 (e.g. 1X and 2A chromosomes) are destined to become male, and *Sxl* remains inactive. In turn, *Sxl* controls expression of the *transformer* (*tra*) and *msl-2* genes that modulate somatic sex determination and dosage compensation respectively. In embryos destined to become female, the activity of *Sxl* produces a particular isoform of *tra*, which together with Tra-2 modulates downstream expression of *double-sex* (*dsx*) and *fruitless* (*fru*), with *dsx* responsible for female somatic determination. By contrast, embryos destined to become male lack *Sxl* activity, producing a different isoform of *tra*, and in turn male-specific transcripts of *dsx* and *fru* are produced determining male sex differentiation and behaviour (Penalva and Sanchez, 2003) (Fig.5.21).

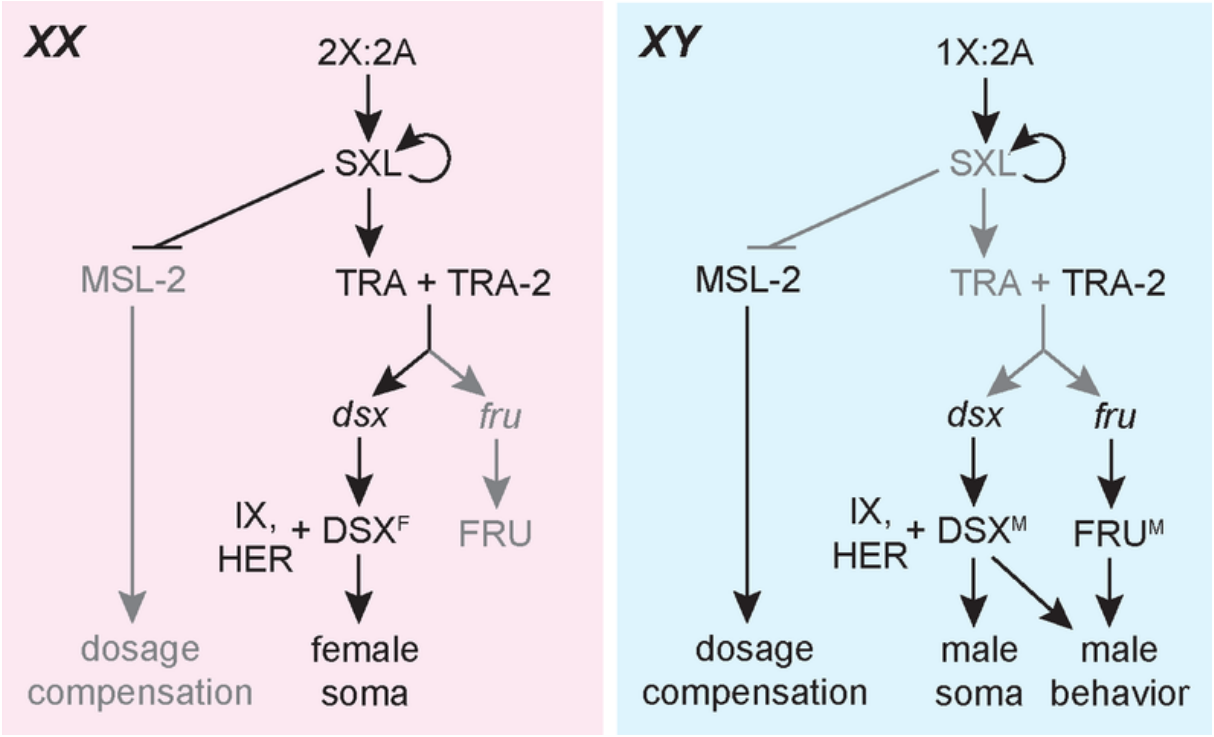


FIGURE 5.21: **Sex Determination in *D.melanogaster***. The above illustrates the cascade resulting in somatic sexual determination for XX (X:A ratio 1) and XY (X:A ratio 0.5) individuals. Each step is modulated by the previous step in the cascade, resulting in differing splicing and resulting protein isoforms that in turn modulate the activity of the next step in the sequence. Figure from Robinett et al., 2010.

In the context of anomalous *Rcd-1r* expression, it may be that unreported expression occurs in the early embryo, prior to expression of *dsx* during embryo development at approximately 3-10 hours. Indeed, *in situ* hybridisation date supports this notion. At this point embryos have not undergone somatic sexual determination, and as such *Rcd-1r* expression may not be sex-specific. If this is the case there may be a degree of expression in both males and females, which has resulted in drive activity in both male and female $Rcd\text{-}1r^{d}$ individuals. The fact this was not observed in $Rcd\text{-}1r^{d,CFP}$ females may be a combined consequence of stronger expression in

males, and cryptic splicing of the *ftz* intron diminishing expression in both sexes. It is however important to caveat this hypothesis with the understanding that *Rcd-1r* is expressed in the germline, and *dsx* is expressed in the soma. This limits the possibility that such an interaction is occurring.

An important observation from (Quezada-Díaz et al., 2010), is that insertion of a *P*-element into the 5'-UTR of *Rcd-1r* resulted in an atypical expression pattern. Whereas previously *Rcd-1r* expression was only observed in the testes, following knock-in expression was observed in the carcass of mutant individuals. It was hypothesised that cryptic regulatory regions in the *P*-element were responsible for this unexpected change in spatial expression. Insertion of the *Rcd-1r* drive component does not make any alterations to the 5'UTR (though SNPs may become modified), whilst *P*-element insertion resulted in clear modification of the transcriptional start site - these observations suggest that the unexpected activity of the *Rcd-1r* drive component in females of the *Rcd-1r*^{d} line is unlikely to be a consequence of inadvertent modification of the endogenous *Rcd-1r* regulatory region. Moreover, expression was not observed in *Rcd-1r*^{d,CFP} females, suggesting the cause is most likely linked to excision of the Cas9 intronic cassette. This must be caveated with the understanding that a multifactor interaction may be responsible of which excision of the intronic reporter cassette is just one factor.

As stated above, excision of the 3xP3:CFP:SV40 cassette from the *ftz* intron contained within the SpCas9 coding sequence of the *Rcd-1r* drive may have played a role in the anomalous bi-sexual expression of *Rcd-1r*. It has been previously noted that the presence/and or splicing of introns can be sufficient to augment transgene expression, independent of the presence of intronically-encoded enhancer elements (Duncker, Davies, and Walker, 1997). While it is still debated as to whether the presence alone of introns is sufficient to augment gene expression, it is known that in some mammalian and plant models introns (or their associated activity) are able to mediate increased transcription and translation (Shaul, 2017). It has been previously demonstrated that the presence of the *ftz* intron was sufficient to increase expression of a reporter cassette in *D.melangoaster* (Haley, Foys, and Levine, 2010). While it is known that splice-independent mechanisms can be responsible for intron-mediated enhancement (IME) (Rose, 2019), it is possible that that following excision of the 3xP3:CFP:SV40 reporter cassette splicing efficiency of the *ftz* intron increased, resulting in augmented expression from the *Rcd-1r* locus and thus increased drive activity. In *Arabidopsis thaliana*, it has been found that an efficiently spliced intron is required to be within 1kb of the transcriptional start site in order to modulate an increase in mRNA levels of a given gene (Rose, 2004). This does not necessitate that the same is true in this case, but it does emphasise the importance of context when investigating IME. Given that IME activity is contextual for a range of parameters including proximity to regulatory regions, intron size, splice efficiency, and exonic junctions, it may be interesting to compare drive activity when using identical constructs and loci, whilst varying the SpCas9 intron to gain an initial understanding of its effect (or lack of one).

Analysis of a Markerless IGD Drive Component

A particular challenge of this chapter was conducing homing assays using the *Rcd-1r*^{d} drive line, which was rendered markerless as a consequence of Cre-Lox recombination. In so doing this had the effect of decreasing the complexity of the drive component, and decreasing the size of the *ftz* intron, contained within the SpCas9 cassette. Benefits of this were: 1) removed additional cryptic splice sites that may have resulted in aberrant splicing, reducing drive efficacy, and impacting on expression of the endogenous *Rcd-1r* locus; 2) reduction of molecular resources required to express the IGD drive component.

Moreover, as identified in Chapter 3, it is possible for unwanted repair events such as incomplete HDR, to result in decoupling of drive elements. Should this happen in the context of a reporter element, it may lead to false negative readings for the drive when conducting laboratory, and field research. This is particularly important given the predicted insidiousness of gene drives (Noble et al., 2018), and the subsequent need for rigorous post-release surveillance.

To this end we proposed molecular genotyping as a means to robustly identify individuals carrying drive alleles, and developed an experimental pipeline for the extraction, amplification, and analysis of said alleles. While it was possible to extract and analyse gDNA from >2000 individuals, it is clear that the process has limitations. Namely, the amount of laboratory resource and labour required to perform screening of this kind is non-trivial, and likely to be challenging for research teams operating in malaria-endemic regions, given the low economic resources available in these areas. Chelex resin had been selected for use in this pipeline given its applicability as a low-cost crude gDNA extraction reagent, stable for long periods without refrigeration. However, a consequence of using Chelex resin is the suspsension of impurities within samples, that can lead to PCR inhibition. Protocols using a sodium-acetate and isopropanol mix have been developed to increase purity of extracted samples (Singh, Kumari, and Iyengar, 2018), and may be useful going forward, however they will certainly add to the labour-intensive nature of this screening methodology.

5.7.4 Sex-Specific IGD Modelling

Work conducted within our research team had resulted in the generation of IGD drive lines in *An. gambiae* that exhibited differential fitnesses depending on sex (David Ellis, Personal Communication). Simple expansion of the existing IGD model permitted for the attribution of individual parameters on the basis of sex, allowing for consideration of such dynamics. In contrast with results shown in Chapter 3 of this thesis, where drive rates were significantly higher, the lower experimentally determined transmission rates found in Chapter 5 resulted in a reduction in the persistence and frequency of the drive allele. Moreover, the decrease in the rate of frameshift mutations as a consequence of NHEJ events reduced the benefits of the IGD paradigm, also impacting on the persistence of the drive. This is because indel mutations were biased towards forming functional-resistance alleles, which have no associated fitness cost, and as such compete with the drive allele.

5.7.5 Limitations of the IGD Paradigm

It is important to underline existing shortcomings in IGD, as examined within the context of this chapter. While it was possible to successfully engineer an IGD line, carrying a functional drive component, significant numbers of microinjections performed by a dedicated specialist were required before transgenesis was achieved. In this case, this may be a consequence of attempting to make large in-frame insertions into sensitive germline genes. The need to make precise, scarless insertions with an endogenous locus forgoes the use of strategies with higher efficiencies such as site-specific recombination, that have historically been used to generate transgenic drive lines (Simoni et al., 2014; Hammond et al., 2015).

The need for scarless integration of the transgene within the target locus in the host not only limits transgenesis methodologies, but also restricts cloning strategies. Modular cloning strategies such as Golden Gate, and Gateway cloning facilitate rapid assembly of construct elements, but rely on recombination sites that leave molecular scars. These scars would disrupt the open reading frame of the drive and endogenous gene, were they used to produce IGD constructs. As a consequence, all IGD constructs must be engineered using a seamless method - in this case Gibson assembly. Such an approach means that all overhangs used for assembly are context specific, and decreases the modularity of any given part. This decreases the throughput with which novel IGD drives can be produced and tested.

Moreover, when compared to conventional gene drive paradigms where constructs are integrated into 'neutral' genomic loci, it is clear that IGD has a significant disadvantage. Namely, that the choice of target site is restricted by the availability of characterised genes, and suitable gRNA target sites amenable to modifying said genes. Here, drive modules have been inserted proximal to the 5' transcriptional start site of candidate loci - it may be possible to expand the number of loci amenable to modification by also considering gRNA sites proximal to the 3' transcriptional stop site. This approach would not however, benefit from one of the key advantages of the IGD strategy, specifically IGD's engineered intolerance to the majority of frameshift mutations. An alternative approach would be to conduct targeted data-mining of collated resources such as FlyBase (Thurmond et al., 2019), or in the case of work in the Anopheline mosquito, VectorBase (Giraldo-Calderón et al., 2015) in order to identify putative loci that would be amenable to modification, and which may have suitable spatiotemporal expression.

A further consideration in relation to the above, is the difficulty that may be encountered when optimising expression levels of the drive component. In a conventional strategy, there are a number of approaches that may be used to systematically produce libraries of promoters that can be characterised and selected for a given property spatio-temporal quality. Indeed, the work conducted in Chapter 4 of this thesis is an example of a platform that may be amenable to testing of molecular components. In IGD, a user is limited to the known (and thus characterised) loci available for hijacking in a given organism - if further targets are required, the user must characterise them, as is the case with conventional drive strategies. In this context, IGD

does not confer an advantage over conventional design paradigms, when selecting promoter elements.

An important caveat of all work conducted within this thesis is that while the genome of *Drosophila melanogaster* has a high degree of homology to Anopheline mosquitoes, and has frequently been used as an analog for other species in prior work (Chan et al., 2011; Chan et al., 2013a; Simoni et al., 2014; Gantz and Bier, 2015; Guichard et al., 2019), there are still significant differences between the two and as such all findings must be viewed with this understanding.

Chapter 6

Discussion and Future Perspectives

6.1 Conclusions

Gene drives face significant biological and regulatory hurdles if they are ever to be successfully deployed in the field. The aim of this thesis was to explore an entirely novel gene drive paradigm, which we have termed integral gene drive. By using a split drive architecture and integrating drive and effector components into endogenous host loci, we aimed to mitigate a number of these aforementioned hurdles. Here we have demonstrated the theoretical proof, prototyping, and initial experimental proof-of-concept of an integral gene drive system for population replacement.

Initially we constructed a deterministic, discrete-time model (Nash et al., 2019), using an established framework (Beaghton et al., 2017) and existing parameters to evaluate the performance of the IGD paradigm, and make comparisons to a 'conventional' drive architecture. In so doing, we also aimed to evaluate key parameters that would contribute to the success of an IGD intervention, when used to combat malaria. Modelling demonstrated that the IGD paradigm is theoretically capable of outperforming conventional replacement strategies, by increasing resilience to resistance alleles. Similarly, a parameter screen identified the generation of pre-existing resistance as being one of the factors with the strongest negative effect on drive efficacy. This emphasises the importance of developing genetic control strategies that can combat molecular resistance.

In order to begin experimental validation of IGD, it was logical to begin with the drive component, and to make use of an insect vector with significant genetic homology to the Anopheline mosquito, but which is more amenable to rapid prototyping. A critical question in implementing the IGD drive component *in vivo*, was the issue of the intronic gRNA cassette, proposed as part of the design paradigm. Whilst limited examples of igRNAs had been demonstrated in mammalian tissues (Kiani et al., 2014) and rice plants (Ding et al., 2018), no work had been conducted in insect tissues, nor had igRNA variants been evaluated for relative performance. We performed an initial pilot study with a two-component screening system for use in *Drosophila* S2 cell culture, based around the QUAS binary transactivation system (Potter et al., 2010). It was possible to use flow cytometry to quantify relative performance of different igRNA cassettes within the well characterised *ftz* intron. Performance was quantified using

relative fluorescence readings of a transiently expressed fluorophore. While this was sufficient to rapidly screen igRNA variants for use in the IGD strategy, the use of a suitable fluorescent standard (e.g. fluorescein) would permit direct comparison of any igRNA variant, in any intron, by generating absolute fluorescent measurements. This represents a natural expansion of the platform to aid in optimising igRNA cassette design for future iterations of the drive, driving increased drive performance.

Having inferred that in all screened conditions, the 'Full'-variant of the igRNA cassette elicited the strongest relative response, the next research question was whether it was possible to generate transgenic *Drosophila* lines with functional, in-frame insertions of drive components, within germline loci. In so doing, igRNA cassettes were used, providing an initial validation of their efficacy *in vivo*. This is particularly important given that *in vitro* findings are not always reflected when moving to the whole organism. Review of the existing literature regarding germline promoters in *D.melanogaster* and Anopheline mosquitoes proffered a number of potential targets (Chan et al., 2011; Chan et al., 2013a; Gantz et al., 2015; Hammond et al., 2016; Champer et al., 2018b). Microinjections were performed with three progressive drive construct iterations. Lines were successfully generated at the *Rcd-1r* and *bam* loci, though not for any other germline locus. This observation suggests that while it may be possible to generate transgenic lines using the Integral Replacement paradigm, many loci may not be amenable to disruption - even if the knock-in in question is in-frame, and designed to have minimal effect on the endogenous gene. As discussed in chapter 5, this highlights a potential limitation of Integral Replacement, and raises the question of whether it may be possible to further ameliorate the design of drive components to reduce their effect when hijacking endogenous germline loci.

In order to understand the effect that integration of drive components had on fitness and fecundity of transgenic lines, developmental and oviposition assays were performed. While developmental assays had limitations pertaining to the presence of rescue alleles in offspring, no significant decrease in developmental traits was observed. This observation was expanded upon by examining oviposition and hatching rates using homozygous lines, lacking rescue alleles with the *Rcd-1r*^{d,CFP} and *Rcd-1r*^{d} drive components. Neither exhibited significant differences when compared to the control line. Given existing findings identifying *Rcd-1r* as essential for male fertility, these results suggest that the *Rcd-1r* drive component has a minimal effect on fecundity and fertility. It is important to caveat this finding with the understanding that it is possible that *Rcd-1r* is merely strongly tolerant to non-functional mutations generated by drive activity. Amplicon sequencing of a population trial across multiple generations may grant some understanding of how non-functional mutations accrue, and their influence on fitness.

A critical question then arose, as to whether drive components at the *bam* and *Rcd-1r* sites were capable of homing. Homing assays were performed, and whilst the *bam* drive exhibited no significant activity, the *Rcd-1r*^{d,CFP} drive exhibited statistically significant levels of transmission in males. This finding was sufficient to provide evidence for an initial proof-of-concept for

integral replacement. Importantly, this finding also re-validated findings from the *in vitro* pilot study, establishing that it is feasible to express gRNAs intronically using the chosen cassette *in vivo*.

In order to produce a truly minimal drive component, Cre-mediated recombination was used to remove the marker cassette within the *Rcd-1r*^{d,CFP} drive intron. The effect was two-fold - an increase in transmission, and unexpected expression in females carrying the *Rcd-1r*^{d} component. While the drive component functioned, it is important to recognise that the transmission rate remained relatively low - many published studies using conventional drive constructs have illustrated levels of transmission exceeding 95% (Gantz et al., 2015; Hammond et al., 2016. This illustrates that while Integral Replacement may be feasible on a conceptual level, meaningful refinement of the strategy is required if it is to be successful.

As discussed in the previous chapter, the logistics surrounding the genotyping of markerless is non-trivial, and this posits questions around the applicability of such a strategy in resource-poor locales. If this proves to be a practical challenge to the adoption of Integral Replacement, it may be necessary to retain the reporter cassette, and instead focus on improving drive efficacy. Such challenges may prove an area for future work.

6.2 Future Work

In the immediate future, there are a number of steps that can be made to further build upon the work conducted here, and grant additional clarity to extant questions. These relate primarily to 1) the nature of *Rcd-1r* gene expression and, 2) The efficacy of the *Rcd-1r* IGD line over multiple generations.

6.2.1 Investigating Expression of *Rcd-1r*

An unexpected result of quantifying homing efficacy of the *Rcd-1r* drive in *D.melanogaster* has been the implication that expression of *Rcd-1r* may not be limited to the testes in the *Rcd-1r*^{d} drive line. This observation was made whilst quantifying the homing efficacy of the *Rcd-1r* drive in transgenic heterozygous females. By comparison, the *Rcd-1r*^{d,CFP} line exhibited activity only in males. Previous publications have used qPCR and RT-PCR (Quezada-Díaz et al., 2010) to determine tissue specificity of *Rcd-1r* expression in wild-type and mutant males. High-throughput micro-array data is available for *Rcd-1r*, indicating high-levels of expression in the testes, with limited expression observed in other tissues (Tomancak et al., 2002b). By contrast, *in-situ* hybridisation and FISH studies (Lécuyer et al., 2007; Wilk et al., 2016) indicate expression in the early embryo. This conflict requires further elucidation in order to understand the outcome of the aforementioned homing assay.

To investigate this issue further, one could conduct qPCR quantification of expression levels of *Rcd-1r* in the *Rcd-1r* gene drive line compared to a wild-type control line, looking at testes, ovaries, carcasses, as well as embryos of varying ages. In the context of these embryos, an

important consideration will be at what developmental stage *Rcd-1r* expression is detectable (if at all) - should detectable expression occur before sex determination, this may contribute to the expression observed in *Rcd-1r* females. This work is currently being conducted within our research team, with preliminary results implying expression in the early embryo - these must be further validated and reproduced before a conclusive remark can be made.

6.2.2 Setup of a Population Trial

The *Rcd-1r*^{d} line has now been demonstrated to exhibit homing in both transgenic males and females. The next logical step in investigating the homing capacity of this line, and its applicability as a genetic control strategy is to conduct a population trial. A trial has been initiated, taking design cues from Simoni et al., 2014 and Buchman et al., 2018. Three release conditions have been selected, based on forecasts provided by the above IGD model, seeding heterozygous transgenic *Rcd-1r*$^{CFP-}$ males at 5%, 15%, and 25% within each population. The net population size is 200 individuals, with the remainder of the flies being w^{1118} males and females, such that the population at timepoint 0, is composed of an equal ratio of males and females. All three conditions have been initiated in four biological replicates, and maintained at 25°C to ensure high levels of Cas9 activity and shorter generation times, without compromising the viability of each population. All experiments are being performed in 250 mL bottles, with acetate plugs, filled with 50 mL of standard yeast-based fly food.

Following seeding, flies were allowed to mate for five days, before being recovered using CO_2, and stored at -80°C for subsequent analysis. Three days after G_1 individuals had eclosed, adults were retrieved from each bottle and counted. 200 individuals were selected at random, and placed into a fresh bottle to begin the next generation. Any excess flies were frozen at -80°C for later analysis. This cycle will repeat until at least 10 generations have elapsed, and surviving adults have been collected for analysis, which will be conducted using the same workflow as was previously used for markerless lines. In order to avoid the shortcomings of using Chelex resin, it may be reasonable to use an alternative crude gDNA extraction methodology, with the understanding that this may incur additional costs that would be unsustainable for many research teams in malaria-endemic regions. It will be subsequently possible to determine the efficacy of the *Rcd-1r*^{d} drive over multiple generations, and to compare it to forecasts produced by the existing IGD model. Analysis of the wild-type *Rcd-1r* target site in non-drive G_{10} progeny should also give insight as to nature of resistance alleles that may have accrued over previous generations.

6.2.3 Development of Suitable Effector Components

A clear next step in the context of prototyping IGD in *Drosophila*, is the development of an analogous effector component. *D.melanogaster* are not human disease vectors, and as such do not propagate *Plasmodium* - it is therefore not possible to interrogate the efficacy of an antimalarial

transgene in this system. Such work needs to be conducted, first *in vitro*, using *Plasmodium falciparum* cultures, and then *in vivo*, in Anopheline mosquitoes, with particular emphasis placed on *An.gambiae*. Fortunately such work is currently being conducted within our research team.

Analogous effector components can instead be trialled in *D.melanogaster*, propagating fluorescent reporter genes in place of antimalarial effectors. *D.melanogaster* do not require bloodmeals, and so it is unlikely that the target locus to be hijacked by the effector will have a useful ortholog in Anopheline mosquitoes. Instead, constitutively expressed somatic loci, expressed in the cytoskeleton such as actin (*act5c*) or myofilin (*Mf*) would suffice as an approximate analog for the investigation of drive dynamics.

6.2.4 Improving Gene Drive Performance

The IGD prototype described in this thesis acts as a proof-of-principle study, providing evidence that this drive architecture can home at super-Mendelian rates. However, when compared to other examples of gene drives in the literature in a range of organisms (Gantz and Bier, 2015; Gantz et al., 2015; Hammond et al., 2016; Kyrou et al., 2018), it is clear that improvements could be made to increase the efficacy of the IGD strategy.

Numerous strategies have been proposed to ameloriate performance of gene drives (Burt and Koufopanou, 2004; Hammond and Galizi, 2018). Multiplexing of gRNAs to reduce the effect of resistance as a result of NHEJ events has been trialled in yeast (Yan and Finnigan, 2018) and *D.melanogaster* (Oberhofer, Ivy, and Hay, 2018), using CRISPR-Cas9 and HEGs respectively. While this approach is better suited to suppression strategies, use of multiple gRNAs at independent loci may improve the persistence of replacement gene drives.

Another important improvement relates to the understanding of putative microhomology-mediated end-joining events (MMEJ) (Wang and Xu, 2017). This error-prone repair pathway causes large excisions, and is liable to producing non-functional indel mutations of target sites. By selecting target sites that are prone to MMEJ, should homing not occur following cleavage, MMEJ may result in non-functional mutant alleles (Hammond and Galizi, 2018). In the context of IGD, this may be a useful mechanism to limit the propagation of detrimental in-frame indels, that are resilient to further cleavage.

As previously mentioned, considerable effort has been attributed to characterising novel promoters to overcome issues associated with maternal deposition of Cas9 protein into the early embryo, resulting in high rates of NHEJ and MMEJ (Champer et al., 2017). Recent work conducted in *An.gambiae* has demonstrated that using promoters with germline-restricted expression such as *nanos*, *zpg*, and the newly characterised *exu* can significantly mitigate resistance rates (Hammond et al., 2018). In the context of IGD, improvements will be made by identifying suitable germline loci, with tightly regulated expression patterns.

An improvement that is particularly amenable to IGD, and which has been explored in Chapter 3 of this thesis, is the release of multiple, autonomous drive components. In the context of IGD, this strategy may allow for initial release of components targeting multiple endogenous loci, or it may facilitate staggered release of drive components, acting to 'top-up' the drive as resistance alleles begin to accrue at each locus. Our modelling work has already demonstrated that this latter strategy is an effective means to extend levels of protection, when faced with functional resistance alleles.

Characterisation of novel endonucleases, with improved cleavage efficiency and lower off-target activity, such as Cpf1 (Zetsche et al., 2015) and FokI may also improve gene drive efficacy. The dimeric nature of FokI in particular, reduces the frequency of off-target activity, though at the cost of lower cleavage rates. This drawback has however been addressed in part by the systematic engineering of FokI nucleases, to increase cleavage activity, whilst retaining high fidelity for intended target sites (Havlicek et al., 2017).

A key consideration in the use of gene drives is their safety. This thesis has already touched upon the importance of ecological monitoring and evaluation, to understand the environmental impact any vector control methodology. The invasiveness of replacement gene drives is a double-edged sword, and has been struck upon extensively (Noble et al., 2018). In order to mitigate the risk of malicious or accidental release of gene drives, a number of strategies have been developed. These include Daisy-Chain (Noble et al., 2019), CATCHA, ERACR, and CHACR, which have been described in Chapter 1 of this thesis (Wu, Luo, and Gao, 2016; Gantz and Bier, 2016). The development of refinement of such strategies is necessary in order to ensure protection against unwanted release events, as well as providing the means to halt or reverse population modifications, if biological surveillance finds the modifications to have an undesriable effect (i.e. increased fitness, higher pathogenicity, modified behaviour).

6.2.5 Ethics of Gene Drive Technologies

A key consideration when considering future work involving genetic drive technologies pertains to bioethical and regulatory challenges, and any prospective strategy employing IGD would be no different. Aside from disease control, a range of beneficial applications have been theorised (Callies, 2019). For example, there are a number of hypothetical applications in wildlife conservation, such as the control of non-indigenous invasive species or reduction of zoonotic disease vectors. Similarly, gene drives could be used to address common herbicide-resistant flora that impact on agricultural output.

Whilst gene drives have been primarily discussed as a means to control insect populations, applications to mammalian vectors is also being considered. Such applications posit necessary concern for the long term fate of this technology (Grunwald et al., 2019). Proponents of controls on gene drive technologies are able to point to historical instances of ecological intervention by humans ending in disaster. For example, the introduction of the cane toad to Australia, in order to control cane beetle populations (Phillips and Shine, 2004 negatively affecting a range

of indigenous species. The question here is one of hubris, and the selection of a technological solution over social change as a means to combat malaria (Callies, 2019).

Conversely, technological innovations which aim to solve questions of such enormity may be considered to be intrinsically conceited - that does not negate the individual suffering felt by those afflicted, and may be considered a natural consequence of humanity's attempts to exert control over the natural world. Rather, our past arrogance regarding large-scale ecological interventions should serve as a lesson for future regulatory frameworks, ensuring suitable care is taken in order to avoid repetition of historic mistakes. Moreover, when Alvin Weinberg coined the term 'techno-fix' in 1966, he made explicit reference to the expediency of technological solutions as a means to solve pressing societal challenges, without the need for 'social engineering'. Given the impact of malaria on human health - socially, economically, and on the level of the individual - it is not unreasonable to suggest that this situation demands prompt, and decisive action, within the confines of a suitable regulatory framework.

However, the regulation of gene drive technologies remains ambiguous, with an historic tendency towards self-regulation (Oye et al., 2014). More recently the American National Academies of Sciences, Engineering and Medicine (NASEM) has made a number of recommendations as to the continuing nature of gene drive research (Collins, 2018). It was concluded that further evidence was required prior to field release of GMOs carrying gene drives; nevertheless, a series of phased steps was devised, representative of best practice:

1. Research planning and preparation
2. Laboratory research
3. Field research
4. Staged environmental release
5. Post-release surveillance

(NASEM, 2016)

The report identifies that, given the selective competitiveness gene drive populations are selected to have, confinement and containment are essential at every phase. Moreover, there is an emphasis on the need for robust ecological risk assessment in order to identify any impacts (and causes) of gene drive release, as well evaluating the probability of particular outcomes (NASEM, 2016).

A consideration highlighted by the NASEM report is the importance of engaging all stakeholders. This is to gain insight into the specific complexities of each given release site, to ensure suitable transparency, and to mediate mutual knowledge exchange. In so doing, experts aiming to facilitate field-releases of gene drive vectors can establish trust within local communities,

whilst learning key lessons that may be critical to the success of a given gene drive intervention. There are a number of considerations that must be made if the IGD, and gene drives more broadly are to be implemented into the field. Advances must be made, not only in the laboratory, but in a broader social and cultural context. If not, we risk creating a solution to a problem of significant importance, that will never be deployed.

Appendix A

Supplementary Information

The following appendix details additional materials relevant to the work performed in Chapter 5 of this thesis.

A.0.1 Additional Sequencing Primers

Primer ID	Description	Sequence	Forward/Reverse
pC55	pBID-*vasa*-eGFP-SpCas9-*ftz*	CGGATCTTCGAGATCAGCA	Forward
pC56	pBID-*vasa*-eGFP-SpCas9-*ftz*	GATGCTCTGGTGGATCAGG	Forward
pC57	pBID-*vasa*-eGFP-SpCas9-*ftz*	TTATCGCTGTTCCTCTTGGG	Forward
pC58	pBID-*vasa*-eGFP-SpCas9-*ftz*	GCCGCCAGTAGTTCTTCATC	Forward
pC59	pBID-*vasa*-eGFP-SpCas9-*ftz*	TACCTGCAGGACTAGAGCGA	Forward
pC60	pBID-*vasa*-eGFP-SpCas9-*ftz*	ATCTTTTCCCGGTTGTCCTT	Forward
pC61	pBID-*vasa*-eGFP-SpCas9-*ftz*	GTTGATGGGGTTTTCCTCG	Forward
pC62	pBID-*vasa*-eGFP-SpCas9-*ftz*	AGTCTCCGTCGTGGTCCTTA	Forward
pC63	pBID-*vasa*-eGFP-SpCas9-*ftz*	GTAGGTGGCATCGCCCTC	Forward
pC64	pBID-*vasa*-eGFP-SpCas9-*ftz*	ATCTACGTAGGGAGGAAGGTTTG	Forward
pC65	pBID-*vasa*-eGFP-SpCas9-*ftz*	CCCTACTTTTCTCCAACAGGG	Reverse
pC66	pBID-*vasa*-eGFP-SpCas9-*ftz*	CCCTGCCGCCTTCAAGTA	Reverse
pC67	pBID-*vasa*-eGFP-SpCas9-*ftz*	GCATGCCCCAAGTGAATATC	Reverse
pC68	pBID-*vasa*-eGFP-SpCas9-*ftz*	ACAACAAGGTGCTGACCAGA	Reverse
pC69	pBID-*vasa*-eGFP-SpCas9-*ftz*	CCTGTTCGACGACAAAGTGA	Reverse
pC70	pBID-*vasa*-eGFP-SpCas9-*ftz*	GACCTTCGACAACGGCAG	Reverse
pC71	pBID-*vasa*-eGFP-SpCas9-*ftz*	ACCTGAACCCCGACAACAG	Reverse
pC72	pBID-*vasa*-eGFP-SpCas9-*ftz*	CTGTACAAGCGCCGTAAGAG	Reverse
pC73	pBID-*vasa*-eGFP-SpCas9-*ftz*	CAAGGGCGAGGAGCTGTT	Reverse
pC74	pBID-*vasa*-eGFP-SpCas9-*ftz*	AGCCAATGGATCACCAAATC	Reverse
pC107	3xP3:CFP:SV40	ACGGACCCCTTACTCTCGTC	Forward
pC108	3xP3:CFP:SV40	TCTTGGCTGAGTAGGTGGCG	Reverse
pC158	*ftz* intron	GTAAGTATCACACACGATTAACAACC	Forward
pC159	*ftz* intron	CTGTAAGCATAAGCAAAGAAAAAATG	Reverse

TABLE A.1: The above table details the sequencing primers for the initial two iterations of the *vasa* drive construct. Primers were designed to ensure complete coverage of the drive component in both the 5′ and 3′ directions when using Sanger Sequencing.

Primer ID	Description	Sequence	Forward/Reverse
pC114	*β2-tubulin*	CGAGCGCTTTTGAATCTGTT	Forward
pC115	*β2-tubulin*	CAATTATACACACTGATTTGGTCG	Forward
pC116	*β2-tubulin*	TCGGTAGACTCCGGTTGTGT	Reverse
pC117	*β2-tubulin*	AGATAATTAAAATATAGGTGGCCTAA	Reverse
pC118	*nanos*	CAGTTGTTGCCTGCTCCC	Forward
pC119	*nanos*	AAAATTCGAGATGGGATGGA	Forward
pC120	*nanos*	ATTCTCGTGGACCTGGAGC	Reverse
pC121	*nanos*	GGAGTAAAATCCAGGACATCCA	Reverse
pC122	*zpg*	TTGTCTAAGGGACATACGATTTTT	Forward
pC123	*zpg*	TTGAAAATCTTTGCGCAGGT	Forward
pC124	*zpg*	TATTCGCGATACTTCTTTATCTAAT	Reverse
pC125	*zpg*	GCGCACTGTTTTTGGCTATT	Reverse
pC126	*Rcd-1r*	TCTGAAATGAAGAATCCGTGC	Forward
pC127	*Rcd-1r*	TTTGACGAGAAAATAGCGGG	Forward
pC128	*Rcd-1r*	AAGGAATGGAGCTGCACAGT	Reverse
pC130	SpCas9-*ftz*-T2A	CGAAGATTTAACATGTGCAACG	Reverse
pC131	SpCas9-*ftz*-T2A	AAAGTCAGGGCGTAGCACAC	Forward
pC132	SpCas9-*ftz*-T2A	AAATGGCCAGACTAATTGCG	Forward
pC133	SpCas9-*ftz*-T2A	GTCGGTACCACCTGCACC	Reverse
pC134	SpCas9-*ftz*-T2A	GTCCTTCTTTCTGGCGATCA	Forward
pC135	SpCas9-*ftz*-T2A	TTCTGGTCAGCACCTTGTTG	Forward
pC136	SpCas9-*ftz*-T2A	GCTGCTTCATCACTTTGTCG	Forward
pC137	SpCas9-*ftz*-T2A	ATCTTTTCCAGGATGGGCTT	Forward
pC138	SpCas9-*ftz*-T2A	GATGTTGCCGAAGATGGG	Forward
pC139	SpCas9-*ftz*-T2A	GAAAAAGACCGAGGTGCAGA	Reverse
pC140	*β2-tubulin*	CCAGGAACTGGACATCAACC	Reverse
pC141	*β2-tubulin*	TATCGTGCTGACCCTGACAC	Reverse
pC142	*bam*	ACCAGAGCAAGAACGGCTAC	Reverse
pC143	*bam*	ATCTTCAGCAACGAGATGGC	Reverse
pC144	*nanos*	TCGTGGTCCTTATAGTCCATG	Forward
pC145	*nanos*	CCAGGCAAAAAAGAAAAAG	Reverse
pC146	*Rcd-1r*	TCGTGGTCCTTATAGTCCATATG	Forward
pC147	*Rcd-1r*	CCAGGCAAAAAAGAAAAAG	Reverse

TABLE A.2: The above table details the sequencing primers for final iterations of the drive construct for all targeted germline loci. Primers were designed to ensure complete coverage of the drive component in both the 5′ and 3′ directions when using Sanger Sequencing. Where the name of the locus is given the corresponding primer was used to sequence homology arms unique to that locus. For shared components, a common set of primers was used.

Component	Sequence (5′-3′)	Size (bp)
ftz intron	GTAAGTATCACACACGATTAACAACCCCTAAAAATACACTTTG AAAATATTGAAAATATGGCTAGCTCTAGTTTCGAATTTGTATA CATTTTTGTATACATTTTTGATATTTTCAAACAATACGCAGTTATA AAACTCATTAGCTAACCCATTTTTTCTTTGCTTATGCTTACAG	175
Minimal intron	GTAAGTGGTCCGCTAGCTCTAGTTTCGAATA CTAACTCTCGAGTCTTCTTTTTTTTTTTCACAG	64

TABLE A.3: The above gives details of the sequence of the modified *ftz* and Minimal mammalian introns used in Chapter 3.

A.0.2 IGD Sex-Specific Model Parameters

The parameters used for the expanded, sex-specific iteration of the IGD model are listed below. The model runs in accordance with the same logic as described in Chapter 3 and contains no other significant deviations. Baseline parameter values are given as before. For both sets of parameters, the following definitions are valid as before; HDR - homology-directed repair; NHEJ - non-homologous end joining. Baseline parameter values used for all modeling in this chapter were obtained from existing work on mosquitoes Hammond et al., 2016).

TABLE A.4: Baseline Parameter Values for the Sex-Specific IGD Model - Male Parameters

Symbol	Parameter^{+}	Baseline Value
cc_0M	Initial release frequency of homozygous drive and effector construct	10-4 x (1/3)
dd_0M	Initial release frequency of homozygous drive construct	0
ee_0M	Initial release frequency of homozygous effector construct	0
$rr1_0M$	Initial frequency of resistance (locus 1)	0
$rr2_0M$	Initial frequency of resistance (locus 2)	0
$rr3_0M$	Initial frequency of resistance (locus 3)	0
d_1M	Transmission rate of drive (locus 1)	0.99
$d_{2a}M$	Transmission rate of drive (locus 2) if drive component present in heterozygosity	0.99
$d_{2b}M$	Transmission rate of drive (locus 2) if drive component present in homozygosity	0.99
$d_{3a}M$	Transmission rate of drive (locus 3) if drive component present in heterozygosity (I)	0.99
$d_{3b}M$	Transmission rate of drive (locus 3) if drive component present in homozygosity (I)	0.99
l_1M	Resistance arising from HR at locus 1	10-4 x (1/3)
l_2M	Resistance arising from HR at locus 2	10-4 x (1/3)
l_3M	Resistance arising from HR at locus 3 (I)	10-4 x (1/3)
u_1M	Resistance arising from NHEJ at locus 1	0.5 x (1/3)
u_2M	Resistance arising from NHEJ at locus 2	0.5 x (1/3)
u_3M	Resistance arising from NHEJ at locus 3 (I)	0.5 x (1/3)
L_1M	Loss of gene function arising from HR at locus 1	10-4 x (2/3)
L_2M	Loss of gene function arising from HR at locus 2	10-4 x (2/3)
L_3M	Loss of gene function arising from HR at locus 3 (I)	10-4 x (2/3)
U_1M	Loss of gene function arising from NHEJ at locus 1	0.5 x (2/3)
U_2M	Loss of gene function arising from NHEJ at locus 2	0.5 x (2/3)
U_3M	Loss of gene function arising from NHEJ at locus 3 (I)	0.5 x (2/3)

Continued on next page

Table A.4 – continued from previous page

Symbol	Parameter[+]	Baseline Value
s_mM	Cost of loss of gene function	1
s_{d1}M	Cost of hijacking target locus 1 (Drive)	0
s_{d2}M	Cost of hijacking target locus 2 (Effector)	0
s_{d3}M	Cost of hijacking target locus 3 (Effector) (I)	0
s_nM	Cost of expressing nuclease at locus 1	0.05
s_{e2}M	Cost of expressing effector at locus 2	0.1
s_{e3}M	Cost of expressing effector at locus 3 (I)	0.1
h_mM	Dominance coef. for loss of gene function	0.2
h_{d1}M	Dominance coef. for hijacking at locus 1	0.5
h_{d2}M	Dominance coef. for hijacking at locus 2	0.5
h_{d3}M	Dominance coef. for hijacking at locus 3 (I)	0.5
h_nM	Dominance coef. for expressing nuclease	0.5
h_{e2}M	Dominance coef. for expressing effector (locus 2)	0.5
h_{e3}M	Dominance coef. for expressing effector (locus 3) (I)	0.5
h_{rc1}M	Dominance coef. for refractoriness (1 effector allele)	1
h_{rc2}M	Dominance coef. for refractoriness (2 effector alleles) (I)	1
h_{rc3}M	Dominance coef. for refractoriness (3 effector alleles) (I)	1
r_cM	Homozygous degree of refractoriness	1

Baseline parameters for females in the sex-specific IGD model are given below -

TABLE A.5: Baseline Parameter Values for the Sex-Specific IGD Model - Female Parameters

Symbol	Parameter[+]	Baseline Value
cc_0F	Initial release frequency of homozygous drive and effector construct	10-4 x (1/3)
dd_0F	Initial release frequency of homozygous drive construct	0
ee_0F	Initial release frequency of homozygous effector construct	0
$rr1_0$F	Initial frequency of resistance (locus 1)	0
$rr2_0$F	Initial frequency of resistance (locus 2)	0
$rr3_0$F	Initial frequency of resistance (locus 3)	0
d_1F	Transmission rate of drive (locus 1)	0.99
d_{2a}F	Transmission rate of drive (locus 2) if drive component present in heterozygosity	0.99
d_{2b}F	Transmission rate of drive (locus 2) if drive component present in homozygosity	0.99
d_{3a}F	Transmission rate of drive (locus 3) if drive component present in heterozygosity (I)	0.99
d_{3b}F	Transmission rate of drive (locus 3) if drive component present in homozygosity (I)	0.99
l_1F	Resistance arising from HR at locus 1	10-4 x (1/3)
l_2F	Resistance arising from HR at locus 2	10-4 x (1/3)
l_3F	Resistance arising from HR at locus 3 (I)	10-4 x (1/3)
u_1F	Resistance arising from NHEJ at locus 1	0.5 x (1/3)
u_2F	Resistance arising from NHEJ at locus 2	0.5 x (1/3)
u_3F	Resistance arising from NHEJ at locus 3 (I)	0.5 x (1/3)
L_1F	Loss of gene function arising from HR at locus 1	10-4 x (2/3)

Continued on next page

Table A.5 – continued from previous page

Symbol	Parameter[+]	Baseline Value
L_2F	Loss of gene function arising from HR at locus 2	10-4 x (2/3)
L_3F	Loss of gene function arising from HR at locus 3 (I)	10-4 x (2/3)
U_1F	Loss of gene function arising from NHEJ at locus 1	0.5 x (2/3)
U_2F	Loss of gene function arising from NHEJ at locus 2	0.5 x (2/3)
U_3F	Loss of gene function arising from NHEJ at locus 3 (I)	0.5 x (2/3)
s_mF	Cost of loss of gene function	1
$s_{d1}F$	Cost of hijacking target locus 1 (Drive)	0
$s_{d2}F$	Cost of hijacking target locus 2 (Effector)	0
$s_{d3}F$	Cost of hijacking target locus 3 (Effector) (I)	0
s_nF	Cost of expressing nuclease at locus 1	0.05
$s_{e2}F$	Cost of expressing effector at locus 2	0.1
$s_{e3}F$	Cost of expressing effector at locus 3 (I)	0.1
h_mF	Dominance coef. for loss of gene function	0.2
$h_{d1}F$	Dominance coef. for hijacking at locus 1	0.5
$h_{d2}F$	Dominance coef. for hijacking at locus 2	0.5
$h_{d3}F$	Dominance coef. for hijacking at locus 3 (I)	0.5
h_nF	Dominance coef. for expressing nuclease	0.5
$h_{e2}F$	Dominance coef. for expressing effector (locus 2)	0.5
$h_{e3}F$	Dominance coef. for expressing effector (locus 3) (I)	0.5
$h_{rc1}F$	Dominance coef. for refractoriness (1 effector allele)	1
$h_{rc2}F$	Dominance coef. for refractoriness (2 effector alleles) (I)	1
$h_{rc3}F$	Dominance coef. for refractoriness (3 effector alleles) (I)	1
r_cF	Homozygous degree of refractoriness	1

A.0.3 Transgenic Lines and Target Loci

All the transgenic lines obtained from external sources are listed here (TableA.6). In each case, where a designated name has been used in this thesis in place of a genotype, that name is given.

Line ID	Genotype	Source
51323	*y^1 M{vas-Cas9}ZH-2A w^{1118}*	Bloomington
54591	*y^1 P(nos-cas9, w+) M(3xP3-RFP.attP)ZH-2A w^**	Bloomington
13-20	*y w M(eGFP, vas-int, dmRFP)ZH-2A; P{CaryP}attP40*	Cambridge Fly Facility
w^{1118};if/Sm6a	*w^{1118};if/Sm6a*	Southall Lab
1092	*y^1 w^{67c23}; sna[Sco]/CyO, P{w[+mC]=Crew}DH1*	Bloomington
1501	*y^1 w^{67c23}; MKRS, P{ry[+t7.2]=hsFLP}86E/TM6B, P{w[+mC]=Crew}DH2, Tb^1*	Bloomington

TABLE A.6: **Externally Sourced *Drosophila* Lines**. The above details all transgenic *Drosophila* lines that were sourced from external institutes. In each case, a line designation is given, along with complete genotype information and the source.

The below depicts the target site for each gRNA sequence, within each candidate germline locus (FigA.1).

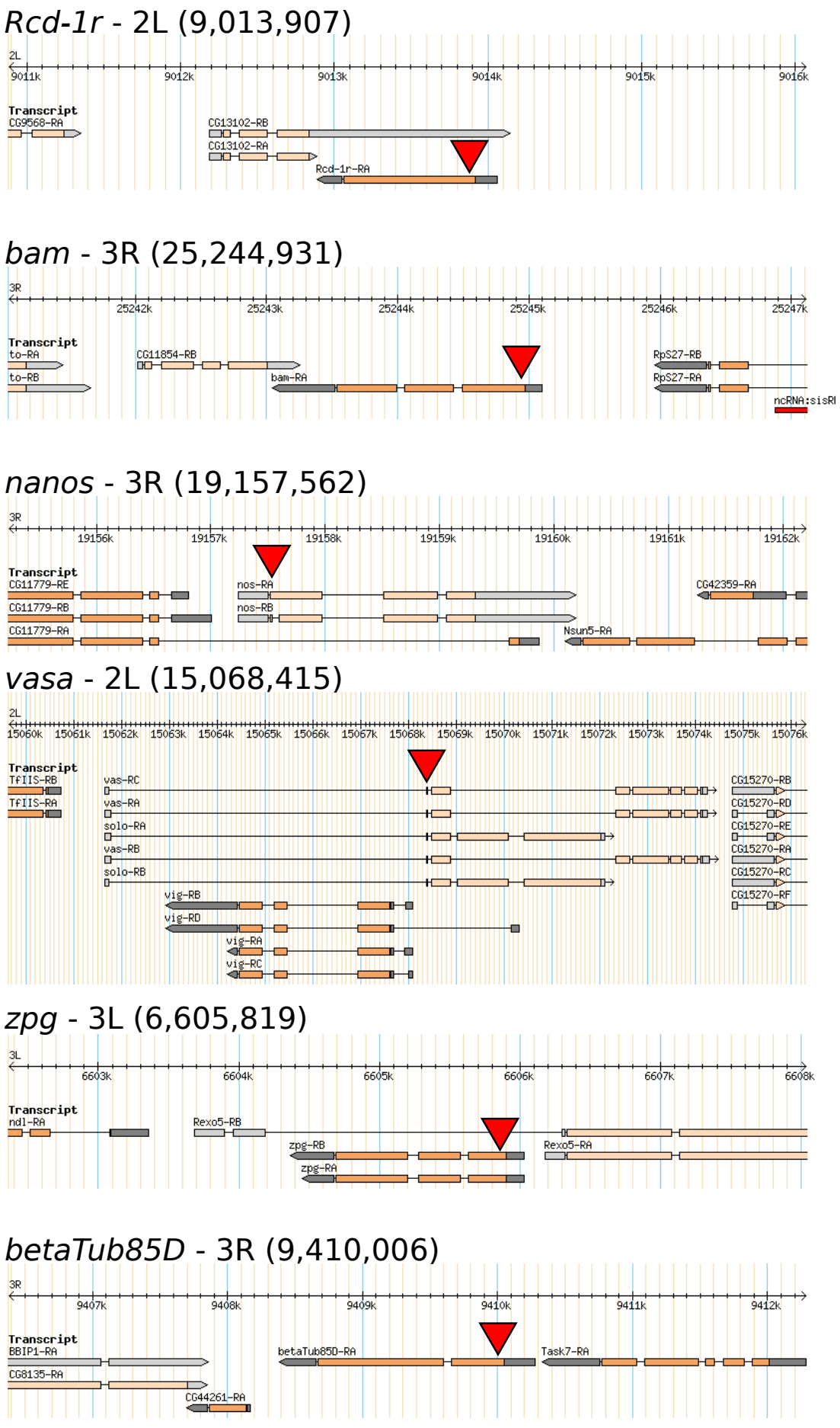


FIGURE A.1: **Gene Drive Targets**. Each target insertional site is listed for each endogenous germline gene, along with the chromosomal and base-pair location at which the insertion is intended to occur. Red triangles indicate the approximate position of each insertional point relative to the gene model.

A.0.4 Additional Amplicon Sequencing Information

Additional analysis was performed on the amplicon sequencing data described in Chapter 5. It examined the frequency of alleles per sample that were represented at $\geq$ 0.5% of reads, or $\geq$ 10 reads at the *Rcd-1r* gRNA target site. 0.5% was chosen given that each experimental sample contained 50 individuals, which were the offspring of heterozygous *Rcd-1r*^{d,CFP} and *w*^{1118} individuals. Thus, 50 of all alleles examined in amplicon sequencing were derived from transgenic individuals, whom have two alleles each. This corresponds to 0.5% of the total sequenced alleles.

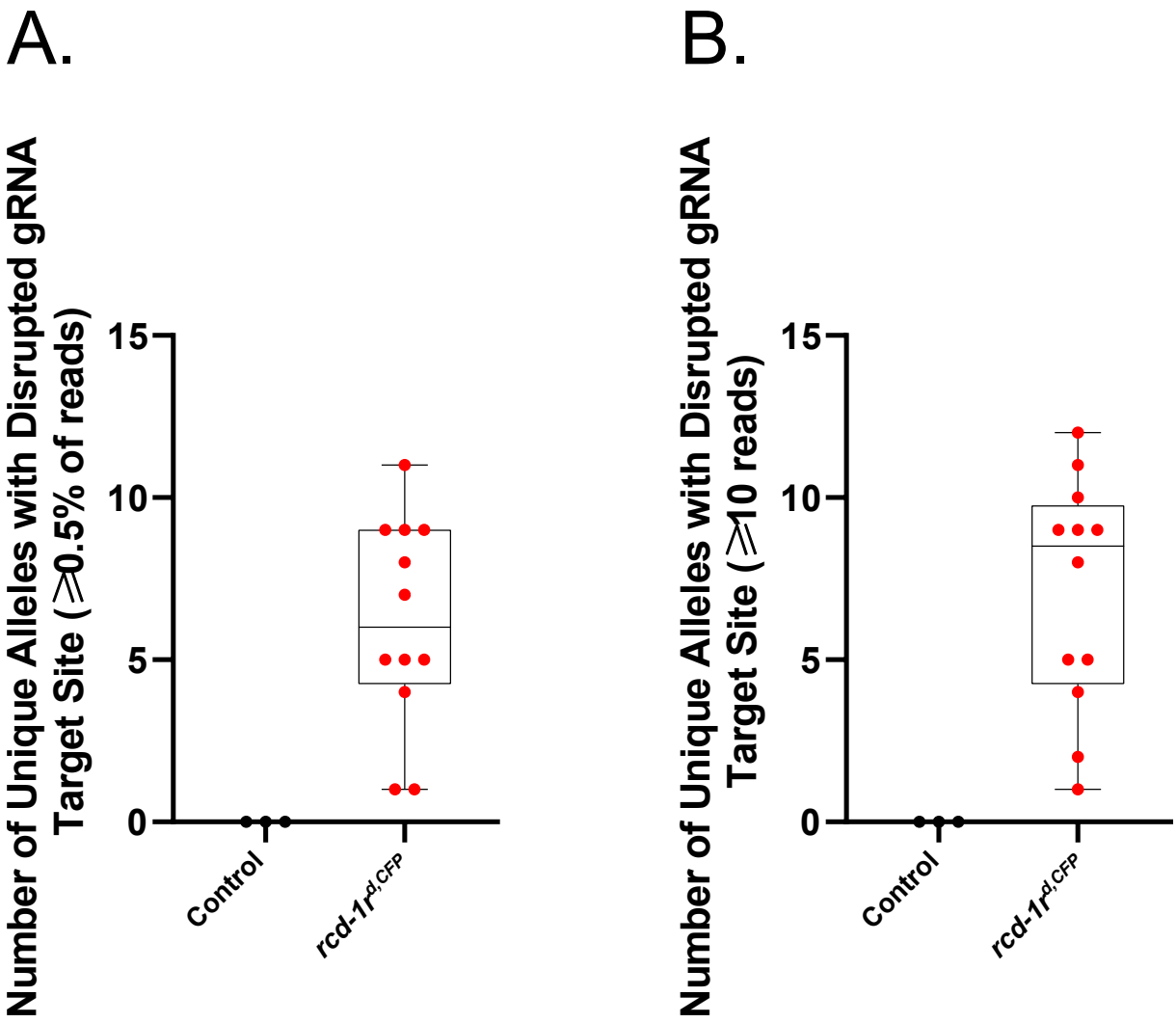


FIGURE A.2: **Additional Amplicon Sequencing Analysis**.The number of unique alleles representing disruption of the gRNA target site, represented by A) $\geq$ 0.5% or B) the $\geq$ 10 net reads was analysed. No such alleles were found for the wild-type controls, whilst an average of 6, and 7 respectively were found for the experimental samples.

When examining numbers of unique mutant alleles amongst the control offspring, none were found when using cut offs of $\geq$0.5% of reads, or represented at $\geq$10 reads. By contrast, an average of 6 unique mutant alleles were found in F_1 *Rcd-1r*^{d,CFP} offspring. This number increases to an average of 7 alleles when looking at those alleles represented by $\geq$10 reads.

This implies that in the *Rcd-1r*^{d,CFP} indels are formed at a relatively low rate, though it could equally be an indicator of low drive efficacy.

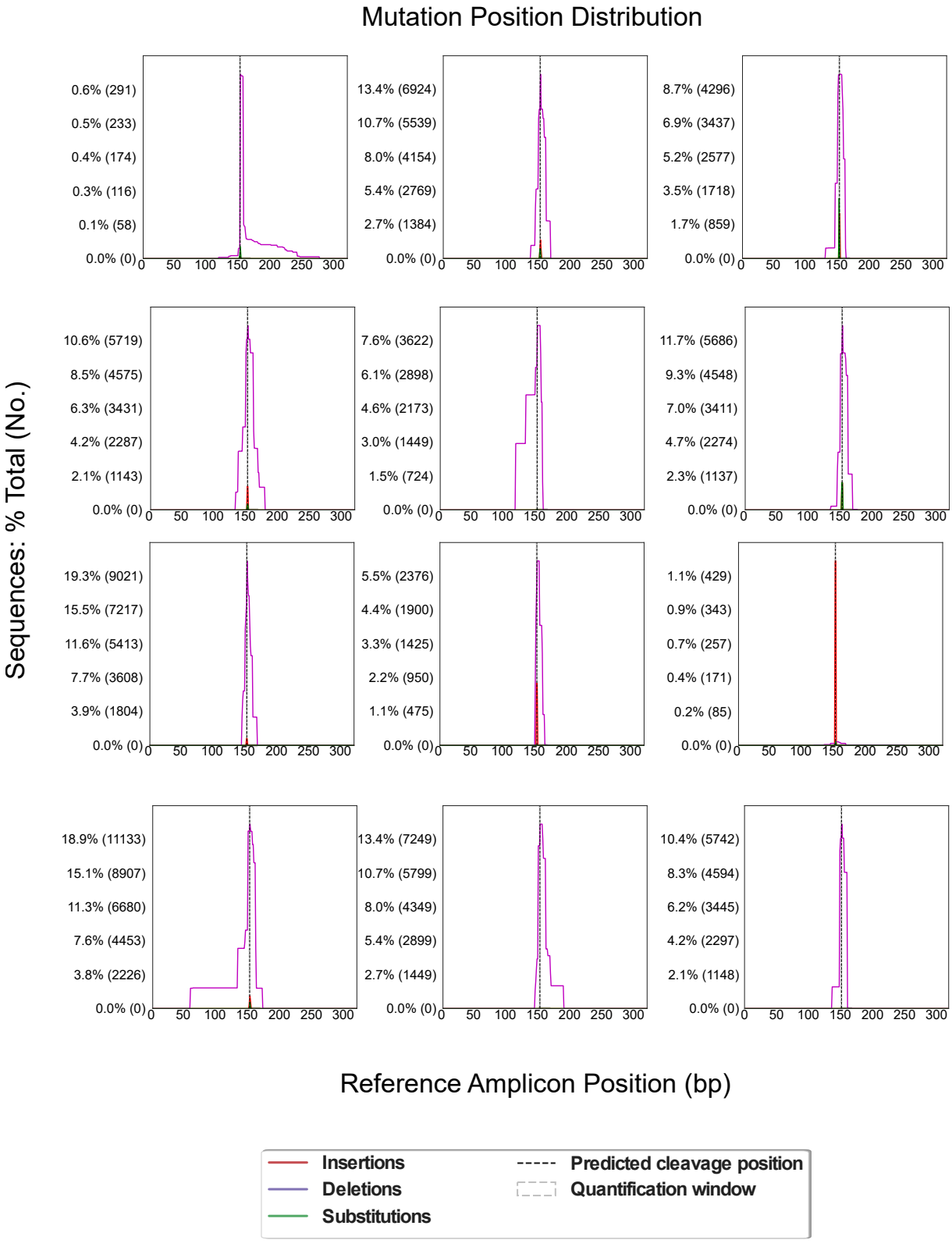


FIGURE A.3: **Amplicon Sequencing - Mutation Distributions by Replicate**. Each graph represents the distribution of insertions, deletions, and mutations for each experimental replicate. In all but one case, deletions are substantially favoured over all over mutation types.

The above (A.3) illustrates the distribution of mutation events relative to the sgRNA target site in the *Rcd-1r* locus. All twelve replicates are illustrated, and with the exception of a single replicate, deletions are the most common mutation found.

Bibliography

Abd-Alla, Adly M.M. et al. (2013). "Improving Sterile Insect Technique (SIT) for tsetse flies through research on their symbionts and pathogens". In: *Journal of Invertebrate Pathology* 112.2, S2–S10. ISSN: 00222011. DOI: 10.1016/j.jip.2012.07.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022202X15370834https://linkinghub.elsevier.com/retrieve/pii/S0022201112001784.

Adams, M. D. (2000). "The Genome Sequence of Drosophila melanogaster". In: *Science* 287.5461, pp. 2185–2195. ISSN: 00368075. DOI: 10.1126/science.287.5461.2185. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.287.5461.2185.

Adli, Mazhar (2018). "The CRISPR tool kit for genome editing and beyond". In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-04252-2.

Ahanger, Sajad H. et al. (2013). "Conserved boundary elements from the Hox complex of mosquito, Anopheles gambiae". In: *Nucleic Acids Research* 41.2, pp. 804–816. ISSN: 03051048. DOI: 10.1093/nar/gks1178.

Akbari, Omar S. et al. (2013). "A synthetic gene drive system for local, reversible modification and suppression of insect populations". In: *Current Biology* 23.8, pp. 671–677. ISSN: 09609822. DOI: 10.1016/j.cub.2013.02.059. URL: http://dx.doi.org/10.1016/j.cub.2013.02.059.

Akbari, Omar S et al. (2014). "Novel synthetic Medea selfish genetic elements drive population replacement in Drosophila; a theoretical exploration of Medea-dependent population suppression." In: *ACS synthetic biology* 3.12, pp. 915–28. ISSN: 2161-5063. DOI: 10.1021/sb300079h. URL: http://www.ncbi.nlm.nih.gov/pubmed/23654248http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3742681.

Alkan, Ferhat et al. (2018). "CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters." In: *Genome biology* 19.1, p. 177. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1534-x. URL: http://www.ncbi.nlm.nih.gov/pubmed/30367669http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6203265.

Alphey, Luke (2014). "Genetic Control of Mosquitoes". In: *Annual Review of Entomology* 59.1, pp. 205–224. ISSN: 0066-4170. DOI: 10.1146/annurev-ento-011613-162002. URL: http://www.ajtmh.org/content/journals/10.4269/ajtmh.1972.21.764http://www.annualreviews.org/doi/10.1146/annurev-ento-011613-162002.

Alphey, Luke et al. (2010). "Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis". In: *Vector-Borne and Zoonotic Diseases* 10.3, pp. 295–311. ISSN: 1530-3667. DOI: 10.1089/vbz.2009.0014. URL: https://www.liebertpub.com/doi/10.1089/vbz.2009.0014.

Alphey, Nina, Luke Alphey, and Michael B. Bonsall (2011). "A model framework to estimate impact and cost of genetics-based sterile insect methods for dengue vector control". In: *PLoS ONE* 6.10. ISSN: 19326203. DOI: 10.1371/journal.pone.0025384.

Altrock, Philipp M. et al. (2010). "Using underdominance to bi-stably transform local populations". In: *Journal of Theoretical Biology* 267.1, pp. 62–75. ISSN: 00225193. DOI: 10.1016/j.jtbi.2010.08.004. URL: http://www.ncbi.nlm.nih.gov/pubmed/20691703https://linkinghub.elsevier.com/retrieve/pii/S0022519310004030.

Aly, Ahmed S I and Kai Matuschewski (2005). "A malarial cysteine protease is necessary for Plasmodium sporozoite egress from oocysts". In: *Journal of Experimental Medicine* 202.2, pp. 225–230. ISSN: 00221007. DOI: 10.1084/jem.20050545.

Aly, Ahmed S.I., Ashley M. Vaughan, and Stefan H.I. Kappe (2009). "Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host". In: *Annual Review of Microbiology* 63.1, pp. 195–221. ISSN: 0066-4227. DOI: 10.1146/annurev.micro.091208.073403.

Amo, Víctor López del et al. (2019). "Split-gene drive system provides flexible application for safe laboratory investigation and potential field deployment Split-gene drive system provides flexible application for safe laboratory investigation and potential field deployment". In: *bioRxiv* June. DOI: 10.1101/684597.

Anopheles gambiae 1000 Genomes Consortium et al. (2017). "Genetic diversity of the African malaria vector Anopheles gambiae." In: *Nature* 552.7683, pp. 96–100. ISSN: 1476-4687. DOI: 10.1038/nature24995. URL: http://www.ncbi.nlm.nih.gov/pubmed/29186111http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6026373.

Anraku, Yasuhiro, Ryuta Mizutani, and Yoshinori Satow (2005). *Protein splicing: Its discovery and structural insight into novel chemical mechanisms*. DOI: 10.1080/15216540500215499.

Anton, Tobias et al. (2014). "Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system". In: *Nucleus* 5.2, pp. 163–172. ISSN: 19491042. DOI: 10.4161/nucl.28488.

Ariey, Frédéric et al. (2014). "A molecular marker of artemisinin-resistant Plasmodium falciparum malaria". In: *Nature* 505.7481, p. 50. ISSN: 1476-4687. DOI: 10.1038/NATURE12876. URL: http://www.ncbi.nlm.nih.gov/pubmed/24352242http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5007947.

Ashley, E. A. et al. (2014). "Spread of artemisinin resistance in Plasmodium falciparum malaria". In: *New England Journal of Medicine* 371.5, pp. 411–423. ISSN: 15334406. DOI: 10.1056/NEJMoa1314981.

Autino, Beatrice et al. (2012). *Pathogenesis of malaria in tissues and blood*. DOI: 10.4084/MJHID.2012.061.

Bae, Sangsu et al. (2014). *Microhomology-based choice of Cas9 nuclease target sites*. DOI: 10.1038/nmeth.3015.

Bai, Yongsheng et al. (2007). "Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in Drosophila". In: *Genome Biology* 8.1. ISSN: 14747596. DOI: 10.1186/gb-2007-8-1-r11.

Baldini, Francesco et al. (2014). "Evidence of natural Wolbachia infections in field populations of Anopheles gambiae". In: *Nature Communications* 5. ISSN: 20411723. DOI: 10.1038/ncomms4985.

Barr, Philip J. et al. (1991). "Recombinant pfs25 protein of plasmocdium falciparum elicits malaria transmission-blocking immunity in experimental animals". In: *Journal of Experimental Medicine* 174.5, pp. 1203–1208. ISSN: 15409538. DOI: 10.1084/jem.174.5.1203.

Barrangou, Rodolphe and Luciano A. Marraffini (2014). "CRISPR-cas Systems: Prokaryotes Ppgrade to Adaptive Immunity". In: *Molecular Cell* 54.2, pp. 234–244. ISSN: 10974164. DOI: 10.1016/j.molcel.2014.03.011. URL: http://dx.doi.org/10.1016/j.molcel.2014.03.011.

Baton, Luke Anthony et al. (2013). "wFlu: Characterization and Evaluation of a Native Wolbachia from the Mosquito Aedes fluviatilis as a Potential Vector Control Agent". In: *PLoS ONE* 8.3. ISSN: 19326203. DOI: 10.1371/journal.pone.0059619.

Baum, J A, R Geever, and N H Giles (1987). "Expression of qa-1F activator protein: identification of upstream binding sites in the qa gene cluster and localization of the DNA-binding domain." In: *Molecular and cellular biology* 7.3, pp. 1256–66. ISSN: 0270-7306. DOI: 10.1128/mcb.7.3.1256. URL: http://www.ncbi.nlm.nih.gov/pubmed/2951591http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC365200.

Bax, Nicholas J and Ronald E Thresher (2009). "Ecological, behavioral, and genetic factors influencing the recombinant control of invasive pests". In: *Ecological Applications* 19.4, pp. 873–888. ISSN: 1051-0761. DOI: 10.1890/07-1588.1. URL: http://doi.wiley.com/10.1890/07-1588.1.

Beaghton, Andrea, Pantelis John Beaghton, and Austin Burt (2017). "Vector control with driving y chromosomes: Modelling the evolution of resistance". In: *Malaria Journal* 16.1, pp. 1–14. ISSN: 14752875. DOI: 10.1186/s12936-017-1932-7.

Beaghton, Andrea et al. (2017). "Requirements for Driving Antipathogen Effector Genes into Populations of Disease Vectors by Homing." In: *Genetics* 205.4, pp. 1587–1596. ISSN: 1943-2631. DOI: 10.1534/genetics.116.197632. URL: https://doi.org/10.1534/genetics.116.197632http://www.ncbi.nlm.nih.gov/pubmed/28159753http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5378115.

Beckmann, John F., Judith A. Ronau, and Mark Hochstrasser (2017). "A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility". In: *Nature Microbiology* 2.March, p. 17007. ISSN: 2058-5276. DOI: 10.1038/nmicrobiol.2017.7. URL: http://www.nature.com/articles/nmicrobiol20177.

Beeman, R. W, K. S Friesen, and R. E Denell (1992). *Maternal-effect selfish genes in flour beetles.* DOI: 10.1126/science.1566060. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1566060.

Bellen, H. J. et al. (1989). "P-element-mediated enhancer detection: a versatile method to study development in Drosophila." In: *Genes & development* 3.9, pp. 1288–1300. ISSN: 08909369. DOI: 10.1101/gad.3.9.1288.

Belton, Peter and Robert A Costello (1979). "Flight Sounds of the Females of Some Mosquitoes of Western Canada". In: *Entomologia Experimentalis et Applicata* 26.1, pp. 105–114. ISSN: 00138703.

DOI: 10.1111/j.1570-7458.1979.tb02904.x. URL: https://doi.org/10.1111/j.1570-7458.1979.tb02904.xhttp://doi.wiley.com/10.1111/j.1570-7458.1979.tb02904.x.

Ben-David, Eyal, Alejandro Burga, and Leonid Kruglyak (2017). "A maternal-effect selfish genetic element in Caenorhabditis elegans". In: *Science* 356.6342, pp. 1051–1055. ISSN: 10959203. DOI: 10.1126/science.aan0621.

Benedict, Mark Q. and Alan S. Robinson (2003). "The first releases of transgenic mosquitoes: An argument for the sterile insect technique". In: *Trends in Parasitology* 19.8, pp. 349–355. ISSN: 14714922. DOI: 10.1016/S1471-4922(03)00144-2.

Benelli, Giovanni, Claire L. Jeffries, and Thomas Walker (2016). "Biological control of mosquito vectors: Past, present, and future". In: *Insects* 7.4, pp. 1–18. ISSN: 20754450. DOI: 10.3390/insects7040052.

Bershteyn, Anna et al. (2018). "Implementation and applications of EMOD, an individual-based multi-disease modeling platform". In: *Pathogens and Disease* 76.5. DOI: 10.1093/femspd/fty059.

Bhat, Krishna Moorthi (1999). "The posterior determinant gene nanos is required for the maintenance of the adult germline stem cells during drosophila oogenesis". In: *Genetics* 151.4, pp. 1479–1492. ISSN: 00166731.

Bhatt, S. et al. (2015). "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015". In: *Nature* 526.7572, pp. 207–211. ISSN: 14764687. DOI: 10.1038/nature15535.

Bian, Guowu et al. (2013). "Wolbachia Invades Anopheles stephensi". In: *Science* 340.May, pp. 748–751. ISSN: 1095-9203. DOI: 10.1126/science.1236192.

Bikard, David et al. (2013). "Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system". In: *Nucleic Acids Research* 41.15, pp. 7429–7437. ISSN: 03051048. DOI: 10.1093/nar/gkt520.

Bischof, Johannes et al. (2007). "An optimized transgenesis system for Drosophila using germline-specific C31 integrases". In: *Proceedings of the National Academy of Sciences* 104.9, pp. 3312–3317. ISSN: 0027-8424. DOI: 10.1073/pnas.0611511104. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1805588&tool=pmcentrez&rendertype=abstracthttp://www.pnas.org/cgi/doi/10.1073/pnas.0611511104.

Boch, Jens and Ulla Bonas (2010). "Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function ". In: *Annual Review of Phytopathology* 48.1, pp. 419–436. ISSN: 0066-4286. DOI: 10.1146/annurev-phyto-080508-081936.

Bordenstein, Sarah R. and Seth R. Bordenstein (2016). "Eukaryotic association module in phage WO genomes from Wolbachia". In: *Nature Communications* 7. ISSN: 20411723. DOI: 10.1038/ncomms13155.

Bosco, Giovanni et al. (2007). "Analysis of Drosophila Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species". In: *Genetics* 177.3, pp. 1277–1290. ISSN: 0016-6731. DOI: 10.1534/genetics.107.075069. URL: http://www.ncbi.nlm.nih.gov/pubmed/18039867http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2147996http://www.genetics.org/lookup/doi/10.1534/genetics.107.075069.

Brand, A H and N Perrimon (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." In: *Development (Cambridge, England)* 118.2, pp. 401–15. ISSN: 0950-1991. URL: http://www.ncbi.nlm.nih.gov/pubmed/8223268.

Brennan, Mark D., Robert G. Rowan, and W. J. Dickinson (1984). "Introduction of a functional P element into the germ-line of Drosophila hawaiiensis". In: *Cell* 38.1, pp. 147–151. ISSN: 00928674. DOI: 10.1016/0092-8674(84)90535-X.

Brown, James B. et al. (2014). "Diversity and dynamics of the Drosophila transcriptome". In: *Nature* 512.7515, pp. 393–399. ISSN: 14764687. DOI: 10.1038/nature12962.

Buchman, Anna et al. (2018). "Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii." In: *Proceedings of the National Academy of Sciences of the United States of America* 115.18, pp. 4725–4730. ISSN: 1091-6490. DOI: 10.1073/pnas.1713139115. URL: http://www.ncbi.nlm.nih.gov/pubmed/29666236http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5939061.

Buckler IV, Edward S. et al. (1999). "Meiotic drive of chromosomal knobs reshaped the maize genome". In: *Genetics* 153.1, pp. 415–426. ISSN: 00166731. DOI: 10.1007/s10681-006-9155-y.

Bull, James J. and Harmit S. Malik (2017). "The gene drive bubble: New realities". In: *PLOS Genetics* 13.7. Ed. by Gregory P. Copenhaver, e1006850. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006850. URL: http://dx.plos.org/10.1371/journal.pgen.1006850.

Burt, Austin (2003). "Site-specific selfish genes as tools for the control and genetic engineering of natural populations". In: *Proceedings of the Royal Society B: Biological Sciences* 270.1518, pp. 921–928. ISSN: 0962-8452. DOI: 10.1098/rspb.2002.2319. URL: http://www.ncbi.nlm.nih.gov/pubmed/12803906%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1691325http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2002.2319.

Burt, Austin and Vassiliki Koufopanou (2004). "Homing endonuclease genes: The rise and fall and rise again of a selfish element". In: *Current Opinion in Genetics and Development* 14.6, pp. 609–615. ISSN: 0959437X. DOI: 10.1016/j.gde.2004.09.010.

Burt, Austin and Robert Trivers (2006). *The Biology of Selfish Genetic Elements*. Harvard University Press. ISBN: 9780674017139. URL: www.jstor.org/stable/j.ctvjhzrc6.

Callies, Daniel Edward (2019). "The ethical landscape of gene drive research". In: *Bioethics* 33.9, pp. 1091–1097. ISSN: 0269-9702. DOI: 10.1111/bioe.12640. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/bioe.12640.

Calvo, Eric et al. (2005). "Nanos (nos) genes of the vector mosquitoes, Anopheles gambiae, Anopheles stephensi and Aedes aegypti". In: *Insect Biochemistry and Molecular Biology* 35.7, pp. 789–798. ISSN: 09651748. DOI: 10.1016/j.ibmb.2005.02.007.

Carballar-Lejarazú, Rebeca and Anthony A. James (2017). "Population modification of Anopheline species to control malaria transmission". In: *Pathogens and Global Health* 111.8, pp. 424–435. ISSN: 20477732. DOI: 10.1080/20477724.2018.1427192.

Castranova, Vincent et al. (2016). "The Dawn of Active Genetics". In: *BioEssays* 38.1, pp. 1922–2013. DOI: 10.1080/10937404.2015.1051611.INHALATION.

Catteruccia, Flaminia, Jason P. Benton, and Andrea Crisanti (2005). "An Anopheles transgenic sexing strain for vector control". In: *Nature Biotechnology* 23.11, pp. 1414–1417. ISSN: 10870156. DOI: 10.1038/nbt1152.

Celniker, Susan E et al. (2009). "Unlocking the secrets of the genome". In: *Nature* 459.7249, pp. 927–930. ISSN: 0028-0836. DOI: 10.1038/459927a. URL: http://tinyurl.com/modENCODEhttp://www.nature.com/doifinder/10.1038/459927a.

Centers for Disease Control and Prevention (2019). *CDC - Malaria - About Malaria - Biology*.

Champer, Jackson, Anna Buchman, and Omar S. Akbari (2016). "Cheating evolution: engineering gene drives to manipulate the fate of wild populations". In: *Nature Reviews Genetics* 17.3, pp. 146–159. ISSN: 1471-0056. DOI: 10.1038/nrg.2015.34. URL: http://dx.doi.org/10.1038/nrg.2015.34.

Champer, Jackson et al. (2017). "Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations". In: *PLOS Genetics* 13.7. Ed. by Harmit S. Malik, e1006796. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006796. URL: https://doi.org/10.1371/journal.pgen.1006796https://dx.plos.org/10.1371/journal.pgen.1006796.

Champer, Jackson et al. (2018a). "Multiple loci of small effect confer wide variability in efficiency and resistance rate of CRISPR gene drives". In: *bioRxiv*. DOI: http://dx.doi.org/10.1101/447615.

Champer, Jackson et al. (2018b). "Reducing resistance allele formation in CRISPR gene drive". In: *Proceedings of the National Academy of Sciences* 115.21, pp. 5522–5527. ISSN: 0027-8424. DOI: 10.1073/pnas.1720354115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1720354115.

Champer, Jackson et al. (2019a). "A toxin-antidote CRISPR gene drive system for regional population modification". In: *bioRxiv*, p. 628354. DOI: 10.1101/628354. URL: https://www.biorxiv.org/content/10.1101/628354v1.full.

Champer, Jackson et al. (2019b). "Molecular safeguarding of CRISPR gene drive experiments". In: *eLife* 8. ISSN: 2050084X. DOI: 10.7554/eLife.41439.

Champer, Jackson et al. (2019c). "Performance analysis of novel toxin-antidote CRISPR gene drive systems". In: *bioRxiv*, p. 628362. DOI: 10.1101/628362. URL: http://biorxiv.org/content/early/2019/05/05/628362.abstract.

Chan, Yuk Sang et al. (2011). "Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in Drosophila melanogaster". In: *Genetics* 188.1, pp. 33–44. ISSN: 00166731. DOI: 10.1534/genetics.111.127506.

Chan, Yuk Sang et al. (2013a). "Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The Drosophila melanogaster Experience". In: *PLoS ONE* 8.1. ISSN: 19326203. DOI: 10.1371/journal.pone.0054130.

Chan, Yuk-Sang et al. (2013b). "The design and in vivo evaluation of engineered I-OnuI-based enzymes for HEG gene drive." In: *PloS one* 8.9, e74254. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074254. URL: http://www.ncbi.nlm.nih.gov/pubmed/24040217http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3769382.

Charlesworth, Scott (2008). *Mosquitoes*. URL: https://extension.entm.purdue.edu/publichealth/insects/mosquito.html.

Chase, Jonathan M. and Rachel S. Shulman (2009). "Wetland isolation facilitates larval mosquito density through the reduction of predators". In: *Ecological Entomology* 34.6, pp. 741–747. ISSN: 03076946. DOI: 10.1111/j.1365-2311.2009.01128.x.

Chavez, Alejandro et al. (2015). "Highly efficient Cas9-mediated transcriptional programming". In: *Nature Methods* 12.4, pp. 326–328. ISSN: 1548-7091. DOI: 10.1038/nmeth.3312. URL: http://www.nature.com/doifinder/10.1038/nmeth.3312.

Chavez, Alejandro et al. (2016). "Comparison of Cas9 activators in multiple species". In: *Nature Methods* 13.7, pp. 563–567. ISSN: 1548-7091. DOI: 10.1038/nmeth.3871. URL: http://www.nature.com/doifinder/10.1038/nmeth.3871.

Chen, Chun-hong et al. (2007). "A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila." In: *Science (New York, N.Y.)* 316.5824, pp. 597–600. ISSN: 1095-9203. DOI: 10.1126/science.1138595. URL: http://www.ncbi.nlm.nih.gov/pubmed/17395794.

Chen, Sidi et al. (2012). "Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene". In: *EMBO Journal* 31.12, pp. 2798–2809. ISSN: 02614189. DOI: 10.1038/emboj.2012.108.

Chevalier, B. S. (2001). "Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility". In: *Nucleic Acids Research* 29.18, pp. 3757–3774. ISSN: 13624962. DOI: 10.1093/nar/29.18.3757. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.18.3757.

Chitra, Thangadurai et al. (2013). "Laboratory and field efficacy of Pedalium murex and predatory copepod, Mesocyclops longisetus on rural malaria vector, Anopheles culicifacies". In: *Asian Pacific Journal of Tropical Disease* 3.2, pp. 111–118. ISSN: 22221808. DOI: 10.1016/S2222-1808(13)60054-7.

Cho, Seung Woo et al. (2014). "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases". In: *Genome Research* 24.1, pp. 132–141. ISSN: 10889051. DOI: 10.1101/gr.162339.113.

Choi, Leslie, Joseph Pryce, and Paul Garner (2019). *Indoor residual spraying for preventing malaria in communities using insecticide-treated nets*. DOI: 10.1002/14651858.CD012688.pub2.

Chrostek, Ewa and Michael Gerth (2019). "Is anopheles gambiae a natural host of wolbachia?" In: *mBio* 10.3. ISSN: 21507511. DOI: 10.1128/mBio.00784-19.

Clement, Kendell et al. (2019). *CRISPResso2 provides accurate and rapid genome editing sequence analysis*. DOI: 10.1038/s41587-019-0032-3.

Coates, Craig J et al. (1998). "Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti". In: *Proceedings of the National Academy of Sciences* 95.7, pp. 3748–3751. ISSN: 0027-8424.

Collins, James P. (2018). *Gene drives in our future: Challenges of and opportunities for using a self-sustaining technology in pest and vector management*. DOI: 10.1186/s12919-018-0110-4.

Cong, L. et al. (2013). "Multiplex Genome Engineering Using CRISPR/Cas Systems". In: *Science* 339.6121, pp. 819–823. ISSN: 0036-8075. DOI: 10.1126/science.1231143. URL: http://www.

sciencemag.org/cgi/doi/10.1126/science.1225053http://www.sciencemag.org/cgi/doi/10.1126/science.1231143http://arxiv.org/abs/1106.5958http://dx.doi.org/10.1103/PhysRevA.85.033634http://link.aps.org/doi/10.1103/PhysRevA.85.033634.

Connell, Marie Sue (2013). "Myosin Dynamics in Drosophila Neuroblasts Lead to Asymmetric Cytokinesis". PhD thesis. University of Oregon.

Corby-Harris, Vanessa et al. (2010). "Activation of akt signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in anopheles stephensi mosquitoes". In: *PLoS Pathogens* 6.7, pp. 1–10. ISSN: 15537366. DOI: 10.1371/journal.ppat.1001003.

Cowman, Alan F et al. (2016). "Leading Edge Review Malaria: Biology and Disease". In: DOI: 10.1016/j.cell.2016.07.055. URL: http://dx.doi.org/10.1016/j.cell.2016.07.055.

Craig, G. B., W. A. Hickey, and R. C. Vandehey (1960). "An inherited male-producing factor in Aedes aegypti". In: *Science* 132.3443, pp. 1887–1889. ISSN: 00368075. DOI: 10.1126/science.132.3443.1887.

Crispino, J D et al. (1996). "Cis-acting elements distinct from the 5′ splice site promote U1-independent pre-mRNA splicing." In: *RNA (New York, N.Y.)* 2.7, pp. 664–73. ISSN: 1355-8382. URL: http://www.ncbi.nlm.nih.gov/pubmed/8756409http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1369405.

Curtis, C. F. (1968). "Possible use of translocations to fix desirable genes in insect pest populations". In: *Nature* 218.5139, pp. 368–369. DOI: 10.1038/218368a0.

Dafa'alla, Tarig H. et al. (2006). "Transposon-free insertions for insect genetic engineering". In: *Nature Biotechnology* 24.7, pp. 820–821. ISSN: 10870156. DOI: 10.1038/nbt1221.

Daniels, Richard W. et al. (2014). "Expression of multiple transgenes from a single construct using viral 2A peptides in Drosophila". In: *PLoS ONE* 9.6. ISSN: 19326203. DOI: 10.1371/journal.pone.0100637.

Davidson, G. et al. (1970). "Field attempt to assess the mating competitiveness of sterile males produced by crossing 2 member species of the Anopheles gambiae complex." In: *Bulletin of the World Health Organization* 42.1, pp. 55–67. ISSN: 0042-9686. URL: http://www.ncbi.nlm.nih.gov/pubmed/5309520http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2427523.

Davis, Kevin M. et al. (2015). "Small molecule-triggered Cas9 protein with improved genome-editing specificity". In: *Nature Chemical Biology* 11.5, pp. 316–318. ISSN: 15524469. DOI: 10.1038/nchembio.1793.

Davis, Stephen, Nicholas Bax, and Peter Grewe (2001). "Engineered Underdominance Allows Efficient and Economical Introgression of Traits into Pest Populations". In: *Journal of Theoretical Biology* 212.1, pp. 83–98. DOI: 10.1006/jtbi.2001.2357. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022519301923574.

Depinay, Jean-Marc O et al. (2004). "A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission." In: *Malaria journal* 3.1, p. 29. ISSN: 1475-2875. DOI: 10.1186/1475-2875-3-29. URL: http://www.malariajournal.com/content/3/1/29%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=514565&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=514565%5C&tool=pmcentrez%5C&rendertype=abstrac.

Deredec, Anne, H. Charles J. Godfray, and Austin Burt (2011). "Requirements for effective malaria control with homing endonuclease genes". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.43. ISSN: 00278424. DOI: 10.1073/pnas.1110717108.

Dhole, Sumit, Alun L. Lloyd, and Fred Gould (2019). "Tethered homing gene drives: A new design for spatially restricted population replacement and suppression". In: *Evolutionary Applications* 12.8, pp. 1688–1702. ISSN: 17524571. DOI: 10.1111/eva.12827. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12827.

Dhole, Sumit et al. (2018). "Invasion and migration of spatially self-limiting gene drives: A comparative analysis." In: *Evolutionary applications* 11.5, pp. 794–808. ISSN: 1752-4571. DOI: 10.1111/eva.12583. URL: http://www.ncbi.nlm.nih.gov/pubmed/29875820http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5978947.

Di Nocera, P P and I B Dawid (1983). "Transient expression of genes introduced into cultured cells of Drosophila." In: *Proceedings of the National Academy of Sciences of the United States of America* 80.23, pp. 7095–8. ISSN: 0027-8424. DOI: 10.1073/pnas.80.23.7095. URL: http://www.ncbi.nlm.nih.gov/pubmed/6417663http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC389999.

Diao, Fengqiu et al. (2015). "Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes." In: *Cell reports* 10.8, pp. 1410–21. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.01.059. URL: http://www.ncbi.nlm.nih.gov/pubmed/25732830http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4373654.

DiCarlo, James E. et al. (2015). "Safeguarding CRISPR-Cas9 gene drives in yeast". In: *Nature Biotechnology* 33.12, pp. 1250–1255. ISSN: 15461696. DOI: 10.1038/nbt.3412. URL: http://dx.doi.org/10.1038/nbt.3412.

Ding, Dan et al. (2018). "Engineering Introns to Express RNA Guides for Cas9- and Cpf1-Mediated Multiplex Genome Editing". In: *Molecular Plant* 11.4, pp. 542–552. ISSN: 17529867. DOI: 10.1016/j.molp.2018.02.005. URL: https://doi.org/10.1016/j.molp.2018.02.005.

Dondorp, Arjen M et al. (2009). "Artemisinin Resistance in Plasmodium falciparum Malaria". In: *New England Journal of Medicine* 361.5, pp. 455–467. ISSN: 0028-4793. DOI: 10.1056/NEJMoa0808859. URL: https://doi.org/10.1056/NEJMoa0808859.

Dong, Yuemei et al. (2011). "Engineered Anopheles Immunity to Plasmodium Infection". In: *PLoS Pathogens* 7.12. Ed. by Kenneth D. Vernick, e1002458. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002458. URL: http://dx.plos.org/10.1371/journal.ppat.1002458.

Donnelly, M L et al. (2001). "The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences." In: *The Journal of general virology* 82.Pt 5, pp. 1027–41. ISSN: 0022-1317. DOI: 10.1099/0022-1317-82-5-1027. URL: http://www.ncbi.nlm.nih.gov/pubmed/11297677.

Duncker, B. P., P. L. Davies, and V. K. Walker (1997). "Introns boost transgene expression in Drosophila melanogaster". In: *Molecular and General Genetics* 254.3, pp. 291–296. ISSN: 00268925. DOI: 10.1007/s004380050418.

Dylan Shropshire, J. et al. (2018). "One prophage WO gene rescues cytoplasmic incompatibility in Drosophila melanogaster". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.19, pp. 4987–4991. ISSN: 10916490. DOI: 10.1073/pnas.1800650115.

Eckhoff, Philip A. et al. (2017). "Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics". In: *Proceedings of the National Academy of Sciences* 114.2, E255–E264. ISSN: 0027-8424. DOI: 10.1073/pnas.1611064114. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1611064114http://www.ncbi.nlm.nih.gov/pubmed/28028208http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5240713.

Edgington, Matthew P. and Luke S. Alphey (2019). "Modeling the mutation and reversal of engineered underdominance gene drives". In: *Journal of Theoretical Biology* 479, pp. 14–21. ISSN: 10958541. DOI: 10.1016/j.jtbi.2019.06.024.

Engels, W.R (2001). "P Elements". In: *Encyclopedia of Genetics*. Elsevier, pp. 1403–1407. DOI: 10.1006/rwgn.2001.0967. URL: https://linkinghub.elsevier.com/retrieve/pii/B0122270800009678.

Esvelt, Kevin M et al. (2014). "Concerning RNA-guided gene drives for the alteration of wild populations". In: *eLife* 3, e03401. ISSN: 2050-084X. DOI: 10.7554/eLife.03401. URL: http://elifesciences.org/lookup/doi/10.7554/eLife.03401.

Fackenthal, J D et al. (1995). "Structural analysis of mutations in the Drosophila beta 2-tubulin isoform reveals regions in the beta-tubulin molecular required for general and for tissue-specific microtubule functions." In: *Genetics* 139.1, pp. 267–86. ISSN: 0016-6731. URL: http://www.ncbi.nlm.nih.gov/pubmed/7705629http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1206324.

Fasulo, Barbara et al. (2012). "The Drosophila Mi-2 Chromatin-Remodeling Factor Regulates Higher-Order Chromatin Structure and Cohesin Dynamics In Vivo". In: *PLoS Genetics* 8.8. ISSN: 15537390. DOI: 10.1371/journal.pgen.1002878.

Flagg, Raymond O. (1988). *Carolina Drosophila Manual*. Carolina Biological Supply Company, p. 3. ISBN: 0-89278-027-4.

Flick, Karen E. et al. (1998). "DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-Ppol". In: *Nature* 394.6688, pp. 96–101. ISSN: 00280836. DOI: 10.1038/27952.

Forbes, Alexandria and Ruth Lehmann (1998). "Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells". In: *Development* 125.4, pp. 679–690. ISSN: 09501991.

Fox-Walsh, Kristi L. et al. (2005). "The architecture of pre-mRNAs affects mechanisms of splice-site pairing". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.45, pp. 16176–16181. ISSN: 00278424. DOI: 10.1073/pnas.0508489102.

Frangioni, Giuliano and Gianfranco Borgioli (1993). "Consequences of splenectomy on the compensatory mechanism of cutaneous respiration in the newt". In: *Journal of Experimental Zoology* 267.2, pp. 130–136. ISSN: 1097010X. DOI: 10.1002/jez.1402670206.

Fraser, Malcolm J. (2012). "Insect Transgenesis: Current Applications and Future Prospects". In: *Annual Review of Entomology* 57.1, pp. 267–289. ISSN: 0066-4170. DOI: 10.1146/annurev.ento.54.110807.090545.

Fu, Yanfang et al. (2013). "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells". In: *Nature Biotechnology* 31.9, pp. 822–826. ISSN: 10870156. DOI: 10.1038/nbt.2623.

Fu, Yanfang et al. (2014). "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs". In: 32.3. DOI: 10.1038/nbt.2808.

Fukuto, T R (1990). "Mechanism of action of organophosphorus and carbamate insecticides." In: *Environmental health perspectives* 87, pp. 245–54. ISSN: 0091-6765. DOI: 10.1289/ehp.9087245. URL: http://www.ncbi.nlm.nih.gov/pubmed/2176588http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1567830.

Galizi, Roberto et al. (2014). "A synthetic sex ratio distortion system for the control of the human malaria mosquito." In: *Nature communications* 5, p. 3977. ISSN: 2041-1723. DOI: 10.1038/ncomms4977. URL: http://www.nature.com/ncomms/2014/140610/ncomms4977/full/ncomms4977.html.

Galizi, Roberto et al. (2016). "A CRISPR-Cas9 sex-ratio distortion system for genetic control." In: *Scientific reports* 6.April, p. 31139. ISSN: 2045-2322. DOI: 10.1038/srep31139. URL: http://www.ncbi.nlm.nih.gov/pubmed/27484623%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4971495.

Gantz, Valentino M and Ethan Bier (2015). "The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations". In: *Science* 348.6233, pp. 442–444. ISSN: 0036-8075. DOI: 10.1126/science.aaa5945. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.aaa5945http://www.ncbi.nlm.nih.gov/pubmed/24651067%5Cnhttp://www.sciencemag.org/content/early/2015/03/18/science.aaa5945.abstract.

Gantz, Valentino M. and Ethan Bier (2016). "The dawn of active genetics". In: *BioEssays* 38.1, pp. 50–63. ISSN: 15211878. DOI: 10.1002/bies.201500102.

Gantz, Valentino M. et al. (2015). "Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi". In: *Proceedings of the National Academy of Sciences* 112.49, E6736–E6743. ISSN: 0027-8424. DOI: 10.1073/pnas.1521077112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1521077112.

Gari, Taye and Bernt Lindtjørn (2018). "Reshaping the vector control strategy for malaria elimination in Ethiopia in the context of current evidence and new tools: Opportunities and challenges 11 Medical and Health Sciences 1108 Medical Microbiology 11 Medical and Health Sciences 1117 Public He". In: *Malaria Journal* 17.1, pp. 1–8. ISSN: 14752875. DOI: 10.1186/s12936-018-2607-8. URL: https://doi.org/10.1186/s12936-018-2607-8.

Ghosh, Anil K., Paulo E.M. Ribolla, and Marcelo Jacobs-Lorena (2001). "Targeting Plasmodium ligands on mosquito salivary glands and midgut with a phage display peptide library". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.23, pp. 13278–13281. ISSN: 00278424. DOI: 10.1073/pnas.241491198.

Gibson, Daniel G. et al. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases". In: *Nature Methods* 6.5, pp. 343–345. ISSN: 15487091. DOI: 10.1038/nmeth.1318.

Gilbert, Luke A et al. (2014). "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes". In: *Cell* 154.2, pp. 442–451. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.06.044.CRISPR-Mediated.

Gilbert, Scott F (2000). "The Origins of Anterior-Posterior Polarity". In: URL: https://www.ncbi.nlm.nih.gov/books/NBK10039/.

Gill, G and M Ptashne (1988). "Negative effect of the transcriptional activator GAL4 University, Cambridge, Massachusetts 02138". In: *Nature* 334, pp. 721–724.

Gilles, H M and D A Warrell (1996). *Bruce-Chwatt's essential malariology*. English. Ed. 3. London: Edward Arnold (Publisher) Ltd., xix + 340 pp. ISBN: 0340677139.

Giraldo-Calderón, Gloria I. et al. (2015). "VectorBase: An updated Bioinformatics Resource for invertebrate vectors and other organisms related with human diseases". In: *Nucleic Acids Research* 43.D1, pp. D707–D713. ISSN: 13624962. DOI: 10.1093/nar/gku1117.

Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017*. URL: http://ghdx.healthdata.org/gbd-results-tool.

Golic, Mary M. et al. (1997). "FLP-mediated DNA mobilization to specific target sites in Drosophila chromosomes". In: *Nucleic Acids Research* 25.18, pp. 3665–3671. ISSN: 03051048. DOI: 10.1093/nar/25.18.3665.

Gong, Peng et al. (2005). "A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly". In: *Nature Biotechnology* 23.4, pp. 453–456. ISSN: 10870156. DOI: 10.1038/nbt1071.

Gossen, M. and H. Bujard (1992). "Tight control of gene expression in mammalian cells by tetracycline-responsive promoters." In: *Proceedings of the National Academy of Sciences* 89.12, pp. 5547–5551. ISSN: 0027-8424. DOI: 10.1073/pnas.89.12.5547. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.89.12.5547.

Gratz, Scott J. et al. (2013). "Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease". In: *Genetics* 194.4, pp. 1029–1035. ISSN: 00166731. DOI: 10.1534/genetics.113.152710.

Gratz, Scott J. et al. (2014). "Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila". In: *Genetics* 196.4, pp. 961–971. ISSN: 19432631. DOI: 10.1534/genetics.113.160713.

Griffin, Jamie T et al. (2010). "Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies." In: *PLoS medicine* 7.8. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1000324. URL: http://www.ncbi.nlm.nih.gov/pubmed/20711482http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2919425.

Grossman, G L et al. (2001). "Germline transformation of the malaria vector, Anopheles gambiae, with the piggyBac transposable element". In: *Insect molecular biology* 10.6, pp. 597–604. ISSN: 0962-1075.

Grunwald, Hannah A. et al. (2019). "Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline". In: *Nature* 566.7742, pp. 105–109. DOI: 10.1038/s41586-019-0875-2.

Gubler, Duane (1998). "Resurgent Vector-Borne Diseases as a Global Health Problem". In: *Emerging Infectious Diseases* 4.3, pp. 442–450. ISSN: 10806040. DOI: 10.3201/eid0403.980326. URL: http://www.cdc.gov/ncidod/eid/vol4no3/gubler.htm.

Guichard, Annabel et al. (2019). "Efficient allelic-drive in Drosophila". In: *Nature Communications* 10.1. ISSN: 20411723. DOI: 10.1038/s41467-019-09694-w.

Gurwitz, David (2014). "Gene drives raise dual-use concerns". In: *Science* 345.6200, 1010 LP – 1010. URL: http://science.sciencemag.org/content/345/6200/1010.2.abstract.

Guyant, Philippe et al. (2015). "Past and new challenges for malaria control and elimination: the role of operational research for innovation in designing interventions". In: *Malaria Journal* 14.1, pp. 1–11. ISSN: 14752875. DOI: 10.1186/s12936-015-0802-4.

Hafen, Ernst, Atsushi Kuroiwa, and Walter J. Gehring (1984). "Spatial distribution of transcripts from the segmentation gene fushi tarazu during Drosophila embryonic development". In: *Cell* 37.3, pp. 833–841. ISSN: 0092-8674. DOI: 10.1016/0092-8674(84)90418-5. URL: https://www.sciencedirect.com/science/article/pii/0092867484904185.

Hales, Simon and Wilbert Van Panhuis (2005). "A new strategy for dengue control". In: *Lancet* 365.9459, pp. 551–552. ISSN: 01406736. DOI: 10.1016/S0140-6736(05)70772-8.

Haley, Benjamin, Bryon Foys, and Michael Levine (2010). "Vectors and parameters that enhance the efficacy of RNAi-mediated gene disruption in transgenic Drosophila." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.25, pp. 11435–11440. ISSN: 0027-8424. DOI: 10.1073/pnas.1006689107.

Hamilton, W. D. (1967). "Extraordinary Sex Ratios". In: *Science* 156.3774, pp. 477–488. ISSN: 0036-8075. DOI: 10.1126/science.156.3774.477. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.156.3774.477.

Hammond, Andrew et al. (2015). "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae". In: *Nature Biotechnology* 34.1, pp. 78–83. ISSN: 1087-0156. DOI: 10.1038/nbt.3439. URL: http://www.nature.com/doifinder/10.1038/nbt.3439.

— (2016). "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae". In: *Nature Biotechnology* 34.1, pp. 78–83. ISSN: 15461696. DOI: 10.1038/nbt.3439.

Hammond, Andrew M. and Roberto Galizi (2018). "Gene drives to fight malaria: current state and future directions". In: *Pathogens and Global Health* 7724, pp. 1–12. ISSN: 2047-7724. DOI: 10.1080/20477724.2018.1438880. URL: http://www.ncbi.nlm.nih.gov/pubmed/29457956%0Ahttps://www.tandfonline.com/doi/full/10.1080/20477724.2018.1438880.

Hammond, Andrew M et al. (2018). "Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations". In: *bioRxiv*, p. 360339. DOI: 10.1101/360339. URL: https://www.biorxiv.org/content/10.1101/360339v1.

Hammond, Andrew Marc et al. (2017). "The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito". In: *bioRxiv*. URL: http://biorxiv.org/content/early/2017/06/12/149005.1.abstract.

Handler, Alfred M. (2001). "A current perspective on insect gene transformation". In: *Insect Biochemistry and Molecular Biology* 31.2, pp. 111–128. ISSN: 09651748. DOI: 10.1016/S0965-1748(00)00159-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965174800001594.

Handler, Alfred M., Grazyna J. Zimowska, and Carsten Horn (2004). "Post-integration stabilization of a transposon vector by terminal sequence deletion in Drosophila melanogaster". In: *Nature Biotechnology* 22.9, pp. 1150–1154. ISSN: 10870156. DOI: 10.1038/nbt1002.

Harris, Angela F. et al. (2012). "Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes". In: *Nature Biotechnology* 30.9, pp. 828–830. ISSN: 10870156. DOI: 10.1038/nbt.2350.

Hauck, Eric S. et al. (2013). "Overexpression of phosphatase and tensin homolog improves fitness and decreases Plasmodium falciparum development in Anopheles stephensi". In: *Microbes and Infection* 15.12, pp. 775–787. ISSN: 12864579. DOI: 10.1016/j.micinf.2013.05.006. URL: http://www.ncbi.nlm.nih.gov/pubmed/23774695http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3788859https://linkinghub.elsevier.com/retrieve/pii/S1286457913001160.

Havlicek, Steven et al. (2017). "Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells". In: *Molecular Therapy* 25.2, pp. 342–355. ISSN: 15250024. DOI: 10.1016/j.ymthe.2016.11.007.

Hay, Simon I et al. (2010). "Developing global maps of the dominant anopheles vectors of human malaria". eng. In: *PLoS medicine* 7.2, e1000209–e1000209. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1000209. URL: https://www.ncbi.nlm.nih.gov/pubmed/20161718https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2817710/.

Heffel, Matthew G. and Gregory C. Finnigan (2019). "Mathematical modeling of self-contained CRISPR gene drive reversal systems". In: *Scientific Reports* 9.1, pp. 1–10. ISSN: 20452322. DOI: 10.1038/s41598-019-54805-8.

Heinrich, J. C. and M. J. Scott (2000). "A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program". In: *Proceedings of the National Academy of Sciences* 97.15, pp. 8229–8232. ISSN: 0027-8424. DOI: 10.1073/pnas.140142697.

Hendrichs, J, G Franz, and P Rendon (1995). "Increased effectiveness and applicability of the sterile insect technique through male-only releases for control of Mediterranean fruit flies during fruiting seasons". In: *Journal of Applied Entomology* 119.1-5, pp. 371–377. ISSN: 0931-2048. DOI: 10.1111/j.1439-0418.1995.tb01303.x. URL: https://doi.org/10.1111/j.1439-0418.1995.tb01303.x.

Herold, N. et al. (2009). "Conservation of the Protein Composition and Electron Microscopy Structure of Drosophila melanogaster and Human Spliceosomal Complexes". In: *Molecular and Cellular Biology* 29.1, pp. 281–301. ISSN: 0270-7306. DOI: 10.1128/mcb.01415-08.

Hertig, Marshall and S Burt Wolbach (1924). "Studies on rickettsia-like micro-organisms in insects". In: *The Journal of medical research* 44.3, p. 329.

Hesselberth, Jay R. (2013). "Lives that introns lead after splicing". In: *Wiley Interdisciplinary Reviews: RNA* 4.6, pp. 677–691. ISSN: 17577004. DOI: 10.1002/wrna.1187. URL: http://doi.wiley.com/10.1002/wrna.1187.

Hickey, W A and G B Craig (1966). "Genetic distortion of sex ratio in a mosquito, Aedes aegypti." In: *Genetics* 53.6, pp. 1177–96. ISSN: 0016-6731. URL: http://www.ncbi.nlm.nih.

gov/pubmed/5958915http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1211089.

Hilgenboecker, Kirsten et al. (2008). "How many species are infected with Wolbachia?–a statistical analysis of current data". In: *FEMS microbiology letters* 281.2, pp. 215–220. ISSN: 0378-1097.

Hilton, Isaac B. et al. (2015). "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers". In: *Nature Biotechnology* 33.5, pp. 510–517. ISSN: 15461696. DOI: 10.1038/nbt.3199.

Hoess, Ronald H. and Kenneth Abremski (1985). "Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system". In: *Journal of Molecular Biology*. ISSN: 00222836. DOI: 10.1016/0022-2836(85)90224-4.

Holdren, John P. et al. (2015). *Improving Transparency and Ensuring Continued Safety in Biotechnology*. URL: https://obamawhitehouse.archives.gov/blog/2015/07/02/improving-transparency-and-ensuring-continued-safety-biotechnology.

Hollingdale, M R et al. (1984). "Inhibition of entry of Plasmodium falciparum and P. vivax sporozoites into cultured cells; an in vitro assay of protective antibodies." In: *Journal of immunology (Baltimore, Md. : 1950)* 132.2, pp. 909–13. ISSN: 0022-1767. URL: http://www.ncbi.nlm.nih.gov/pubmed/6317752.

Horn, Carsten et al. (2002). *Fluorescent transformation markers for insect transgenesis*. DOI: 10.1016/S0965-1748(02)00085-1.

Howard, Ken and Philip Ingham (1986). "Regulatory interactions between the segmentation genes fushi tarazu, hairy, and engrailed in the Drosophila blastoderm". In: *Cell* 44.6, pp. 949–957. ISSN: 0092-8674. DOI: 10.1016/0092-8674(86)90018-8. URL: https://www.sciencedirect.com/science/article/pii/0092867486900188.

Hsu, Patrick D. et al. (2013). "DNA targeting specificity of RNA-guided Cas9 nucleases". In: *Nature Biotechnology* 31.9, pp. 827–832. ISSN: 10870156. DOI: 10.1038/nbt.2647.

Hu, Peinan et al. (2018). "Comparison of various nuclear localization signal-fused Cas9 proteins and Cas9 mRNA for genome editing in Zebrafish". In: *G3: Genes, Genomes, Genetics* 8.3, pp. 823–831. ISSN: 21601836. DOI: 10.1534/g3.117.300359.

Hughes, Grant L et al. (2011). "Wolbachia infections are virulent and inhibit the human malaria parasite Plasmodium falciparum in Anopheles gambiae." In: *PLoS pathogens* 7.5, e1002043. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002043. URL: http://www.ncbi.nlm.nih.gov/pubmed/21625582http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3098226.

Impoinvil, Daniel E et al. (2008). "Abundance of immature Anopheles and culicines (Diptera: Culicidae) in different water body types in the urban environment of Malindi, Kenya." eng. In: *Journal of vector ecology : journal of the Society for Vector Ecology* 33.1, pp. 107–116. ISSN: 1081-1710 (Print).

Isaacs, Alison T et al. (2011). "Engineered Resistance to Plasmodium falciparum Development in Transgenic Anopheles stephensi". In: *PLoS Pathogens* 7.4. Ed. by Nora J. Besansky, e1002017. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002017. URL: http://dx.plos.org/10.1371/journal.ppat.1002017.

Isaacs, Alison T. et al. (2012). "Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.28. ISSN: 00278424. DOI: 10.1073/pnas.1207738109.

Ishino, Yoshizumi et al. (1987). "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product." In: *Journal of bacteriology* 169.12, pp. 5429–5433. ISSN: 0021-9193.

Ito, Junitsu et al. (2002). "Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite". In: *Nature* 417.6887, pp. 452–455. ISSN: 0028-0836. DOI: 10.1038/417452a. URL: http://www.nature.com/doifinder/10.1038/417452a.

James, Stephanie et al. (2018). "Pathway to deployment of gene drive mosquitoes as a potential biocontrol tool for elimination of malaria in sub-Saharan Africa: Recommendations of a scientific working group". In: *American Journal of Tropical Medicine and Hygiene* 98.Suppl 6, pp. 1–49. ISSN: 00029637. DOI: 10.4269/ajtmh.18-0083.

James, Stephanie L et al. (2020). "Toward the Definition of Efficacy and Safety Criteria for Advancing Gene Drive-Modified Mosquitoes to Field Testing". In: *Vector-borne and Zoonotic Diseases*. DOI: 10.1089/vbz.2019.2606. URL: https://fnih.org/sites/default/files/final/.

Jang, S K et al. (1988). "A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation." In: *Journal of virology* 62.8, pp. 2636–43. ISSN: 0022-538X. URL: http://www.ncbi.nlm.nih.gov/pubmed/2839690http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC253694.

Jansen, Ruud et al. (2002). "Identification of genes that are associated with DNA repeats in prokaryotes". In: *Molecular microbiology* 43.6, pp. 1565–1575. ISSN: 0950-382X.

Jasinskiene, Nijole et al. (1998). "Stable transformation of the yellow fever mosquito, Aedes aegypti, with the Hermes element from the housefly". In: *Proceedings of the National Academy of Sciences* 95.7, pp. 3743–3747. ISSN: 0027-8424.

Jin, Li et al. (2013). "Engineered female-specific lethality for control of pest lepidoptera". In: *ACS Synthetic Biology* 2.3, pp. 160–166. ISSN: 21615063. DOI: 10.1021/sb300123m.

Jinek, Martin et al. (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". In: *Science* 337.6096, pp. 816–821. ISSN: 10959203. DOI: 10.1126/science.1225829.

John, H. A., M. L. Birnstiel, and K. W. Jones (1969). "Eradicating Mosquitoes using Translocations". In: *Nature* 223, pp. 582 –587.

Josling, Gabrielle A and Manuel Llinás (2015). "Sexual development in Plasmodium parasites: knowing when it's time to commit". In: DOI: 10.1038/nrmicro3519. URL: www.nature.com/reviews/micro.

Joung, J. Keith and Jeffry D. Sander (2013). "TALENs: A widely applicable technology for targeted genome editing". In: *Nature Reviews Molecular Cell Biology* 14.1, pp. 49–55. DOI: 10.1038/nrm3486.

Jurica, M. S. and B. L. Stoddard (1999). *Homing endonucleases: Structure, function and evolution*. DOI: 10.1007/s000180050372.

Kambris, Zakaria et al. (2010). "Wolbachia stimulates immune gene expression and inhibits plasmodium development in anopheles gambiae". In: *PLoS Pathogens* 6.10. ISSN: 15537366. DOI: 10.1371/journal.ppat.1001143.

Kandul, Nikolay P. et al. (2019). "Transforming insect population control with precision guided sterile males with demonstration in flies". In: *Nature Communications* 10.1, p. 84. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07964-7. URL: http://www.nature.com/articles/s41467-018-07964-7.

Kats, Lee B and Ryan P Ferrer (2003). "Alien predators and amphibian declines: review of two decades of science and the transition to conservation". In: *Diversity and distributions* 9.2, pp. 99–110. ISSN: 1366-9516.

Kessler, Curt C (2001). "Eradication of feral goats and pigs and consequences for other biota on Sarigan Island, Commonwealth of the Northern Mariana Islands". In: URL: https://www.researchgate.net/publication/267362995.

Kiani, Samira et al. (2014). "CRISPR transcriptional repression devices and layered circuits in mammalian cells". In: *Nature Methods* 11.7, pp. 723–726. ISSN: 1548-7091. DOI: 10.1038/nmeth.2969. URL: http://www.ncbi.nlm.nih.gov/pubmed/24797424http://www.nature.com/doifinder/10.1038/nmeth.2969.

Kim, Jin Hee et al. (2011). "High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice". In: *PLoS ONE* 6.4, pp. 1–8. ISSN: 19326203. DOI: 10.1371/journal.pone.0018556.

Kim, Sojung et al. (2014). "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins". In: *Genome Research* 24.6, pp. 1012–1019. ISSN: 15495469. DOI: 10.1101/gr.171322.113.

Kim, Won et al. (2004). "Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium." In: *Journal of medical entomology* 41.3, pp. 447–55. ISSN: 0022-2585. DOI: 10.1603/0022-2585-41.3.447. URL: https://academic.oup.com/jme/article-lookup/doi/10.1603/0022-2585-41.3.447http://www.ncbi.nlm.nih.gov/pubmed/15185949.

Knipling, E. F. (1959). "Sterile-Male Method of Population Control: Successful with some insects, the method may also be effective when applied to other noxious animals". In: *Science* 130.3380, pp. 902–904. ISSN: 0036-8075. DOI: 10.1126/science.130.3380.902. URL: http://www.ncbi.nlm.nih.gov/pubmed/14410136http://www.sciencemag.org/cgi/doi/10.1126/science.130.3380.902.

Kobayashi, Satoru et al. (1996). "Essential role of the posterior morphogen nanos for germline development in Drosophila". In: *Nature* 380.6576, pp. 708–711. ISSN: 00280836. DOI: 10.1038/380708a0.

Kocak, D. Dewran et al. (2019). "Increasing the specificity of CRISPR systems with engineered RNA secondary structures". In: *Nature Biotechnology* 37.6, pp. 657–666. ISSN: 15461696. DOI: 10.1038/s41587-019-0095-1.

Krause, H M, R Klemenz, and W J Gehring (1988). "Expression, modification, and localization of the fushi tarazu protein in Drosophila embryos." In: *Genes & development* 2.8, pp. 1021–

36. ISSN: 0890-9369. DOI: 10.1101/gad.2.8.1021. URL: http://www.ncbi.nlm.nih.gov/pubmed/3049237.

Krstic, Dimitrije, Werner Boll, and Markus Noll (2013). "Influence of the White locus on the courtship behavior of Drosophila males." In: *PloS one* 8.10. ISSN: 19326203. DOI: 10.1371/journal.pone.0077904.

Kuhar, Ryan et al. (2014). "Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks". In: *Nucleic Acids Research* 42.1, e4. DOI: 10.1093/nar/gkt872.

Kuscu, Cem et al. (2014). "Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease". In: *Nature Biotechnology* 32.7, pp. 677–683. ISSN: 15461696. DOI: 10.1038/nbt.2916.

Kyrou, Kyros et al. (2018). "A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes". In: *Nature Biotechnology* September. ISSN: 1087-0156. DOI: 10.1038/nbt.4245. URL: http://www.nature.com/doifinder/10.1038/nbt.4245.

Labow, M A et al. (1990). "Conversion of the lac repressor into an allosterically regulated transcriptional activator for mammalian cells." In: *Molecular and Cellular Biology* 10.7, pp. 3343–3356. ISSN: 0270-7306. DOI: 10.1128/MCB.10.7.3343. URL: http://mcb.asm.org/lookup/doi/10.1128/MCB.10.7.3343.

Lai, Sen-Lin and Tzumin Lee (2006). "Genetic mosaic with dual binary transcriptional systems in Drosophila". In: *Nature Neuroscience* 9.5, pp. 703–709. ISSN: 1097-6256. DOI: 10.1038/nn1681. URL: http://www.ncbi.nlm.nih.gov/pubmed/16582903http://www.nature.com/articles/nn1681.

Larracuente, Amanda M and Daven C Presgraves (2012). "The selfish Segregation Distorter gene complex of Drosophila melanogaster." In: *Genetics* 192.1, pp. 33–53. ISSN: 1943-2631. DOI: 10.1534/genetics.112.141390. URL: http://www.ncbi.nlm.nih.gov/pubmed/22964836http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3430544.

Laughon, Allen and Matthew P. Scott (1984). "Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins". In: *Nature* 310.5972, pp. 25–31. ISSN: 0028-0836. DOI: 10.1038/310025a0. URL: http://www.nature.com/articles/310025a0.

Laven, H. (1967). "Eradication of Culex pipiens fatigans through Cytoplasmic Incompatibility". In: *Nature* 216.5113, pp. 383–384. ISSN: 0028-0836. DOI: 10.1038/216383a0. URL: doi:10.1038/216383a0%5Cnfile:///y:/1149.pdfhttp://www.nature.com/articles/Art216383a0.

Lécuyer, Eric et al. (2007). "Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function". In: *Cell* 131.1, pp. 174–187. ISSN: 00928674. DOI: 10.1016/j.cell.2007.08.003.

Lee, Pei Tseng et al. (2018). "A gene-specific T2A-GAL4 library for drosophila". In: *eLife* 7. ISSN: 2050084X. DOI: 10.7554/eLife.35574.

LePage, Daniel P. et al. (2017). "Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility". In: *Nature* 543.7644, pp. 243–247. ISSN: 0028-0836. DOI: 10.1038/nature21391. URL: http://www.nature.com/doifinder/10.1038/nature21391.

Levin, B R (1979). "Problems and promise in genetic engineering in its potential applications to insect management". In: *Genetics in Relation to Insect Management* 170.

Li, Daisy, Hong Zhou, and Xiao Zeng (2019). "Battling CRISPR-Cas9 off-target genome editing". In: *Cell Biology and Toxicology*. DOI: 10.1007/s10565-019-09485-5.

Li, Fengwu et al. (2004). "Plasmodium ookinete-secreted proteins secreted through a common micronemal pathway are targets of blocking malaria transmission". In: *Journal of Biological Chemistry*. Vol. 279. 25, pp. 26635–26644. DOI: 10.1074/jbc.M401385200.

Li, Ming et al. (2019). "Development of a Confinable Gene-Drive System in the Human Disease Vector, Aedes aegypti". In: *bioRxiv*, p. 645440. DOI: 10.1101/645440. URL: https://www.biorxiv.org/content/10.1101/645440v1.full.

Li, X, J C Heinrich, and M J Scott (2001). "piggyBac-mediated transposition in Drosophila melanogaster: an evaluation of the use of constitutive promoters to control transposase gene expression". In: *Insect Molecular Biology* 10.5, pp. 447–455. ISSN: 0962-1075. DOI: 10.1046/j.0962-1075.2001.00283.x. URL: https://doi.org/10.1046/j.0962-1075.2001.00283.x.

Lim, L P and C B Burge (2001). "A computational analysis of sequence features involved in recognition of short introns." In: *Proceedings of the National Academy of Sciences of the United States of America* 98.20, pp. 11193–8. ISSN: 0027-8424. DOI: 10.1073/pnas.201407298. URL: http://www.pnas.org/content/98/20/11193.short.

Lin, Shuailiang et al. (2015). "In Vivo Transcriptional Activation Using CRISPR/Cas9 in Drosophila". In: *Genetics* 201.2, pp. 433–442. ISSN: 0016-6731. DOI: 10.1534/genetics.115.181065. URL: http://www.genetics.org/cgi/doi/10.1534/genetics.115.181065.

Liu, Ziqing et al. (2017). "Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector". In: *Scientific Reports* 7.1, pp. 1–9. ISSN: 20452322. DOI: 10.1038/s41598-017-02460-2.

Long, Tristan A. F. et al. (2009). "A Cost of Sexual Attractiveness to High-Fitness Females". In: *PLoS Biology* 7.12. Ed. by Russell Bonduriansky, e1000254. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1000254. URL: http://dx.plos.org/10.1371/journal.pbio.1000254.

Lyon, Mary F. (2003). "Transmission Ratio Distortion in Mice". In: *Annual Review of Genetics* 37.1, pp. 393–408. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.37.110801.143030. URL: http://www.annualreviews.org/doi/10.1146/annurev.genet.37.110801.143030.

Lyttle, T W (1977). "Experimental population genetics of meiotic drive systems. I. Pseudo-Y chromosomal drive as a means of eliminating cage populations of Drosophila melanogaster." In: *Genetics* 86.2 Pt. 1, pp. 413–45. ISSN: 0016-6731. URL: http://www.ncbi.nlm.nih.gov/pubmed/407130http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1213685.

Macias, Vanessa, Johanna Ohm, and Jason Rasgon (2017). "Gene Drive for Mosquito Control: Where Did It Come from and Where Are We Headed?" In: *International Journal of Environmental Research and Public Health* 14.9, p. 1006. ISSN: 1660-4601. DOI: 10.3390/ijerph14091006. URL: http://www.mdpi.com/1660-4601/14/9/1006.

Majumdar, Sharmistha and Donald C. Rio (2015). "P Transposable Elements in Drosophila and other Eukaryotic Organisms". In: *Mobile DNA III*. Vol. 176. 2. American Society of Microbiology, pp. 727–752. ISBN: 5106421071. DOI: 10.1128/microbiolspec.MDNA3-0004-2014. URL: https://linkinghub.elsevier.com/retrieve/pii/S0031938416312148http://www.asmscience.org/content/book/10.1128/9781555819217.chap33.

Makarova, Kira S., Yuri I. Wolf, and Eugene V. Koonin (2018). "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" In: *The CRISPR Journal* 1.5, pp. 325–336. ISSN: 2573-1599. DOI: 10.1089/crispr.2018.0033.

Marshall, John M. (2009). "The effect of gene drive on containment of transgenic mosquitoes". In: *Journal of Theoretical Biology* 258.2, pp. 250–265. ISSN: 00225193. DOI: 10.1016/j.jtbi.2009.01.031.

Marshall, John M. and Bruce A. Hay (2011). "Inverse Medea as a Novel Gene Drive System for Local Population Replacement: A Theoretical Analysis". In: *Journal of Heredity* 102.3, p. 336. DOI: 10.1093/JHERED/ESR019.

— (2014). "Medusa: A novel gene drive system for confined suppression of insect populations". In: *PLoS ONE* 9.7. ISSN: 19326203. DOI: 10.1371/journal.pone.0102694.

Marshall, John M. et al. (2011). "Semele: A killer-male, rescue-female system for suppression and replacement of insect disease vector populations". In: *Genetics* 187.2, pp. 535–551. ISSN: 00166731. DOI: 10.1534/genetics.110.124479.

Marshall, John M. et al. (2017). "Overcoming evolved resistance to population-suppressing homing-based gene drives". In: *Scientific Reports* 7.1, pp. 1–12. ISSN: 20452322. DOI: 10.1038/s41598-017-02744-7.

Massey, Jonathan H et al. (2019). "The yellow gene influences Drosophila male mating success through sex comb melanization". In: *eLife* 8. ISSN: 2050-084X. DOI: 10.7554/eLife.49388. URL: https://elifesciences.org/articles/49388.

Matera, A. Gregory and Zefeng Wang (2014). *A day in the life of the spliceosome*. DOI: 10.1038/nrm3742.

Matuschewski, Kai and Herwig Schüler (2008). "Actin/Myosin-Based Gliding Motility in Apicomplexan Parasites". In: *Sub-cellular Biochemistry* 47, pp. 110–120. DOI: 10.1007/978-0-387-78267-6{_}9. URL: http://link.springer.com/10.1007/978-0-387-78267-6_9.

Mauri, Federico et al. (2014). "The conserved discs-large binding partner banderuola regulates asymmetric cell division in drosophila". In: *Current Biology* 24.16, pp. 1811–1825. ISSN: 09609822. DOI: 10.1016/j.cub.2014.06.059.

McKearin, D. M. and A. C. Spradling (1990). "Bag-of-marbles: A Drosophila gene required to initiate both male and female gametogenesis". In: *Genes and Development* 4.12 B, pp. 2242–2251. ISSN: 08909369. DOI: 10.1101/gad.4.12b.2242.

McNaughton, Darlene (2012). "The Importance of Long-Term Social Research in Enabling Participation and Developing Engagement Strategies for New Dengue Control Technologies". In: *PLoS Neglected Tropical Diseases* 6.8. ISSN: 19352727. DOI: 10.1371/journal.pntd.0001785.

Meredith, Janet M. et al. (2011). "Site-specific integration and expression of an anti-malarial gene in transgenic Anopheles gambiae significantly reduces Plasmodium infections". In: *PLoS ONE* 6.1, pp. 1–9. ISSN: 19326203. DOI: 10.1371/journal.pone.0014587.

Michiels, F. et al. (1989). "A 14 bp promoter element directs the testis specificity of the Drosophila beta 2 tubulin gene." In: *The EMBO Journal* 8.5, pp. 1559–1565. ISSN: 0261-4189. DOI: 10.1002/j.1460-2075.1989.tb03540.x.

Miles, Alistair et al. (2017). "Genetic diversity of the African malaria vector anopheles gambiae". In: *Nature* 552, pp. 96–100. ISSN: 14764687. DOI: 10.1038/nature24995.

Miller, Barry R. et al. (1988). "Microinjection of DNA into Aedes Triseriatus Ova and Detection of Integration". In: *The American Journal of Tropical Medicine and Hygiene* 39.5, pp. 502–510. ISSN: 0002-9637. DOI: 10.4269/ajtmh.1988.39.502. URL: http://www.ajtmh.org/content/journals/10.4269/ajtmh.1988.39.502.

Miller, J, A D McLachlan, and A Klug (1985). "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes." In: *The EMBO journal* 4.6, pp. 1609–14. ISSN: 0261-4189. URL: http://www.ncbi.nlm.nih.gov/pubmed/4040853http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC554390.

Miller, L. et al. (1987). "Stable integration and expression of a bacterial gene in the mosquito Anopheles gambiae". In: *Science* 237.4816, pp. 779–781. ISSN: 0036-8075. DOI: 10.1126/science.3039658. URL: http://science.sciencemag.org/content/237/4816/779.abstracthttp://www.sciencemag.org/cgi/doi/10.1126/science.3039658.

Moiroux, Nicolas et al. (2012). "Changes in Anopheles funestus biting behavior following universal coverage of long-lasting insecticidal nets in Benin." In: *The Journal of infectious diseases* 206.10, pp. 1622–9. ISSN: 1537-6613. DOI: 10.1093/infdis/jis565. URL: http://www.ncbi.nlm.nih.gov/pubmed/22966127.

Mojica, F. J.M. et al. (2009). "Short motif sequences determine the targets of the prokaryotic CRISPR defence system". In: *Microbiology* 155.3, pp. 733–740. ISSN: 13500872. DOI: 10.1099/mic.0.023960-0.

Mojica, Francisco J M et al. (2000). "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria". In: *Molecular microbiology* 36.1, pp. 244–246. ISSN: 0950-382X.

Mordecai, Erin A. et al. (2013). "Optimal temperature for malaria transmission is dramatically lower than previously predicted". In: *Ecology Letters* 16.1, pp. 22–30. ISSN: 1461023X. DOI: 10.1111/ele.12015.

Moreira, L A et al. (2000). "Robust gut-specific gene expression in transgenic Aedes aegypti mosquitoes." In: *Proceedings of the National Academy of Sciences of the United States of America* 97.20, pp. 10895–8. ISSN: 0027-8424. DOI: 10.1073/pnas.97.20.10895. URL: http://www.ncbi.nlm.nih.gov/pubmed/11005862http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC27120.

Moreira, Luciano A. et al. (2002). "Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes". In: *Journal of Biological Chemistry* 277.43, pp. 40839–40843. ISSN: 00219258. DOI: 10.1074/jbc.M206647200.

Moretti, R. and M. Calvitti (2013). "Male mating performance and cytoplasmic incompatibility in a wPip Wolbachia trans-infected line of Aedes albopictus (Stegomyia albopicta)". In: *Medical and Veterinary Entomology* 27.4, pp. 377–386. ISSN: 0269283X. DOI: 10.1111/j.1365-2915.2012.01061.x.

Morris, Alison C, Paul Eggleston, and Julian M Crampton (1989). "Genetic transformation of the mosquito Aedes aegypti by micro-injection of DNA". In: *Medical and Veterinary Entomology* 3.1, pp. 1–7. ISSN: 0269-283X. DOI: 10.1111/j.1365-2915.1989.tb00467.x. URL: https://doi.org/10.1111/j.1365-2915.1989.tb00467.xhttp://doi.wiley.com/10.1111/j.1365-2915.1989.tb00467.x.

Moure, Carmen M., Frederick S. Gimble, and Florante A. Quiocho (2002). "Crystal structure of the intein homing endonuclease pi-scei bound to its recognition sequence". In: *Nature Structural Biology* 9.10, pp. 764–770. ISSN: 10728368. DOI: 10.1038/nsb840.

Mutero, C M et al. (2000). "Water management for controlling the breeding of Anopheles mosquitoes in rice irrigation schemes in Kenya." In: *Acta tropica* 76.3, pp. 253–63. ISSN: 0001-706X. URL: http://www.ncbi.nlm.nih.gov/pubmed/10974166.

Nam, Vu Sinh et al. (2005). "Elimination of dengue by community programs using Mesocyclops (copepoda) against Aedes aegypti in central Vietnam". In: *American Journal of Tropical Medicine and Hygiene* 72.1, pp. 67–73. ISSN: 00029637.

NASEM (2016). *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. The National Academies Press. ISBN: 978-0-309-43787-5. DOI: 10.17226/23405. URL: https://doi.org/10.17226/23405.

Nash, Alexander et al. (2019). "Integral gene drives for population replacement." In: *Biology open* 8.1, bio037762. ISSN: 2046-6390. DOI: 10.1242/bio.037762. URL: http://www.ncbi.nlm.nih.gov/pubmed/30498016http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6361204.

Nirmala, Xavier et al. (2006). "Functional characterization of the promoter of the vitellogenin gene, AsVg1, of the malaria vector, Anopheles stephensi". In: *Insect Biochemistry and Molecular Biology* 36.9, pp. 694–700. ISSN: 0965-1748. DOI: 10.1016/J.IBMB.2006.05.011. URL: https://www.sciencedirect.com/science/article/pii/S0965174806001184?via%3Dihub.

Nissim, Lior et al. (2014). "Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells". In: *Molecular Cell* 54.4, pp. 698–710. ISSN: 10974164. DOI: 10.1016/j.molcel.2014.04.022. URL: http://dx.doi.org/10.1016/j.molcel.2014.04.022.

Noble, Charleston et al. (2017). "Current CRISPR gene drive systems are likely to be highly invasive in wild populations". In: *bioRxiv*, pp. 9–12.

Noble, Charleston et al. (2018). "Current CRISPR gene drive systems are likely to be highly invasive in wild populations". In: *eLife* 7. ISSN: 2050084X. DOI: 10.7554/eLife.33423.

Noble, Charleston et al. (2019). "Daisy-chain gene drives for the alteration of local populations". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.17, pp. 8275–8282. ISSN: 10916490. DOI: 10.1073/pnas.1716358116.

Nolan, Tony et al. (2011). "Analysis of Two Novel Midgut-Specific Promoters Driving Transgene Expression in Anopheles stephensi Mosquitoes". In: *PLoS ONE* 6.2. Ed. by Pedro Oliveira, e16471. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0016471. URL: https://dx.plos.org/10.1371/journal.pone.0016471.

North, Ace R., Austin Burt, and H. Charles J. Godfray (2019). "Modelling the potential of genetic control of malaria mosquitoes at national scale". In: *BMC Biology* 17.1, p. 26. ISSN: 17417007. DOI: 10.1186/s12915-019-0645-5. URL: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0645-5.

Nuckolls, Nicole L et al. (2017). "wtf genes are prolific dual poison-antidote meiotic drivers." In: *eLife* 6. ISSN: 2050-084X. DOI: 10.7554/eLife.26033. URL: http://www.ncbi.nlm.nih.gov/pubmed/28631612http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5478261.

Oberhofer, Georg, Tobin Ivy, and Bruce A Hay (2018). "Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs." In: *Proceedings of the National Academy of Sciences of the United States of America* 115.40, E9343–E9352. ISSN: 1091-6490. DOI: 10.1073/pnas.1805278115. URL: https://www.biorxiv.org/content/early/2018/03/27/289546http://www.ncbi.nlm.nih.gov/pubmed/30224454.

Oberhofer, Georg, Tobin Ivy, and Bruce A. Hay (2019). "Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.13, pp. 6250–6259. ISSN: 10916490. DOI: 10.1073/pnas.1816928116.

Oberstein, Adam et al. (2005). "Site-specific transgenesis by Cre-mediated recombination in Drosophila". In: *Nature Methods* 2.8, pp. 583–585. ISSN: 15487091. DOI: 10.1038/nmeth775.

O'Brochta, David A. and Peter W. Atkinson (1996). "Transposable elements and gene transformation in non-drosophilid insects". In: *Insect Biochemistry and Molecular Biology* 26.8-9, pp. 739–753. ISSN: 09651748. DOI: 10.1016/S0965-1748(96)00022-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965174896000227.

Ooi, Siew Loon et al. (2001). "RNA Lariat Debranching Enzyme". In: pp. 233–248. DOI: 10.1016/S0076-6879(01)42548-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0076687901425481.

Oye, Kenneth a et al. (2014). "Regulating gene drives". In: *Science* 345.6197, pp. 626–628. ISSN: 0036-8075. DOI: 10.1126/science.1254287. URL: http://www.ncbi.nlm.nih.gov/pubmed/25035410http://www.sciencemag.org/cgi/doi/10.1126/science.1254287.

Pai, Athma A. et al. (2017). "The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture". In: *eLife* 6. ISSN: 2050084X. DOI: 10.7554/eLife.32537.

Papathanos, Philippos A et al. (2009). "The vasa regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies". In: *BMC Molecular Biology* 10.1, p. 65. ISSN: 1471-2199. DOI: 10.1186/1471-2199-10-65. URL: http://bmcmolbiol.biomedcentral.com/articles/10.1186/1471-2199-10-65.

Parma, David H., Paul E. Bennett, and Robert E. Boswell (2007). "Mago Nashi and Tsunagi/Y14, respectively, regulate Drosophila germline stem cell differentiation and oocyte specification". In: *Developmental Biology* 308.2, pp. 507–519. ISSN: 00121606. DOI: 10.1016/j.ydbio.2007.06.007.

Pattanayak, Vikram et al. (2013). "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity". In: *Nature Biotechnology* 31.9, pp. 839–843. ISSN: 10870156. DOI: 10.1038/nbt.2673.

Pelletier, Jerry and Nahum Sonenberg (1988). "Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA". In: *Nature* 334.6180, pp. 320–325. ISSN: 00280836. DOI: 10.1038/334320a0.

Penalva, L. O. F. and L. Sanchez (2003). "RNA Binding Protein Sex-Lethal (Sxl) and Control of Drosophila Sex Determination and Dosage Compensation". In: *Microbiology and Molecular Biology Reviews* 67.3, pp. 343–359. ISSN: 1092-2172. DOI: 10.1128/mmbr.67.3.343-359.2003.

Phillips, Ben L. and Richard Shine (2004). "Adapting to an invasive species: Toxic cane toads induce morphological change in Australian snakes". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.49, pp. 17150–17155. ISSN: 00278424. DOI: 10.1073/pnas.0406440101.

Phuc, Hoang et al. (2007). "Late-acting dominant lethal genetic systems and mosquito control". In: *BMC Biology* 5.1, p. 11. ISSN: 17417007. DOI: 10.1186/1741-7007-5-11. URL: http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-5-11.

Port, Fillip et al. (2014). "Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.29. ISSN: 10916490. DOI: 10.1073/pnas.1405500111.

Potter, Christopher J. et al. (2010). "The Q system: A repressible binary system for transgene expression, lineage tracing, and mosaic analysis". In: *Cell* 141.3, pp. 536–548. ISSN: 00928674. DOI: 10.1016/j.cell.2010.02.025. URL: http://dx.doi.org/10.1016/j.cell.2010.02.025.

Pougach, Ksenia et al. (2010). "Transcription, processing and function of CRISPR cassettes in Escherichia coli." In: *Molecular microbiology* 77.6, pp. 1367–79. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2010.07265.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/20624226http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2939963.

Praetorius, Christian (2016). *Homemade PCR Test for Mycoplasma Contamination - Bitesize Bio*. URL: https://bitesizebio.com/23682/homemade-pcr-test-for-mycoplasma-contamination/.

Protopopoff, Natacha et al. (2018). "Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two factorial design trial." In: *Lancet (London, England)* 391.10130, pp. 1577–1588. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(18)30427-6. URL: http://www.ncbi.nlm.nih.gov/pubmed/29655496http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5910376.

Provataris, Panagiotis et al. (2018). "Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola". In: *Genome Biology and Evolution* 10.4,

pp. 1185–1197. ISSN: 1759-6653. DOI: 10.1093/gbe/evy066. URL: https://doi.org/10.1093/gbe/evy066.

Qi, Lei S et al. (2013). "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression". In: *Cell* 152.5, pp. 1173–1183. DOI: 10.1016/j.cell.2013.02.022.Repurposing.

Quezada-Díaz, Jorge Ernesto et al. (2010). "Drcd-1 related: a positively selected spermatogenesis retrogene in Drosophila". In: *Genetica* 138.9, pp. 925–937. ISSN: 00166707. DOI: 10.1007/s10709-010-9474-8.

Raag, R and M Whitlow (1995). "Single-chain Fvs." In: *The FASEB Journal* 9.1, pp. 73–80. ISSN: 0892-6638. DOI: 10.1096/fasebj.9.1.7821762. URL: http://www.fasebj.org/doi/10.1096/fasebj.9.1.7821762.

Raghavendra, Kamaraju et al. (2011). "Malaria vector control: From past to future". In: *Parasitology Research* 108.4, pp. 757–779. ISSN: 09320113. DOI: 10.1007/s00436-010-2232-0.

Reese, M G et al. (1997). "Improved splice site detection in Genie." In: *Journal of computational biology : a journal of computational molecular cell biology* 4.3, pp. 311–23. ISSN: 1066-5277. DOI: 10.1089/cmb.1997.4.311.

Reeves, R. Guy et al. (2014). "First Steps towards Underdominant Genetic Transformation of Insect Populations". In: *PLoS ONE* 9.5. Ed. by Alexander W. E. Franz, e97557. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0097557. URL: https://dx.plos.org/10.1371/journal.pone.0097557.

Riabinina, Olena and Christopher J Potter (2016). "The Q-System: A Versatile Expression System for Drosophila." In: *Methods in molecular biology (Clifton, N.J.)* 1478, pp. 53–78. ISSN: 1940-6029. DOI: 10.1007/978-1-4939-6371-3{_}3. URL: http://www.ncbi.nlm.nih.gov/pubmed/27730575http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5270762.

Riabinina, Olena et al. (2015). "Improved and expanded Q-system reagents for genetic manipulations". In: *Nature Methods* 12.3, pp. 219–222. ISSN: 1548-7091. DOI: 10.1038/nmeth.3250. URL: http://www.ncbi.nlm.nih.gov/pubmed/25581800http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4344399http://www.nature.com/articles/nmeth.3250.

Riehle, Michelle M. et al. (2008). "Anopheles gambiae APL1 Is a Family of Variable LRR Proteins Required for Rel1-Mediated Protection from the Malaria Parasite, Plasmodium berghei". In: *PLoS ONE* 3.11. Ed. by Matthew W. Hahn, e3672. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0003672. URL: https://dx.plos.org/10.1371/journal.pone.0003672.

Robinett, Carmen C. et al. (2010). "Sex and the single cell. II. there is a time and place for sex". In: *PLoS Biology* 8.5. ISSN: 15449173. DOI: 10.1371/journal.pbio.1000365.

Robinson, A S (1976). "Progress in the use of chromosomal translocations for the control of insect pests." In: *Biological reviews of the Cambridge Philosophical Society* 51.1, pp. 1–24. ISSN: 1464-7931. URL: http://www.ncbi.nlm.nih.gov/pubmed/3229.

Rode, Nicolas O et al. (2013). "Cryptic microsporidian parasites differentially affect invasive and native Artemia spp." In: *International journal for parasitology* 43.10, pp. 795–803. ISSN:

1879-0135. DOI: 10.1016/j.ijpara.2013.04.009. URL: http://www.ncbi.nlm.nih.gov/pubmed/23851079.

Rojas-Fernandez, Alejandro et al. (2015). "Rapid generation of endogenously driven transcriptional reporters in cells through CRISPR/Cas9". In: *Scientific Reports* 5, pp. 1–6. ISSN: 20452322. DOI: 10.1038/srep09811.

Rose, Alan B. (2004). "The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis". In: *Plant Journal* 40.5, pp. 744–751. ISSN: 09607412. DOI: 10.1111/j.1365-313X.2004.02247.x.

— (2019). "Introns as Gene Regulators: A Brick on the Accelerator". In: *Frontiers in Genetics* 9.FEB, p. 672. ISSN: 1664-8021. DOI: 10.3389/fgene.2018.00672. URL: https://www.frontiersin.org/article/10.3389/fgene.2018.00672/full.

Roux, Kyle J et al. (2012). "A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells." In: *The Journal of cell biology* 196.6, pp. 801–10. ISSN: 1540-8140. DOI: 10.1083/jcb.201112098. URL: http://www.ncbi.nlm.nih.gov/pubmed/22412018http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3308701.

Roux, Olivier and Vincent Robert (2019). *Larval predation in malaria vectors and its potential implication in malaria transmission: An overlooked ecosystem service?* DOI: 10.1186/s13071-019-3479-7.

Rowan, Kathryn H. et al. (2004). "Tn5 as an insect gene vector". In: *Insect Biochemistry and Molecular Biology* 34.7, pp. 695–705. ISSN: 09651748. DOI: 10.1016/j.ibmb.2004.03.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965174804000736.

Roy, Scott William and Walter Gilbert (2006). *The evolution of spliceosomal introns: Patterns, puzzles and progress*. DOI: 10.1038/nrg1807.

RTSS Clinical Trials Partnership (2015). "Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial". In: *The Lancet* 386.9988, pp. 31–45. DOI: 10.1016/S0140-6736(15)60721-8.

Rubin, Gerald M. and Allan C. Spradling (1982). "Genetic transformation of Drosophila with transposable element vectors". In: *Science* 218.4570, pp. 348–353. ISSN: 00368075. DOI: 10.1126/science.6289436.

Rupp, H R (1996). "Adverse assessments of Gambusia affinis: an alternate view for mosquito control practitioners." In: *Journal of the American Mosquito Control Association* 12.2 Pt 1, pp. 155–159. ISSN: 8756-971X.

Sano, Hiroko, Akira Nakamura, and Satoru Kobayashi (2002). "Identification of a transcriptional regulatory region for germline-specific expression of vasa gene in Drosophila melanogaster". In: *Mechanisms of Development* 112.1-2, pp. 129–139. ISSN: 09254773. DOI: 10.1016/S0925-4773(01)00654-2.

Scavarda, N. J. and D. L. Hartl (1984). "Interspecific DNA transformation in Drosophila". In: *Proceedings of the National Academy of Sciences of the United States of America* 81.23 I, pp. 7515–7519. ISSN: 00278424. DOI: 10.1073/pnas.81.23.7515.

Schneider, Imogene (1972). "Cell lines derived from late embryonic stages of Drosophila melanogaster". In: *Development* 27.2.

Schüpbach, Trudi and Eric Wieschaus (1986). "Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo". In: *Roux's Archives of Developmental Biology* 195.5, pp. 302–317. ISSN: 0930035X. DOI: 10.1007/BF00376063.

Searles, Lillie L. et al. (1982). "Molecular cloning of sequences from a drosophila RNA polymerase II locus by P element transposon tagging". In: *Cell* 31.3, pp. 585–592. ISSN: 00928674. DOI: 10.1016/0092-8674(82)90314-2. URL: https://linkinghub.elsevier.com/retrieve/pii/0092867482903142.

Senecoff, J. F., R. C. Bruckner, and M. M. Cox (1985). "The FLP recombinase of the yeast 2-μm plasmid: Characterization of its recombination site". In: *Proceedings of the National Academy of Sciences of the United States of America* 82.21, pp. 7270–7274. ISSN: 00278424. DOI: 10.1073/pnas.82.21.7270.

Serebrovsky, A. S. (1969). *On the possibility of a new method for the control of insect pests.* English.

Shaukat, Ayesha M., Joel G. Breman, and F. Ellis McKenzie (2010). *Using the entomological inoculation rate to assess the impact of vector control on malaria parasite transmission and elimination.* DOI: 10.1186/1475-2875-9-122.

Shaul, Orit (2017). *How introns enhance gene expression*. DOI: 10.1016/j.biocel.2017.06.016.

Shaw, W. Robert et al. (2016). "Wolbachia infections in natural Anopheles populations affect egg laying and negatively correlate with Plasmodium development". In: *Nature Communications* 7. ISSN: 20411723. DOI: 10.1038/ncomms11772.

Sherrard-Smith, Ellie et al. (2018). "Systematic review of indoor residual spray efficacy and effectiveness against Plasmodium falciparum in Africa". In: *Nature Communications* 9.1. ISSN: 20411723. DOI: 10.1038/s41467-018-07357-w.

Siegal, M L and D L Hartl (1996). "Transgene Coplacement and high efficiency site-specific recombination with the Cre/loxP system in Drosophila." In: *Genetics* 144.2, pp. 715–26. ISSN: 0016-6731. URL: http://www.ncbi.nlm.nih.gov/pubmed/8889532http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1207562.

Simoni, Alekos et al. (2014). "Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster". In: *Nucleic Acids Research* 42.11, pp. 7461–7472. ISSN: 13624962. DOI: 10.1093/nar/gku387.

Simões, Maria L. et al. (2017). "The Anopheles FBN9 immune factor mediates Plasmodium species-specific defense through transgenic fat body expression". In: *Developmental & Comparative Immunology* 67, pp. 257–265. ISSN: 0145-305X. DOI: 10.1016/J.DCI.2016.09.012. URL: https://www.sciencedirect.com/science/article/pii/S0145305X16302907?via%3Dihub.

Sinden, R E (1974). "Excystment by sporozoites of malaria parasites." In: *Nature* 252.5481, p. 314. ISSN: 0028-0836. DOI: 10.1038/252314a0. URL: http://www.ncbi.nlm.nih.gov/pubmed/4431453.

Sinden, R E and K Matuschewski (2005). *Molecular Approaches to Malaria*. Washington DC: ASM Press, pp. 169–190.

Singh, Utkarsha A., Mukta Kumari, and Soumya Iyengar (2018). "Method for improving the quality of genomic DNA obtained from minute quantities of tissue and blood samples using Chelex 100 resin". In: *Biological Procedures Online* 20.1. ISSN: 14809222. DOI: 10.1186/s12575-018-0077-6.

Sinka, Marianne E et al. (2012). "A global map of dominant malaria vectors". In: *Parasites & Vectors* 5.1, p. 69. ISSN: 1756-3305. DOI: 10.1186/1756-3305-5-69. URL: https://doi.org/10.1186/1756-3305-5-69https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-5-69.

Sinkins, Steven P. and Fred Gould (2006). "Gene drive systems for insect disease vectors". In: *Nature Reviews Genetics* 7.6, pp. 427–435. ISSN: 1471-0056. DOI: 10.1038/nrg1870. URL: http://www.nature.com/doifinder/10.1038/nrg1870.

Smith, Ryan C., Joel Vega-Rodríguez, and Marcelo Jacobs-Lorena (2014). "The Plasmodium bottleneck: Malaria parasite losses in the mosquito vector". In: *Memorias do Instituto Oswaldo Cruz* 109.5, pp. 644–661. ISSN: 16788060. DOI: 10.1590/0074-0276130597.

Sorek, Rotem, C. Martin Lawrence, and Blake Wiedenheft (2013). "CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea". In: *Annual Review of Biochemistry* 82.1, pp. 237–266. ISSN: 0066-4154. DOI: 10.1146/annurev-biochem-072911-172315.

Souza-Moreira, Tatiana M. et al. (2018). "Screening of 2A peptides for polycistronic gene expression in yeast". In: *FEMS Yeast Research* 18.5. ISSN: 15671364. DOI: 10.1093/femsyr/foy036.

Spradling, Allan C. et al. (1999). "The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes". In: *Genetics* 153.1, pp. 135–177. ISSN: 00166731.

Stark, Kenneth R. and Anthony A James (1996). "Salivary Gland Anticoagulants in Culicine and Anopheline Mosquitoes (Diptera: Culicidae)". In: *Journal of Medical Entomology* 33.4, pp. 645–650. ISSN: 00222585. DOI: 10.1093/jmedent/33.4.645.

Steffan, Wallace a and Neal L Evenhuis (1981). "Biology of Toxorhynchites". In: *Annual Review of Entomology* 26.1, pp. 159–181. ISSN: 0066-4170. DOI: 10.1146/annurev.en.26.010181.001111. URL: http://www.annualreviews.org/doi/10.1146/annurev.en.26.010181.001111.

Stoddard, Barry L. (2005). "Homing endonuclease structure and function". In: *Quarterly Reviews of Biophysics* 38.1, pp. 49–95. DOI: 10.1017/S0033583505004063.

Sturm, Angelika et al. (2006). "Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids". In: *Science* 313.5791, pp. 1287–1290. ISSN: 00368075. DOI: 10.1126/science.1129720.

Sturtevant, A. H. (1915). "Experiments on sex recognition and the problem of sexual selection in Drosoophilia". In: *Journal of Animal Behavior* 5.5, pp. 351–366. ISSN: 00959928. DOI: 10.1037/h0074109.

Sweeny, T L and A R Barr (1978). "Sex Ratio Distortion Caused by Meiotic Drive in a Mosquito, Culex pipiens L". eng. In: *Genetics* 88.3, pp. 427–446. ISSN: 0016-6731. URL: https://www.ncbi.nlm.nih.gov/pubmed/17248804https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1224591/.

Szymczak, Andrea L and Dario A A Vignali (2005). "Development of 2A peptide-based strategies in the design of multicistronic vectors." In: *Expert opinion on biological therapy* 5.5, pp. 627–38. ISSN: 1744-7682. DOI: 10.1517/14712598.5.5.627. URL: http://www.ncbi.nlm.nih.gov/pubmed/15934839.

Szymczak, Andrea L. et al. (2004). "Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector". In: *Nature Biotechnology* 22.5, pp. 589–594. ISSN: 10870156. DOI: 10.1038/nbt957.

Takeuchi, Ryo et al. (2011). "Tapping natural reservoirs of homing endonucleases for targeted gene modification." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.32, pp. 13077–82. ISSN: 1091-6490. DOI: 10.1073/pnas.1107719108. URL: http://www.ncbi.nlm.nih.gov/pubmed/21784983http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3156218.

Takken, Willem and Niels O Verhulst (2013). "Host Preferences of Blood-Feeding Mosquitoes". In: *Annual Review of Entomology* 58.1, pp. 433–453. ISSN: 0066-4170. DOI: 10.1146/annurev-ento-120811-153618. URL: http://www.annualreviews.org/doi/10.1146/annurev-ento-120811-153618.

Talerico, M and S M Berget (1994). "Intron definition in splicing of small Drosophila introns." In: *Molecular and Cellular Biology* 14.5, pp. 3434–3445. ISSN: 0270-7306. DOI: 10.1128/MCB.14.5.3434. URL: http://mcb.asm.org/lookup/doi/10.1128/MCB.14.5.3434.

Talhouarne, Gaëlle J S and Joseph G Gall (2018). "Lariat intronic RNAs in the cytoplasm of vertebrate cells." In: *Proceedings of the National Academy of Sciences of the United States of America* 115.34, E7970–E7977. ISSN: 1091-6490. DOI: 10.1073/pnas.1808816115. URL: http://www.ncbi.nlm.nih.gov/pubmed/30082412http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6112733.

Tang, Xu et al. (2017). "A CRISPR – Cpf1 System for Efficient Genome Editing and Transcriptional Repression in Plants". In: *Nature Plants* 17018.February, pp. 1–5. DOI: 10.1038/nplants.2017.18.

Tazuke, Salli I. et al. (2002). "A germline-specific gap junction protein required for survival of differentiating early germ cells". In: *Development* 129.10.

Terns, Michael P. and Rebecca M. Terns (2011). "CRISPR-based adaptive immune systems". In: *Current Opinion in Microbiology* 14.3, pp. 321–327. DOI: 10.1016/j.mib.2011.03.005.

Tetreau, Guillaume et al. (2013). "Monitoring resistance to Bacillus thuringiensis subsp. israelensis in the field by performing bioassays with each Cry toxin separately". In: *Memorias do Instituto Oswaldo Cruz* 108.7, pp. 894–900. ISSN: 00740276. DOI: 10.1590/0074-0276130155.

Thomas, Dean D. et al. (2000). "Insect population control using a dominant, repressible, lethal genetic system". In: *Science* 287.5462, pp. 2474–2476. ISSN: 00368075. DOI: 10.1126/science.287.5462.2474.

Thomsen, Edward K. et al. (2017). "Mosquito behavior change after distribution of bednets results in decreased protection against malaria exposure". In: *Journal of Infectious Diseases*. Vol. 215. 5. Oxford University Press, pp. 790–797. DOI: 10.1093/infdis/jiw615.

Thurmond, Jim et al. (2019). "FlyBase 2.0: the next generation". In: *Nucleic Acids Research* 47.D1, pp. D759–D765. ISSN: 0305-1048. DOI: 10.1093/nar/gky1003. URL: https://academic.oup.com/nar/article/47/D1/D759/5144957.

Thyme, Summer B et al. (2014). "Reprogramming homing endonuclease specificity through computational design and directed evolution." In: *Nucleic acids research* 42.4, pp. 2564–76. ISSN: 1362-4962. DOI: 10.1093/nar/gkt1212. URL: http://www.ncbi.nlm.nih.gov/pubmed/24270794http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3936771.

Thyme, Summer B. et al. (2016). "Internal guide RNA interactions interfere with Cas9-mediated cleavage". In: *Nature Communications* 7. ISSN: 20411723. DOI: 10.1038/ncomms11750.

Tilgner, Hagen et al. (2012). "Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs". In: *Genome Research* 22.9, pp. 1616–1625. ISSN: 10889051. DOI: 10.1101/gr.134445.111.

Tomancak, Pavel et al. (2002a). "(No Title)". In: *Genome Biology* 3.12, research0088.1. ISSN: 14656906. DOI: 10.1186/gb-2002-3-12-research0088. URL: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2002-3-12-research0088.

Tomancak, Pavel et al. (2002b). "Systematic determination of patterns of gene expression during Drosophila embryogenesis." In: *Genome biology* 3.12. ISSN: 14656914. DOI: 10.1186/gb-2002-3-12-research0088.

Travassos, Mark and Miriam K Laufer (2019). *Antimalarial drugs: An overview - UpToDate*. URL: https://www.uptodate.com/contents/antimalarial-drugs-an-overview#H12.

Triezenberg, S J, R C Kingsbury, and S L McKnight (1988). "Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression." In: *Genes & Development* 2.6, pp. 718–729. ISSN: 0890-9369. DOI: 10.1101/gad.2.6.718. URL: http://www.genesdev.org/cgi/doi/10.1101/gad.2.6.718.

Truong, Dong Jiunn Jeffery et al. (2015). "Development of an intein-mediated split-Cas9 system for gene therapy". In: *Nucleic Acids Research* 43.13, pp. 6450–6458. ISSN: 13624962. DOI: 10.1093/nar/gkv601.

Tsai, Shengdar Q. et al. (2015). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases". In: *Nature Biotechnology* 33.2, pp. 187–198. ISSN: 15461696. DOI: 10.1038/nbt.3117.

Unckless, Robert L., Andrew G. Clark, and Philipp W. Messer (2017). "Evolution of resistance against CRISPR/Cas9 gene drive". In: *Genetics* 205.2, pp. 827–841. ISSN: 19432631. DOI: 10.1534/genetics.116.197285.

Uphoff, Cord C and Hans G Drexler (2002). "Comparative PCR analysis for detection of mycoplasma infections in continuous cell lines." In: *In vitro cellular & developmental biology. Animal* 38.2, pp. 79–85. ISSN: 1071-2690. DOI: 10.1290/1071-2690(2002)038<0079:CPAFDO>2.0.CO;2. URL: http://www.ncbi.nlm.nih.gov/pubmed/11928999.

— (2011). "Detecting mycoplasma contamination in cell cultures by polymerase chain reaction." In: *Methods in molecular biology (Clifton, N.J.)* 731, pp. 93–103. ISSN: 1940-6029. DOI: 10.1007/978-1-61779-080-5{_}8. URL: http://www.ncbi.nlm.nih.gov/pubmed/21516400.

Urnov, Fyodor D. et al. (2010). "Genome editing with engineered zinc finger nucleases". In: *Nature Reviews Genetics* 11.9, pp. 636–646. DOI: 10.1038/nrg2842.

Vella, Michael R. et al. (2017). "Evaluating strategies for reversing CRISPR-Cas9 gene drives". In: *Scientific Reports* 7.1, pp. 1–8. ISSN: 20452322. DOI: 10.1038/s41598-017-10633-2. URL: http://dx.doi.org/10.1038/s41598-017-10633-2.

Volohonsky, Gloria et al. (2015). "Tools for anopheles gambiae transgenesis". In: *G3: Genes, Genomes, Genetics* 5.6, pp. 1151–1163. ISSN: 21601836. DOI: 10.1534/g3.115.016808.

Volohonsky, Gloria et al. (2017). "Transgenic Expression of the Anti-parasitic Factor TEP1 in the Malaria Mosquito Anopheles gambiae". In: *PLOS Pathogens* 13.1. Ed. by Kenneth D Vernick, e1006113. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1006113. URL: https://dx.plos.org/10.1371/journal.ppat.1006113.

Wakimoto, B.T., F.R. Turner, and T.C. Kaufman (1984). "Defects in embryogenesis in mutants associated with the antennapedia gene complex of Drosophila melanogaster". In: *Developmental Biology* 102.1, pp. 147–172. ISSN: 0012-1606. DOI: 10.1016/0012-1606(84)90182-9. URL: https://www.sciencedirect.com/science/article/pii/0012160684901829.

Walker, K and M Lynch (2007). "Contributions of Anopheles larval control to malaria suppression in tropical Africa: review of achievements and potential." eng. In: *Medical and veterinary entomology* 21.1, pp. 2–21. ISSN: 0269-283X (Print). DOI: 10.1111/j.1365-2915.2007.00674.x.

Walker, T. et al. (2011). "The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations". In: *Nature* 476.7361, pp. 450–455. ISSN: 00280836. DOI: 10.1038/nature10355.

Wang, Charlotte and Ruth Lehmann (1991). "Nanos is the localized posterior determinant in Drosophila". In: *Cell* 66.4, pp. 637–647. ISSN: 00928674. DOI: 10.1016/0092-8674(91)90110-K.

Wang, Hailong and Xingzhi Xu (2017). "Microhomology-mediated end joining: new players join the team". In: *Cell & Bioscience* 7.1, p. 6. ISSN: 2045-3701. DOI: 10.1186/s13578-017-0136-8. URL: http://cellandbioscience.biomedcentral.com/articles/10.1186/s13578-017-0136-8.

Wang, Sibao and Marcelo Jacobs-Lorena (2013). "Genetic approaches to interfere with malaria transmission by vector mosquitoes". In: *Trends in Biotechnology* 31.3, pp. 185–193. ISSN: 01677799. DOI: 10.1016/j.tibtech.2013.01.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167779913000024.

Werther, Rachel et al. (2017). "Crystallographic analyses illustrate significant plasticity and efficient recoding of meganuclease target specificity." In: *Nucleic acids research* 45.14, pp. 8621–8634. ISSN: 1362-4962. DOI: 10.1093/nar/gkx544. URL: http://www.ncbi.nlm.nih.gov/pubmed/28637173http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5737575.

White, Steven M., Pejman Rohani, and Steven M. Sait (2010). "Modelling pulsed releases for sterile insect techniques: Fitness costs of sterile and transgenic males and the effects on mosquito dynamics". In: *Journal of Applied Ecology* 47.6, pp. 1329–1339. ISSN: 00218901. DOI: 10.1111/j.1365-2664.2010.01880.x.

Wienert, Beeke et al. (2019). "Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq." In: *Science (New York, N.Y.)* 364.6437, pp. 286–289. ISSN: 1095-9203. DOI: 10.1126/science.aav9023. URL: http://www.ncbi.nlm.nih.gov/pubmed/31000663http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6589096.

Wilk, Ronit et al. (2016). "Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs". In: *Genes and Development* 30.5, pp. 594–609. ISSN: 15495477. DOI: 10.1101/gad.276931.115.

Will, Cindy L. and Reinhard Luhrmann Max (2002). "Spliceosome Structure and Function". In: *Cold Spring Harb Perspect Biol* 3.7, pp. 2875–2881. DOI: 10.1039/DT9930002875.

Windbichler, Nikolai, Philippos Aris Papathanos, and Andrea Crisanti (2008). "Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae". In: *PLoS Genetics* 4.12. ISSN: 15537390. DOI: 10.1371/journal.pgen.1000291.

Windbichler, Nikolai et al. (2007). "Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos". In: *Nucleic Acids Research* 35.17, pp. 5922–5933. ISSN: 03051048. DOI: 10.1093/nar/gkm632.

Windbichler, Nikolai et al. (2011). "A synthetic homing endonuclease-based gene drive system in the human malaria mosquito". In: *Nature* 473.7346, pp. 212–215. ISSN: 0028-0836. DOI: 10.1038/nature09937. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3093433/http://www.nature.com/doifinder/10.1038/nature09937.

World Health Organisation (2017). *WHO | Malaria*. URL: http://www.who.int/mediacentre/factsheets/fs094/en/.

— (2019). "WHO | Q&A on the malaria vaccine implementation programme (MVIP)". In: *WHO*.

World Health Organization (2017a). "A global brief on vector-borne diseases". In: *World Health Organization*, p. 9. DOI: WHO/DCO/WHD/2014.1. URL: http://apps.who.int/iris/bitstream/10665/111008/1/WHO_DCO_WHD_2014.1_eng.pdf.

— (2017b). "Global Vector Control Response 2017 – 2030: A strategic approach to tackle vector-borne diseases". In: *Who*, pp. 0–3. ISSN: 00111643. URL: http://www.who.int/vector-control/publications/WHO_HTM_GVCR_2017.01/en/.

— (2018). *World Malaria Report 2018*, pp. 978–92. ISBN: 9789241565653. DOI: ISBN9789241564403. URL: www.who.int/malaria.

World Health Organization: Global Malaria Programme (2013). "Test procedures for insecticide resistance monitoring in malaria vector mosquitoes: Second edition". In: *World Health Organisation Technical Report Series*, p. 22. ISSN: 15206106. DOI: 10.1021/jp052915b.

Wu, Bing, Liqun Luo, and Xiaojing J. Gao (2016). "Cas9-triggered chain ablation of cas9 as a gene drive brake". In: *Nature Biotechnology* 34.2, pp. 137–138. ISSN: 15461696. DOI: 10.1038/nbt.3444. URL: http://dx.doi.org/10.1038/nbt.3444.

Wu, Xuebing et al. (2014). "Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells". In: *Nature Biotechnology* 32.7, pp. 670–676. ISSN: 15461696. DOI: 10.1038/nbt.2889.

Xiang, Guanghai et al. (2017). "Temperature effect on CRISPR-Cas9 mediated genome editing". In: *Journal of Genetics and Genomics* 44.4, pp. 199–205. ISSN: 1673-8527. DOI: 10.1016/J.JGG.2017.03.004. URL: https://www.sciencedirect.com/science/article/pii/S1673852717300565?via%3Dihub.

Yakob, Laith, Luke Alphey, and Michael B. Bonsall (2008). "Aedes aegypti control: The concomitant role of competition, space and transgenic technologies". In: *Journal of Applied Ecology* 45.4, pp. 1258–1265. ISSN: 00218901. DOI: 10.1111/j.1365-2664.2008.01498.x.

Yan, Yao and Gregory C. Finnigan (2018). "Development of a multi-locus CRISPR gene drive system in budding yeast". In: *Scientific Reports* 8.1. ISSN: 20452322. DOI: 10.1038/s41598-018-34909-3.

Yang, Jianying and Michael Reth (2012). "Drosophila S2 Schneider Cells: A Useful Tool for Rebuilding and Redesigning Approaches in Synthetic Biology". In: Humana Press, pp. 331–341. DOI: 10.1007/978-1-61779-412-4{_}20. URL: http://link.springer.com/10.1007/978-1-61779-412-4_20.

Yasuoka, Junko et al. (2006). "Community-based rice ecosystem management for suppressing vector anophelines in Sri Lanka." In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100.11, pp. 995–1006. ISSN: 0035-9203. DOI: 10.1016/j.trstmh.2005.12.014. URL: http://www.ncbi.nlm.nih.gov/pubmed/16730765.

Young, Joshua et al. (2019). "CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-target Activity and Its Relevance in Crop Improvement." In: *Scientific reports* 9.1, p. 6729. ISSN: 2045-2322. DOI: 10.1038/s41598-019-43141-6. URL: http://www.ncbi.nlm.nih.gov/pubmed/31040331http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6491584.

Yue, Bai Gong and Göran Akusjärvi (1999). "A downstream splicing enhancer is essential for in vitro pre-mRNA splicing". In: *FEBS Letters* 451.1, pp. 10–14. ISSN: 00145793. DOI: 10.1016/S0014-5793(99)00521-9.

Zetsche, Bernd et al. (2015). "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System". In: *Cell* 163.3, pp. 759–771. ISSN: 00928674. DOI: 10.1016/j.cell.2015.09.038. URL: http://www.cell.com/article/S0092867415012003/fulltext.

Zhang, Dong et al. (2019). "Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design." In: *Proceedings of the National Academy of Sciences of the United States of America* 116.18, pp. 8693–8698. ISSN: 1091-6490. DOI: 10.1073/pnas.1820523116. URL: http://www.ncbi.nlm.nih.gov/pubmed/30988204http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6500161.

Zhou, Hong et al. (2017). "Whole genome analysis of CRISPR Cas9 sgRNA off-target homologies via an efficient computational algorithm." In: *BMC genomics* 18.Suppl 9, p. 826. ISSN: 1471-2164. DOI: 10.1186/s12864-017-4225-1. URL: http://www.ncbi.nlm.nih.gov/pubmed/29219081http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5773898.